Proc. of the Intl. Conf. on Advances In Bio-Informatics And Environmental Engineering - ICABEE 2014.

ISBN: 978-1-63248-019-4 doi: 10.15224/ 978-1-63248-019-4-12

# *Classification DNA Sequences of Bacterias using Multi Library Wavelet Networks*

Abdesselem DAKHLI[1], Wajdi BELLIL[2], Chokri BEN AMAR[3]

*Abstract*— **Genomic sequences allow to classify organisms into different categories and classes which have significant biological knowledge and can justify the evolution and identification of unknown organisms. Also they study mutual relations between organisms. The purpose of this classification is to study living organisms. Our system consists in three phases. The first phase is called transformation wich is composed of three steps; binary codification of DNA sequence, Fourier Transform and Power Spectrum Signal Processing. The second phase is called approximation. This phase is empowered by the use of Multi Library Wavelet Neural Networks (MLWNN).The third phase is called classification wich is realized by applying the algorithm of hierarchical classification. The results of this contribution are more interesting in comparison with some others works, in terms of rate classification using bacteria database.**

***Keywords*— Classification, DNA, wavelet networks, power spectrum, Multi Library Wavelet Neural Networks.**

# I. Introduction

The biologists have discovered several unknown organisms that can be classified in the taxonomic hierarchy. These discoveries help to understand biological organisms during life time. Computer bodies lying in the DNA sequences explain and redirect the functions of inherited charactertics of different generations of organisms. These sequences can be processed from the raw material by biological methods of DNA sequencing. The DNA sequence is formed by a chain comprising serie of nucleotides. Each nucleotide is composed of three subunits: a phosphate group, a sugar and nucleic bases (A, T, C, G). Classification of organisms has been studied by several researchers. Sandberg et al. [4] proposed a method based on Bayesian approach. The mean accuracy obtained was 85%. Francisca Z. et al, used Markov Model to classify proteins of microbes, eukaryotes and Archea. This classification had followed accuracy equal to 83.51%, 82.12% and 66.63% respectively for Eukaryota, Microbes and Archaea [5].

[1]Department of Computer Science, REGIM, University of Gabes, Tunisia

[2]Department of Electrical Engineering, REGIM, University of Gafsa, Tunisia

[3]Department of Electrical Engineering, REGIM, University of Sfax, Tunisia

Narasimhan S. et al. applied Principal Component Analysis (PCA) to extract features from the genomic sequence to classify organisms. They obtained some effective results [6]. This paper is organized as follow: in section 2 we describe an overview of the proposed approach. Section 3 presents the theory of Beta wavelet. This function will be used at Wavelet Network. Section 5 presents the simulation results of the proposed DNA sequences classification method and section 6 closes with a conclusion and discussion.

# II. Proposed approach

This paper presents a new approach of classification of DNA sequence based on wavelet network using Multi Library Wavelet Neural Networks (MLWNN) to approximate f (x) of a DNA sequence. This approach is divided in three stages: transformation of DNA, approximation of the input signal and classification of compact signature DNA sequences using algorithm of hierarchical clustering.

## A. *Transformation of DNA sequence*

### 1) Binary codification of DNA sequence and Feature Extraction

The proposed classification of species in class is according to DNA sequence components. This sequence is formed by four basic nucleotides, adenine (A), guanine (G), cytosine (C) and thymine (T), and each organism is identified by its DNA sequence [9]. The representation of multidimensional data is an important question when we have to process data with neural networks in the field of the artificial intelligence. To reduce the complexity and to have a simple data representation we have to extract the characteristics[8]. Linear feature extraction can be viewed as finding a set of vectors which represent effectively information content of an observation while reducing the dimensionality [10,8]. The method of indicator translates the data into digital format which can be used for DNA signal spectrum analysis. This method uses the binary number and its indicates 1 or 0 for the existence or not of a specific nucleotide at DNA sequence level [1].

TABLE I. BINARY ENCODING OF NUCLEOTIDES

| Nucleotides | 4-bit binary encoding |
|---|---|
| **A** | 1000 |
| **C** | 0100 |
| **G** | 0010 |
| **T** | 0001 |

The binary indicator sequence is formed by replacing the individual nucleotides with values either 0 or 1. 1 stands for

Proc. of the Intl. Conf. on Advances In Bio-Informatics And Environmental Engineering - ICABEE 2014.

ISBN: 978-1-63248-019-4 doi: 10.15224/ 978-1-63248-019-4-12

presence and 0 for absence of a particular nucleotide in specified location in DNA signal [4, 16].

### *2)* Fourier transform and power spectrum signal processing

After the genomic data have been converted into these indicator sequences, they can be manipulated with mathematical methods. The discrete Fourier Transform is applied to each indicator sequence $x(n)$ and a new sequence of complex numbers, called $f(x)$ is obtained:

$$f(x) = \sum_{n=0}^{N-1} x(n) e^{-j\pi n/N}, k = 0,1,2,..N-1 \quad (1)$$

It is easier to work with sequence Power Spectrum, rather than original discrete Fourier Transform. The power spectrum Se[k] for frequencies k= 0, 1, 2, . . . ,N-1 is defined as,

$$Se[k] = \left| f(x) \right|^2 \quad (2)$$

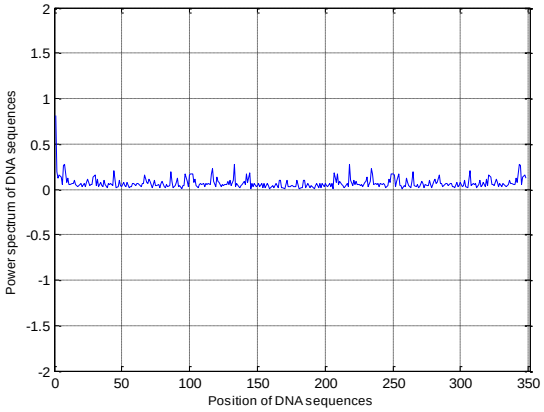


Figure 1. Signal of a DNA sequence using Power spectrum

## *B. Approximation of DNA signal*

In this paper, a classifier that classifies the DNA sequences using Fourier Transform, Power Spectrum to process the signal and the application of Beta wavelet networks as a classification model. This classifier solves the classification problems for DNA sequences. Initially, the approach can bring the learning index defined by the 1D wavelet network to develop a compact signature DNA sequences. This signature is formed by the wavelet coefficients and that will be used to match the DNA test sequences with all sequences in the training set. Then, for classification, test DNA sequence is projected onto the wavelet networks of the learning DNA sequences and new coefficients specific to this sequence are calculated (Fig.2). Finally, we compare the coefficients of the learning DNA sequences with the coefficients of the test DNA sequences by computing the Correlation Coefficient. In this step we can apply the principle of hierarchical clustering to classify the characteristics of sequences DNA. to approximate f (x) of a DNA sequence must select the optimal wavelet to obtain signal representation with minimal error rate. To solve the approximation problem we use the library wavelet which contains a family wavelet. This library is called Multi Library Wavelet Neural Network Model (MLWNN). In our approach the second phase is to build the library wavelet and to approximate the function f (x) of a DNA sequence. We intend to construct a several wavelets families library for the network construction. Eatch wavelet has different dilations following different inputs. The library size is very important.

## *C. Learning Wavelet Network using Multi Library Wavelet Neural Network(MLWNN)*

In this section we will show how we can learn a wavelet network using library wavelet.

### *1)* Proposed Learning Algorithm

**Step 1**: Build a library of candidate wavelet to be choose to construct the wavelet network. This wavelet is used as activation function of network. This step includes the following items:

1) Choose the mother wavelet covering all the support of the signal of DNA sequence to analyze.
2) Build a library that contains wavelets of the discret wavelet transform using dyadic sampling.
3) Choose the lowest frequency wavelet of library. This wavelets allow a coarse approximation of the signal of DNA sequence to be analyzed is introduced the first.
4) Set as a stop learning condition an error Emin between the signal f and the output of the network or a number I of wavelet used for the learning or a number j of neuron in the hidden layer of the network.
5) Each time we choose the next wavelet of the library and iterate the following steps:

**Step 2**: Compute the dual basis formed by the activation wavelets of the hidden layer of the network and the new selected wavelet.

**Step 3**: the wavelet is used as an activation function of a new neuron in the hidden layer when it creates a basic orthogonal or bi-orthogonal with the (n-1) activation wavelet of the network; else it will update the (n-1) old weights of network.

**Step 4**: we compute the output of the network by using the wavelet of hidden layers and the weights of connection which are already calculated.

**Step 5**: if the error Emin or the number of wavelets used I or the number of neuron j are reached then it's the end of learning, else another wavelet of the library is choose and we return to step2.

### *2)* Creation of the Library Wavelet

To build the library of wavelets to join our wavelet network, a sampling on a dyadic grid of dilation and translation parameters is proceded.

## *D. The Hierarchical Ascending Classification*

### *1)* Presentation of the Algorithm

This algorithm includes the following steps:

Proc. of the Intl. Conf. on Advances In Bio-Informatics And Environmental Engineering - ICABEE 2014.

ISBN: 978-1-63248-019-4 doi: 10.15224/ 978-1-63248-019-4-12

**Step 1**: Start the input by preparing a list of DNA sequences signatures and the number of classes that wants to obtain. These signatures are the outputs of the approximation 1D using wavelet networks.

**Step 2**: Create an empty matrix (Classes_signature) which has to contain the groups of DNA sequences.

**Step 3:** Starts with as each DNA sequence signature in its own cluster. This procedure starts with n classes (each DNA sequence signature forms a class containing only itself).Classes_signature.Add(newclass(DNA_signature[i])); where i=1…DNA_number.

**Step 4:** Compute the similarity between classes. The covariance matrix is used to measure the similarity between the DNA sequences signatures.

Matsim(i,j)=sim(Classes_signature(i),Classes_signature(j)) ,where i=1…Classes_signature_Length and j=i+1…Classes_signature_Length-1.

**Step 5:** Research minimum similarity.

**Step 6:** Find the two classes s1 and s2 with the minimum similarity to each other.

**Step 7:** Merge the clusters s1 and s2 and replace s1 with the new class. Delete s2 and recalculate all similarities, which have been affected by the merge.

**Step 8:** Repeat step (6) and (7) until the total number of classes become one.

# III. The beta wavelet family

The function beta is defined by $\beta(x)=\beta_{x0,x1,p,q}(x)$[17,18,20], $x_0$ and $x_1$ are real parameters. Where $x_0<x_1$

$$\beta(x,p,q,x_0,x_1)=\begin{cases}\left(\frac{x-x_0}{x_c-x_0}\right)^p\left(\frac{x-x_1}{x_1-x_c}\right)^q & if \quad x\in[x_0,x_1]\\ 0 & \text{otherwise}\end{cases} \qquad (3)$$

Where $x_c=\frac{p\,x_1-x_0}{p+q}$, We have proved in [20,21,22,23] that all derivatives of Beta function ∈ L2(ℜ) and are of class C∞. The general form of the nth derivative of Beta function is:

$$\psi_n(x)=\frac{d^n\beta(x)}{dx^n}=\left[(-1)^n\frac{n!\,p}{(x-x_0)^{n+1}}+\frac{n!\,q}{(x_1-x)^{n+1}}\right]\beta(x)+P_n(x)P_1(x)\beta(x)+$$
$$\sum_{i=1}^{n}C_n^i\left[(-1)^n\frac{(n-i)!\,p}{(x-x_0)^{n+1-i}}+\frac{(n-i)!\,q}{(x_1-x)^{n+1-i}}\right]\times P_1(x)\beta(x) \qquad (4)$$

$$\text{where : } P_1(x)=\frac{P}{x-x_0}-\frac{q}{x_1-x}; \qquad (5)$$

$$P_n(x)=(-1)^n\frac{n!\,p}{(x-x_0)^{n+1}}-\frac{n!\,q}{(x_1-x)^{n+1}} \qquad (6)$$

If p = q, for all n ∈ N and 0 < n < p, the functions $\Psi_n(x) = d^n\beta(x)/dx^n$ are wavelets [21,22,23]. The first, second and third derivatives of Beta wavelet.

# IV. Wavelet Network

The combination of wavelet transform and artificial neuron networks defines the concept of wavelet networks. This network uses the wavelet functions instead of the traditional sigmoid function as the transfer function of each neuron. This type is composed of two layers (input layer and hidden layer). It has the same structure as architecture radial function. The salaries of weighted outputs are using an adder. Each neuron is connected to the other of the following layer. Wavelet network is defined by pondering a set of wavelets dilated and translated from one mother wavelet with weight values to approximate a given signal f.

$$Y=\sum_{i=1}^{N_w}\omega(a,b)\Psi\left(\frac{x-b}{a}\right)+\sum_{k=0}^{N_i}a_k\,x_k \qquad (7)$$

Where y is there the output of the network, ($x_1$, $x_2$,…,$x_{Ni}$ ) is the vector of the input and Nw is number of wavelets.

# V. Experiment and Results

To evaluate the performance of our approach, we have developed different experiments, each consisting of a different subset of test data. The DNA sequences of Bacterias organisms belonging to 4 categories [24] which contains 30 DNA sequences four bacteria data, that is, bacillus-subtilis, aeropyrum-pernix, aquifex-aeolicus and buchnera-sp.The categories of taxonomical Hierarchy was obtained from NCBI Organelle database [23]. From these 1417 data, 709 are used for training and 708 are used for testing. In this section, we present some experimental results of classification of DNA sequences by using the Fourier transform, Power Spectrum and applying the Beta wavelet networks on approximating three 1-D functions. In the phase of learning our system gets ready to distinguish the various classes by means of the examples of learning of DNA sequences. Our system builds a model for every DNA sequence of learning. At the beginning, during the phase of approximation our model tried to decompose the input signal for every sequence and at the end it tried to reconstruct the input signal. The estimation of the performance of this phase we measured by the Mean Square Error (MSE). TAB. II shows that the Mean Square Error(MSE) obtained are low (0.000793441) and the run time increases relatively with size of the DNA sequence. The result shows that the size of DNA sequence increases the time of the training phase. This time is related to the size of a DNA sequence. When the size is equal to 907 the training time equal to 82.805 seconds. To solve the approximation problem we use the library wavelet which contains a family wavelet. Our approach used to decompose the input signal for every sequence of DNA sequences of bacterias and at the end it tried to reconstruct the input signal. The estimation of the performance of this phase we measured by the Mean Square Error (MSE). TAB. III shows that the Mean Square Error (MSE) obtained are low (0.649223) and the run time increases relatively with size of the DNA sequence. The result shows

Proc. of the Intl. Conf. on Advances In Bio-Informatics And Environmental Engineering - ICABEE 2014.

ISBN: 978-1-63248-019-4 doi: 10.15224/ 978-1-63248-019-4-12

that the size of DNA sequence increases the time of the training phase. This time believes according to the size of a DNA sequence. When the size is equal to 907 the training time is equal to 82.805 seconds (TAB.IV).

To solve the approximation problem we use the library wavelet which contains a family wavelet. This library is called Multi Library Wavelet Neural Network Model (MLWNN) [28]. The library contains 6 mother wavelets ( $Beta^1$, $Beta^2$, $Beta^3$, $Mexican^4$ hat, $Polywog^5$ and $Slog^6$)

TAB. V shows the selected mother wavelets and Normalized Root Mean Sequare Error (NRMSE) of test given by equation after 100 trainings iteration for DNA sequence signals for each class.

TABLE VI.SELECTED MOTHER WAVELETS AND NORMALIZED ROOT MEAN SEQUARE ERROR(NRMSE)

| DNA sequence for each Class | Size | $Beta^1$ wavelet | $Beta^2$ wavelet | $Beta^3$ wavelet | $Mexhat^4$ wavelet | $Slog^5$ wavelet | $Polywog^6$ wavelet | NSRMSE |
|---|---|---|---|---|---|---|---|---|
| Bacillus-subtilis (S1) | 697 | 1 | 3 | 2 | 1 | 1 | 2 | 0.774212 |
| Aeropyrum-pernix (S2) | 219 | 0 | 2 | 1 | 2 | 3 | 2 | 0.710603 |
| Aquifex-aeolicus (S3) | 432 | 2 | 2 | 0 | 1 | 0 | 5 | 0.66832 |
| Buchnera-sp (S4) | 907 | 0 | 2 | 1 | 2 | 1 | 4 | 0.649223 |

TABLE VII. MSE OF APPROXIMATION OF THE SIGNAL FOR DNA SEQUENCE.

| DNA sequence for each Class | Size | MSE (Mean Square Error) | Training Time(sec) |
|---|---|---|---|
| Bacillus-subtilis | 697 | 0.0095726 | 55.611 |
| Aeropyrum-pernix | 219 | 0.000793441 | 27.971 |
| Aquifex-aeolicus | 432 | 0.00367709 | 44.335 |
| Buchnera-sp | 907 | 0.0361482 | 82.805 |

TABLE IV. ACCURACY FOR EACH CLASS PROVIDED BY ANN, SVM AND WNN WITH 317 OF DNA SEQUENCE

| Class | Size | Precision obtained using Support Vector Machine (SVM )(%) | Precision obtained using Artificial Neural Network (ANN)(%) | Precision obtained using Wavelet Neural Network (WNN)(%) |
|---|---|---|---|---|
| Bacillus-subtilis | 337 | 98.4 | 92.9 | 98.98 |
| Aeropyrum-pernix | 182 | 92.5 | 90.2 | 98.8 |
| Aquifex-aeolicus | 256 | 96.3 | 80.4 | 97.5 |
| Buchnera-sp | 201 | 89.7 | 41.7 | 87.9 |
| Total (%) | | 94.2 | 76.3 | 95,795 |

TAB. IV shows the result of classification of DNA sequences. The result indicated in this table proves the performance of our approach compared with the other methods to solve the classification problem of DNA sequences. The results show that our method has higher precision (95.795%) than of the SVM and ANN (TAB. IV). S1 signal is reconstructed with a Normalized Root

Proc. of the Intl. Conf. on Advances In Bio-Informatics And Environmental Engineering - ICABEE 2014.

ISBN: 978-1-63248-019-4 doi: 10.15224/ 978-1-63248-019-4-12

Mean Sequare Error (NRMSE) of 0.774212 using 10 wavelets in hidden layer the best regressors for Library Wavelet Neural Network Model (MLWNN) are: 1 wavelet from Beta1, 3 wavelets from Beta2, 2 wavelets from Beta3, 1 wavelet from the Mexican hat mother wavelet, 1 wavelets from Slog5, and 2 wavelet from Polywog6 wavelet (TAB.VIII). The best Normalized Root Mean Sequare Error (NRMSE) of reconstruction for the signal S4 of DNA sequence of bacteria called Buchnera. The NRMSE is equal to 0.649223 using Library Wavelet Neural Network Model (MLWNN). The seleleted mother wavelets are: 0 wavelet from Beta1, 2 wavelets from Beta2, 1 wavelets from Beta3, 2 wavelet from the Mexican hat mother wavelet, 1 wavelets from Slog5, and 4 wavelet from Polywog6 wavelet. The results of reconstruction of the signals S1, S2, S3, and S5 are of almost the same.

The comparison of our approach with other methods for example Support Vector Machine (SVM) and Artificial Neural Network (ANN) shows the performance of our method to classify bacteria in several classes (TAB.IV).

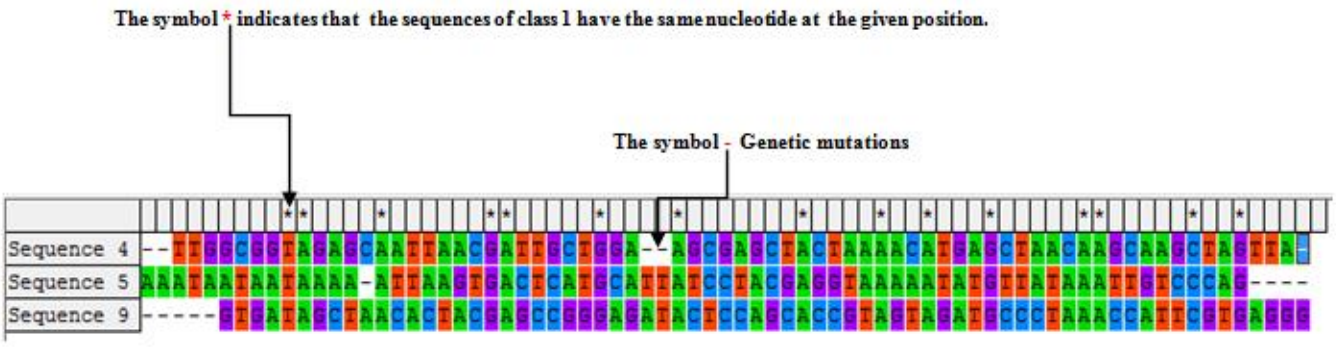


Figure 2. alignement of DNA groups of class 1

Fig.3 shows that the sequences of class DNA have biologic similarity.this resemblance can assure a similar biologic function to the level of the cells of the organisms Bacterias.

The results obtained is very positve and suggests that the proposed techniques can be successfully used in resolution of problem of classification of DNA sequences. This approach allows to calalogue and to characterize any group of organisms Bacterias. from one such alignment of DNA sequences we can estimate the place of every body in the family tree to be alive ( the tree of life ). The similarity of sequences allows to improve in a significant biology way on clusterings of the family of current DNA sequences which are incapable to affect directly the considerable mass of these data. Organization of DNA sequences into groups (clusters) so that similar sequences are in the same group: This similarity (intra-group) between the DNA sequences in a group is very high and the similarity(extra-group)between the groups is very high. The results show high biological relations between these sequences of bacteria. These relations can prove biologically that bacteria derived from a common ancestor and their diversity can be obtained by genetic mutation. This mutation occurs when a DNA gene is damaged or changed in such a way as to alter the genetic message carried by that gene. Genetic mutations change the sequence of bases in a DNA gene in a number of ways:

- Replication abnormal of the DNA sequences.
- Remove (Deletion) parts of the nucleotide.
- Add of a nucleotide in a DNA sequence.

# VI. Conclusion

In this paper, we used a new method of training called Library Wavelet Neural Network Model (MLWNN) which is used to construct Wavelet Neural Network (WNN). The WNN is used to approximate the function f (x) of a DNA sequence signal. Our method depends on Binary codification, Fourier

ISBN: 978-1-63248-019-4 doi: 10.15224/ 978-1-63248-019-4-12

Transform and Power Spectrum to process the DNA sequence signal. Applying this hierarchical classification allows us to group the similar DNA sequences according to certain criteria. In our approach we used the Correlation Coefficient or Pearson Correlation Coefficient which is applied to measure of association between two vectors of DNA sequences signal. Our approach allows us to classify organisms into different categories and classes which have significant biological knowledge and can justify the evolution and identification of unknown organisms. Also they study mutual relations between organisms. In this article we noticed that our approach gives rates of classification (95,795%) better than that given by other approaches proposed by other researchers. Simulation results are demonstrated to validate the generalization ability and efficiency of the proposed Wavelet Neural Network Model.

## *Acknowledgment*

I would like to to thank Research Group on Intelligent Machines (REGIM).

## *References*

[1] M. Ahmad, A. Abdullah and K. Buragga, "A novel optimized approach for gene identification in DNA sequence" ,Asian Network for Scientific Information, Journal of Applied Sciences 11 (5), p.806-814, 2011.

[2] S. Brunak, J. Engelbrecht, and S.Knudsen, "Prediction of human mRNA donor and acceptor sites from the dna sequence", Journal of Molecular Biology, vol. 220, pp. 49-65, Jul. 1991.

[3] V. A. Emanuele II, T. T. Tran, and G. Tong Zhou "A forurier product method for detecting approximate tandom repeats in dna", Scholl of Electrical and Computer Engineering Georgia Institute of Technology, Atlanta, GA 30332-0250 USA, p. 1390 - 1395, Jul. 2005.

[4] R. Sandberg, G. Winberg, C.-I. Bränden,A. Kaske ,I. Ernberg and Coster, "Capturing Whole - Genome characteristics in short sequences using a naive Bayesian classifer", *Genome Res*., Vol. 11, pp. 1404-09, May 2001

[5] F. Zanoguera and M. de Francesco, "Protein classification into domains of life using Markov chain models", *Proceedings of the 2004 IEEE Computational Systems Bioinformatics Conference*, 0-7695-2194-0/04 , 2004..

[6] S.Narasimhan , S.Sen and Konar, "Species identification based on mitochondrial genomes", *ICCR 2005, International Conference of Cognition and Recognition, Mysore, India*, 22-23 Dec. 2005.

[7] K. Vijayan, V. Nair and D.P. Gopinath "Classification of Organisms using Frequency-Chaos Game Representation of Genomic Sequences and ANN", *10th National Conference on Technological Trends (NCTT09)* 6-7 Nov 2009

[8] C. Wu, M. Berry, Y.-S. Fung and J. McLarty, "Neural Networks For Molecular Sequence Classification" , *Proc Int Conf Intell Syst Mol Biol*., p. 429-437, 1993.

[9] S. Kumar and N.Duraipandian ," Artificial Neural Network Based Method for Classification of Gene Expression Data of Human Diseases along with Privacy Preserving", *International Journal of Computers & Technology*, Volume 4 No. 2, March-April, 2013, ISSN 2277-3061.

[10] L. Valim de Freitas et A. P. Barbosa Rodrigues de Freitas , «*L'analyse multivariée dans la gestion, l'ingénierie et les sciences*" , livre édité par, ISBN 978-953-51-0921-1, parution: 9 Janvier 2013 sous licence CC BY 3.0

[11] K. Vijayan, V. Nair and P.Deepa Gopinath "Classification of Organisms using Frequency-Chaos Game Representation of Genomic Sequences and ANN", *10th National Conference on Technological Trends (NCTT09)* 6-7 Nov 2009

[12] S. Kumar and N.Duraipandian ,"An Effective Identification of Species from DNA Sequence: A Classification Technique by Integrating DM and ANN", *(IJACSA) International Journal of Advanced Computer Science and Applications*, Vol. 3, No.8,p.104-114, 2012.

[13] S. B. Arniker and H. K. Kwan, "Advanced Numerical Representation of DNA Sequences", *International Conference on Bioscience, Biochemistry and Bioinformatics IPCBEE* vol.3 1,p.1-5, 2012

[14] W. Bellil, C. Ben Amar and A.M. Alimi, "Beta Wavelet Based Image Compression*", International Conference on Signal, System and Design, SSD03, Tunisia*, vol. 1, pp. 77-82, Mars, 2003

[15] W. Bellil, C. Ben Amar and M.A Alimi, "Synthesis of wavelet filters using wavelet neural networks", *Transactions on Engineering, Computation and Technology*, vol. 13 . ISSN 1305-5313, pp 108-111, 2006.

[16] S. Sitharama Iyengar, E.C. Cho and V.V..Phoha, "*Foundation of Wavelet Nework and Application*", Chapman and Hall/CRC Press, June 2002.

[17] C. Ben Amar, M. Zaied and M. A. Alimi, "*Beta wavelets. Synthesis and application to lossy image compression"*, Journal of Advances in Engineering Software, Elsevier Edition, Vol. 36, N7, pages 459 – 474, 2005.

[18] W. Bellil, C. Ben Amar and M.A Alimi, "*Synthesis of wavelet filters using wavelet neural networks*", Transactions on Engineering, Computation and Technology, vol. 13 . ISSN 1305-5313, pp 108-111, 2006.

[19] C. Ben Amar, W. Bellil, M.A. Alimi, "*Beta Function and its Derivatives: A New Wavelet Family*", Transactions on Systems, Signals and Devices, Vol.1, Number 3, p.275-293, 2005-2006.

[20] W. Bellil, C. Ben Amar C. And A.M. Alimi, "Beta wavelets networks for function approximation", *International Conference on Adaptative and Natural Computing Algorithms, ICANNGA05, Coimbra Portugal, SpringerWien NewYork*, p. 18-21, 2005.

[21] V. V . Nair, L. P. Anto and A. Nair, "Naive Bayesian Classification of unknown sequence fragments based on chaos game representation of mitochondrial genomes", *Communications of SIWN*, vol 7, pp:27-33, May 2009.

[22] Q. Wang, G. M. Garrity, J. M. Tiedje and J. R. Cole ,"Naive Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy", *applied and environmental mictobiologyGY*, Vol. 73, No. 16, p. 5261–5267,Aug. 2007.

[23] www.ncbi.nlm.nih.gov

[24] www.ncbi.nlm.nih.gov/Genbank/genomes/bacteria

Proc. of the Intl. Conf. on Advances In Bio-Informatics And Environmental Engineering - ICABEE 2014.

ISBN: 978-1-63248-019-4 doi: 10.15224/ 978-1-63248-019-4-12

**Abdesselem DAKHLI** continued these academic studies to FSEG Sfax ,Tunisia. He obtained his teacher's cerificate in data processing applied to management in June 2001. He continued his degree of Masters in ISIMG of Gabes, Tunisia in 2008. In 2010, he received his Master degree in Information system in the same Institute. In August 2005 he was assigned to the ISG Gabes, Tunisia. Currently, he is a teacher at ISG. In 2012, now he prepares a Phd thesis of bioinformatic in ENIS. Abdesselem dakhli he also participated in one internationnale conference. His areas of research are: Tomography, Bioinformatics.

**Wajdi BELLIL** received the B.S. degree in Electrical Engineering from the National Engineering School of Sfax (ENIS) in 2000, the M.S. and PhD degrees in Electrical Engineering from the National Engineering School of Sfax (ENIS), in 2003 and 2009, respectively. He spent five years at the ISET Gafsa, Tunisia, as a technologic assistant and researcher before joining the faculty of Science of Gafsa, Tunisia, as Assistant. He joined the Higher Institute of Applied Sciences and Technology, Gafsa University, where he is currently an assistant professor in the Department of computer science. He was a member of the REsearch Group on Intelligent Machines (REGIM). He is a junior member of IEEE.

**Chokri BEN AMAR** received the B.S. degree in Electrical Engineering from the National Engineering School of Sfax (ENIS) in 1989, the M.S. and PhD degrees in Computer Engineering from the National Institute of Applied Sciences in Lyon, France, in 1990 and 1994, respectively. He spent one year at the University of "Haute Savoie" (France) as a teaching assistant and researcher before joining the higher School of Sciences and Techniques of Tunis as Assistant Professor in 1995. In 1999, he joined the Sfax University (USS), where he is currently a professor in the Department of Electrical Engineering of the National Engineering School of Sfax (ENIS), He is a senior member of IEEE, and the chair of the IEEE SPS Tunisia Chapter since 2009. He was the chair of the IEEE NGNS'2011 (IEEE Third International Conference on Next Generation Networks and Services) and the Workshop on Intelligent Machines.