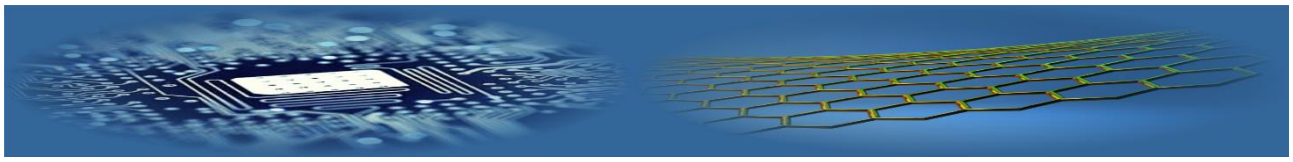

A comparative study between artificial immune system and incremental neural network for digits handwritten recognition

H Khelil[1*], A Benyettou[1], A Kacem[2]

[1]*Laboratoire SIMPA, Université des Sciences et de la Technologie d'Oran USTO-MB, Oran, Algérie*

[2]*Laboratoire LATICE&GE, Université de Tunis, Tunisie*

**Corresponding author's e-mail: hibakhelil@yahoo.fr*

Received 15 August 2016, www.cmnt.lv

Abstract

The artificial immune systems and the incremental neural networks are among novel paradigms used in artificial intelligence and pattern recognition. In this paper, we use MNIST database in order to compare these two approaches and to extract advantages and disadvantages of each one. This work is an introduction to improve the artificial immune system using principle of the incremental neural network.

Keywords:

Artificial Immune Systems,
Incremental Neural Network,
CLONCLAS,
I2GNG,
Digit Handwritten Recognition

1 Introduction

The digit handwritten recognition is an active topic in OCR applications and pattern recognition research [1- 2]. In OCR applications, digit recognition is dealt with in postal mail sorting, bank check processing, form data entry, etc. For these applications, the performance (accuracy and speed) of digit recognition is crucial to the overall performance. While in pattern classification and machine learning communities, the problem of handwritten digit recognition is a good example to test the classification performance.

Different classifiers were tested for digit handwriting recognition as linear classifiers, K-nearest neighbors, Boosted Stumps, Non-Linear Classifiers, SVM, neural nets, etc. [3]. Statistical techniques [4] and neural networks [5 - 6] have been widely used for classification due to the implementation efficiency. Structural [7] and stochastical [4] were also experimented on well-known databases so as to be compared with previous results.

The bio-inspired methods are one of the most used in artificial intelligence and pattern recognition. The artificial immune system and incremental neural network are examples of bio-inspired methods.

In this paper, we present CLONCLAS and I2GNG algorithms for digit handwritten recognition, we compare the obtained results and we give the advantages and disadvantages of each method.

2 MNIST database

To test our systems, we have carried out several experiments on the MNIST benchmark database. MNIST contains material for training and testing of handwritten digit recognition. It includes digits 0 to 9 distributed into training and test sets. The training set contains 60000 examples and the test set is composed of 10000 examples. MNIST data base is created from two bases SD-3 and SD-1. The SD-3 was collected among Census Bureau Employees and SD-1 was collected among High-School students. All pictures were normalized to images of 20X20 pixels, ready to be used in the training system.

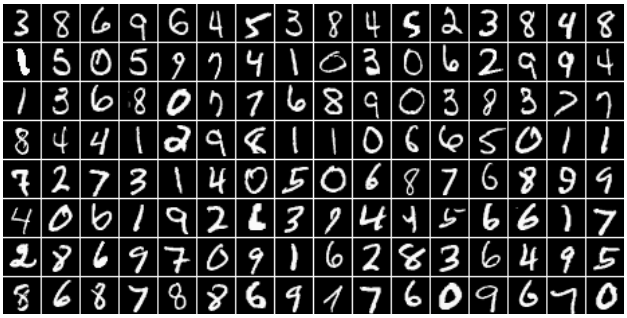

FIGURE 1 some examples of MNIST database

3 Natural immune system

As explained in [1], the immune system of vertebrates is composed of a great variety of molecules, cells, and organs spread all over the body. There is no central organ controlling the functioning of the immune system, and there are several elements in transit and in different compartments performing complementary roles. The main task of the immune system is to survey the organism in the search for malfunctioning cells from their own body (e.g., cancer and tumour cells), and foreign disease causing elements (e.g., viruses and bacteria). Every element that can be recognised by the immune system is called an antigen (Ag). The cells that originally belong to our body and are harmless to its functioning are termed self (or self antigens), while the disease causing elements are named non-self (or non-self antigens). The immune system, thus, has to be capable of distinguishing between what is self from what is non-self; a

process called self/non-self discrimination, and performed basically through pattern recognition events.

The Natural Immune System (NIS) is a distributed novel-pattern detection system with several functional components positioned in strategic locations throughout the body. Immune system regulates mechanism of the body by means of innate and adaptive immune responses (see Fig. 2).

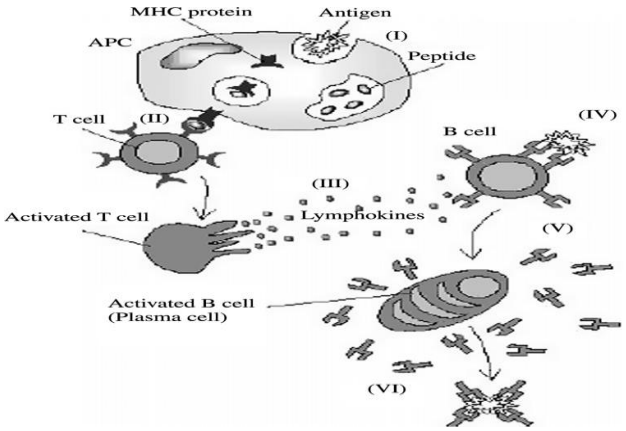

FIGURE 2 Natural Immune

The adaptive immune response is more important for us because it contains metaphors like recognition, memory acquisition, diversity, self-regulation, etc. The main architects of adaptive immune response are lymphocytes, which are of two classes as T and B lymphocytes (cells), each having its own function. Especially B cells have a great importance because of their secreted antibodies (Abs) that take very critical roles in adaptive immune response [8 - 9].

To build the Artificial Immune System, different models were proposed. B cell modeling is the most encountered representation type, among these, CLONCLAS representation. In the next section, we will briefly present the CLONCLAS algorithm and we give results of its application to digit handwritten recognition.

4 Artificial immune system for pattern recognition

As Neural Networks, the Artificial Immune System was inspired from the biological phenomenon. It simulates the Natural Immune System of the living body.

This algorithm typically exploits the features of the innate immune system in terms of learning and memorization as a method for solving problem [10].

Natural Immune Systems have several interesting properties such as the distinction between self and non-self, detection of the alteration, memorization, adaptability, and resource management [11].

In this way, the Artificial Immune System is used in several disciplines as intrusion detection in computer network [12], the illnesses diagnostic [13], handwriting recognition [14 - 18], etc.

We present below the different types of Artificial Immune Systems, their applications and we explore CLONCLAS algorithm for pattern recognition.

4.1 DIFFERENT TYPES OF ARTIFICIAL IMMUNE SYSTEM

4.1.1 Negative Selection

The Negative Selection is used for self and non-self detection. The fundamental principle is to from a set of detectors in order to identifying if a set of channels (self set) has undergone a change or no [19].

4.1.2 Clonal Selection

The Clonal Selection uses an artificial storage mechanism of immune systems. The developed algorithms are usually dedicated to the optimization or research [19- 22].

4.1.3 Artificial Immune Networks (AINET)

It is an algorithm that combines the theory of immune networks and clonal selection. This algorithm is proposed by De Castro and Von Zuben, in 2001 [23]. It uses the notion of internal image to represent groupings of data in a network. AINET can be used to data compression.

In this paper we interest to the classification problem, so we choose CLONCLAS as algorithm of clonal selection.

In the next section, we will explain the principle of this method and give the training and test steps.

4.2 CLONCLAS

In clonal selection, each antibody represents a solution to the problem. Each solution is evaluated by an affinity function. The solution opted is the one with the greatest affinity.

CLONALG is one of the clonal selection algorithms. Proposed by De Castro and Von Zuben in 2002 [20, 26]. This algorithm is designed for pattern recognition. It is very limited because it accepts just one training example per class. This limit is a disadvantage for this method.

To overcome the drawbacks of CLONALG, White and Garrett propose, in 2003 [21], a new algorithm called CLONCLAS; this algorithm is also classified in clonal selection category. It is adapted to solve the classification problem [10, 24].

4.3 CLONCLAS TRAINING AND TEST ALGORITHMS

CLONCLAS is one of the most used algorithms in artificial immune system. CLONCLAS can be used to solve several problems as optimization and pattern recognition.

In this session, we will give the CLONCLAS algorithm for pattern recognition. The objective of CLONCLAS is to train the input data in order to determine the class of the unknown forms later. The data training are named antigens, and CLONCLAS must use these antigenic forms in order to generate antibodies (named also memory cells). These antibodies are used in the classification step. Each class has one antibody, which is a solution of each class [17, 25].

In the test algorithm, each unknown form is represented to antibodies and this unknown data will be affected to the class of the nearest antibody.

We give then the training and test algorithms of CLONCLAS [10].

Algorithm 1: CLONCLAS training algorithm
Require: Training data
Ensure: set of *Abm*

1- Load the training data in Ag
2- Generate randomly the initial population of antibodies Ab, this population will be devised into two sets: memory cells Abm and a population reservoir Abr
3- Select one antigen Ag_i from Ag
4- For G generation do
4.1- For each element from Ab, calculate its affinity with the antigenic form Ag_i
4.2- Select n antibodies which have the best affinity values. Clone these antibodies and put it in the set C_i
4.3- Mutate all elements of C_i and put these clones in C_i*
4.4- Calculate the affinity between the antigenic form Agi and all the elements of C_i*
4.5- Select the candidate clone which has the best affinity value
4.6- If the affinity of this candidate clone is better than the affinity of Abm then replace Abm
4.7- Replace antibodies of Abr by the best clones of C_i*
4.8- Replace d antibodies of Abr by antibodies generated randomly.
4.9- Clear clones in C_i and C_i*
4.10- Return to 4 if G is not completed
5- Return to 3 if there are more antigenic forms.

Once the training program was finished, we will obtain the memory cells (*Abm*) of antibodies (*Ab*). These cells will be used in classification algorithm in order to classify the unknown antigenic form.

We give the classification steps as following:

Algorithm 2: CLONCLAS test algorithm
Require: Set of Abm for all classes.
The Unknown antigenic form Ag.
Ensure: Class of the Unknown Ag

1- Calculate the affinity values between all memory cells Abm and the unknown form Ag
2- Affect Ag to the same class of the memory cell which gives the best affinity value.
3- Return to step 1 if there is more antigenic form

As given in the algorithm 2, the unknown antigenic form will be assigned to the class of the nearest memory cell.

5 CLONCLAS for digit handwriting recognition

This section addresses the steps required to recognize digits using CLONCLAS.

5.1 CODING AND PREPARING DATABASE

In this section we will explain how applying CONCLAS for digit handwriting recognition. The first step is to prepare the dataset, so we binarize pictures. All pictures will be coded to 0 for white and 1 for black.

Antigens and antibodies must have the same structure. All the representative vectors have size of 400 elements. All antigens represent the training data and the antibodies were generated following the training algorithm in order to be operated for the classification.

5.2 CLONCLAS APPLICATION RESULTS ON MNIST DATABASE

In this session, we will give results of application of CLONCLAS to MNIST database.

First, we note that each digit has 6000 training samples and 1000 test samples. As given in the following table:

TABLE 1 Data Distribution

Data set	Training set (antigenic forms)	Test set (the new antigenic forms)
0	6000	1000
1	6000	1000
2	6000	1000
3	6000	1000
4	6000	1000
5	6000	1000
6	6000	1000
7	6000	1000
8	6000	1000
9	6000	1000

We run CLONCLAS training program for 10 iterations. The obtained results are 85.68% for precision rate and 84.76% for recall rate. These results can be considered important because they were obtained after only 10 iterations and using only 10 generated antibodies. Tables 2, 3 and 4 give the recall, precision and F-Measure rates (F-measure = 2 * ((precision_rate * recall_rate) / (precision_rate + recall_rate))) of MNIST database recognition using CLONCLAS algorithm:

TABLE 2 The recall rates

Data set	Recall rates
0	93.20
1	97.80
2	82.20
3	86.50
4	84.80
5	86
6	85.80
7	87.10
8	69
9	75.20

TABLE 3 The precision rates

Data set	Precision rates
0	87.26
1	71.85
2	91.43
3	90.38
4	83.13
5	79.85
6	92.15
7	79.90
8	92.12
9	88.67

TABLE 4 F-Measures rates

Data set	Precision rates
0	90,13
1	82,84
2	86,57
3	88,4
4	83,95
5	82,81
6	88,86
7	83,34
8	78,9
9	81,38

According to the recall rates, CLONCLAS can recognize better ‘0’ and ‘1’ than the other digits, but it confused the classification of the other digits with ‘0’ and ‘1’ which is confirmed by the decrease of the precision rates (87.26% and 71.85%).

We observe also that the precision rates of ‘2’, ‘3’, ‘6’ and ‘8’ are better comparing with ‘0’, ‘1’, ‘4’, ‘5’, ‘7’ and ‘9’, this can be explained that CLONCLAS misclassify

more the new antigenic forms in '0', '1', '4', '5', '7' and '9'.

We note that these results are obtained after 10 iterations only. Table 5 represents a comparison between the previous results and those obtained after 100 iterations.

TABLE 5 Influence of iteration number on recall and precision rates

Number of iterations	Recall rates	Precision rates	F-Mesure rates
10 iterations	84.76	85.68	85,21
100 iterations	87.88	90.24	89,04

According to these results, the recall, precision and F-measure rates have been respectively increased by 3.12% and 4.56% and 3,83%. Thus, the AIS give results better than the linear classifier.

We note by increasing the number of iterations the recognition rates was increased.

6 Incremental neural networks

The objective of this paper is to compare two bio-inspired systems: the artificial immune system (CLONCLAS) and the incremental neural network (IGNG, I2GNG). In this session we give the training and test algorithms of IGNG algorithm and we compare after the obtained results of session V with those obtained by Hatem in [27]

6.1 IGNG TRAINING AND TEST ALGORITHMS

GNG, IGNG and I2GNG are incremental neural networks which can start without any neuron and learn from incoming data. These dynamic neurons can add and/or remove neurons depending on the evolution of the data over the time. They have been successfully applied in incremental classification tasks as classification of images or synthetic data [27].

I2GNG is an improvement of IGNG and GNG. It gives better results than IGNG and GNG.

The training algorithm of IGNG is given as following:

Algorithm 3: IGNG training algorithm
Require: Training data
Ensure: IGNG neural network

Initialization of parameters: S, ε_b, ε_n, α_{edge}, $\alpha_{neuronmax}$.
While (a stopping criteria is not found) do
begin
Take an input signal E and find its nearest neuron n_1.
If (the network is empty or $d(E, n_1)>S$) then create a new embryon neuron $w_{new} = E$
else begin
Find the second nearest neuron n_2
if (There is only one neuron in the network
or $d(E, n_2)>S$) then
begin
create a new embryon neuron $W_{new} = E$
create a new edge between n_1 and E
end
else begin
increment the age of all edges coming from n_1.
n_1 += ε_b . d(E,n_1)
n_m += ε_n .d(E, n_m), (m are the neighbors of n_1)
if (n_1 and n_2 are connected) then $age_{n1 \to n2}$ =0
else begin create an edge between n_1 and n_2 end
Increment the age of all the neighbors neurons of n_1
for each embryon neuron do
begin
if (age(neuron)>$\alpha_{neuronmax}$) then
embryon neuron becomes mature
end
end
for each edge do
begin
if (age(edge) > α_{edge}) then remove edge
end
done
end

Once learning has finished, we obtain our neuronal network; where each class has its representative neurons. It is not necessary to have the same number of neurons in all classes, but neurons are generated according the distribution of input data. In this stage, we can apply the test algorithm to classify the unknown data.

Algorithm 4: IGNG test algorithm
Require: IGNG neural network, test data
Ensure: test data classified

for data = 1 to nb_test_data do
- Select the nearest neuron from the neural network.
- The new data will be assigned to the same class of this nearest neuron
end for

According the IGNG test algorithm, we classify the unknown data to the class of the nearest neuron.

6.2 IGNG APPLICATION RESULTS ON MNIST DATABASE

The MNIST handwritten digit database was used by multiple authors. Prudent and Ennaji [28] applied the GNG and IGNG for the recognition of these images. The images are not used in their raw state, but a four-level pyramid decomposition was performed on these images before obtaining representation vectors.

The results obtained are in the range of 91.44% to the GNG and 91.71% for IGNG. The foundation for learning and testing has been partially used for these tests (2626 examples for training and 2619 examples for testing). The obtained results are given in table 6:

TABLE 6 Recognition rates of the three networks in the portion of the NIST database [28]

Number of iterations	Cycles	Offline
LVQ	50	89.65%
GNG	50	91.44%
IGNG	10	91.71%

We observe here that IGNG gives better results than GNG and LVQ.

I2GNG is an improvement of IGNG. It has the same training steps as IGNG, the only difference is about parameter S. the threshold S is adapted to data volume.

$$S = m + \alpha\,\sigma, \quad (1)$$

where m is the average of distances between all data and neuron, and σ is the standard deviation of distances.

Hatem [27, 29] proposes to use MNIST database entire (60000 training samples gradually). In each experience, he use n*10000 samples for I2GNG training algorithm (n varies from 1 to 6). A single iteration is sufficient to apply the I2GNG. The test set is composed of 10000 samples as given in [3]. The recognition rates are given in table 7:

TABLE 7 Recognition rates obtained after applying one iteration of I2GNG to MNIST database.

Training samples	Recognition rates
10000	88.45%
20000	91.02%
30000	92.58%
40000	93.66%
50000	94.06%
60000	94.29%

The advantages of IGNG and I2GNG that they are fastest. We do not need to reiterate training for new data integrated. We need to store the only neurons in memory in order to be used next time. IGNG and I2GNG present a lot of advantages as well as the recognition rates and training time.

7 Comparison between CLONCLAS and I2GNG

In this session, we will study the advantages and disadvantages of each method and we will propose after an improvement of CLONCLAS. We can note several points as following:

- I2GNG training is faster than CLONCLAS, because it is not an obligation to reiterate training steps, a single iteration is sufficient for training.
- If there are more training data, I2GNG can continue training using the memorized neurons but for CLONCLAS we must restart training from the beginning using all the training data.
- In CLONCLAS we use only the affinity between antigens and antibodies in order to select the best antibody, but in I2GNG we use the threshold Si and the age of neurons in order to add or remove neurons. This threshold Si is in function of all data and neurons so IGNG takes account the distribution of all data on the contrary of CLONCLAS the affinity value is in function of the specific data and antibody.
- At the end of CLONCLAS training we obtain antibodies (memory cells) for each class. One representative is not sufficient for a large data training set. It is important here to think about a solution to create more antibodies for each class.

From this comparison we can optimize the CLONCLAS algorithm. These notes can open the new perspectives for next works.

8 Conclusions

In this paper, we have used two bio-inspired methods for digit handwritten recognition, so we have used the artificial immune system and the incremental neural network. We give training and test steps for each algorithm and we have compared results of each method.

We observe also that I2GNG can give satisfied results comparing with the artificial immune system. We note that, I2GNG improves the recognition rate on the rage of 6.41% comparing with CLONCLAS. We have study also the advantages and disadvantages of CLONCLAS and I2GNG; and we think in the next works to propose an amelioration of CLONCLAS using the incremental approaches.

References

[1] De Castro L N, Timmis J 2002 *Artificial Immune Systems: A Novel Paradigm to Pattern Recognition Artificial Neural Networks in Pattern Recognition* J. M. Corchado, L. Alonso, and C. Fyfe (eds.), SOCO-2002, University of Paisley, UK, p 67

[2] LeCun Y 1995 Comparison of learning algorithms for handwritten digit recognition F. Fogelman-Soulié, P. Gallinari (Eds.) *Proceedings of the International Conference on Artificial Neural Networks, Nanterre, France*, p 53

[3] http://yann.lecun.com/exdb/mnist/

[4] Ben Amara N 1999 *Utilisation des modèles de Markov cachés planaires en reconnaissance de l'écriture arabe imprimée* PHD thesis, University of sciences, techniques and medicine of Tunis

[5] Ben Cheikh I, Kacem A 2007 Neural Network for the Recognition of Handwritten Tunisian City Names *International Conference on Document Analysis and Recognition ICDAR Curitiba, Brazil*, pp 23

[6] Souici L, Zmirli Z, Sellami M 1997 *Système connexionniste pour la reconnaissance de l'arabe manuscrit scientific and technical conferences* JST FRANCIL Avignon, France, p 383

[7] Anigbogu J 1992 *Reconnaissance de textes imprimés mutifontes* PHD thesis, University of Nancy

[8] Kemal P, Salih G 2006 *Hepatitis disease diagnosis using a new hybrid system based on feature selection (FS) and artificial immune recognition system with fuzzy resource allocation* ELSEVIER, Digital Signal Processing

[9] Emilie P 2006 *Organisation du system immunitaire felin* PhD thesis, National school Lyon, France

[10] Denech A, Meshoul S, Batouch M 2005 *Une approche hybride pour la reconnaissance de forme en utilisant les systèmes immunitaires artificiels* Computer Vision Group, LIRE Laboratory, Mentouri University, Constantine, Algeria

[11] Denech A 2006 *Approches bio inspirées pour la reconnaissance de formes* Master report, Mentouri university, Constantine, Algeria

[12] Jungwon Kim, Peter J B 2001 The Artificial Immune System for Network Intrusion Detection: An Investigation of Clonal Selection with a Negative Selection Operator *Evolutionary Computation* p 1244

[13] Khelil H, Benyettou A, Kacem A 2012 *Novel Method To Choose Training Immune Parameters* ICARIS Italia

[14] Bouafif S F, Snoussi M, Hamrouni K 2005 Segmentation de chèques bancaires arabes *3rd international conference: Sciences of Electronic, Technologies of Information and Telecommunications SETIT'05, Tunis, Tunisia*

[15] Olivier C Miled H, Romeo K, Lecourtier Y 1996 Segmentation and coding of Arabic Handwritten words *ICPR* **3** 264, Vienna, Austria

[16] Casey R G, Lecolinet E 1995 Strategies in Characater Segmentation: a Survey *3rd International Conference on Document Analysis and Recognition ICDAR* **2** 1028, Montréal

[17] Christian G, Mokdong C 2016 Number Plate Detection with a Multi-Convolutional Neural Network Approach with Optical Character Recognition for Mobile Devices *Journal of Information Processing Systems* **12**(1) 100

[18] Yuefeng C 2010 A handwritten character recognition algorithm based on artificial immune International conference on computer application and system modelling *ICCASM* **12** 273

[19] Esponda F, Forrest S 2003 The crossover closure and partial match detector international conference on artificial immune systems *ICARIS'03* 249, Edinburgh, UK.

[20] De Castro L N, Von Zuben F J 2002 Learning and optimization using the clonal selection principle *IEEE Transactions on Evolutionary Computation*

[21] White J A, Garrett S M 2003 Improved pattern recognition with artificial clonal selection international conference on artificial immune systems *ICARIS'03* Edinburgh, UK

[22] Bentley K 2002 A model of Gene Library Evolution in the Dynamic Clonal Selection Algorithm international conference on artificial immune systems *ICARIS'02* Canterbury, UK

[23] De Castro L N, Von Zuben F J 2001 Immune and Neural Network Models: Theoretical and Empirical Comparisons *International Journal of Computational Intelligence and Applications (IJCIA)* **1**(3) 239

[24] Ben Amara N, Bouslama F 2003 Classification of Arabic script using multiple sources of information: State of the art and perspectives *International Journal on Document Analysis and Recognition IJDAR'03* **5**(3) 195

[25] Brownlee J 2005 *Immunos-81 the Misunderstood Artificial Immune System Master report (Information Technology)* University of Swinburne of Technologies

[26] Khelil H, Benyettou A, Belaîd A, Kacem A 2012 Le système immunitaire artificiel pour la reconnaissance de noms propres

manuscrits arabes *15th International Conference on electronic document - CIDE, Tunis, Tunisia*
[27] Hatem H, Belaîd Y, Belaîd A, Bidyut B C 2008 Incremental classification of invoice documents *19th International Conference on Pattern Recognition (ICPR 2008), Tampa, Florida, USA*
[28] Prudent Y, Ennaji A 2005 A new learning algorithm for incremental self-organizing maps, *ESANN European symposium on artificial neural networks Bruges (Belgium)*
[29] Hatem H 2008 *Application du raisonnement à partir de cas à l'analyse de documents administratifs* PHD thesis, university of Nancy 2

AUTHORS

Khelil Hiba, 13.07.1982, Oran - Algeria

Current position, grades: Ph. D. Computer Sciences student at the USTO-MB University (Université des Sciences et de la Technologie d'Oran- Mohamed BOUDIAF). Assistant Lecturer at USTO-MB, Algeria; Scientific interest: Handwritten recognition, Bio-inspired algorithms, Artificial intelligence

Benyettou Abdelkader, 1959, Oran - Algeria

Current position, grades: Professor in department of Computer Sciences at the USTO-MB University (Université des Sciences et de la Technologie d'Oran- Mohamed BOUDIAF). Assistant Lecturer at USTO-MB, Algeria; director of the Signal-Image-Image SIMPA Laboratory, Department of computer sciences, Faculty of Mathematics and Computer Sciences, since 2002; PhD in electrical engineering from Metz University and Nancy-1 University, in 1993

Kacem Afef, 1971, Tunis – Tunisia

Current position, grades: Senior Leturer at University of Tunis, Tunisia; Head of LaTICE Laboratory, University of Tunis; PhD in computer science from university of Tunis, in 2001