International Journal of Medical, Medicine and Health Sciences
ISSN: 2517-9969
Vol:1, No:10, 2007

Eukaryotic Gene Prediction by an Investigation of Nonlinear Dynamical Modeling Techniques on EIIP Coded Sequences

Mai S. Mabrouk, Nahed H. Solouma, Abou-Bakr M. Youssef, and Yasser M. Kadah

***Abstract*—**Many digital signal processing, techniques have been used to automatically distinguish protein coding regions (exons) from non-coding regions (introns) in DNA sequences. In this work, we have characterized these sequences according to their nonlinear dynamical features such as moment invariants, correlation dimension, and largest Lyapunov exponent estimates. We have applied our model to a number of real sequences encoded into a time series using EIIP sequence indicators. In order to discriminate between coding and non coding DNA regions, the phase space trajectory was first reconstructed for coding and non-coding regions. Nonlinear dynamical features are extracted from those regions and used to investigate a difference between them. Our results indicate that the nonlinear dynamical characteristics have yielded significant differences between coding (CR) and non-coding regions (NCR) in DNA sequences. Finally, the classifier is tested on real genes where coding and non-coding regions are well known.

***Keywords*—**gene prediction**,** nonlinear dynamics**, c**orrelation dimension, Lyapunov exponent.

I. INTRODUCTION

GENE prediction typically refers to the area of computational biology that is concerned with algorithmically identifying stretches of sequence, usually genomic DNA, that are biologically functional. This especially includes protein-coding genes, but may also include other functional elements such as RNA genes and regulatory regions. Gene finding is one of the first and most important steps in understanding the genome of a species once it has been sequenced. The algorithms for identification of genes make use of one or more of the several available coding measures. These coding measures incorporate a unique feature or character of the coding sequence, based on which accurate identification of the sequence can be done.

In eukaryotic DNA, genes generally consist of coding regions (exons) and non coding regions (introns). Proteins are translated from a copy of the gene where introns have been removed and exons are joined together, a process called splicing, It is therefore of importance to identify reliably the start of a gene, its exons and introns (if present) as well as the end of the gene, whereas in prokaryotes these regions are continuous.

A single strand of DNA is a bio-molecule consisting of many linked, smaller components called nucleotides. Each nucleotide is one of four possible types designated by the letters A, T, C, and G and has two distinct ends, the 5' end and the 3' end, so that the 5' end of a nucleotide is linked to the 3' end of another nucleotide by a strong chemical bond, thus forming a long, one-dimensional chain (backbone) of a specific directionality. The position of nucleotides – adenine, guanine, cytosine and thymine – in a DNA sequence depends on the distribution of nucleotides on the entire chromosome. Base pairs in DNA do not occur in a completely random fashion, especially outside the coding regions, in the so called "junk DNA" area that does not encode directly information about protein synthesis. Classical techniques have been used to address the problem of identifying coding regions (exons) from non-coding regions (introns) such as frequency domain analysis, time domain analysis, and Hidden markov model [1,2]. Other techniques used filtering Methods [3]. Even though a good results have been obtained using such techniques, but they seem to provide only a limited amount of information about the DNA sequence because they ignore the underlying nonlinear sequence dynamics.

In the last two decades, there has been an increasing interest in applying techniques from the domains of nonlinear analysis and chaos theory in different fields of research. In this paper, we have applied the nonlinear analysis and chaos theory to genomic systems in order to differentiate between coding and non coding regions in DNA sequences. In the field of chaotic dynamical system theory, several features can be used to describe system dynamics including moment invariants, correlation dimension (D_2) and Lyapunov exponents (LLE). In this work, these features have been used to explain different genomic sequences encoded to its signal behavior by several studies. Also, we have addressed the problem of characterizing the nonlinear dynamics of real DNA sequences. The implementation details to automatically compute three

M. S.Mabrouk is with the Biomedical Department, Misr University for Science and Technology (MUST), Egypt, E-mail: msm_eng@ k-space.org.
N. H. Solouma is with the National Institute of Laser Enhanced Sciences, Cairo University, Egypt, E-mail: nsolouma@k-space.org.
A. M. Youssef is with the Biomedical Engineering Department, Cairo University, Egypt. E-mail: youssefa@internetegypt.com
Y. M. Kadah is with the Biomedical Engineering Department, Cairo University, Egypt. E-mail: ymk@k-space.org.

important chaotic system parameters namely, the moment invariants, correlation dimension and largest Lyapunov exponent, are discussed. Our method is based on training and testing a K-means clustering classifier. We simulate the EIIP encoded coding and non-coding regions for training, apply the chaos theory to sequences of those regions, extract their feature vectors and train classifiers to discriminate coding from non coding regions. The classifier is tested on a real DNA sequence where the coding and non coding regions are identified.

II. Numerical Sequence Representation

In order to apply digital signal processing (DSP) techniques, the character sequence of DNA are first converted into numeric sequence. In this work we have used the EIIP sequence indicators; the energy of delocalized electrons in amino acids and nucleotides has been calculated as the Electron-ion interaction pseudopotential (EIIP). The EIIP values of amino acids have already been used in Resonant Recognition Models (RRM) to substitute for the corresponding amino acids in protein sequences, whose power spectrum is taken to extract the information contents [4]. The EIIP values for the DNA nucleotides are given below in Table I. For example, if $x[n]$ = [A A T G C A T C A], then using the values from Table I, $X_e[n]$ = [0.1260 0.1260 0.1335 0.0806 0.1340 0.1260 0.1335 0.1340 0.1260].

TABLE I
Electron Ion Interaction Pseudo Potentials of Nucleotides

EIIP	Nucleotide
0.1260	A
0.1260	G
0.0806	C
0.1335	T

III. Phase Space Trajectory Reconstruction.

The mathematical description of a dynamical system consists of two parts: the state which is a snapshot of the process at a given instant in time and the dynamics which is the set of rules by which the states evolve over time. To study the dynamics of our system, we first need to reconstruct the state space trajectory. Phase space reconstruction is the fundamental for analyzing nonlinear signals, by which a time series can be embedded to n-dimensional space. In this section we briefly demonstrate basic steps of the reconstruction of the phase space. We start first by encoding the different segments of coding (CR) and non-coding regions (NCR) into a time series signal using EIIP sequence indicators. A good choice for a delay time is yielded by using the first minimum of the auto mutual information function. The first minimum of the auto mutual information can be found at four. Now we need to know the minimal embedding dimension for coding and non-coding time series signals. We use Cao's method with a delay time of four, a maximal dimension of eight, three nearest neighbors and reference point depending on the length of each signal. There is a kink produced by Cao's method at 3. So we need a time delay reconstruction of coding and non-coding time series signals with embedding dimension of 3 and delay of 4. Finally, we obtained the phase space trajectory for coding and non-coding time series signals. The step following obtaining the phase trajectory of coding and non-coding time series signals is the step of feature extraction.

IV. Feature Extractions

Now all the coding and non coding region time series were transformed into phase space, and ready to extract their nonlinear dynamical features. This can be done by applying the following three methods, moment invariants, correlation dimension and largest lyapunov exponent. In all methods, the Matlab-based software package for nonlinear time series analysis TSTOOL was used to compute the features based on the phase space of the numerical sequence representation [5].

A. Moment Invariants

In this section, the feature set based on central moments of the phase space are considered using moment invariants [6]. The basic idea is to construct the phase space usually done for nonlinear dynamical feature extraction [7] and then try to utilize invariant moments to describe the resultant trajectory. Features obtained by moment invariants are simple calculated features that do not change under translation, scaling or rotation [8]. The n-dimensional moments of order p of a function of intensity $\rho(x_1, \ldots, x_n) = \rho(x)$ are defined in terms of Rieman integral as,

$$m_{p_1 \cdots p_n} = \int \cdots \int x_1^{p_1} \cdots x_n^{p_n} \rho(x) dx_1 \cdots dx_n \quad . \tag{1}$$

Here $p_i + \ldots p_n = p$, $0 < p < \infty$. It is assumed that $\rho(x)$ is piecewise continuous and therefore bounded function, and it can have nonzero values only in a finite part of the R^n; then the moments of all orders exist. The central moments are defined as,

$$\mu_{P_1 \ldots P_n} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (x_1 - \bar{x}_1)^{P_1} \ldots (x_n - \bar{x}_n)^{P_n} \rho(x) dx_1 \ldots dx_n \,. \tag{2}$$

Here,

$$\bar{x}_1 = \frac{m_{1\ldots0}}{m_{0\ldots0}}, \ldots\ldots, \bar{x}_n = \frac{m_{0\ldots1}}{m_{0\ldots0}} \quad . \tag{3}$$

The moment invariant features are given as [5],

$$\phi_1 = \frac{1}{\mu^4} \begin{vmatrix} \mu_{2\ldots0} & \cdots & \mu_{1\ldots1} \\ \cdots & \cdots & \cdots \\ \mu_{1\ldots1} & \cdots & \mu_{0\ldots2} \end{vmatrix} \quad , \tag{4}$$

$$\phi_2 = \frac{1}{\mu^4} (\mu_{20}\mu_{02} - \mu^2{}_{11}) \quad , \tag{5}$$

$$\phi_3 = \frac{1}{\mu^{10}} ((\mu_{30}\mu_{03} - \mu_{21}\mu_{12})^2 - 4(\mu_{30}\mu_{12} - \mu_{21}{}^2)(\mu_{21}\mu_{03} - \mu^2{}_{12})) \quad , \tag{6}$$

$$\phi_4 = \frac{1}{\mu^6}(\mu_{40}\mu_{04} - 4\mu_{31}\mu_{13} - 3\mu^2{}_{22}) \ , \quad (7)$$

$$\phi_5 = \frac{1}{\mu^9}(\mu_{40}\mu_{22}\mu_{04} + 2\mu_{31}\mu_{22}\mu_{13} - \mu_{40}\mu^2{}_{13} - \mu^2{}_{31}\mu_{04} - \mu^3{}_{22}) \ , \quad (8)$$

$$\phi_7 = \frac{1}{\mu^5}(\mu_{200}\mu_{020}\mu_{002} + 2\mu_{110}\mu_{101}\mu_{011} - \mu_{200}\mu^2{}_{011} - \mu^2{}_{110}\mu_{002} - \mu^2{}_{101}\mu_{020}) \ , \quad (9)$$

and,

$$\phi_8 = (\mu^2{}_{20}\mu_{04}) - 4\mu_{20}\mu_{11}\mu_{13} + 2\mu_{20}\mu_{02}\mu_{22} + 4\mu^2{}_{11}\mu_{22} - 4\mu_{11}\mu_{02}\mu_{31} + \mu^2{}_{02}\mu_{40} \ . \quad (10)$$

B. Correlation Dimension Estimation (D_2)

The correlation dimension provides a straight forward way to measure the spatial organization and hence the predictability (finite dimensionality) and dimensionality of a given signal (time series). That is, the measure of correlation dimension provides a way to determine whether the signal (time series) has fractional dimension i.e. chaotic attractors. However, in order to estimate the correlation dimension it is required to reconstruct the state space trajectory of the time series. This can be accomplished using the delay time embedding theorem through the creation of larger dimensional geometric object by embedding into a larger m-dimensional embedding space. The embedding dimension m should be large enough for delay time embedding to work. When a suitable value for m is used, the orbits of the system do not cross each other [7].

However, we have selected the first minimum of the mutual information function as a suitable value for the embedding time lag. The embedding dimension has been estimated using the Cao's method [9]. The Grassberger–Procaccia algorithm [10] uses a correlation integral C(r) to represent the object, which is defined as the average number of neighbors each point in the reconstructed phase space has within a given distance r, given as:

$$C(r) = \frac{1}{N_P}\sum_{i,j}\theta[r - \|x(i) - x(j)\|] \ . \quad (11)$$

Here, ||.|| symbolized the Euclidean distance between reconstructed state vectors *x(i)* and *x(j)*, $N_p=k(k-1)/2$ is the number of distinct pairs of reconstructed state vectors, θ is the Heaviside unit step function (i.e., $\theta(x)=0$ when $x<0$ and $\theta(x)=1$ when $x \geq 0$). The correlation dimension D_2 is defined as the slope of the linear region of the plot of log ($C(r)$) versus log(r) for small values of r [7]. That is,

$$D_2 = \lim_{r\to 0}\frac{\log C(r)}{\log r} \quad (12)$$

Nevertheless, we have computed the correlation dimension using Taken's estimator provided with the TSTOOL software package [5].

C. Lyapunov exponent (LLE)

The notion of Lyapunov exponents is a generalization of the idea of the eigenvalues as a measure of stability of a fixed point (characteristic exponent) as it provides a measure of stability of a periodic orbit. That is, Lyapunov spectrum (exponents) characterizes the behavior (contraction or expansion) of the trajectories close to a fixed point. Therefore, these exponents provide a mean to measure the sensitivity to perturbed initial conditions. For a system to undergo chaotic dynamics, it must have at least one positive Lyapunov exponent. The largest Lyapunov exponent (λ1), nevertheless, may be regarded as an estimator to the dominant chaotic behavior of the system [7]. However, in this work we have used the TSTOOL largest Lyapunov estimation algorithm (larglyap). This algorithm is similar to Wolf's algorithm and provides an efficient estimation of the largest Lyapunov exponent through the calculation of the scaling (rate of increase) of the prediction error (separation of nearby trajectories) versus the prediction time. The algorithm used is as follows,

1. Compute the distance d_0 of two, very close points in the reconstructed phase space orbit.
2. Follow both points as they travel a short distance along the orbit. The distance d_1 between them is calculated.
3. If d_1 become too large, one of the points is kept and an appropriate replacement for the other point is chosen.
4. The two points are now allowed to evolve again following steps 1-3.
5. After S propagation steps, the largest Lyapunov exponent λ_1 is estimated as:

$$\lambda_1 = \frac{1}{t_S - t_0}\sum_{k=1}^{S}\log_2\left(\frac{d_1(t_k)}{d_0(t_{k-1})}\right) \quad (13)$$

V. Experimental Results

A. DNA Sequence Database

We have applied our work to a set of DNA sequences coded in EIIP indicator sequences. Coding and non coding regions of these sequences are extracted. The length of these regions has been chosen to be 1000 to 2000 bp. Then, In order to discriminate between coding and non coding DNA regions of the given genes, the phase space trajectory was first reconstructed for each time series of both of them. Then feature extraction techniques as moment invariants, correlation dimension and largest Lyapunov exponent were applied to coding and non coding regions. Finally, we have compared the nonlinear features extracted from coding regions to those extracted from non-coding regions to investigate any significant difference between them. In this paper, we have checked our work on several coding and non coding segments of eukaryotic genes in a number of organisms using EIIP indicators on two data sets. One is the dataset prepared by Burset and Guigó [11] and the other is HMR195 [12] prepared by Sanja Rogic.

B. Results of Moment Invariants

The coding and non-coding encoded segments were transformed into phase space, and some features based on moment invariants were computed using this phase space. A significance test (t-test) was performed on the proposed features to assess the use of such parameters for discriminating between the different coding and non-coding regions. Considering the p-value, if $p< 0.05$ there is a significant difference, if $p> 0.05$ there is no significant difference. The seven features are arranged as (ϕ_1, ϕ_2, ϕ_3, ϕ_4, ϕ_5, ϕ_7, and ϕ_8), result of comparing the average features extracted based on moment invariants for coding regions (CR) and non-coding regions (NCR) is shown in Fig. 1. Also, result of the t-test for all seven features is shown in Table II.

C. Results of Correlation Dimension (D2)

We have applied the Taken's estimator to the EIIP encoded sequences in order to obtain their correlation dimension estimates. Table III shows a sample of the fractional (chaotic) correlation dimension estimates of a set of coding regions (CR) and non-coding regions (NCR) of different genes. We have calculated the average correlation dimension (D_2) of all coding and non-coding regions; they are 0.4933 and 0.4667 respectively.

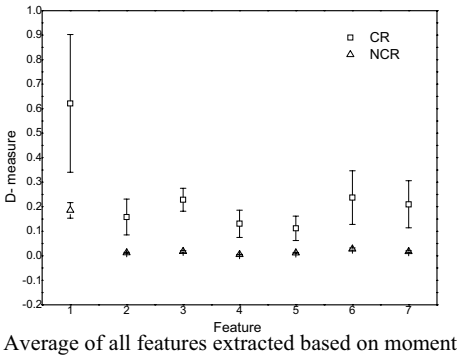

Fig. 1 Average of all features extracted based on moment invariants for coding regions (CR) and non-coding regions (NCR)

TABLE II
P-VALUE OF t-TEST ON A SET OF CR`S AND NCR`S FOR ALL FEATURES EXTRACTED BASED ON MOMENT INVARIANTS

Feature	p-Value
Feature 1 (ϕ_1)	0.0329
Feature 2 (ϕ_2)	0.0357
Feature 3 (ϕ_3)	0.0009
Feature 4(ϕ_4)	0.0063
Feature 5 (ϕ_5)	**0.0970**
Feature 6 (ϕ_7)	0.0357
Feature 7 (ϕ_8)	0.0216

We have also generated a random DNA sequence of length 1500 bp consisting of symbols from the alphabet {A,T,C,G}.The correlation dimension D_2 of this random sequence is 0.4564. Also, a significance test (t-test) was performed on the proposed D_2 feature to assess the use of such parameter for discriminating between the different coding and non-coding regions. The p-value of the t-test was calculated and it is 0.5147; which is too large. So, Considering the D_2 value of a generated random sequence (0.4564) and the t-test p-value (0.5147), we can conclude that for all CR and NCR encoded EIIP sequences, the Correlation dimensions D_2 are close to each other and close to the correlation dimension of the generated random sequences.

D. Results of Largest Lyapunov Exponent (LLE)

The largest Lyapunov exponent has been estimated from the scaling (linear increase) of the prediction error versus the prediction time using the TSTOOL package. Table IV, depicts the LLE estimates of a sample of a set of coding and non-coding regions of different genes. It is calculated by largelyap algorithm; an algorithm very similar to the Wolf algorithm, where it computes the average exponential growth of the distance of neighboring orbits via the prediction error. The increase of the prediction error vs. the prediction time allows an estimation of the largest Lyapunov exponent.

TABLE III
CORRELATION DIMENSION (D_2) ESTIMATES FOR A SET OF CR`S AND NCR`S

CR (D_2)	NCR (D_2)
0.4316	0.4419
0.4331	0.4490
0.4411	0.4560
0.4436	0.4580
0.4493	0.4624
0.4580	0.4643
0.4718	0.4653
0.4748	0.4672
0.4851	0.4898
0.8450	0.5134

A significance test (t-test) was performed on the proposed LLE feature to assess the use of such parameter for discriminating between the different coding and non-coding regions. The p-value of the t-test was calculated and it is 4.8858e-015. We have also generated a random DNA sequence of length 1500 bp consisting of symbols from the alphabet {A,T,C,G}.The largest Lyapunov exponent (LLE) of this random sequence is 1.4046; which is far from the average LLE of both CR (0.1055) and NCR (3.5518). So, Considering the LLE value of a generated random sequence (1.4046) and the t-test p-value (4.8858e-015), we can conclude that for all CR'S and NCR's encoded EIIP sequences the largest Lyapunov exponent (LLE) estimates are far from each other and from the generated random sequences.

VI. CLASSIFICATION

We have evaluated the accuracy of a test that predicts outcomes. They are used to describe a diagnostic test and how well a test discriminates between the two coding and non-

coding regions. The feature vectors were fed into classification process based on K-means clustering classifier. In this work, we have three feature vectors defined as, V1= $\{\phi_1, \phi_2, \phi_3, \phi_4, \phi_5, \phi_7, \phi_8\}$, V2= {LLE}, and V3= $\{\phi_1, \phi_2, \phi_3, \phi_4, \phi_7, \phi_8$, LLE$\}$. The results of this step are shown in Table V.

TABLE IV
LARGEST LYAPUNOV EXPONENT (LLE) ESTIMATES OF THE GIVEN CR`S AND NCR`S

CR (LLE)	NCR (LLE)
0.0712	3.1859
0.1018	3.7143
0.1100	3.4767
0.1122	4.0441
0.1180	3.2161
0.1212	2.8924
0.1188	3.0470
0.0954	4.2767
0.1018	3.8462
0.1042	3.8188

TABLE V
ACCURACY OF THE PROPOSED FEATURE EXTRACTION METHODS TO K-MEANS CLUSTERING CLASSIFIER

	V1	V2	V3
CR	80%	100	100
NC	100	%	%
R	%	100	100
		%	%

VII. CONCLUSIONS

In this work, we have characterized the genomic sequences based on their nonlinear dynamical behavior. That is, we have established a nonlinear dynamical model consists of moment invariant, correlation dimension (D_2), largest Lyapunov exponent (LLE) estimates of plain integer mapping EIIP encoded sequences. The pattern of these model parameters varied considerably between different coding and non-coding DNA regions. Furthermore, we have used the K-means clustering algorithm in order to classify these different regions into their respective genes. Experiments were performed on a dataset obtained from [11, 12] to evaluate the reliability of the proposed nonlinear dynamical model. The proposed model has yielded reasonable classification accuracy between the coding (CR) sequences and the non-coding (NCR) sequences. Also, we have found that due to the p-values of the t-tests performed on the fifth nonlinear feature (ϕ_5) extracted based on moment invariants and the eighth nonlinear feature based on correlation dimension (D_2), that there is no significant difference between coding and non-coding regions. On the other hand, the other nonlinear features performed well with high accuracy. We have also examined and compared statistical measures like average and standard deviation of coding and non-coding regions with the all nonlinear features. In conclusion, throughout this work we have found that the natural nonlinear dynamics that the genomic sequences undergo differ between coding and non coding DNA regions. Therefore, it is rather encouraging to predict genes and discriminate their coding from non coding regions according to the nonlinear dynamical characteristics of their respective translated sequences.

REFERENCES

[1] M. Akhtar, "Comparison of Gene and Exon Prediction Techniques for Detection of Short Coding Regions," *International Journal of Information Technology,* Vol. 11, No.8, 2005.
[2] A. Krogh, I. Saira Mian, and D. Haussler, "A hidden Markov Model that *Finds Genes in E. Coli DNA*,"*Nucleic Acids Rsearch*, Vol. 22 pp. 4768-4778, 1994.
[3] P. P. Vaidyanathan, B.-J. Yoon, "Digital filters for gene prediction applications," *IEEE Asilomar Conference on Signals, and Computers,* Monterey, U.S.A., Nov. 2002.
[4] A. S. Nair, S. P. Sreenadhan, "A Coding Measure Scheme Employing Electron-Ion Interaction Pseudopotential (EIIP),"*Bioinformation, vol. 1, no. 6, pp. 197- 202,* 2006.
[5] http://www.physik3.gwdg.de/tstool/.
[6] A. G. Mamistvalov, "n-Dimensional Moment Invariants and Conceptual Mathematical Theory of Recognition n-Dimensional Solids," *IEEE Trans. on Pattern Recogn. Mach. Intell.* , Vol. 20, no. 8, pp. 819-831, 1998.
[7] M. I. Owis, A. H. Abou-Zied, A. M. Youssef, and Y. M. Kadah, "Study of features based on nonlinear dynamical modeling in ECG arrhythmia detection and classification," *IEEE. Trans. Biomedical Engineering*, vol. 79, pp. 733-736, July 2002.
[8] R. C. Gonzalez, and R. E. Woods, *Digital Image Processing, 2nd ed.*, Pearson Education, New York, 2001.
[9] L. Cao, A. Mees, K. Judd, and G. Froyland, "Determining of the minimum embedding dimensions of input-output time series data," *Intl. Journal. Bifurcation and chaos*, vol. 8, pp. 1491-1504, 1997.
[10] W. S. Pritchard, D.W. Duke, "Measuring chaos in the brain: A tutorial review of EEG dimension estimation," *Brain Cogn.*, vol. 27, no. 3, pp. 353–397, 1995.
[11] M. Burset, R. Guigo, "Evaluation of Gene Structure Prediction Prgrams,"*Genomics,*http://genome.Imim.es/datasets/genomics96. 1996.
[12] S. Rogic, "Evaluation of Gene- Finding Programs," *University of British Columbia,* http:/ /www.cs.ubc.ca/~rogic/evaluation.

Mai Said Mabrouk received her B.Sc. degree in Systems and Biomedical Engineering Department, Cairo University, Giza, Egypt, in 2000. She completed her M.Sc. and Ph.D. degrees in Biomedical Engineering from the same school in 2004 and 2008 respectively. She is a lecturer with the Biomedical Engineering Department, Misr University for Science and Technology (MUST) since August 2008. She has several research articles in the area of image processing and bioinformatics. Her research interests include medical image processing, bioinformatics, and digital signal processing applied to genomic signal processing.

Nahed H. Solouma received her B.Sc. and M.Sc. degrees from the Biomedical Engineering Department at Cairo University in 1990 and 1996 respectively. She received another MSc in Laser applications in medicine and biology from the National Institute of Laser Enhanced Sciences, Cairo University in 1997. Her Ph.D. degree was from the Biomedical Engineering department, Cairo University in 2001. She worked as a Lecturer and then an Associate Professor in NILES, Cairo University since 2001 till 2007. She is currently in sabbatical leave in Imperial College as an academic visitor in the Biomedical Engineering Institute. Her research interest is in biological modeling and medical data processing.

Abou-Bakr M. Youssef was born in Cairo Egypt 1951. He graduated from Cairo University—Faculty of Engineering—Electrical Engineering & Communication Department with Distinction with Honor in 1974. Then, he joined the Faculty of Medicine and worked in parallel on engineering research and received M.S.E.E. degree in 1978, and the Ph.D. degree in biomedical engineering (MIT fellow) in 1982—Cairo University. While continuing his engineering career, he was also able to graduate from the Medical School,

Cairo University 1980, and earned the MBBH, Distinction with Honors, and the foeign equivalent, ECFMG, in 1981. In 1984, he received the M.S. degree in radiology from Cairo University, and the MD degree from Heildelberg University, Heildelberg, Germany, in 1989. He conducted medical research in the German cancer research center (DKFZ) from 1987–1993. He contributed to the establishment of the Biomedical Engineering Department at Cairo University in 1977, developed courses of biomedical equipment and clinical engineering as well as medical imaging, and joined in the establishment of the Technology Transfer Focus in Medical School-Cairo University, in cooperation with the National Science Foundation in 1978–1980. He gained clinical experience in his research and educational life span by working as the Head of the Ultrasound Department in major hospitals in Cairo as well as within private clinics. In 1990–1992, he participated in the Health Care Development Plan, in cooperation with the Egyptian Scientific Research Academy, Cairo. Many of his researches were published in international and local scientific magazines. Dr. Youssef was registered as a Diagnostic Medical Sonographer RDMS in 1995. As a member of the American Institute of Ultrasonography in Medicine (AIUM), in 1997 he was elected as a Senior Member. He established the Biomedical Division of International Electronics, as a practical and technical service provider for medical industry and health care community in 1991. He was appointed CEO for International Electronic Company Biomedical group, Cairo. He participated with the Information & Decision Support Center—Egyptian Council of Ministers, in the release of medical software and expert systems in 1994. He was awarded from the government twice for the scientific research—Cairo University 1998 and the national award for advanced technological sciences—2001.

Yasser M. Kadah received his B.Sc. and M.Sc. degrees from the Biomedical Engineering Department at Cairo University in 1989 and 1992 respectively. He received his Ph.D. in Biomedical Engineering from the University of Minnesota in 1997. He worked as a research assistant with the Department of Radiology, University of Minnesota Medical School in 1996-1997 and as a post-doctoral fellow at the Center of Magnetic Resonance Research at the University of Minnesota in 1998. He became an Assistant Professor of Biomedical Engineering at Cairo University from 1998 to 2003. He also worked during the same period as the director of research and development at IBE technologies, Egypt. He also was a Research Associate with the Biomedical Imaging Technology Center at the Biomedical Engineering Department, Emory University/Georgia Institute of Technology, from 2002 till 2004. He is currently an associate professor of Biomedical Engineering at Cairo University and joined Nile University in August 2008 as an associate professor of Medical Imaging. Along his career, he received several awards and recognitions including the record for highest undergraduate GPA in the department of Biomedical Engineering at Cairo University, the Islamic Development Bank Merit Scholarship (1993-1996), his biography was selected to appear in Marquis Who's Who in the World in 2002, as well as winning the E.K. Zaviosky stipend from the International Society of Magnetic Resonance in Medicine in 2002. He was also elected a Senior Member of the Institute of Electrical and Electronic Engineers (IEEE) in 2003. He received the State Encouragement Award in Engineering Sciences for the year 2005 from the Academy of Scientific Research and Technology, Egypt. He was also selected to receive the 2007 Professor of the Year award in biomedical engineering from the Cairo University Faculty Club. He was awarded of the 2008 Cairo University Prize for Academic Excellence. He was the Scientific Program Chair for the Third Cairo International Biomedical Engineering Conference (CIBEC'06), which was held in Cairo in December 2006. He is currently an active member of the International Society of Magnetic Resonance in Medicine (ISMRM) and the International Society of Optical Engineering (SPIE). He published over 120 peer-reviewed journal and conference articles in the areas of medical imaging and in particular MRI and ultrasound imaging, and multi-dimensional signal processing for biomedical applications.