ANNALS of Faculty Engineering Hunedoara
– International Journal of Engineering
Tome XIII [2015] – Fascicule 2 [May]
ISSN: 1584-2665 [print]; ISSN: 1584-2673 [online]
a free-access multidisciplinary publication of the Faculty of Engineering Hunedoara

[1.] Antara GHOSH, [2.] Soma BARMAN

REALIZATION OF AN EVD MODEL IN LABVIEW ENVIRONMENT FOR IDENTIFICATION OF CANCER AND HEALTHY HOMO SAPIENS GENES

[1-2.] Institute of Radio Physics & Electronics, University of Calcutta, INDIA

Abstract: Homo-sapiens cancer and healthy genes is identified here by estimating correlation coefficient of genes. A model has been realized in Lab VIEW environment for classification of genes based on eigen value decomposition (EVD) technique. Breast, Prostate and Colon genes are taken as sample database for analysis of genes. EVD method is used to reduce the dimensionality as well as preserving the class discriminatory information of genes. Feature Vector is calculated based on highest eigen value of data set which is treated as measurement metric to differentiate cancer genes from healthy genes. The EVD model successfully screen out breast cancer genes from breast healthy genes, prostate cancer genes from prostate healthy genes and also colon cancer genes from colon healthy genes.
Keywords: Amino-acid, Cancer, EIIP, Eigen Value Decomposition, Lab VIEW

1. INTRODUCTION

Genomics based research has evolved rapidly because of large sequencing data of complete genome is available in public domain. The research areas of genomics are mainly DNA sequence analysis and disease diagnosis [Vaidyanathan, 2004] which is based on information extraction and data analysis. DNA sequence analysis is a well developed research area used to reveal hidden features present in protein coding regions whereas diagnosis of disease is used to find out abnormalities present in DNA sequence because almost all the genetic diseases, such as Parkinson, Alzheimer, Cancer and development of abnormalities are characterized by the presence of genetic variations [Golub et al. 1999]. Among all, cancer is responsible for one in eight deaths in the world and the complexity and heterogeneity encourages the aggressive growth of cancer cells leading to significant mortality in patients. Therefore, understanding the mechanism of cancer development, accurate detection and classification of cancer is a research topic of significant importance. Two major goals of functional genomics are: to use genomic signals to classify disease on molecular level and to screen for genes that determine specific cellular disease and model their activity in such a way that normal and abnormal behavior can be differentiated.

It is well known that the amino acids of a DNA sequence are responsible for the formation of protein [Barman et al. 2011]. Amino acids are essential to form antibodies to combat bacteria and viruses, they are the part of the enzyme and hormonal system, play a significant role in cancer progression. Controlled Amino Acid Therapy (CAAT) is an efficient medical treatment used to impair the development of cancer gene [Vaidyanathan, 2004]. Understanding the complex function of amino acids in cancer, authors have concentrated their study in amino acids levels to model the breast, prostate and colon genes and an attempt has been made to differentiate healthy and abnormal genes. The model is tested on Homo sapiens genes and Eigen Value Decomposition technique is used for estimating the correlation between genes. The model is realized in NI Lab VIEW environment to find out principal eigen values that are describe the correlation between healthy and cancer genes. The National Institutes of Health (NIH) protein database [http://www.ncbi.nlm.nih] used for sample collection. The eigen decomposition techniques produces feature vector which is used as indicator for identification of genes. The Eigen analysis technique is successfully screen out cancer genes from healthy genes and percentage of false detection is computed. The paper is organized into number of sections: Introduction, Materials & Methods, Results and Discussion and Conclusions.

2. MATERIALS AND METHODS

2.1. Brief background

The story of life starts from genome and DNA is important to all genomes and organisms because an organism could do nothing without DNA [Alberts et al. 1998]. A mutation is the permanent change in DNA can arise spontaneously with apparent cause and can lead to cell death, cell alteration, cell formation or in some cases development of cancer genes [Qie et al. 2007]. DNA, an informational molecule encoding the genetic information, encoded as a sequence of nucleotide bases: guanine (G), adenine (A),

thymine (T) and cytosine (C) and divided into two regions: Genes and Inter-genic spaces. A gene also can be dividing into two sub-regions named Exon (coding region) and Intron (non-coding region) as in Figure 1. Exons of a DNA sequence are the most information bearing part because only the exons take part in protein coding while the introns are spliced off during protein synthesis. In exon region the bases are dividing into three adjacent bases called codon which translated into amino acid and 64 possible codons generate 20 amino acids shown in Table 1. These 20 amino acids are responsible for forming proteins; deficiencies of these may leads to different types of genetic abnormalities [Barman et al. 2011].

The study of amino acids in genes presents a new horizon for cancer classification and prediction. Some amino acids are essential for the growth of tumor genes and restricting them or inhibiting them may be beneficial for curing cancer patients [Barman et al. 2011]. Digital Signal Processing (DSP) can be effectively used in genomics study with great accuracy as Genomic or Proteomic information is digital in nature. In the present study, Discrete Fourier Transform (DFT) is used for spectrum estimation of genes and EVD model is realized for computation of similarity score between healthy and cancer genes. As gene sequences comprise of long chain of amino acids, eigen value decomposition technique is judiciously selected here to reduce the computational complexity of the process.

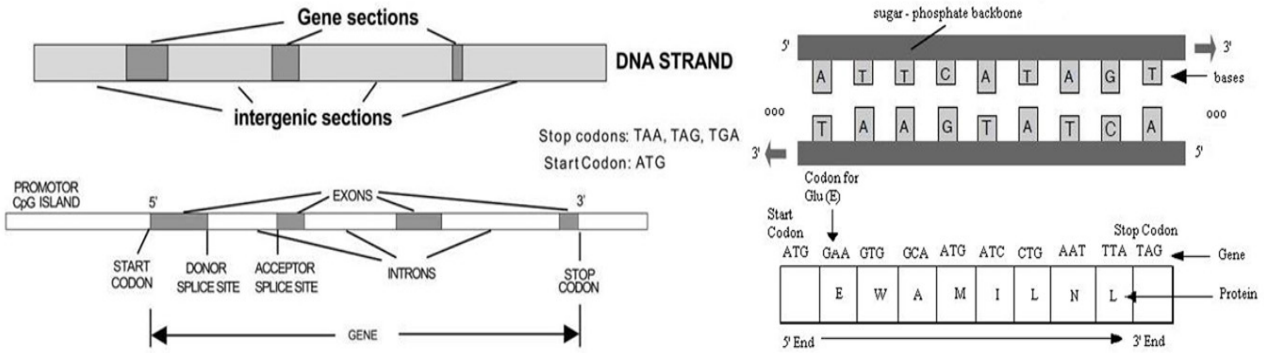

Figure 1. DNA is organized into genes and each gene contains the information to make a protein

2.2. Methodology

Almost all human genetic diseases such as cancer and development of abnormalities are characterized by the presence of genetic variation, it is a challenging problem to the researchers to find out the difference of characteristic between healthy and abnormal gene [Barman et al. 2011]. Digital Signal Processing tools are more effective to process and interpret genomic signal and provide a better description for understanding the biological mechanism of cancers. Discrete Fourier Transform (DFT) technique has found to be very useful for both DNA/RNA, amino acid/protein sequence analysis [Anastassiou 2001; Khare et al. 2011]. Amino acids consist of alphabets A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, Y, hence a numerical conversion technique is required to convert the alphabetic sequence into numerical sequence prior to apply DSP method [Vaidyanathan 2004]. A well known single sequence electron ion interaction pseudo potential (EIIP) mapping rule [Nair et al. 2006; Meher et al. 2011], based on the distribution of free electron's energy along DNA sequence is used for spectral analysis of DNA sequence [Roy and Barman 2011]. The EIIP values of 20 amino acids are displayed in Table 1.

Table 1. List of 20 amino acids with codon and EIIP values

Abbreviation		Amino Acid	Codons	EIIP Value
A	Ala	Alanine	GCA,GCC,GCG,GCT	0.0373
C	Cys	Cystein (has S)	TGC,TGT	0.0829
D	Asp	Aspartic Acid	GAC,GAT	0.1263
E	Glu	Glutamic Acid	GAA,GAG	0.0058
F	Phe	Phenylalanine	TTC,TTT	0.0946
G	Gly	Glycine	GGA,GGC.GGG,GGT	0.0050
H	His	Histidine	CAC,CAT	0.0242
I	Ile	Isoleucine	ATA,ATC,ATT	0.0000
K	Lys	Lysine	AAA,AAG	0.0371
L	Leu	Leucine	TTA,TTG,CTA,CTC,CTG,CTT	0.0000
M	Met	Methionine	ATG	0.0823
N	Asn	Asparagine	AAC,AAT	0.0036
P	Pro	Proline	CCA,CCC,CCG,CCT	0.0198
Q	Gln	Glutamine	CAA,CAG	0.0761
R	Arg	Arginine	AGA,AGG,CGA,CGC,CGG,CGT	0.0959
S	Ser	Serine	AGC,AGT,TCA,TCC,TCG,TCT	0.0829
T	Thr	Threonine	ACA,ACC,ACG,ACT	0.0941
V	Val	Valine	GTA,GTC,GTG,GTT	0.0057
W	Trp	Tryptophan	TGG	0.0548
Y	Tyr	Tyrosine	TAC,TAT	0.0516

Suppose, an amino acid chain of a gene is: x[n] = [M P I G S K E R P T F D];

After EIIP mapping using Table 1: x[n] = [0.0373 0.0198 0.0000 0.0050 0.0829 0.0371 0.0058 0.0959 0.0198 0.0941 0.0946 0.1263];

Spectral estimation of EIIP mapped sequence is obtained by using Discrete Fourier Transform (DFT) technique and an EVD model is realized for discriminate analysis between cancer and healthy genes. EVD has a particularly simple expression for a class of matrices often used in multivariate analysis such as correlation, covariance, or cross-product matrices and the set of eigen values of matrix is

also called its spectrum [Abdi et.al, 2013, Parra et.al 2003 , Yang et. al 2006]. Researchers are used mostly Principal Component Analysis (PCA) tool for classification of protein structures or to detect different features of breast cancers [Christoyianni et. al 2006, Hasan et.al 2010, Melo et.al 2003, Vipsita et. al 2011]. In this present study, EVD model is used as a linear discriminator which reduces the dimensionality but preserving as much as class discriminatory information. The discriminate analysis of genes in the present model is closely related to Principal Component Analysis (PCA).

The block representation of the Eigen Value Decomposition (EVD) model for correlation analysis of genes is depicted in Figure 2 and the algorithm for the analysis is illustrated in the following steps:

1 Convert the scanned amino acid sequence into numerical form using EIIP values and compute DFT of the converted sequence using these equations:

Let, x(n) = a normal prostate/breast gene (taken as reference set) and y(n) = a normal or cancer prostate/breast gene (taken as target set). The DFT of the EIIP sequence is given by:

$$X_S[k] = \sum_n x[n] e^{-j2\pi nk/N} \tag{1}$$

$$X[k] = |Xs[k]| \tag{2}$$

$X_S[k]$ = DFT of reference sequence; $Y_S[k]$= DFT of target sequence; X[k] = Amplitude spectrum of $X_S[k]$;

Similarly, Y[k] = Amplitude spectrum of $Y_S[k]$; N = Length of DNA sequence; k= 0, 1, 2, …, N-1; n= 0, 1, 2, …, N-1;

2 Object is represented as a cloud of points in a multi-dimension space. The centroid of the points is defined by mean. Convert X[k] and Y[k] into column matrix and obtain the mean value of X[k] and Y[k]:

$$X = \overline{X[k]} \quad ; \quad Y = \overline{Y[k]} \tag{3}$$

3 The mean subtracted is the average across each dimension. This produces a data set whose mean is zero.

$$X_{new} = X[k] - X \; ; \; Y_{new} = Y[k] - Y \tag{4}$$

4 The degree to which the sets are correlated is represented by covariance.

$$\text{Covariance matrix} = A = \text{cov}[X_{new}, Y_{new}] = \begin{bmatrix} \text{cov}(x,x) & \text{cov}(x,y) \\ \text{cov}(y,x) & \text{cov}(y,y) \end{bmatrix} \tag{5}$$

5 To maximize class separability by eigen decomposition technique, the principal axes involves the eigen analysis of the cross-product matrix (A), using the characteristic equation:

$$(A - \lambda I)x = 0 \tag{6}$$

where, λ = eigen-values are the solution of characteristic equation and x is the eigen vector associated with eigen value λ.

and $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ = Identity matrix.

6 In this step consideration is given on data compression and dimensionality reduction. Highest eigen-values are the principal component and less significant eigen-values are ignored. Calculate the principal component of the data sets and formed a feature vector (matrix of vector) based on these eigen-values:

$$\text{Feature vector} = M = \begin{bmatrix} eig1 \\ eig2 \end{bmatrix} \tag{7}$$

7 In this step we transformed the data in such a way, that the data are mostly described by a relationship between them and a new Final Data matrix $(1 \times k)$ is created. This is plot with respect to length of the sequence:

$$\text{Final Data (k)} = M^Y * [X_{new}, Y_{new}] \tag{8}$$

The algorithm is tested on several Homo-sapiens databases collected from NCBI Homepage, as shown in Table 2.

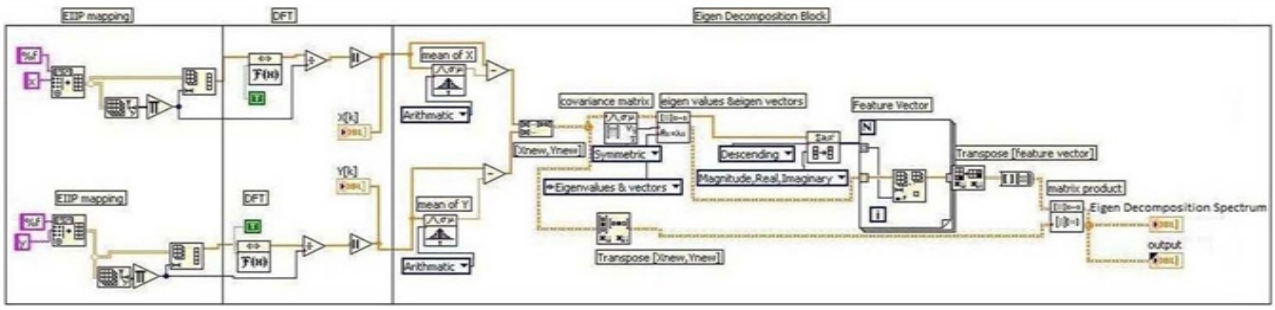

Figure 2. Eigen Decomposition model for correlation study

Table 2. Healthy and Cancerous breast, prostate and colon Homo-sapiens sample genes

Gene Type		Accession Number
Breast Gene	Normal	NM013375.3, NM015407.4, NM015423.2, NM021243.2, NM024666.4, NM024684.2, NM032548.3, NM148912.2
	Cancer	AF012108.1, AF041259.1, AF126008.1, AF308285.1, AY273801.1, NM000059.3, NM007297.3, NM007300.3
Prostate Gene	Normal	AF224278.1, AF331165.1, AF462605.1, M15885.1, M24543.1, M24902.1, NM_007003.2,NM_005984.3
	Cancer	AAQ08976.1, AF304370.1, AF338650.1, AF455138.1, AY008445.1, FJ649644.1, NP001035756.1, NP001231873.1
Colon Gene	Normal	NM_012278.1, NM_024533.4, NM_030754.4, NM_032044.3, NM_138937.2, NM_001127380.2, NM_001159352.1, NM_001159353.1
	Cancer	AB489153.1, AF250731.1, AY217549.1, AY581148.1, NM_031941.3, NM_182762.3, NM_001161345.1, NM_001278623.1

3. RESULTS AND DISCUSSIONS

In this study, initially the authors consider three different types of Homo-sapiens cancer genes breast, prostate and colon. The correlation coefficients obtained from covariance matrix is not varying sufficiently over the different class to differentiate. Eigen decomposition analysis is applied after covariance matrix computation to maximize separability between two classes and a feature vector is computed. The feature vector is used as indicator to discriminate cancer and healthy genes. In this Eigen Value Decomposition (EVD) model, feature vectors have been calculated between (8 healthy × 8 healthy) breast genes and (8 healthy × 8 cancer) breast genes; (8 healthy × 8 healthy) prostate genes and (8 healthy × 8 cancer) prostate genes and (8 healthy × 8 healthy) colon genes and (8 normal × 8 cancer) colon genes. EVD analysis shows all positive feature vectors, when healthy genes of one accession number compared with healthy genes of other accession number.

Therefore, the correlation between them is high or they are similar genes. Whereas when healthy genes compared with the cancerous genes, in case of breast genes out of 64 combinations, 61 show negative value and in case of prostate and colon genes out of 64 combinations, 63 shows negative value means they are orthogonal or uncorrelated samples. Four cross correlation out of 192 cross correlation matrix show unexpected value means total error is only near about 2.6% in this present study. The model proposed in this article will be suitable for preliminary prediction of cancer and healthy genes. The algorithm is tested in Lab VIEW 2012 environment [http://india.ni.com/]. The correlation coefficients obtained from this model for breast healthy vs. healthy and breast healthy vs. cancer genes and prostate healthy vs. healthy and prostate healthy vs. cancer genes and colon healthy vs. healthy and colon healthy vs. cancer genes are depicted in respectively Tables 3- 5, Tables 6-8. Due to space constraint, only some of the simulated plots for the EVD analysis are displayed in Figure 3 to Figure 8.

Table 3. Correlation between breast healthy vs. breast healthy genes

		HEALTHY							
	Accession no.	NM013375.3	NM015407.4	NM015423.2	NM021243.2	NM024666.4	NM024684.2	NM032548.3	NM148912.2
H	NM013375.3	0.06	0.05	0.06	0.05	0.06	0.06	0.06	0.06
E	NM015407.4	0.05	0.05	0.05	0.04	0.05	0.05	0.05	0.06
A	NM015423.2	0.06	0.05	0.06	0.05	0.06	0.06	0.06	0.06
L	NM021243.2	0.05	0.04	0.05	0.05	0.05	0.05	0.06	0.05
T	NM024666.4	0.06	0.05	0.06	0.05	0.06	0.06	0.06	0.06
H	NM024684.2	0.06	0.05	0.06	0.05	0.061	0.05	0.06	0.06
Y	NM032548.3	0.06	0.05	0.06	0.06	0.06	0.06	0.06	0.06
	NM148912.2	0.06	0.05	0.06	0.05	0.06	0.06	0.06	0.06

Table 4. Correlation between breast healthy vs. breast cancerous genes

		CANCER							
	Accession no.	AF012108.1	AF041259.1	AF126008.1	AF308285.1	AY273801.1	NM000059.3	NM007297.3	NM007300.3
H	NM013375.3	-0.07	-0.07	-0.07	-0.07	-0.0	-0.07	-0.075	-0.07
E	NM015407.4	**0.06	**0.05	**0.05	-0.05	-0.06	-0.06	-0.06	-0.06
A	NM015423.2	-0.07	-0.07	-0.07	-0.07	-0.07	-0.07	-0.07	-0.07
L	NM021243.2	-0.06	-0.06	-0.06	-0.06	-0.07	-0.06	-0.06	-0.06
T	NM024666.4	-0.07	-0.06	-0.07	-0.06	-0.07	-0.07	-0.07	-0.07
H	NM024684.2	-0.06	-0.06	-0.06	-0.06	-0.06	-0.07	-0.06	-0.06
Y	NM032548.3	-0.07	-0.07	-0.07	-0.07	-0.07	-0.07	-0.07	-0.07
	NM148912.2	-0.06	-0.06	-0.06	-0.06	-0.06	-0.07	-0.07	-0.07

** marks denote the positive outputs instead of negative outputs i.e. error.

Table 5. Correlation between prostate healthy vs. prostate healthy genes

		HEALTHY							
	Accession no.	AF224278.1	AF331165.1	AF462605.1	M15885.1	M24543.1	M24902.1	NM005984.3	NM007003.2
H	AF224278.1	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06
E	AF331165.1	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06
A	AF462605.1	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06
L	M15885.1	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06
T	M24543.1	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06
H	M24902.1	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06
Y	NM005984.3	0.06	0.06	0.06	0.06	0.05	0.06	0.05	0.06
	NM007003.2	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06

Table 6. Correlation between prostate healthy vs. prostate cancerous genes

	CANCER								
HEALTHY	Accession no.	AAQ08976.1	AF304370.1	AF338650.1	AF455138.1	AY008445.1	FJ649644.1	NP001035756.1	NP001231873.1
	AF224278.1	-0.06	-0.06	-0.06	-0.06	-0.06	-0.06	-0.06	-0.06
	AF331165.1	-0.06	-0.06	-0.06	-0.06	-0.06	-0.06	-0.06	-0.06
	AF462605.1	-0.06	-0.06	-0.06	-0.06	-0.06	-0.06	-0.06	-0.06
	M15885.1	-0.06	-0.06	-0.06	-0.06	-0.06	-0.06	-0.06	-0.06
	M24543.1	-0.06	-0.06	-0.06	-0.063	-0.06	-0.06	-0.06	-0.06
	M24902.1	-0.06	-0.06	-0.06	-0.06	-0.06	-0.06	-0.06	-0.06
	NM005984.3	-0.05	-0.06	-0.06	-0.05	-0.05	**0.06	-0.05	-0.05
	NM007003.2	-0.06	-0.06	-0.06	-0.06	-0.06	-0.06	-0.06	-0.06

** marks denote the positive outputs instead of negative outputs i.e. error.

Table 7. Correlation between colon healthy vs. colon healthy genes

	HEALTHY								
HEALTHY	Accession no.	NM_012278.1	NM_024533.4	NM_030754.4	NM_032044.3	NM_138937.2	NM_001127380.2	NM_001159352.1	NM_001159353.1
	NM_012278.1	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06
	NM_024533.4	0.061	0.06	0.06	0.06	0.06	0.06	0.06	0.06
	NM_030754.4	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06
	NM_032044.3	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06
	NM_138937.2	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06
	NM_001127380.2	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06
	NM_001159352.1	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06
	NM_001159353.1	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06

Table 8. Correlation between colon healthy vs. colon cancerous genes

	CANCER							
HEALTHY	Accession no.	AB489153.1	AF250731.1	AY217549.1	AY581148.1	NM_031941.3	NM_182762.3	NM_001161345.1	NM_001278623.1
	NM_012278.1	-0.06	-0.06	-0.06	-0.06	-0.06	-0.06	**0.06	-0.06
	NM_024533.4	-0.07	-0.07	-0.06	-0.07	-0.07	-0.07	-0.07	-0.07
	NM_030754.4	-0.07	-0.07	-0.07	-0.07	-0.07	-0.07	-0.07	-0.07
	NM_032044.3	-0.07	-0.07	-0.07	-0.07	-0.07	-0.07	-0.07	-0.07
	NM_138937.2	-0.07	-0.07	-0.07	-0.07	-0.07	-0.07	-0.07	-0.07
	NM_001127380.2	-0.07	-0.07	-0.07	-0.07	-0.07	-0.07	-0.07	-0.07
	NM_001159352.1	-0.07	-0.07	-0.07	-0.07	-0.07	-0.077	-0.07	-0.07
	NM_001159353.1	-0.07	-0.07	-0.07	-0.07	-0.07	-0.07	-0.07	-0.07

** marks denote the positive outputs instead of negative outputs i.e. error.

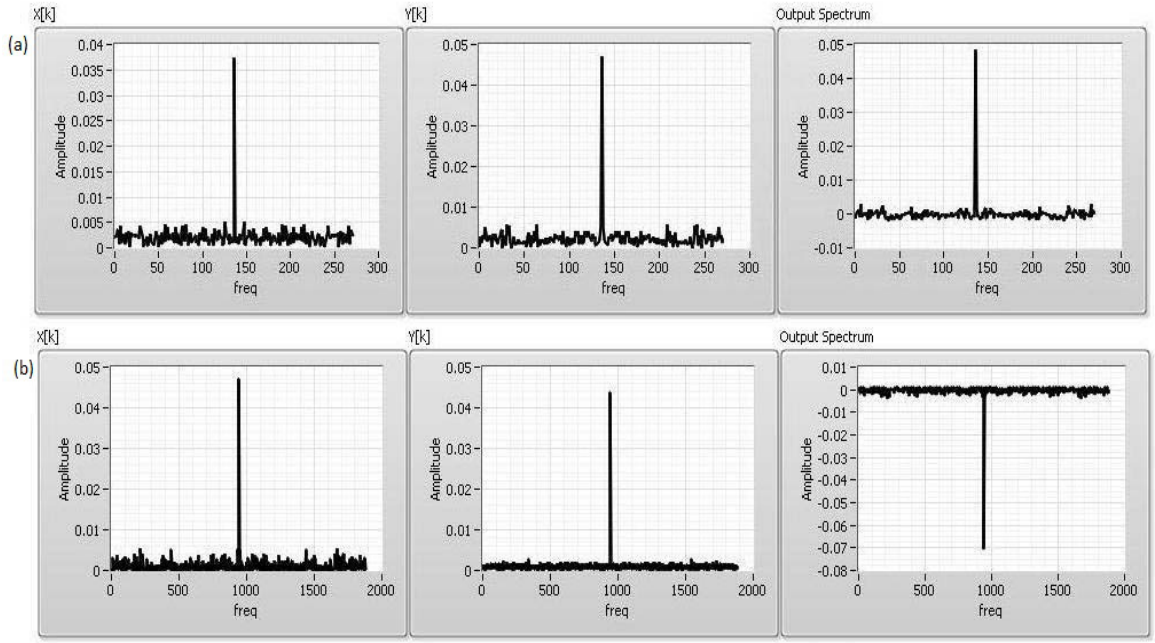

Figure 3. (a) DFT of x(n), DFT of y(n) and EVD of NM_013375.3 (breast healthy) vs. NM_015407.4 (breast healthy). (b) DFT of x(n), DFT of y(n) and EVD of NM_013375.3 (breast healthy) vs. NM_007300.3 (breast cancer).

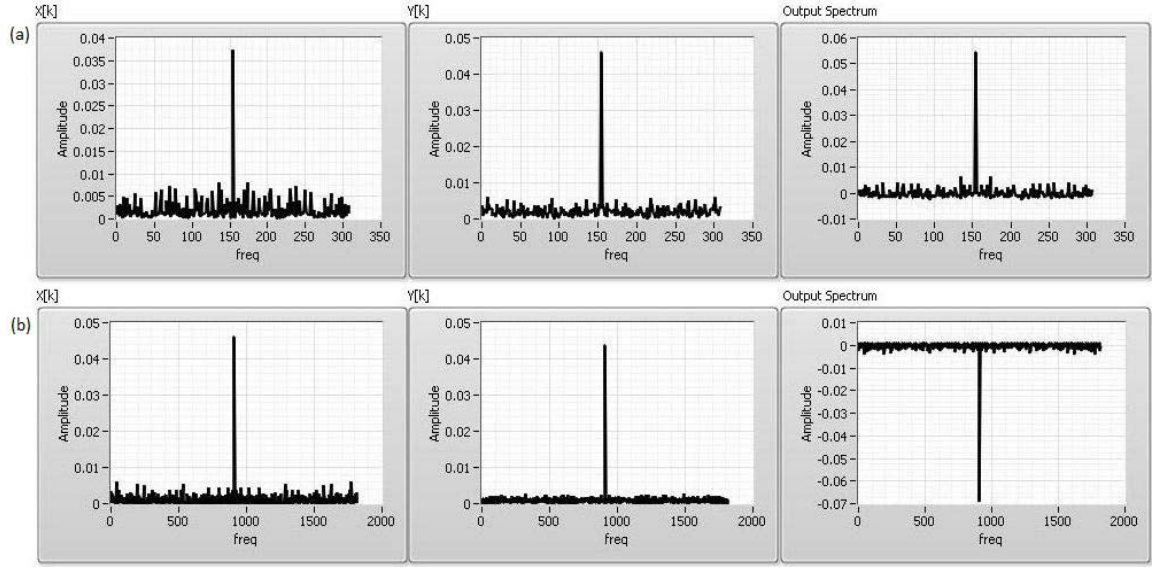

Figure 4. (a) DFT of x(n), DFT of y(n) and EVD of NM015423.2 (breast healthy) vs. NM021243.2 (breast healthy). (b) DFT of x(n), DFT of y(n) and EVD of NM015423.2 (breast healthy) vs. M007297.3 (breast cancer).

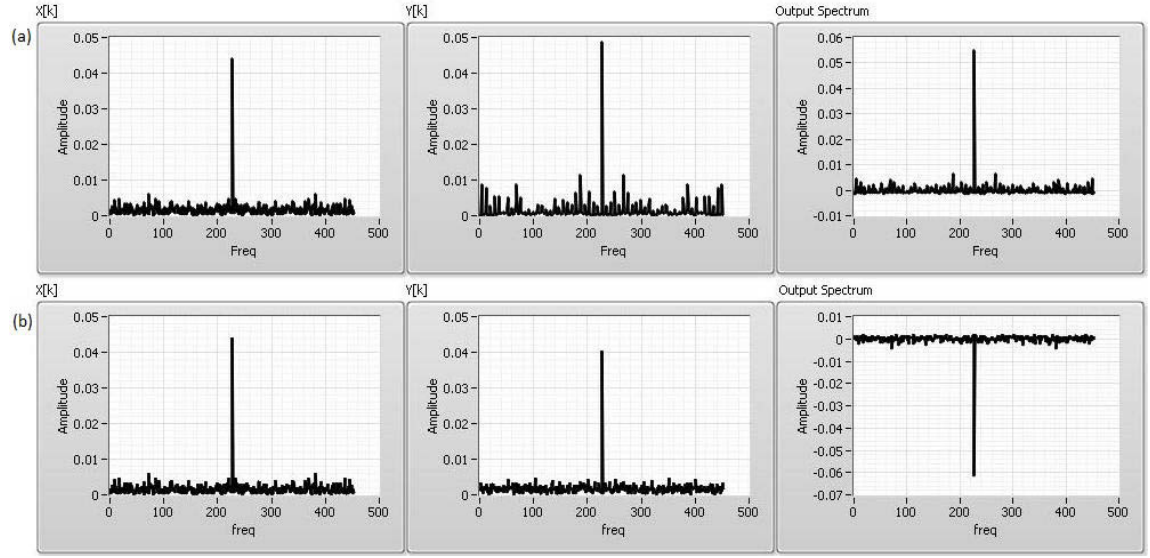

Figure 5. (a) DFT of x(n), DFT of y(n) and EVD of AF224278.1 (prostate healthy) vs. AF331165.1 (prostate healthy). (b) DFT of x(n), DFT of y(n) and EVD of AF224278.1 (prostate healthy) vs. AAQ08976.1 (prostate cancer).

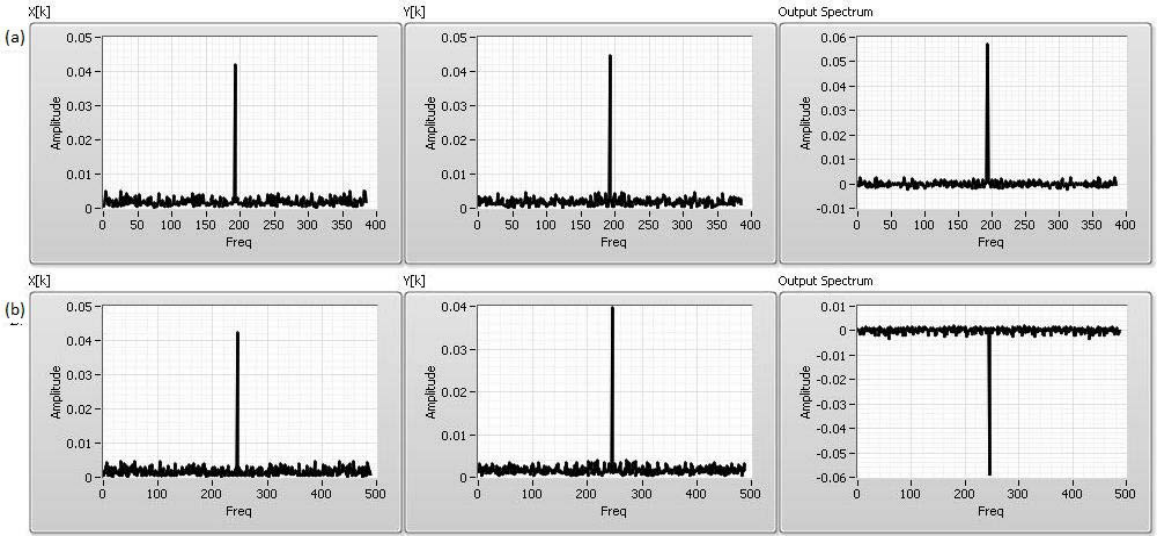

Figure 6. (a) DFT of x(n), DFT of y(n) and EVD of M24543.1 (prostate healthy) vs. M24902.1 (prostate healthy). (b) DFT of x(n), DFT of y(n) and EVD of M24543.1 (prostate healthy) vs. AY008445.1 (prostate cancer).

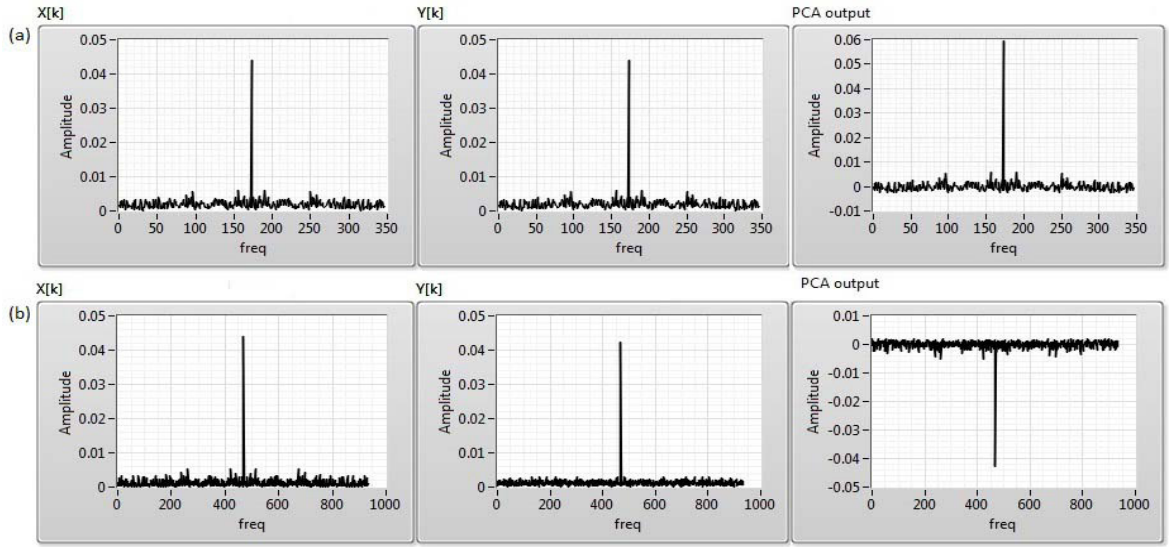

Figure 7. (a) DFT of x(n), DFT of y(n) and EVD of NM_012278.1 (colon healthy) vs. NM_012278.1 (colon healthy). (b) DFT of x(n), DFT of y(n) and EVD of NM_012278.1 (colon healthy) vs. AB489153.1 (colon cancer).

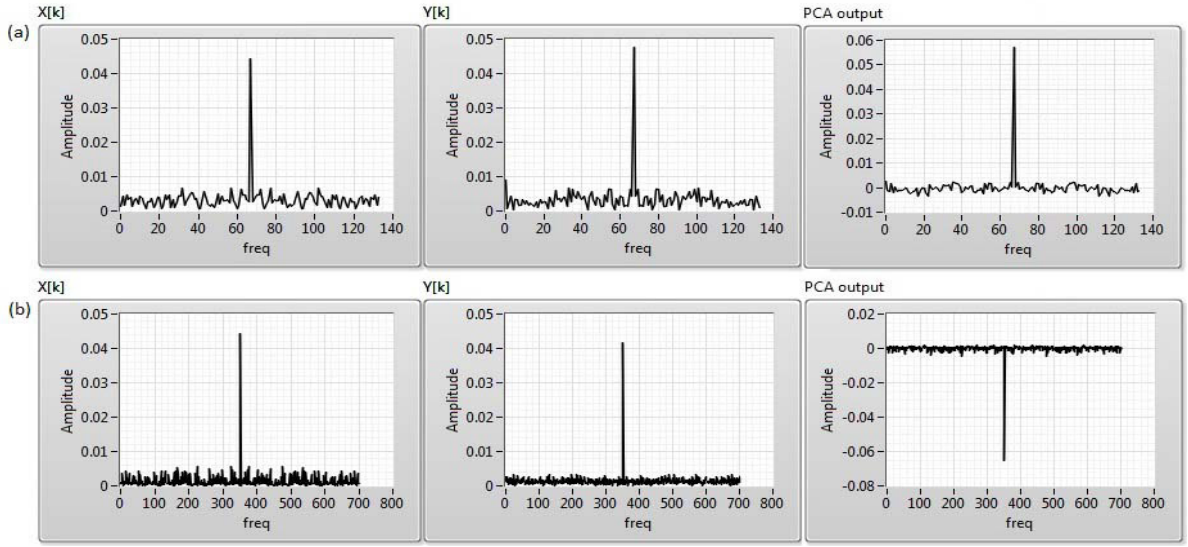

Figure 8. (a) DFT of x(n), DFT of y(n) and EVD of NM_001159353.1 (colon healthy) vs. NM_030754.4 (colon healthy). (b) DFT of x(n), DFT of y(n) and EVD of NM_001159353.1 (colon healthy) vs. NM_031941.3 (colon cancer).

4. CONCLUSIONS

Nowadays Genomics research is not only limited to wet smelly laboratory. Soft databases are available in public domain (NCBI); scientist from different fields may use their expertise for analysis and diagnosis of genetic diseases. Rapid and effective differentiation between healthy and cancer genes is an important challenge for diagnosis and treatment of cancers and different stages of cancer genes. EIIP based Fourier spectra of genes is considered in the EVD model for classification into healthy and cancer genes. The EVD model in this article is tested successfully on breast, prostate and colon genes. This model driven approach can be used as an indicator to sort out disease genes from a large set of genes. The EVD spectrum promise preliminary detection and prediction of genetic abnormalities of genes. In future we plan to explore further studies on classification of different types of cancer genes.

ACKNOWLEDGMENTS

The first author would like to acknowledge the Center for Research in Nanoscience and Nanotechnology (CRNN), University of Calcutta, for providing her scholarship.

REFERENCES

[1.] A P. John Institute for Cancer Research paper on Controlled Amino Acid Therapy (CAAT) works, [online], http://www.apjohncancerinstitute.org (Accessed 30 November 2013).

[2.] Alberts, B; Bray, D; Johnson, A; Lewis, J; Raff, M; Roberts, K and Walter, P: Essential cell biology, Edition (2), New York, Garland Publishing Inc., 1998.

[3.] Anastassiou, D: Genomic Signal Processing, Signal Processing Magazine IEEE, volume(18), 8-20, 2001.

[4.] Abdi, H: The Eigen-Decomposition:Eigenvalues and Eigenvectors, Encyclopedia of Measurement and Statistics, 2007, http://ftp.utdallas.edu/~herve/Abdi-EVD2007-pretty.pdf (Accessed 30 November 2013).
[5.] Barman, S; Saha, S; Mondal, A and Roy; M: Signal Processing Techniques for the Analysis of Human Genome Associated with Cancer Cells, in 2nd Annual international Conf. IEMCON, 570-573, 2011.
[6.] Christoyianni, I; Koutras, A; Dermatas, E and Kokkinakis, G: Computer Aided Diagnosis of Breast Cancer in Digitized Mammograms, Elsevier Science Ltd, Computersied Medical Imaging and Graphics, volume (26), 309-319, 2002.
[7.] Golub, T, R; Slonim, D, K; Tamayo, P; Huard, C; Gaasenbeek, M; Mesirov, J, P; Coller, H; Loh, M, L; Downing, J, R; Caligiuri, M, A; Bloomfield, C, D; and Lander, E, S: Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring, Science, volume (286), 531-537, 1999.
[8.] Hasan, H and Tahir, N: Feature Selection of Breast Cancer Based on Principal Component Analysis, In Sixth Int. Colloquium on Signal Processing and its Applications IEEE, 242-245, 2010.
[9.] Khare, A; Nigam, A and Saxena, M: Identification of DNA Sequences By Signal Processing Tools in Protein-Coding Regions, Search & Research, volume(2), 44-49, 2011.
[10.] Melo, J, C, B; Cavalcanti, G, D, C and Guimaraes, K, S: PCA Feature Extraction for Protein Structure Prediction, In Neural Networks, Proceedings of the International Joint Conference on IEEE, 2952-2957, 2003.
[11.] Meher J, K; Dash G, N; Meher, P, K and Raval, M, K: A reduced computational load protein coding predictor using equivalent amino acid sequence of DNA string with period-3 based time and frequency domain analysis, American Journal of Molecular Biology, volume (1), 79-86, 2011.
[12.] Nair, A, S and Sreenathan, S, P: A coding measure scheme employing electron-ion interaction pseudo-potential (EIIP), Bioinformation, volume(1), 197-202, 2006.
[13.] National Centre for Biotechnology Information (NCBI), [online] http://www.ncbi.nlm.nih (Accessed 3 May 2014).
[14.] National Instrument, [online] http://india.ni.com/ (Accessed 20 November 2013).
[15.] Parra, L and Sajda, P: Blind Source Separation via Generalized Eigen value Decomposition, Journal of Machine Learning Research, volume (4), 1261-1269, 2003.
[16.] Qie, P; Wang, Z, J and Lie, K ,J, R: Genomic Processing for Cancer Classification and Prediction, Signal Processing Magazine IEEE, volume(24), 100-110, 2007.
[17.] Roy, M and Barman, S: Spectral Analysis of Genomic Data by Recursive Winer-Khinchine Theorem using Various Mapping Techniques, In Proc. Of International Conference on Nanotechnology and Biosensors, 276-279, 2011.
[18.] Vaidyanathan, P, P: Genomics and Proteomics: a signal processor's tour, IEEE circuit and system magazine, volume (4), 315-319, 2004.
[19.] Vipsita, S; Shee, B, K and Rath, S, K: Protein Superfamily Classification using Kernel Principal Component Analysis and Probabilistic Neural Networks, In India Conference (INDICON 2011), Annual IEEE, 1-6, 2011.
[20.] Yang, W, H and Dai, D, Q: Generalized Discriminant Analysis for Tumor Classification with Gene Expression Data, In IEEE: Machine Learning and Cybernetics International Conference, Dalian, China, 4322 – 4327, 2006

ANNALS of Faculty Engineering Hunedoara
– International Journal of Engineering