Analysis of Bio-material 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 90058
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 look at all the cells in a sample to determine the degree of malignancy. The fatigue induced by this process result in higher rate of errors and takes a long time. An intelligent data understanding system would quickly identify those areas requiring attention by the skilled pathologists and would greatly increase the accuracy of diagnosis.

We have developed a diagnostic tool (Bio-Analysis system) that uses a novel framework of morphological and molecular assessment of cell tissues biopsies based on multi-spectral signatures. The Bio-Analysis system creates multispectral imagery of cell tissues 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 benign and prostate cancerous cells were created and analyzed. Conclusive results were obtained with high-speed turn-around and 98% percentage of correct classification. System knowledge was gained by biologically inspired neural network learning algorithm generalization of spectral and spatial information. This technology can provide more details, and a human like learning process to help the physicians to inexpensively and accurately diagnose normal vs. abnormal cell tissues. This methodology is applicable to planet surface information gathering and to molecular assessment of bacteria.

1.0 Introduction

Pathologists view biopsies under microscopes to determine degree of malignancy. This process of identifying the cells as normal or tumor is time consuming, exhausting, and fatiguing. The induced exhaustion created by this process causes a large number of human errors. This research describes a Bio-Analysis system that classifies and identifies pathological cells. It provides a novel framework of morphological and molecular assessment of bio-materials (i.e., cell tissues).

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 final sections, we show simulation of results and conclude by analyzing the results.

1.1 Multi-spectral signatures of cells

Each bio-material has unique spectral distribution that can be viewed as its "fingerprint", hence its importance in space exploration or health care. Current bio-analysis system sensor has 22 to 24 channels of data. Future bio-sensors will produce even larger numbers of data channels. Currently, we create our own BioCubes data sets from a tissue micro-array biopsy cells slide obtained from UCLA. Each BioCube encompasses 22 images taken at different wavelength (superimposed sensory channels). The wavelengths range from 470 nano-meters to 710 nano-meters of the electromagnetic spectrum. Zeiss Microscope's epi-fluorescence light energy is continually transmitted through a liquid crystal tunable filter, and H & E biopsy tissue cell. 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 City of Hope and UCLA hospitals; each slide contains benign or malignant prostate cells. The technique of H&E and their transmittance properties are beyond the scope of

this study. See any medical standard clinical practice 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 channel). 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 ranges 450 nm to 710 nm every 5nm band.

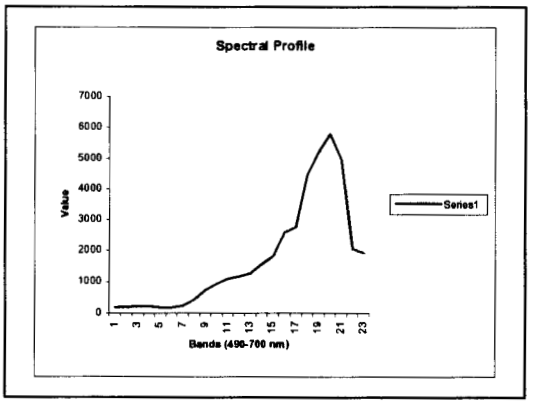

Figure 1 Spectral Profile

1.2 Medical utilization

Multispectral sensors for each 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 deal with these data sets.

Trained pathologists must view all the biopsies to determine the degree (different grades) of malignancy. This process is time consuming and the induced fatigue created by these process results in many human errors. Bio-Analysis system will help pathologists identify anomalous cells in a cost efficient manner.

Figure 2 depicts a typical micro-array biopsy tissue cell image from a slide. They are color images of each tissue spot on the slide. The template with the slide that the image titles corresponds to the tissue location. Dr. Seligson, an 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.)

Diagnostic or detecting places where there are basal cells is problematic. For example, prostate glands are double-layered, and one assumes in general that benign glands have two layers, tumor glands have one layer – though that is not 100% true, it is a good generalization.

We have assumed that the benign glands have a basal cell, since detecting these layers is important to diagnostic imaging separation of 'benign' and 'tumor' cells. A verification of this is in where the physician sees basal cells within the tumor area. The point is a 'tumor' or a pre-tumor cell encased in a basal cell layer is basically PIN. And, a few areas have what is somewhat between PIN and totally invasive tumor – in these areas the pathologists call it tumor and note that there are still a few basal cells. Where one ends and another begins is controversial.

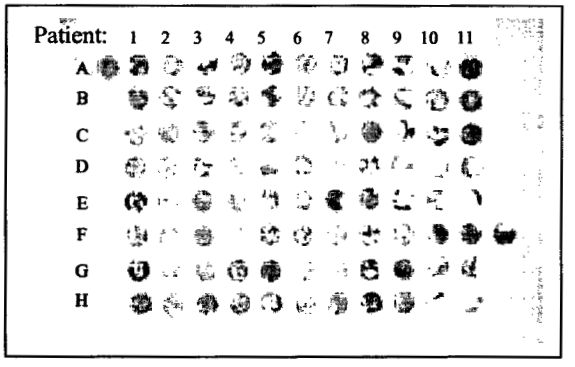

Figure 2 Microarray Biopsy Tissue Cells of Prostate

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 will look.

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.

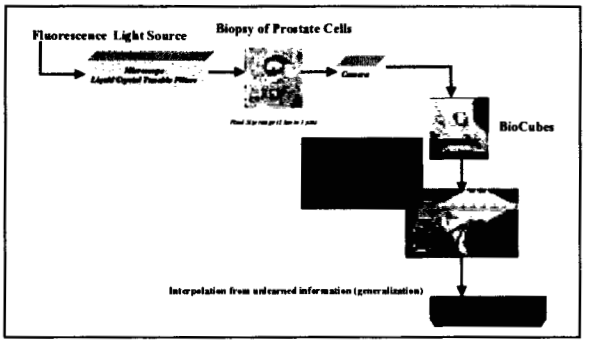

Figure 3 Bio-Analysis System

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.

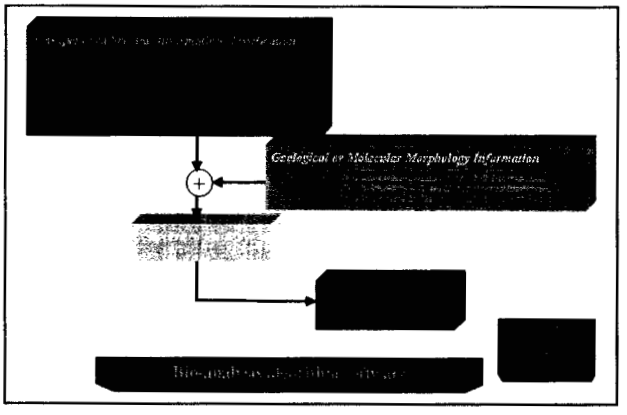

Figure 4 Bio-Analysis Algorithms

2.1 Comparison to Technology used by Physicians

Multispectral image data of tissues provides the information of the absorption and scattering of the cell tissues. The **optical diagnosis** can provide a safe and lower cost health care. In comparison to state-of-art the Ultrasound and Magnetic Resonance Image (MRI) are two widely used instruments by physicians as diagnostic tools. They provide gross tissue structure and some blood flow 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 cell tissues. There is also, no on going 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 normal vs. abnormal cell tissues.

3.0 Learning multi-spectral data

Processing multi-spectral image cubes, in real time, are 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 set. 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 suboptimal 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 is 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 aposteriori 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 percepton 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 biopsy tissues into four molecular classifications: cell walls, nuclei, cytoplasm stoma, 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 test set.

Generally, it takes more than ten years of training to become an expert in this field, and still experts have difficulty in agreeing on a diagnosis. Pathologists or submarine sonar experts are trained for 10 years and more in their respective fields. Figure 5 represents a typical pictorial classification and identification.

Analyzed results were compared with pre-diagnosed biopsy spots / biopsy images. The major features of malignity are 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 from other undesired elements in the cytoplasm. This study is an on going research: 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.

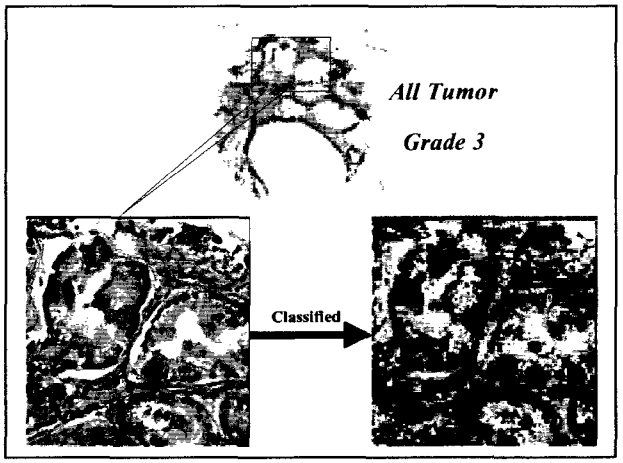

Figure 5 BioCube Classification

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. Imaging spectrometer provides reflectance and transmission spectral signature of molecules within the cell and cell morphology. We have correctly classified and identified bio-materials with high-speed turn around.

As part of 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 cell tissues.

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 sponsored 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.