Cancer Detection using Neural Computing Methodology

Hamid S. Kohen
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, Ca 91109
Kohen@brain.jpl.nasa.gov

Nikzad Toomarian
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, Ca 91109
Nikzad.Toomarian@jpl.nasa.gov

Gregory H. Bearman
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, Ca 91109
Gergory.H.Bearman@jpl.nasa.gov

David B. Seligson
Department of Pathology
University Of California at Los Angeles
Los Angeles, Ca 90095
DSeligson@mednet.ucla.edu

Keywords: Cancer, neural network learning, morphology, molecular assessments, multi-spectral images, and liquid crystal tunable filters.

Abstract

This paper describes a novel learning methodology used to analyze bio-materials. The premise of this research is to help pathologists quickly identify anomalous cells in a cost efficient method. Skilled pathologists must methodically, efficiently and carefully analyze manually histopathologic materials for the presence, amount and degree of malignancy and/or other disease states. The prolonged attention required to accomplish this task induces fatigue that may result in a higher rate of diagnostic errors.. In addition, automated image analysis systems to date lack a sufficiently intelligent means of identifying even the most general regions of interest in tissue based studies and this shortfall greatly limits their utility. An intelligent data understanding system that could quickly and accurately identify diseased tissues and/or could choose regions of interest would be expected to increase the accuracy of diagnosis and usher in truly automated tissue based image analysis.

We have developed a diagnostic tool (Bio-Analysis system) that uses a novel framework of morphological and molecular assessment of human histopathologic tissues based on multi-spectral signatures. The Bio-Analysis system creates multispectral imagery of tissue sections called 'BioCubes' and integrates them with neural network learning processing techniques to provide analysis and diagnosis capabilities. A complex ultra dimensional data set of bio-cubes of both benign and malignant prostate glandular epithelium was created and analyzed. In assigning either a benign or a malignant group designation to each gland, conclusive results were obtained with a high-speed turn-around and a 98% correct classification, as compared to manual analysis. System knowledge was gained by biologically inspired neural network learning algorithm generalization of spectral and spatial information. This technology can provide a human like learning process, which may eventually be utilized to help physicians more inexpensively and accurately diagnose normal vs. abnormal biological tissues. This methodology is also applicable to planet surface information gathering and to molecular assessment of bacteria.

1.0 Introduction

Pathologists view pathologic tissues, typically with brighfield microscopes, to determine the degree of normalcy versus disease. This process is time consuming, and fatiguing. The induced exhaustion created by this process may contribute to diagnostic errors. This research describes a Bio-Analysis system that classifies and identifies pathological tissues in an automated fashion. It provides a novel framework of morphological and molecular assessment of these bio-materials.

In section 1.1, we provide a brief description of multi-spectral signatures of a cell and an analysis of how this concept is related to health care problems. In section 2.0, we discuss our Bio Analysis System setup, and we make a brief comparison to technologies used in health care industry. Later, in section 3.0, we provide a brief description and a motivation for using neural networks by reviewing some of the work in the area as applied to multi-spectral data analysis. In the final sections, we show simulation of results and conclude by analyzing the results.

1.1 Multi-spectral signatures of cells

Every bio-material has a unique spectral distribution that can be viewed as its "fingerprint", hence its importance in space exploration or health care. Current bio-analysis system sensors have 22 to 24 channels of data. Future bio-sensors will produce even larger numbers of data channels. Currently, we create our own BioCube data sets from a prostate Tissue Micro-Array (TMA) slide obtained from UCLA. Each BioCube encompasses 22 images taken at different wavelengths (superimposed sensory channels). The wavelengths range from 470 nano-meters to 710 nano-meters of the electromagnetic spectrum. Epi-fluorescence light energy from a hematoxylin and eosin (H&E) stained TMA is continually transmitted through a Zeiss

microscope and a liquid crystal tunable filter. A sophisticated 12-bit precision camera converts the spectrum of transmitted energy into an image. Figure 1 represents a single pixel spectrum from a BioCube.

The H&E tissue micro-array biopsy slides were produced by UCLA hospitals; each spot in the slides contains benign and malignant prostate cells. The technique of H&E staining and their transmittance properties are beyond the scope of this study. See any standard histotechnology book for more details. For each specimen on the micro-array slide, a 12-bit precision black and white camera was used to capture 20 to 50 images at different wavelengths (sensory channels). These images were laid on top of each other in the shape of a cube, henceforth, the "Bio-cubes". These Bio-cubes contain multispectral imagery of cells in the ranges of from 450 nm to 710 nm, every 5nm band.

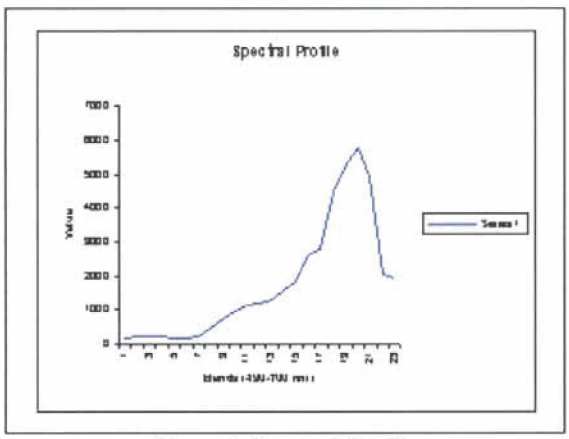

Figure 1 Spectral Profile

1.2 Medical utilization

Multispectral sensors for each bio-material have a unique distribution of signatures that can be viewed as an image (i.e., Figure 5). Processing of such vast amounts of data, especially in real time, are computer intensive [1,2]. Therefore, there is a clear need for advanced bio-inspired computing software and hardware that can effectively handle these significant data sets.

Trained pathologists must carefully view all diagnostic histopathologic materials to determine the presence of tumor, in this example, and the degree (different grades) of malignancy. This process is time consuming and the induced fatigue created by this process may results in human errors. Bio-Analysis systems will help pathologists identify anomalous cells in a precise and cost efficient manner. In addition, the technique is useful as an adjunct to further automated image analysis.

Figure 2 depicts a typical H&E stained TMA tissue spot, imaged from a slide. Each TMA spot inhabits a known coordinate location, and thus tissue and patient database information may be linked to each spot.. Dr. Seligson, a UCLA pathologist has diagnosed each biopsy tissue spot, with the following diagnosis, the top 4 rows are all benign glands, the bottom 4 rows have mostly tumor, but also some intervening normal glands and some PIN (prostatic intraepithelial neoplasia). There are also varying grades, usually from 3 to 5 out of 5 (grade 2 and especially grade 1 are relatively rare). Our test and verification data included a grade on each tumor image. Please note how variable these tumors can look from this selection (Figure 2.)

In general, prostate glands have a double-layered cellular architecture, an outer basal cell layer, and an inner (luminal) glandular cell layer. One assumes in general that benign glands have these two layers, and tumor glands have one layer – though that is not 100% true, it is a reasonable generalization that provides a clue to diagnosis. Deviations from these rules include the fact that the basal cell layer may be slightly discontinuous, allowing for small focal regions of single cell layered benign glands, and indeed, conversely, all dysplasias (eg. prostatic intraepithelial Neoplasia PIN) and some early malignancies remain in regions confined by a basal cell layer, especially where the tumor glands grow along an existing gland or duct. These regions represent carcinoma *in situ*.

For our purposes here, we have assumed that the benign glands have a basal cell layer, since detecting these layers is an important clue to standard diagnostic separation of 'benign' and 'tumor' glandular cells, or more precisely, benign and *in situ* lesions from infiltrative malignancies..

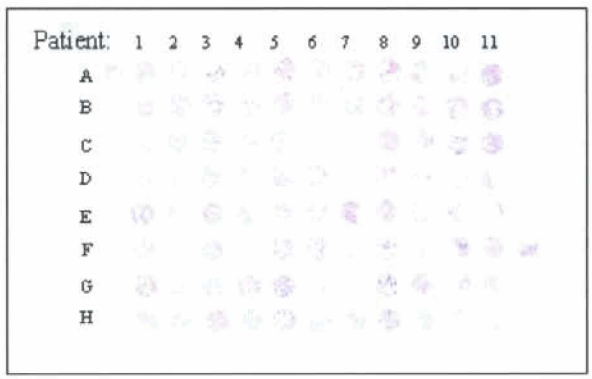

Figure 2 Microarray Biopsy Tissue Cells of Prostate

Please notice, each column (A-H) is from one patient, so it will be interesting to see how different tissue areas in the same patient can generate different spectral signatures.

2.0 Bio-Analysis System Setup

Figure 3, depicts the Bio-Analysis system and Figure 4, shows Bio-Analysis Algorithm, where BioCubes are produced and analyzed. As a first step in analyzing BioCubes, the principle component analysis (PCA) was applied to the entire data set to reduce dimensionality. However, after a review of the final results, it was necessary to use all the bands at this time, because, reduced dimensionality had lost vital information

necessary for classification. All spectral bands of each pixel were used as input into the bio-analysis algorithm.

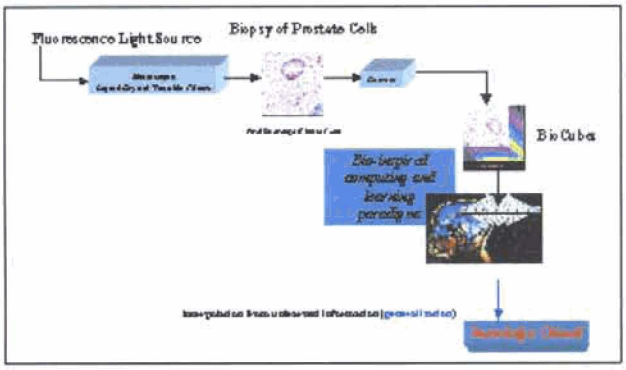

Figure 3 Bio-Analysis System

A self-organizing feature of unsupervised neural networks was applied to the BioCubes imagery [3,4]. This method has partitioned an image cube into individual classes based upon the similarity among the spectral bands of a pixel. Therefore, input BioCube was mapped onto a two-dimensional array of six possible classes and was labeled with the class number (color).

Feature extraction of morphological information was acquired by applying a mathematical filter to capture features such as size, shape, and orientation of cell nuclei [5]. A combination of self-organizing feature classes, molecular morphology, and supervised learning was applied for further identification of classes.

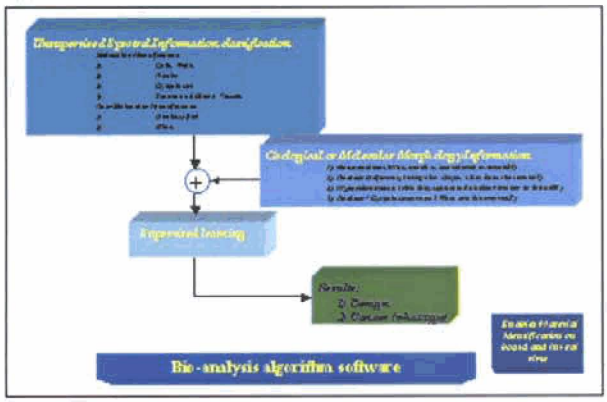

Figure 4 Bio-Analysis Algorithms

2.1 Comparison to Technology used by Physicians

Multispectral image data of tissues provides the information of the absorption and scattering of the cells in the tissues. The **optical diagnosis** can potentially provide an accurate and efficient method as well as lower health care costs. In comparison, Ultrasound and Magnetic Resonance Image (MRI) are two state-of-art techniques widely used by physicians as diagnostic tools that provide information on gross tissue structure and some blood flow data to soft tissues. Both methodologies provide excellent soft tissue differentiation, display and boundary contrast between anatomical structures. Alternatively, imaging technologies such as optical coherence tomography and Computed Tomography (CT) imaging, also known as "CAT scanning" (Computed Axial Tomography) also provide structural information, **but do not provide information on cell biochemistry**. The optical properties of biochemistry of cell tissues has been studied and used for clinical or biomedical applications in recent years. For example, skin cancer is a disease in which malignant cells are found in the outer layer of the skin [6]. Tissue classification of multi-spectral MRI images of brain has been possible [7]. However, the range of wavelengths used is limited and the methodology of classification does not take advantage of morphology information of cells and tissues. There is also, no ongoing learning process. In summary, Bio-analysis technology can provide more details, and a better learning process to help the physicians to inexpensively and accurately diagnose tissue variations, such as normal vs. abnormal cells and tissues.

3.0 Learning multi-spectral data

Processing multi-spectral image cubes, in real time, is computer intensive [1]. There are several different algorithms referenced here that can effectively deal with these data sets. It is shown that bio-inspired neural network learning architectures may offer a valuable alternative to the Bayesian classifier [8,9]. With neural networks, the a posteriori probabilities are computed with no a priori assumptions about the probability distribution functions (PDFs) that generate the classification. The neural classifier learns from training examples and generates a general type of input-output mapping which is then designed optimally to comply with a given set of problem sets. It is demonstrated that a deterministic feedforward network, which is called the Boltzmann perceptron classifier (BPC), can efficiently compute the a posteriori class probabilities. Maximum a posteriori (MAP) classifiers are also constructed as a special case of the BPC. Structural relationships between the BPC and a conventional multilayer perceptron (MLP) are given, and it is demonstrated that rather intricate boundaries between classes can be formed even with a relatively modest number of network units. Simulation results show that the BPC is comparable in performance to a Bayesian classifier [10].

In practical pattern-recognition applications, the Gaussian classifier is often sub-optimal because some features are non-Gaussian or even discrete valued, the class statistics are only estimated, and the covariance matrix inversions can be ill-conditioned. Presently these problems are dealt with by mapping the Gaussian classifier to a sigma-pi neural network, to which it's isomorphic Gaussian classifiers are equivalent to "sigma-pi" networks [11]. Back-propagation learning is then used to improve classifier performance. With a multi-spectral cube of Landsat TM data, it has been shown that feedforward

neural networks yield better result than maximum likelihood method (Bayes optimal discrimination) [12].

In summary, there are several premises that advocate neural networks as the computational framework for the analysis of multi-spectral data. Hertz, Krogh and Palmer [13] have shown that many standard statistical classifiers are special case of neural networks. For instance, Yair and Gersho [10] have pointed out that MAP classifiers (i.e., classifiers that choose the class with the highest a posteriori probability) are a special case of Bolzmann pereceptron network. Yau and Manary [11] have shown the equivalence between Gaussian classifiers and "sigma-pi-" networks. Ruck [14] suggested that multilayer perceptron networks provide an excellent approximation to a Bayes optimal discriminant function. In addition, Benediktsson [15] has found neural networks are distribution free that can detect and exploit nonlinear data patterns and are superior to statistical methods in terms of classification accuracy. This is an advantage over statistical methods, particularly when there is no knowledge of the statistical distribution functions of the data. Herman and Khazenie [16], demonstrated that neural networks perform better or equal to conventional statistical classifiers on multispectral data.

4.0 Results

Bio-analysis algorithm has produced the following results. Unsupervised spectral signature classification module classified the TMA tissues into four molecular classifications: cell walls, nuclei, cytoplasm, stroma, and blood vessels. This module also classifies none producing molecular signatures into background or unknown / unclassified class. Supervised learning module combined spatio-spectral information for further classification and identification (benign or cancerous).

As part of the data collection and verification methodology we have created training and testing sets for comparison analysis. The training and test set contained the spectral information / molecular signatures and morphological information such as shape, orientation, and size of the nuclei of a cell. The supervised learning was performed on using the training set and validated by the test set.

Generally, it takes more than ten years of training to become an expert in the field of surgical pathological diagnosis. And still experts have some difficulty with diagnostic agreement, especially in what could be termed 'grey-area' cases with histologies falling into borderline morphologic categories, and with some inherent variance in perception within the pathology community as to exactly where these categories begin and end. It is possible that Bio-cube analysis could normalize this variance and reduce biased diagnoses. Figure 5 represents a typical pictorial classification and identification.

Analyzed results were compared with pre-diagnosed biopsy spots / biopsy images. One frequent major feature of malignancy is related with the nucleus of the cells, therefore, it was essential to obtain segmentation of the image, in order to isolate these nuclei from the rest of the image, i.e. from the cytoplasm, and. This research study is ongoing: currently we have analyzed 42 cubes out of 88 possible cubes, and we achieved success rate of 98% in classification and identification of the cells.

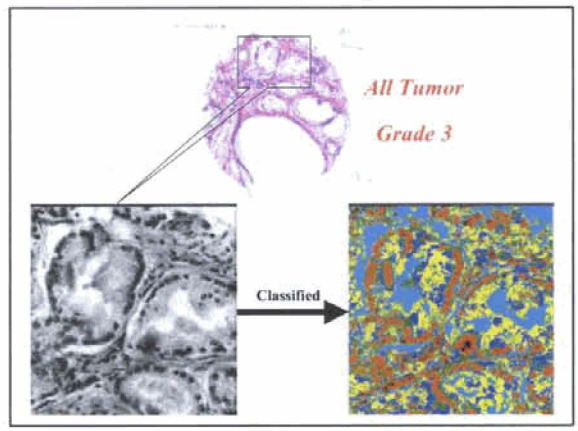

Figure 5 BioCube Classification

5.0 Conclusion

We have developed and brought together multispectral imagery principles and mechanisms of neuromorphic computing found in biological systems, and applied them to diagnostic detection of cancer.

Molecular assessment of the cells was achieved by applying multispectral imagery principles. An imaging spectrometer provides a reflectance and transmission spectral signature of molecules within the cell as well as cell morphology. We have correctly classified and identified bio-materials with high-speed turn around.

As part of our future goals, other methods that can be applied are: converting the bio-cube into its frequency component and then creating a filter that will morphologically classify the cells and tissues. In addition, further morphological distinctions (eg. tumor grading) could potentially be made with further neural network learning methodology. Also, other tissue types will be analyzed.

Acknowledgement

The research described in this paper was performed at the Center for Integrated Space Microsystems (CISM), Jet Propulsion Laboratory, California Institute of Technology, and was contracted by the National Aeronautics and Space Administration.

References

[1] Toomarian, N., and Kohen, H., *Neural Network for Analysis of Hyperspectral Imagery*, Fourth International Conference on Neural Networks and their Applications (Neuroap98), Marseilles, France, 219-224, March 11-13, 1998.

[2] Loch, T., Gouge, J, Bertermann, H, and Lee, F, *Real-time color image analysis (RCIA) by computer: Tissue objectivation?*, Abstracts, 3rd International Symposium on Transrectal Ultrasound of the Prostate, Kiel, Germany, Sept. 8-10, 1988.

[3] Kohonen, T., *Self-Organizing Maps*, Berlin: Springer-Verlag, 1995.

[4] Hepner, G.F., et al., Artificial *Neural Network Classification Using a Minimal Training Set: Comparison to Conventional Supervised Classification*, Photogram. Eng. & Rem. Sens., 56 (4), 469-473, 1990.

[5] Jones, R. and Svalbe, I., *Algorithms for the Decomposition of Color-Scale Morphological Operations*, IEEE Trans. Pattern Anal. Machine Intell, vol 16, No 6, June 1995.

[6] Cao, M., Kang, K.A., Chang C., Bruley, D.F., *Skin cancer detection based on Near Infrared image analysis,* Biomedical Engineering Conference, 431-436, April 1997.

[7] Qian, L. S., Xiang, L., Kun, l., Cheng, Liang, L. S., *Tissue classification of multi-spectral MR images of brain,* Engineering in Medicine and Biology Society, Proceedings of the 20th Annual International Conference of the IEEE, 634 - 636 vol.2, 29 Oct.-1 Nov. 1998.

[8] Lawrence, S., Burns, I., Back, A.D., Tsoi, A.C. and Giles, C. L., *Neural Network Classification and Unequal Prior Class Probabilities*, Tricks of the Trade, Lecture Notes in Computer Science State-of-the-Art Surveys, editor Orr, G., and K.-R. and Caruana, R., Springer Verlag, 299-314, 1998.

[9] Richard, M. D., and Lippmann, R. P. *Neural network classifiers estimate Bayesian a posteriori probabilities*, Neural Computation, vol. 3, 461- 483, 1991.

[10] Yair, E., and Gersho, A., *Maximum a posteriori decision and evaluation of class probabilities by Boltzmann perceptron classifiers*, Proceedings of the IEEE, Volume: 78 Issue: 10, 1620-1628, Oct. 1990.

[11] Yau, H.-C., and Manry, M.T., *Sigma-pi implementation of a Gaussian classifier*, Neural Networks, , Proceeding of IEEE/IJCNN Conference on Neural network ,825-830 vol.3, 1990.

[12] Bischof, H., Leonardis, A., *Finding optimal neural networks for land use classification, Geoscience and Remote Sensing*, IEEE Transactions on , Volume: 36 Issue: 1,337 –341, Jan. 1998.

[13] Hertz, J., Krogh, A., and Palmer, R. G., *Introduction to the Theory of Neural Computation*, Addison-Wesley, Redwood City, CA., 1991.

[14] Ruck, D.W., et al, *The Multilayer Preceptron as an Aproximation to Bayes Optimal Discriminant Function*, IEEE Trans. Neural. Network., 1(4), 296-298, 1990.

[15] Benediktsson, J.S., Swain, P.H. and Ersoy, O.K., *Neural Network Approaches Versus Statistical Methods in Classification Multisource Remote Sensing Data*, IEEE Trans. Geosc. & Rem Ens., 28(4), 540-551, 1990.

[16] Hermann, P.D. and Khazenie, N., *Classification of Multispectral Remote Sensing Data Using a Back-Propagation Neural Network*, IEEE Trans. Geosc. & Rem. Sens., 30 (1), 80-88, 1992.