Copyright © 2012 American Scientific Publishers
All rights reserved
Printed in the United States of America

Journal of
Advanced Microscopy Research
Vol. 7, 1–7, 2012

The Significance of Image Analysis for Cancer Diagnosis

Jona J. Sela[1,*], Alfred Bruckstien[2], Gal Goshen[3], Uri Dubin[3], Nir Karasikov[3], and Juri Kopolovic[4]

[1] *Institute of Dental Sciences, Hebrew University-Hadassah Faculty of Dental Medicine, Jerusalem 91120, Israel*
[2] *Computer Science Department Technion, Technological Institute, Haifa 32000, Israel*
[3] *Ardia, Medical Products, Haifa 31250, Israel*
[4] *Department of Pathology, Hebrew University-Hadassah, Faculty of Medicine Jerusalem 91120, Israel*

Cancer diagnosis is established by pathologists by means of microscopical analysis of cells and tissues. The accuracy of this methodology is dependent upon human skills, training, and judgment. The conventional diagnostic methods are currently re-examined in view of innovative trends in image analysis. The progress in cancer diagnosis by integration of computer assisted analysis is forthcoming. The current study presents several methods of computer assisted histopathological analysis that were recently developed. Biopsy tissue-sections served for the acquisition of microscopical images. These were enhanced and analyzed by dedicated algorithmic functions to evaluate complex and customized features of cell and tissue textures. The methods were designated to differentiate normal from neoplastic features, assisting the pathologist with the diagnosis. Consequently, the assembly of the data by the system, identification of specific cellular and tissue patterns and properties, has been proven to be useful for assessing efficacy of cancer therapy. The technological development was based on models of carcinoma that served for computer analysis of typical neoplastic changes in cells and tissues. The developing of a computerized diagnostic system on one type of cancer was followed effortlessly by its application to other types of cancer. Decision making system, to replace the pathologist, is not proposed. However, a substantial support to the diagnostic process is pertinent. In time, with increase of practice, this methodology is expected to become more accurate than the human eye and mind, in detecting minute deviations in cellular and tissue structures.

Keywords: Histopathological Diagnosis, Cancer, Neoplasia, Microscopy, Computerized Image Analysis, Artificial Intelligence, Fuzzy Logic, Neural Networks.

RESEARCH ARTICLE

1. INTRODUCTION

Routine histopathological analysis is wholly reliant upon human expertise. Classical diagnosis has improved in the second half of the 20th century, mainly throughout the development of immunohistological techniques. Whilst numerous medical diagnostic procedures have integrated computer based image analysis, surgical pathology remained a stronghold of conservative approaches.[1–3] Detailed descriptions of abundant histological features distinguishing neoplastic from normal tissues have been compiled in numerous publications and textbooks. This copious information is studied and memorized by pathologists to serve in combination with personal experience, in the process of diagnosis. It is obvious that the procedure entails an immense responsibility.[4–6] The reliance on human decision requires revision in view of several issues. The diagnostic process is a function of training, skills, experience, fatigue, and other personal traits. These predicaments may impair the competence of the pathologist, every so often arriving at an incorrect decision. Preliminary efforts to implement image expert systems for histopathological analysis have been reported. Bamford and Lovell[1] developed a detection system for nucleus analysis by means of active contour methods. Cai and associates,[2] applied support vector machine in cancer classification with a focus on patterns of gene expression. Comaniciu and Meer[3] proposed a cell segmentation algorithm, based on mean shift for clustering in color space, for image guided decision support system

*Author to whom correspondence should be addressed.

in clinical pathology. Sajda and associates[4] used neural network methodologies comprising image preprocessing and feature extraction. Mangasarian et al.,[5] described a system for analysis of images acquired from fine needle aspiration, taken directly from the breast lump. The methodology involved manual detection and measuring of fundamental features, in particular: area, radius, perimeter, symmetry, number, and size of concavities, fractal dimensions, smoothness, and texture. Mean extreme and standard deviation values of these measurements were used as feature vectors for classification. Carlson and Bergstrand[6] have developed a breast cancer biopsy analysis system, which included image acquisition and processing algorithms. The algorithms were based on neural networks. Further preliminary steps for the utilization of advanced machine vision techniques using quantitative analysis were reported recently.[7–15] The process of cancer diagnosis is based on memorizing and processing of a vast volume of information. In this regard, the recruitment of computer technology to assist the development of expertise data base is imperative.[9] In recent years, image analysis techniques evolved massively in different sciences.[4, 16, 20] Regrettably, histopathological diagnosis of cancer is in arrears. It is advocated that diagnosis could be effectively assisted by a computer based processing. Furthermore, a significant improvement in early detection of cancer is expected to ensue. The poor inter-and intra-diagnostic correlation between individual pathologists[8] rendered further support to the claim, that computerized quantitative systems could provide support, not yet replacement, to the pathologist, and would be of extreme value.

The purpose of the present communication is to describe the successful development of a semi-automated image analysis assistance system that enables qualification and quantification of morphological features, patterns, and data. Following direct acquisition of images from microscopical slides the system is designed to filter insignificant data hence assisting the specialist to concentrate with higher efficiency and reliability on critical issues of diagnosis.

2. MATERIALS AND METHODS

Biopsy specimens of Squamous Cell Carcinoma were used to develop the computerized diagnostic system and the proof of concept. Microscopical sample preparation comprised, fixation, embedding, 6 μm thin slice sectioning, mounting on glass slide and staining with Hematoxylin and Eosin. Computer software technology was assembled in compatibility with standard PC platform for image acquisition, pre-processing, and segmentation. A standard commercially available digital camera was used. This was equipped to interface with a standard microscope (Olympus, Model BH-2, Japan) used for diagnostic pathology. Acquisition of images was followed by an initial scanning of the samples for abnormalities at tissue level. This provided an overview of the biopsy, allowing sampling of healthy and neoplastic tissues and cells. Identification of the suspected areas was followed by image capturing of a series of typical microscopical fields at 10x and 40x magnifications for detection of changes at tissue and cellular levels (Fig. 1). Direct storing of data in a basic personal computer to be analyzed by supplementary tools completed this stage. Further processing consisted improving of accuracy, changing limits of detection, and solution of issues immerging from the variability in the microscopic histological samples. At this stage

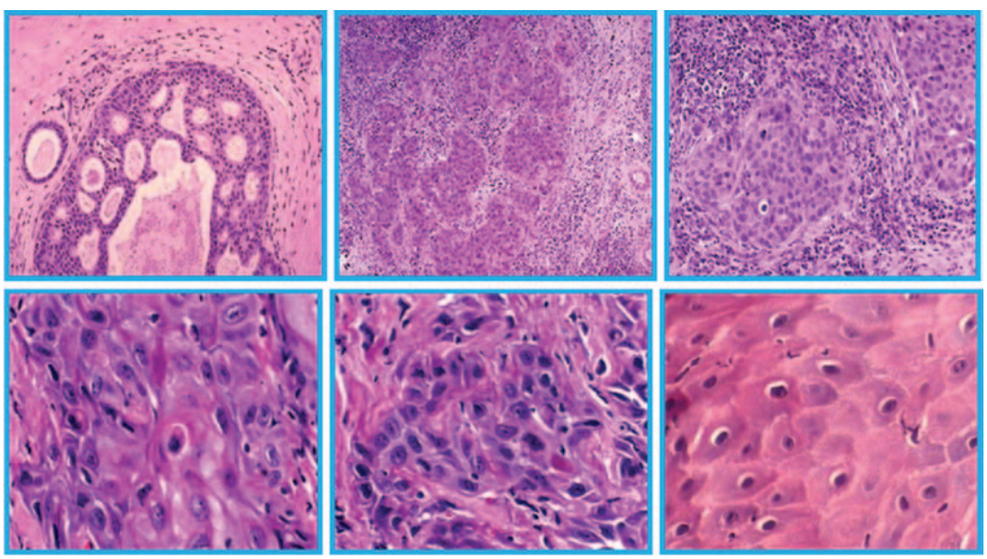

Fig. 1. Typical malignant changes in epithelial tissue in acquired microscopic images of squamous cell carcinoma (hematoxylin and eosin staining; upper 3 photos, magnification of 10×; lower 3 photos, magnification of 40×).

RESEARCH ARTICLE

computer digital pre-processing covered all adjustments for possible variability due to tissue preparation, staining, and camera illumination conditions. The modifications for inconsistencies in concentration and spatial distribution of cells were performed and sample-specific thresholds determined by comparison of normal to suspected tissue images. These consisted extraction of cell structures on the basis of color segmentation, morphological filtering and shape feature extraction. The procedure is based on standard morphological functions and application of proprietary algorithms. These include area, granularity, centroid mean color, color deviation, orientation, Euler number, eccentricity, solidity, convex hull perimeter, texture, and a number of other proprietary features.[5, 17–19] At this stage, each colored region/cell was computer identified by a list of digits that recorded its individual properties. Classification of tissue and cell populations using multidimensional analysis of the measured values for each parameter listed above completed this stage. In addition, calculation of cancer probability using advanced proprietary algorithms and Support Vector Machine (SVM) were performed.

3. RESULTS

Acquisition of images of both normal and suspected tissues from each patient allows the setting of patient-specific thresholds. Differentiation of normal from suspected tissue requires quantification and adjustment for variability in concentration and spatial distribution of cells, to establish sample-specific thresholds. Once corrected for these factors, the field also conveys spatial cell distribution and texture data that can be described as additional features of a lesion. The next phase of analysis consists of image segmentation this is performed using Red-Green-Blue (RGB) based filters (standard output of digital cameras) to assign numeric values to various visual features. Malignancy Index is calculated by a dedicated algorithm that determines the deviation from the normal values. Index values close to zero indicate that the suspected image is normal. The deviation from zero increases the probability of malignancy (Fig. 2).

Demonstration of color filtering of a field of epithelial tissue image is presented in Figure 3. Color filtering does not eliminate background noises and necessitates an additional morphological filter to clear inter-cellular regions of the image not involved in the analysis. Based on this filter, artifacts are removed. When the shape factor constraint is applied to a color-filtered image, we are able to separate cells from the intercellular, or background, space.

Morphological filtering is presented in Figure 4. The image displays the overlay of the segmented and original images. Throughout these processes and in the final segmented image, the software captures data, that when analyzed, have the potential to assign cancer probabilities to suspected tissues.

As an example, 'eccentricity' values measured for each cell are displayed on the labeled image in Figure 5. The data are generated from the image for all the extracted features. Visualization the collected information is received by assigning each measured parameter or feature to an axis in a multi-dimensional space. A calculation of cell area versus cell orientation is demonstrated in Figure 6. In this particular example the orientation feature is not informative due to its even spread. Classification of parameters and features is based on known, well-developed methods like SVM algorithms adapted to the system's specific needs. The classifier training relies on normal and lesion feature data. Partial input of the classification algorithm is showed in Table I. The algorithm classifies each image set using measured feature statistics (columns in the Table I). Up to

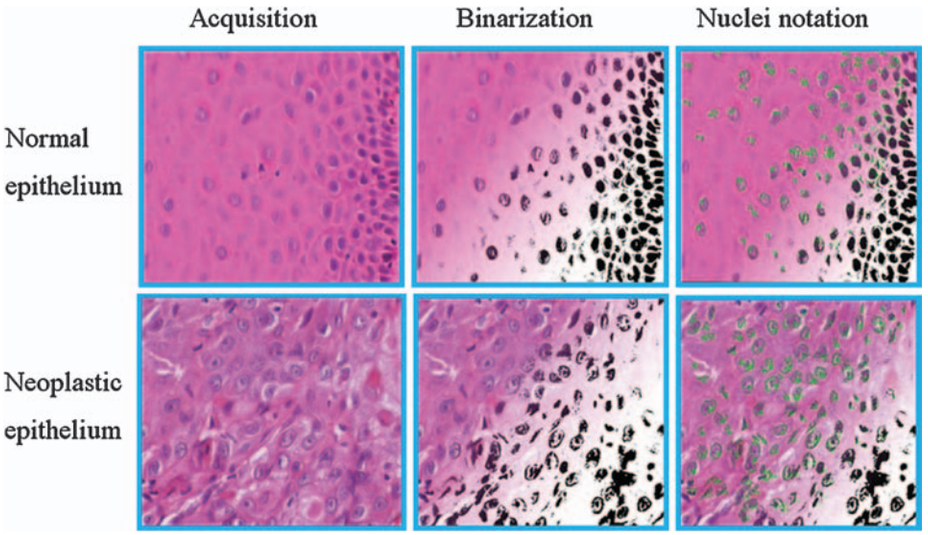

Fig. 2. Data indexing, normal versus suspected tissues.

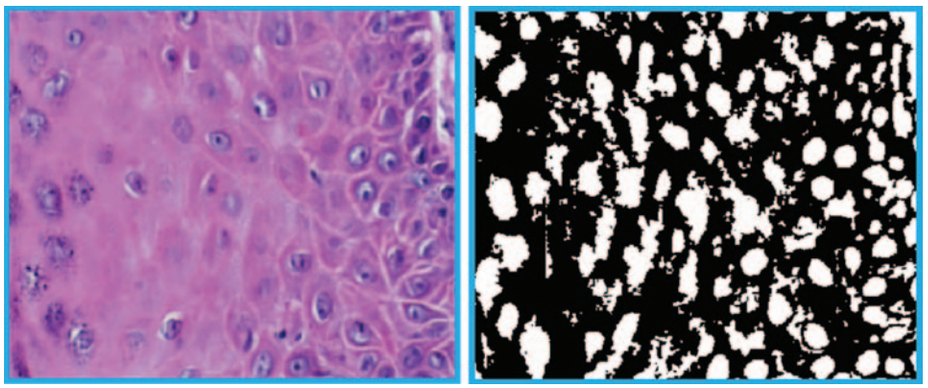

Fig. 3. (left) A field of epithelial tissue at magnification of 40× served for the subsequent computer analysis in Figures 4 and 5. (right) Color-based segmentation of the same field.

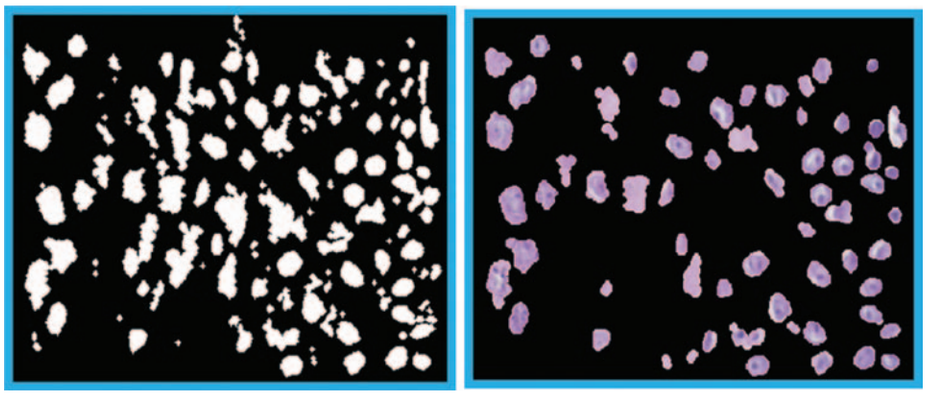

Fig. 4. (left) A color-segmented image of the field shown in Figure 3; (right) shape factor filtered of the same field.

50 features of normal and pathological data were recorded for each site and each patient (set) measure. The columns in Table I display a small part of the entire dataset generated by the image processing algorithm. Since each image set is identified with a particular patient, results ranking can be performed according to their probability to be cancerous tissue.

In the final step, each row (vector of 50 entries which describes the collected information) is fed to the classifier, a software algorithm, which makes the decision. SVM algorithm[18] is used to calculate the probability that tissue changes in the sample are in consistency with cancer. This probability statistic gives the pathologist objective information about the individual slide that can assist in ensuring a correct diagnosis.

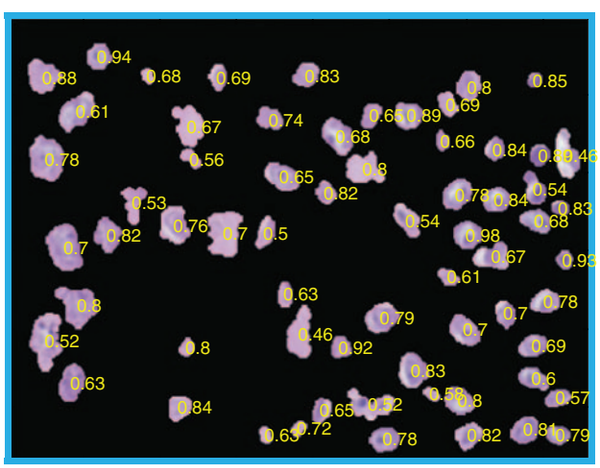

Fig. 5. Sample map of 'eccentricity' measured for each cell, based on the experimental data of the field shown in Figures 3 and 4.

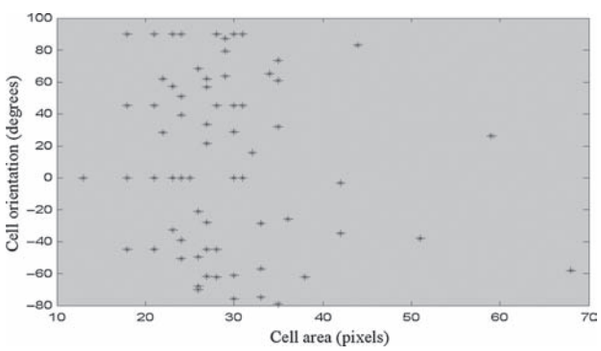

Fig. 6. A sample of detected cell orientations and areas from a single image.

Table I. Features (columns) versus data set and location (rows) partial table example.

Set	Location	Area	Orientation	Color
1	A	76.3	45.1	78
2	B	44.2	23.7	67
3	C	44.9	−35.2	91

Feasibility and accuracy were tested by generating cancer probability features by the computer assisted system and comparing them with the diagnosis determined by the pathologist. Calculation of the sensitivity and specificity showed a high probability statistical prediction of cancer. It should be pointed out that the application of artificial intelligence technology is a key component of the system. As the database of histological samples is in constant growth the plan is to evolve by gathering increasing amounts of data, thus improving the accuracy of the algorithmic performance.

In summary, the software system receives histological images, extracts relevant data, calculates a cancer probability, and returns a cancer probability statistic to the pathologist. The application of the technology for the analysis of a histological sample obtained from a case of intraductal mammary carcinoma is presented in Figures 7–10.

The input image preprocessing consists of elimination of variations in lighting and staining: specifically, light normalization, color normalization, and comparison to normal images. Figure 7 is an example of the input of an acquired image. At this stage the image is prepared for segmentation and structural analysis as shown in Figure 8, the process is based on extraction of structures using color and intensity. The use of morphological operations for image opening is shown in Figure 9, this consists filtering to reduce pixel noise and filling of the remaining structures. Labeling of the connected regions is represented in Figure 10. Each color corresponds to a different cell. The results can be further improved by detecting cells of a grain of certain

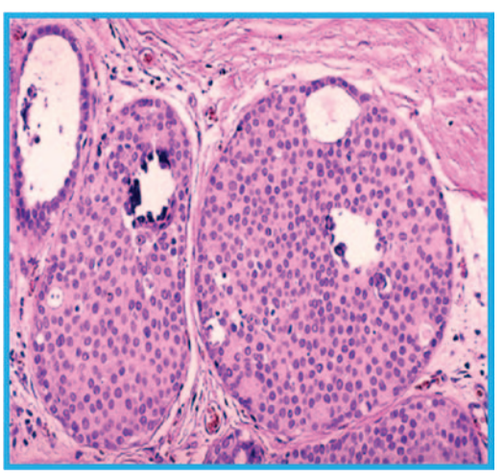

Fig. 7. Input image of intraductal carcinoma, magnification of 40×.

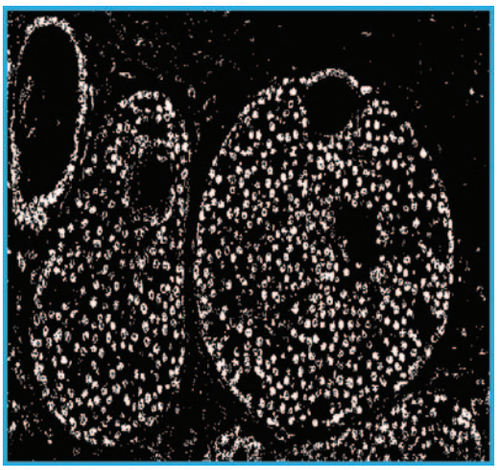

Fig. 8. Color based segmentation of the microphotograph shown in Figure 7.

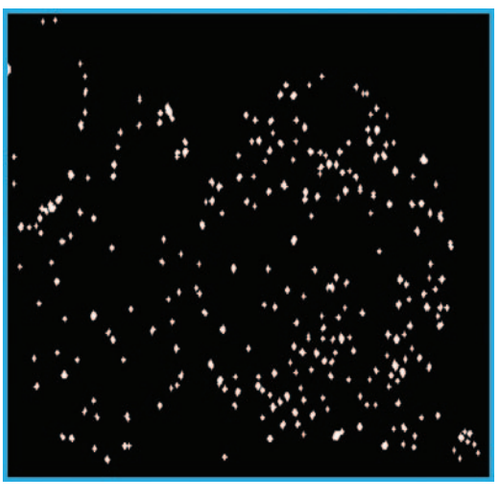

Fig. 9. Image opening result for the microphotograph shown in Figure 7.

Fig. 10. Labeling of the connected regions for the microphotograph shown in Figure 7.

RESEARCH ARTICLE

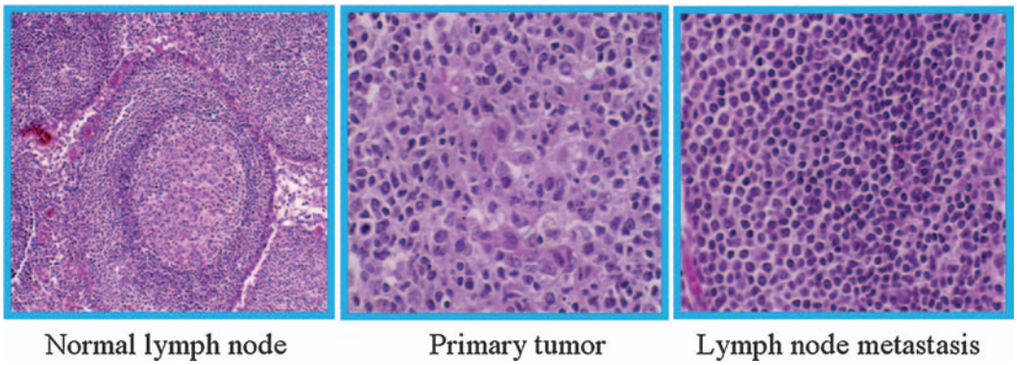

Fig. 11. Metastatic cancer cells in lymph gland.

size. This format of image processing allows a direct assistance to the pathologist for an easy retrieval of catalogue-arranged images and a potential baseline for collaboration among networks of pathologists. Detection of metastatic, cancer cells, in an abundant lymphocytic population of a lymph gland as represented in Figure 11, is recognized as one of the most challenging predicaments in pathological routine. In order to facilitate the identification of the neoplastic cells in a given lymph node, a simple method has been devised. Lymphocytes are automatically characterized by segmentation and feature extraction and deleted as shown in Figure 12. Lymphocyte-free images facilitate the identification of neoplastic cells the characteristics of which are similar to those of the original tumor. Metastasis index can serve in grading the malignant changes as compared to the primary tumor. The computerized system provides an elegant solution for the issue of request of a second opinion from another expert. A web-based

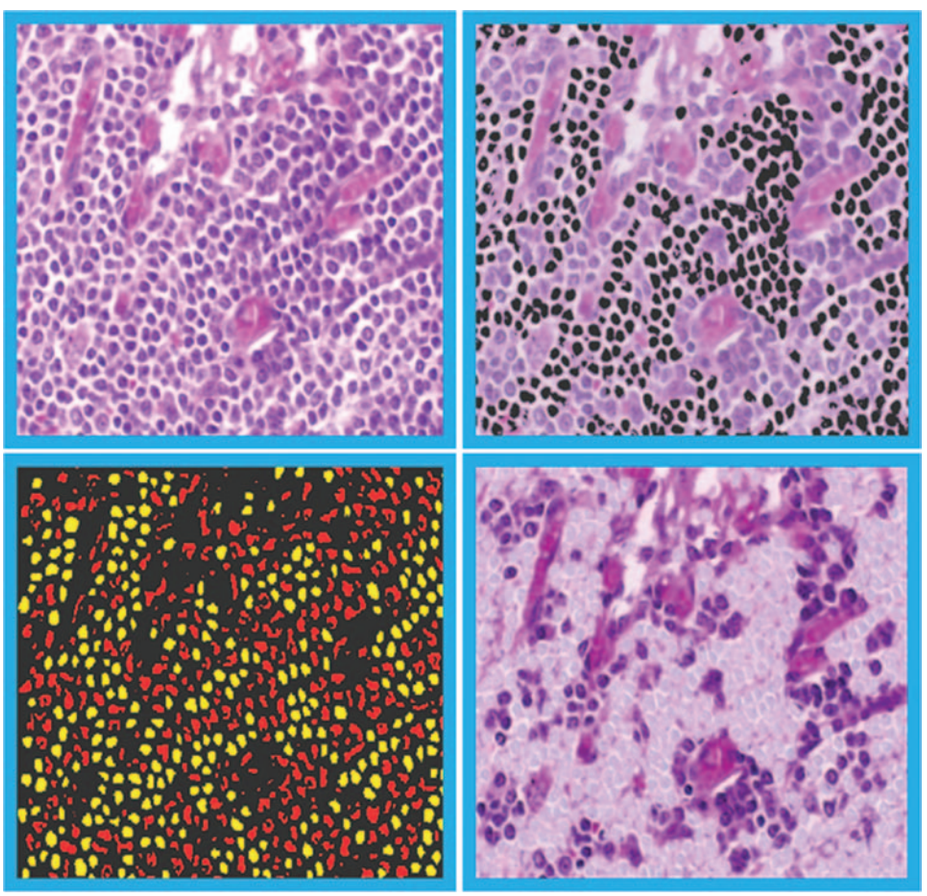

Fig. 12. Lymphocyte identification and deletion.

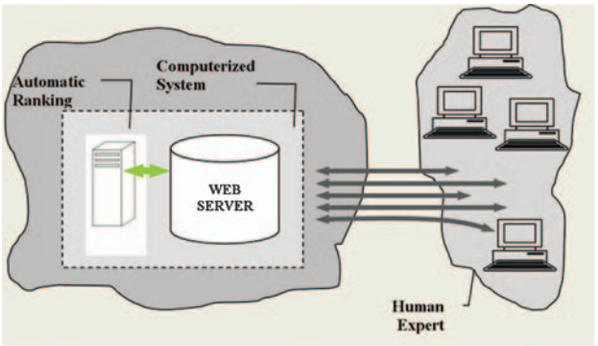

Fig. 13. Web-based expert assistance system.

expert assistance system is describes in Figure 13. Web-based services around the world consist in facilitation of the expert consultation, maintenance of an international database accessible for authorized users. This entails, confidentially coded data samples and corresponding expert opinions, training and tuning of the computer artificial expert system benefiting from a rich database and corresponding human analysis and a follow up assessment of efficacy of cancer treatment.

4. DISCUSSION

The present computerized system was developed in order to assist the pathologist in the process of diagnosis. Analysis by a real time quantitative computer program, proposes a valid assistance for make a decision on malignancy. Squamous cell carcinoma was chosen to initiate the project. The sensitivity and repeatability of the system are expected to match, possibly exceed, the normal diagnostic hit rate. In instances, analysis by the system might assign a high probability statistic identifying atypical changes in a sample that was found normal by the pathologist. The high probability statistic in combination with a negative diagnosis by the pathologist may prompt a more careful analysis or referral to a second opinion. In this way, the computerized system serves as a check on variability in the microscopic diagnosis attributable to human factors. The probability statistic can also provide a tool for pathologist to prioritize workloads and to expedite highly suspected cases. Another achievement of the system lies in the detection of minute tissue changes, undetected by the human eye, thus identifying a pre-malignant lesion. WEB connection allows peer consultation over the Internet on a set of normalized and calibrated images, thus saving precious time in the diagnostic procedure. Based on its features, the system can relate a lesion to other known tumors in the database and by this very early identification, allowing a more efficient treatment with a better prognosis. It is believed that this kind of vision is what this project is all about. At this stage the system development reached an 85% accuracy of diagnosis on 2000 slides of carcinoma of the oral epithelium.

5. CONCLUSIONS

Computer-assisted detection enables pathologist to achieve a more confident and accurate decision by analyzing, identifying and grading suspicious areas on histological slides.

The computerized system, in its current status, is a valuable tool that can assist pathologists in the diagnostic work, research, and education.

Monitoring of cancer therapy by consecutive biopsies allows a continuous follow up by computerized comparison of samples from the same individual to detect cellular changes. This can be done by the system superiorly to any other mode of tentative human eye-brain inspection.

For healthcare providers, the system is designed to improve the quality of patient care, reduce medical errors and related litigation.

References and Notes

1. P. Bamford and B. Lovell, Proc. 9th Brit. Machine Vision Conf. Southampton, UK **(1998)**.
2. J. Cai, A. Dayanik, H. Yu, N. Hasan, T. Terauchi, and W. N. Grundy, *Bioinform.* 16, 906 **(2000)**.
3. D. Comaniciu and P. Meer, Cell Image Segmentation for Diagnostic Pathology, Advances in Pattern Recognition Ser., Springer, NY **(2001)**.
4. P. Sajda, *Ann. Rev. Biomed. Eng.* 8, 537 **(2006)**.
5. O. Mangasarian, W. Street, and W. Wolberg, *Anal. Quant. Cytol. Histol.* 15, 396 **(1993)**.
6. C. Carlsson and K. Bergstrand, Ph.D. Thesis, Cent. App. Autonomous Sensor Sys. Örebro University, Sweden **(2002)**.
7. R. T. Grenko, C. S. Abendroth, E. E. Frauenhoffer, F. M. Ruggiero, and R. J. Zaino, *Am. J. Clin. Pathol.* 114, 735 **(2000)**.
8. A. Renshaw, *Am. J. Clin. Pathol.* 115, 338 **(2001)**.
9. F. T. Bosman, *J. Pathol.* 194, 143 **(2001)**.
10. V. Vapnik, Statistical Learning Theory, J. Wiley and Sons, NY **(1998)**.
11. N. Cristianini and J. Shawe-Taylor, An Introduction To Support Vector Machines (And Other Kernel-Based Learning Methods), Cambridge Univ. Press **(2000)**.
12. B. Schlkopf and A. J. Smola, Learning With Kernels: Support Vector Machines, Regularization, Optimization, and Beyond (Adaptive Computation and Machine Learning), MIT University Press **(2001)**.
13. C. D. Mullins, *Am. J. Manag. Care* 5, S371 **(1999)**.
14. A. Gasteratos, Specialized Hardware Structures For Morphological Image Processing, edited by R. Fisher, On-Line Compendium Computer Vision **(2000)**.
15. S. J. Keenan, J. Diamond, W. G. Mccluggage, H. Bharucha, D. Thompson, P. H. Bartels, and P. W. Hamilton, *J. Pathol.* 192, 351 **(2000)**.
16. P. J. Bickel, Y. Ritov, and A. Zakai, *J. Machin. Learn. Res.* 7, 705 **(2006)**.
17. R. Wootton, D. Springall, and J. Polak, Image Analysis in Histology: Conventional and Confocal Microscopy, 1st edn., Cambridge Univ. Press **(1995)**.
18. W. Skarbek, *Proc. 9th Intern. Conf. CAIP*, Poland, 1st edn. Springer Verlag, NY **(2001)**.
19. Joyce-Loebl, Image Analysis - Principles and Practice: A Technical Handbook, 1st edn., Gateshead Pub. UK **(1985)**.
20. A. Maly, G. Goshen, J. Sela, M. Pinelis, M. Stark, and B. Maly, *J. Human Pathol.* 36, 1074 **(2005)**.

Received: 21 July 2012. Accepted: 2 August 2012.