Proceedings of the 25th Annual International Conference of the IEEE EMBS
Cancun, Mexico • September 17-21, 2003

Ultrasound Backscatter Signal Characterization and Classification Using Autoregressive Modeling and Machine Learning Algorithms

Noushin R.Farnoud[1], Michael Kolios[1,2]
Co-author: Srindhar Krishnan[1]
[1]Department of Electrical Engineering, Ryerson University, Toronto, Canada
[2]Department of Math-Physics and Computer Science, Ryerson University, Toronto, Canada

Abstract— This research explores the possibility of monitoring apoptosis and classifying clusters of apoptotic cells based on the changes in ultrasound backscatter signals from the tissues. The backscatter from normal and apoptotic cells, using a high frequency ultrasound instrument are modeled through an Autoregressive (AR) modeling technique. The proper model order is calculated by tracking the error criteria in the reconstruction of the original signal. The AR model coefficients, which are assumed to contain the main statistical features of the signal, are passed as the input to Linear and Nonlinear machine classifiers (Fisher Linear Discriminant, Conditional Gaussian Classifier, Naive Bayes Classifier and Neural Networks with nonlinear activation functions). In addition, an adaptive signal segmentation method (Least Squares Lattice Filter) is used to differentiate the data from layers of different cell types into stationary parts ready for modeling and classification.

Keywords—**Apoptosis, Ultrasound Backscatter**

I. INTRODUCTION

High frequency ultrasound (US) has been shown to detect the structural changes cells and tissues undergo during cell death. Research has shown that the ultrasound backscatter signals from apoptotic[1] acute myeloid leukemia(AML) cells differ in intensity and frequency spectrum as the result of the change in size, spatial distribution and acoustic impedance of the scattering sources within the cell [1] (Fig. 1). Therefore, we assume that pulse echo data from different cell types contain distinguishable statistical regularities. In this work we attempt to classify normal and apoptotic cancerous cells by tracking the statistics of the ultrasound backscatter signals from tissues by using Autoregressive (AR) method for time series modeling of ultrasound signals.

II. METHODOLOGY

A. Autoregressive (AR) Modeling of US signals

Biomedical signals contain large quantities of data. Moreover these data usually contain some redundancies which make processing and analyzing them more difficult. In such situations signal modeling may help to take out the irrelevant information carried by the signal and simplifies classification and segmentation by using a reduced number of model parameters. Autoregressive (AR) modeling is widely used for speech and biomedical signal processing [2-4]. This model is linear and has been successfully used for high-resolution spectral estimation [5]. An AR model is defined by the difference equation:

$$x(n) = -\sum_{k=1}^{p} a_k x(n-k) + e(n) \quad (1)$$

where $x(n)$ is a wide-sense stationary[3] AR process, $\{a(k)\}$ represent AR coefficients, $e(n)$ is white Gaussian noise and p is the model order which determines the error criterion. In section C, we will present a way to estimate this error and reduce it based on choosing the proper model order (p).

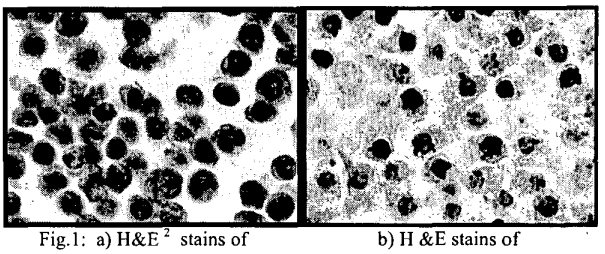

Fig.1: a) H&E [2] stains of Normal Cells b) H &E stains of Apoptotic Cells

B. Data Acquisition

AML cells were grown in suspension and exposed to the chemotherapeutic cisplatin to induce apoptosis. Pellets were made by swing bucket centrifugation. Details on the biological procedure can be found elsewhere (Czernote et al. 1996)[6]. A 20MHz f2.35 or 40 MHz f2 transducer (Visual Sonics[4]) was used to image the pellets of normal and apoptotic cells. RF backscatter data was digitized at 500MHz and stored for later analysis. In one experiment, layers of normal and apoptotic cells were created to emulate a clinical situation.

C. Choosing the proper Model Order

The modeling order (p) controls the error associated with the AR signal approximation. This parameter

[1] Apoptosis is a genetically determined destruction of cells from within due to activation of a stimulus or removal of a suppressing agent or stimuli.

[2] Hematoxylin and Eosin.

[3] A stochastic process is called wide-sense stationary (WSS) if its mean is constant and its autocorrelation depends only on the time difference.

[4] www.visualsonics.com

determines the number of previous samples used to model the original signal. A small model order ignores the main statistical properties of the original signal while a big model order will result in modeling the noise associated with data and over-fitting[5] occurs. A very common method for estimating the proper model order is Akaike Information Criterion (AIC) [7], although applying this method would be very difficult in our work due to nature of US signals. Instead, we used the following parameters based on the statistics of the reconstructed signal and its frequency with different model orders to determine the best modeling order.

a) Ensemble Reconstruction Error

The error(2) shows the total difference of original and reconstructed signals in frequency domain using AR modeling technique:

$$\hat{x}(n) = -\sum_{k=1}^{p} a_k x(n-k) \quad (1)$$

$$E = \sum_{n=1}^{N} \left| f_x(n) - \hat{f}_x(n) \right| \quad (2)$$

where $\hat{x}(n)$ is the approximated signal based on AR modeling with order p, N is the total number of samples within an individual RF line, f and $\hat{f}$ represents the fft of original and estimated signals respectively.

b) Model Noise (error) Variance

The AR process is the output of an all-pole filter invoked by a white noise *e(n)*. This noise, which is also our modeling error, can be viewed as the output of the prediction error filter *A(z)*, as shown in Fig. 2, where *x(n)* is the original signal and *A(z)* is the transfer function of AR modeling.

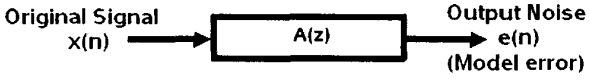

Fig. 2. Block diagram of AR process.

Therefore we expect that after estimating the AR coefficients of our model, if we invoke a filter as shown in fig. 2 with the estimated AR coefficients in A(z) the filter output, *e(n)*, would be a white Gaussian noise. We can verify this by estimating the variance of the output of such a filter and its auto-correlation (which has a jump to one in zero lag and remains zero otherwise).

D. Signal Segmentation

The classification methods we discussed were based on US backscatter from pure apoptotic and normal cell pellets. In patient imaging the data are acquired from tissues which contain different layers or layers with different mixtures on normal and apoptotic cells. The probabilistic behavior of the backscattered US signal from these cells, make the signal non-stationary[6]. This non-stationarity is important from the point of view of AR modeling, as this method is applicable if the signal is stationary[7]. Therefore we must use signal segmentation algorithms to break the signal acquired from tissues into stationary segments and classify each segment respectively. The segmentation algorithms can be classified into fixed [8][8] and adaptive [2,9-11]. Adaptive segmentation algorithms rely on tracking the statistical changes in the signal (such as mean and variance) to set a breaking boundary. We used this method for US signals due to its accuracy, modularity and ease of testing [2].

E. Adaptive signal Segmentation: Recursive-Least Squares Lattice Filter (RLSL)

In adaptive segmentation, the segment length changes dynamically according to the statistical changes in the signal. The main idea of using RLSL filter was to get to a fast convergence by using forward and backward filters. The parameter which expresses the statistical change in the signal is called convergence factor ($\gamma_m(n)$). The convergence factor provides the connecting link between different sets of a priori and posteriori estimation errors in this algorithm and is defined by

$$\gamma_m(n) = \gamma_{m-1}(n) - \frac{b_{m-1}^2(n)}{B_{m-1}(n)} . \quad (4)$$

where m is the order of the lattice filter, $\gamma_m(n)$ is the convergence factor at time sample n in the mth stage of lattice, $b_{m-1}(n)$ and $B_{m-1}(n)$ are the backward prediction error and its power at this stage [2].

IV. RESULTS

a) *Model Order Determination for Autoregressive (AR) Modeling of US signals*

Using the error criteria explained in section C, we calculated the error associated with the frequency of reconstructed and original US signals averaged over 30 normal and apoptotic sample RF lines respectively (Fig. 3). Matlab (version 6.5) was used for all the calculations. Also, as explained in section D, we found the variance of the

[5] When the model do well on training data but poorly on test data.

[6] The statistics of a non-stationary process are variant with respect to any translation among the time axis.

[7] We have determined that US signals from normal and apoptotic cells are quasi-stationary.

[8] Fixed segmentation algorithms are widely used for speech signal processing.

estimated noise generated as the output of a filter with the estimated AR coefficients in its transfer function and the original signal as its input. The result of averaging the variance of this noise over 30 samples is shown in fig. 4. These graphs indicate that model order 15 (p=15) is a good choice for AR modeling order for high frequency US backscatter signals, as we do not see much improvement in ensemble error(the ratio of error between model order 15 and 40 is 2.6 in comparison to 2.9e5 between model order 1 and 15). Furthermore, the variance of the estimated model noise does not change dramatically after this model order. To verify this result, we modeled an US backscatter signal with order 15, reconstructed this signal with the estimated AR coefficients and found the auto-correlation of the model error[9] (noise) .As depicted in Fig. 5; this auto- correlation indicates the similarity of the estimated error to white noise. Therefore we used AR modeling with order 15 for US backscatter signals in the rest of this paper.

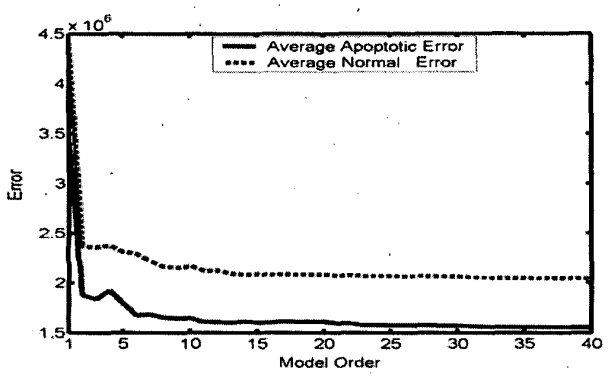

Fig. 3: Average Ensemble Error between the ffts of estimated and original US signal (30 samples of normal and apoptotic signals).

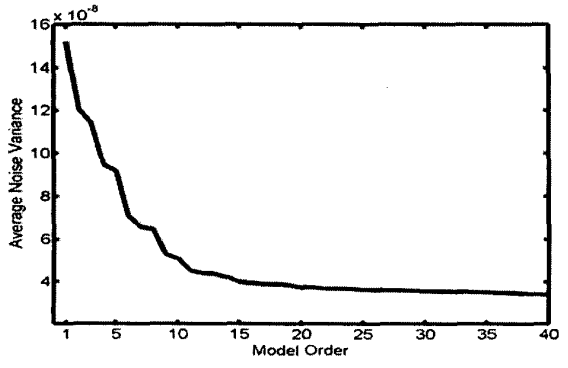

Fig. 4: Average variance of the estimated model noised based on the estimated AR coefficient (30 samples).

[9] This error was assumed to be the absolute difference between original and reconstructed signals.

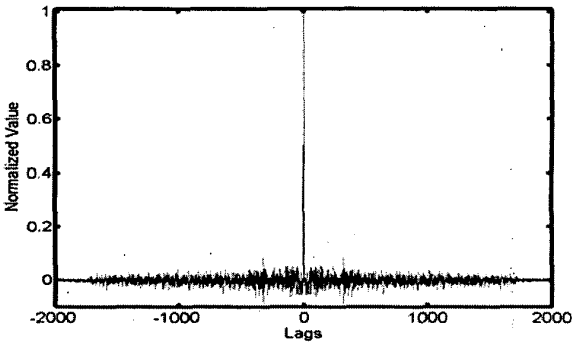

Fig. 5: Auto-correlation of the estimated model error (noise).

b) Ultrasound Signal Classification

After each signal is compressed into 15 autoregressive coefficients, these coefficients are passed as the input of different machine learning classifiers both at training and testing phase [12-17]. Each classifier was trained using AR coefficients of 100 samples of normal and 100 samples of apoptotic US signals and tested with a different set of 100 normal and 100 apoptotic samples The classification results are shown in table 1.

TABLE I
Classification Algorithms Accuracy

Algorithm	Normal Accuracy	Apoptotic Accuracy
Conditional Gaussian Classifier[10]	40%	60%
Naive Bayes Classifier	46%	77%
Fisher's Linear Discriminant	98%	64%
Neural Network with Sigmoid activation function[11]	93.8%	99%
Neural network with tanh activation function	95.5%	99%

This result shows the ability of Neural Networks with non-linear activation functions (in both hidden and output layers) to classify US signals from normal and apoptotic cells. We are still investigating the advantages and disadvantages of each approach.

c) Ultrasound Signal Segmentation

Fig. 5 shows RLSL algorithm applied on a layer on Normal-Apoptotic-Normal cell pellet with the apoptotic layer located between samples 800 and 15000. As long as the input data is stationary, the convergence factor would remain in the same range, but when it drops below a

[10] The priors for each class were equally set (p=0.5).

[11] The network was trained using 50000 iterations.

threshold [12] it indicates a sudden change in statistical properties of the signal which is set to the segment boundary.

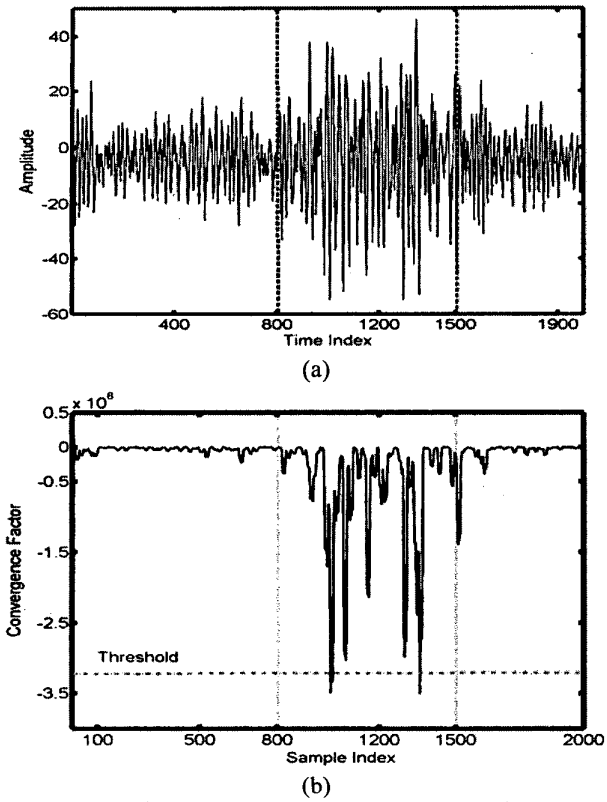

Fig. 5. (a): Original signal from a 3 layer Normal-Apoptotic-Normal cell pellet. (b): Convergence factor as a parameter to detect the layer boundaries (stationary).

These figures indicate that RLSL algorithm can detect the sudden changes in the signal due to the different statistical properties of normal and apoptotic layers and therefore can adaptively found their corresponding boundary in an US backscatter signal. While in Fig. 5.a the difference is evident, in clinical situations it is anticipated that small percentage of apoptotic cells would be surrounded by normal cells.

V. Conclusion

The best model order in using AR technique for US signals was found to be p=15. The accuracy of different classifiers has been studied and it was found that non-linear neural networks were most successful in classification. Because the actual clinical data from patients include US backscatter from layers and mixtures of cells, a method for differentiating these layers was presented which enables the AR modeling to be applicable for US signals.

[12] The threshold in this work is set by visual inspection (however in the future it will be extracted from the signal based on its statistical properties).

Acknowledgment

We should thank Dr. Michael Sherar and Ontario Cancer Institute of the Princess Margaret Hospital for their support, Anoja Giles for helping us with the biological work and Dr. Gregory Czarnota for his scientific input. Noushin R.Farnoud would also like to thank Dr. Sam Roweis at the Computer Science Department of the University of Toronto for his help with the Machine Learning algorithms.

References

[1] MC. Kolios, GJ. Czarnota, M. Lee, JW. Hunt, MD. Sherar, *Ultrasonic spectral parameter characterization of apoptosis*, Ultrasound Med. & Biol. 2002 May, 8(5):589-97.

[2] S. Krishnan, Adaptive *filtering, Modeling, and Classification of Knee Joint Vibroarthrographic Signals*, Master's Thesis, University of Calgary, Alberta, Canada, 1996.

[3] F. Towfiq, C.W. Barenes, E.J. Pisa, *Tissue classification based on autoregressive models for ultrasound pulse echo data*, ACTA Electronica,1984, (.26): 95-110.

[4] A. Nair, BD. Kuban, N. Obuchowski, DG. Vince, *Assessing spectral algorithms to predict atherosclerotic plaque composition with normalized and raw intravascular ultrasound*, Ultrasound in Med. & Biol., 27(10):1319-1331, 2001.

[5] M. Akay, *Time Frequency and Wavelets in Biomedical Signal Processing* (Book style). Piscataway, NJ: IEEE Press, 1998: 123–135.

[6] GJ. Czarnota, MC. Kolios, J. Abraham, M. Portnoy, FP. Ottensmeyer, JW. Hunt, MD. Sherar, *Ultrasound imaging of apoptosis: high-resolution non-invasive monitoring of programmed cell death in vitro*, in situ and in vivo, Br J Cancer. 1999 Oct; 81(3):520-7.

[7] Y. Sakamoto, M. Ishiguro, G. Kitagawa, *Akaike Information Criterion Statistics*, D. Reidel Publishing Company, KTK Scientific Publishers, Tokyo Hardbound, ISBN 90-277-2253-6, November 1986.

[8] J.D. Markel, A.H. Gray, *Jr. Linear Prediction of Speech*. Springer-Verlag, N.Y., New York, 1976.

[9] D. Michael, J. Houchin, *Automatic EEG analysis: A segmentation procedure based on the autocorrelation function*, Electroenceph., Clinical Neurophysiology, (46):.232-235, 1979.

[10] G. Bodenstein, H.M. Praetorious, *Feature extraction from the electroencephalogram by adaptive segmentation*, Proceeding of IEEE, 65(5): 642-652, May 1997.

[11] H.M. Praetorious, G. Bodenstein, O.D. Creutzfeldt, *Adaptive segmentation of EEG records: A new approach to automatic EEG analysis*, Electroencephalogram, Clinical Neurophysiology, Vol.42, pp.84-91, 1977.

[12] T. Mitchell, *Machine Learning*, McGraw Hill, 1997.

[13] C.D. Nugent, J.A. Lopez, A.E. Smith, *Prediction Models in Design of Neural Network based ECG classifiers*, BMC Medical Informatics and Decision Making, 2001.

[14] S. Chakrabarti, N. Bindal, *Robust Radar Target Classifier Using Artificial Neural Networks*, IEEE Transaction on Neural Networks, 6(3), May 1995.

[15] D. Docampo, *Intelligent Methods in Signal Processing and Artificial Communications*, Birkauser Boston, 1997.

[16] D.M.Skapura, *Building Neural Networks Algorithms, Applications, and Programming Techniques*, ACM press, 1998.

[17] J.A. Freeman, D.M. Skapura, *Building Neural Networks*, ACM press, 1998.