

Towards a Method of Automatic Design of Multi-Classifiers System Based Combination

Gasmi I., and Merouani H.

Abstract— At the beginning of research in pattern recognition, the advantage of the developers was to conceive systems of recognition, and to refine them successively in order to improve their performances. Recently, a new approach proposes that we improve the performances by combining results of various less powerful systems, but complementary to lead to better decisions. This new approach known as *multi_algorithms*, *multi_experts*, *multi_classifiers*, fusion or cooperation of classifiers, becomes increasingly interesting sight the improvement which it can make to pattern recognition system.

It is within this framework that our work is placed which presents a method of automatic design of *multi_classifiers* system for the recognition of hepatic pathologies.

At the beginning, a large set of various classifiers is used, all trained by the same training set. Among these classifiers we seek the subset that has the independent behaviors and at the same time providing a great performance. For this purpose, we use a technique based on a genetic algorithm (GA) to select the most effective set of classifiers.

Moreover, we propose to use a function of combination that takes in charge the behavior of the classifiers during the training phase.

Keywords—Pattern recognition, image processing, neural network, k-nearest neighbors classifier, Bayes method, genetic algorithm, combination of classifiers.

I. INTRODUCTION

THE pattern recognition is a branch of the artificial vision which consists to realize operations of automatic perception analogous to the visual human perception system. The field of pattern recognition represents the first stage of a long process of comprehension of our universe, within the general framework of a human-machine communication. This field is interested in the design and the realization of systems (hardware and software) that are able to perceive and to interpret signals collected in the physical world.

Pattern recognition has known a remarkable progression [Plamondon, 2000]. Several methods were developed to satisfy the needs for the existing systems in various applications such as text recognition, the diagnosis assistance, etc....

Gasmi I., Merouani H. are with Equipe SRF, laboratoire LRI, Département d'informatique, Faculté des sciences de l'ingénieur, Université Badji Mokhtar B.P.12, 23000, Annaba, Algérie (e-mail: gasmi_ibtissem@yahoo.fr, hayet_merouani@yahoo.fr).

The pattern recognition systems treat images resulting from the physical world by various sensors (scanner, camera etc). According to the acquisition means and conditions, an image can suffer from the noise; the system must thus minimize it by pretreating the image, then extracting from it the characteristics or the primitives that can be the various types to build the vector of entry of the recognition module [Madhvanath, 2001].

Among the most used methods of classification, connexionists methods are also called neural networks [Towell, 1994]. Their general principle is to find a function which associates a set of starting forms to a set of arrival classes. The performances of a network depend on the variety of the entries which are presented to it during the training phase as well as the pertinence and the quality of these entries.

There is a large variety of characteristics that can be extracted from the same image. The fact of trying to present them all in entry of a network strongly risks increasing the size of the network and to complicate and delay the training of it. As these characteristics are very important to differentiate the images to be analyzed, it would be interesting to benefit from the variety of information which bring.

A solution to this problem is to use several networks having different architectures and train each one in a set of homogeneous features. Recently, the concept of combining several answers, resulting from various classifiers is proposed as a new direction because, used individually, the latter present limited performances [Giacinto, 1998], [Montoliu, 1995], [Gasmi, 2004], [Roli, 2001], [Ruta, 2000], [XU, 1992], [Zouari, 2002].

It is within the framework of classifiers combination that our work is focused. The main objective is to conceive and carry out a method for the automatic design of a multi classifiers system based combination for the recognition of hepatic pathologies.

Our work consists in seeking all the possible classifiers combinations and combining their results by the Bayes method. The results of combination will be then evaluated and the set providing the best result will be selected. The principal difficulty in such approach is the great number of possible classifiers combinations. Consequently, a more effective method is necessary which is able to exclude aside the space of solutions the regions containing the uninteresting solutions. To carry out such research we will use the *genetic algorithm* [Spalanzani, 1999].

The reason to use such approach is the design of a method which does not count on the vague concept of *dependence* or *independence* of classifiers.

The article will present as follows: in section 2 and 3 we respectively recall the problems of the combination and the definition of a multi classifiers system. In section 4 we present the architecture of our system, and the description of the modules which compose it, in section 5 we present the genetic algorithm in the context of our work. Finally we finish by a conclusion.

II. PARALLEL COMBINATION PROBLEM

The difficulty of the parallel classifiers combination can be formed as follow: given a set of K classifiers that have independent answers about the same form to be recognized, how to have a single final response from K provided results (figure1). This problem of parallel combination initially requires remembering the definition of *classifier* within the framework of the combination.

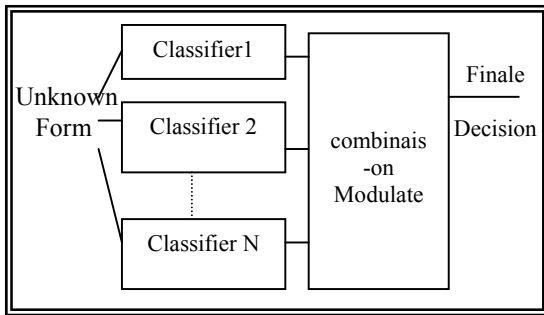


Fig. 1 Diagram of Parallel Classifiers combination

A. Definition and Types of Classifier Output

A classifier is a tool of recognition which receives an entry (form, document, word..etc.), and provides in exit information concerning this unknown form. For a given entry (x), the classifier (e) assigns a class to it (C_i, where i=1...m), among (m) existing classes.

According to the information level provided by the classifier, we can distinguish four categories of exits [Laurent, 1999]:

- *Class Type Output*

It is the most general type but it provides less information. In this case the classifier gives only the proposal class of the entry to be recognized without any other information $e_j(x) = C_i$ ($i \in \{1 \dots m\}$).

- *Set Type Output*

The classifier gives its response as set of candidates classes without specifying its preferences $e_j(x) = \{C_i / i \leq m\}$.

- *Row Type Output*

This type of exit reflects the order of preference of proposals provided by the classifier, that results from the attribution of a row for each class; More the latter is probable minus the row is high.

The list of the proposals can contain all the possible classes or only the best classified. $e_j(x) = [r_1^j, r_2^j, \dots, r_m^j]$ where r_i^j is the row attributed to the class (i) by the classifier (j).

- *Measures Type Output*

This exit is riche in information since the classifier in this case associates for each class a confidence measurement that can be, for example, a probability.

$e_j(x) = [M_1^j, M_2^j, \dots, M_m^j]$ where M_i^j is the measurement attributed to the class (i) by the classifier (j).

B. Performances Measurements of Classifier

For a given entry, a classifier can generate the following answers:

- **Rejection:** To indicate that the classifier could not identify the entry.

$$\text{Rejection rate} = \frac{\text{Number of rejected forms}}{\text{Total numbers of forms}} \quad (1)$$

- **Recognition:** In this case, it identifies the entry well, and it attributes its suitable class to him.

$$\text{Rate of recognition} = \frac{\text{Number of recognized forms}}{\text{Total numbers of forms}} \quad (2)$$

- **Substitution:** The classifier attributes another class to the entry.

$$\text{Rate of substitution} = \frac{\text{Number of badly recognized forms}}{\text{Total numbers of forms}} \quad (3)$$

III. CONSTRUCTION OF MULTI-CLASSIFIERS SYSTEM

Multi-classifiers system (MCS) consists of a set of various classifiers and a decision function to combine their exits. The description of an MCS follows two following phases:

- Generate a set of complementary classifiers that can be combined to get an optimal solution.

- Define the combination function to get a final decision.

The difficulty to choose classifiers, pushed researchers to develop methods that help designers to carry out their choice. Among these methods, that called over produce and choose paradigm. The main idea, of this method, is to produce an initial large set of classifiers, then, to select a subset which is judged most valid to lead optimal performances. For that, the process follows two cycles:

- To construct the set of starting classifiers (over production).
- To choose the most interesting subset.

The choice of the decision function plays an important role, in the design of an MCS. The decision function can be conceived as a *combination function*, consequently the exit of the MCS reflects the decision of all the set by using for example the majority vote, the weighted sum, etc; or as *dynamic selection function* of a classifier, in this case it is necessary to have, at least, one classifier which can be able to correctly classify the shape entry.

Although, the mechanism of selection will be more effective when classifiers *are specialized*, hence the facility of calculation of the selection conditions assigning each entry shape to the most suitable classifier. On the other hand, the mechanism of combination will be more effective when classifiers appear as independent behaviors.

IV. SYSTEMS'S STRUCTURE

In order to conceive an automatic design method of multi classifiers system based combination for the recognition of hepatic pathologies. Our system begins with a large set of different classifiers (see figure2).

We must think of each possible set of classifiers as being an *individual*, its characteristics reflect the presence or the absence of classifiers. For each individual the fitness function reflects the performance of the combination, using Bayes method. Then, a new population will be generated from the previous one according to rules of crossing and mutation. The process stops, when the new generation has a fitness value equal to 100 or when it has a fitness value very close to parents or because the maximum number of generation is reached. Each set of classifiers is represented by a binary string which size is equal to the number of available classifiers. The binary string is the *chromosome*; each byte represents a *gene* in the genetic model. The role of each byte is to indicate the presence (byte equal to 1) or the absence (bit equal to 0) of a particular classifier.

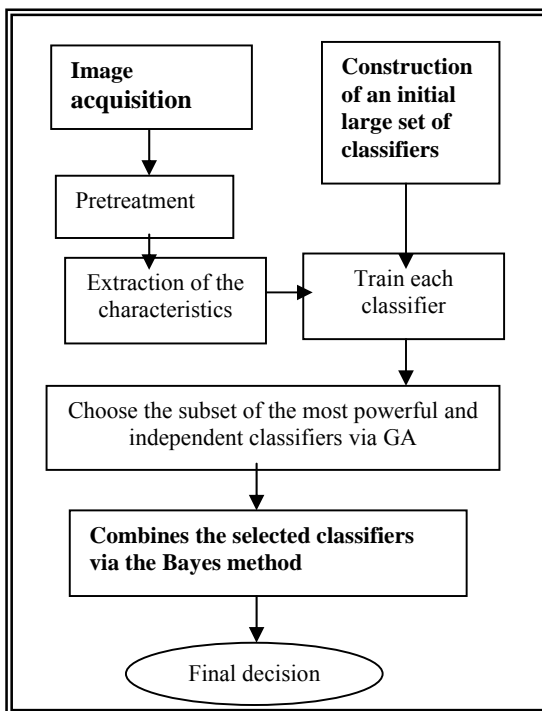


Fig. 2 General diagram of the suggested system

The final answer of the system will be given by the combination, via the Bayes method, of the classifiers constituting the subset having the greatest adaptation function. The answer will be:

- A hepatic pathology accompanies the suitable treatment.
- A rejection.
- An ambiguity (the liver belongs to several class).

A. Image Acquisition

It is the digitalized scanographic image of an abdominal transverse section.

The system can read the scanographic image of the liver from a file on the disc or recuperate it directly from a scanner then stores it in a file.

B. Pretreatment

It is necessary to consider the risk of the noise and the errors introduced at the time of acquisition of image (possibility of data error on the level of the analog-to-digital converter).

This noise is likely to degrade the performances of the recognition; the phase of pretreatment has like objective to prepare the data for the recognition i.e. to select the useful information for the application.

C. Features Extraction

In this phase we extract the characteristics which are sufficient to diagnose the disease of the liver among a group. The choice of these characteristics is in conformity with the results of radiology research [Djebbar, 2004].

These characteristics represent the components of the characteristic vector which are:

- Size.
- The perimeter.
- The acute angle of the liver (the maximum distance between two points A (X0, Y0) and B (X1, Y1) of the liver contour).
- Co-ordinate X0 (coordinate X0).
- Co-ordinate Y0 (ordered Y0).
- Co-ordinate X1 (coordinate X1).
- Co-ordinate Y1 (ordered Y1).
- The homogeneity (a number of regions in the liver).
- Average of gray levels of the liver.

D. Classifiers for Recognition

In this section we describe classifiers used for recognition. We start with an initial large set of classifiers from which we select the most powerful and independent subset. Since we create a set of classifiers, several methods based on the variation of the parameters related to the design of the classifiers are used.

- Various initializations: we create various neural networks using various initial weights and thresholds.
- Various choices of parameters: we create various classifiers of k-nearest neighbors by using various numbers of neighborhoods.
- Various classifiers: we use two types of classifiers trained by the same set of characteristics and the same training set (neural network and k-nearest neighbors classifier).

NEURAL NETWORK

We use the neural networks of the type multilayer perceptron (MLP), of three, four and five layers.

- The number (N) of neurons of the entry layer is equal to the number of used characteristics i.e. nine neurons.
- The number of intermediate neurons is generally contained between N/2 and 2n.
- The neurons of exit represent the 10 most frequent hepatic pathologies.

K-NEAREST NEIGHBORS CLASSIFIER

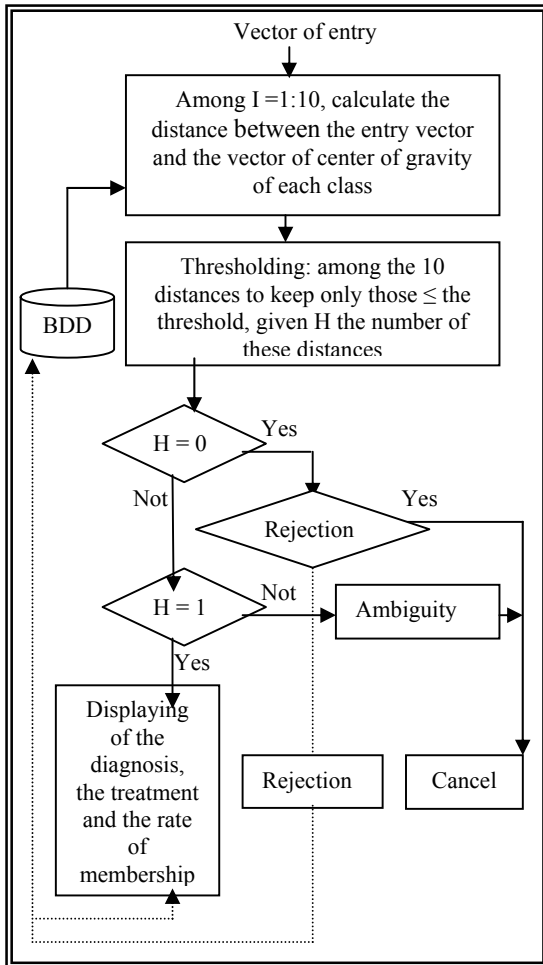


Fig. 3 Flow chart of k-nearest neighbors method

The distance used is the sum of the absolute values also called Minkowski distance of degree 1 or the Hamming distance. In R9 (space of representation) it is equivalent to the other distances (e.g the Euclidean distance,...) but it requires less calculations. Figure 3 shows the various stages of the application of a k-nearest neighbors algorithm [Djebbar, 2004].

After the arrival of a new vector V (liver to be diagnosed), we calculate the M distances Di (Hamming Distance) between this vector (v1, v2,..., v9) and each vector Ci of center of gravity of each class, such as:

$$D_i = \sum_{j=1}^9 |V_i - C_j| \quad (4)$$

i=1..M

When the Di distances are calculated, we compare them with a threshold Th. The threshold Th constitutes the maximum distance which can exist between two vectors of the same liver.

The value of the threshold Th is empirically given by analyzing the Hamming distances calculated (each one) between two scanographic images of the same liver on all the basis of data.

Given H the number of classes that the distance calculated between their vectors and the vector of entry are lower than the threshold Th.

If H = 0, i.e. all the calculated distances Di exceed this threshold then this liver does not resemble any class of reference, therefore it is a case of *rejection* (a new case); in this case there are two choices:

- Cancel the new liver.
- Add the new liver to the base (save its image, its vector, its diagnostic and treatment by the doctor) and to regard it as a new class.

If H ≠ 0, we test:

- If H ≠ 1, i.e. a case of ambiguity (there are several classes correspond to the liver of entry), then *cancel*.
- If H=1, i.e. only one class answers the liver of entry then we obtain the diagnosis and the treatment of this liver thus this last is *recognized*. This answer must be accompanied by a rate by membership. This rate is equal to 100% minus an error rate. This last represents the distance between the vector of entry and the vector of the class which resembles it by the sum of the values of the components characteristic of this last.

We choice to use the multi-layer connexionnists networks of the perceptron type for several reasons. First, these networks have great capacities of classification and showed their aptitudes in pattern recognition. Moreover, compared to the whole of the connexionnists systems, they have the advantage of being based on simple and relatively controllable principles. Contrary to the recurring networks for example, their time of convergence can be relatively fast. It is also possible to detect the moment when the training algorithm is not able to improve the performances, which makes possible the optimization of the training time. The choice to use the k-nearest neighbors algorithm is justified by the fact that in spite of its simplicity, it is a perfect candidate for classifications which must constantly be adjusted or revised and that is due to the absence of training phase.

V. OPTIMIZATION MOTOR

In this section we will define our optimization motor which is the genetic algorithm, his components and his parameters, in the context of our problem.

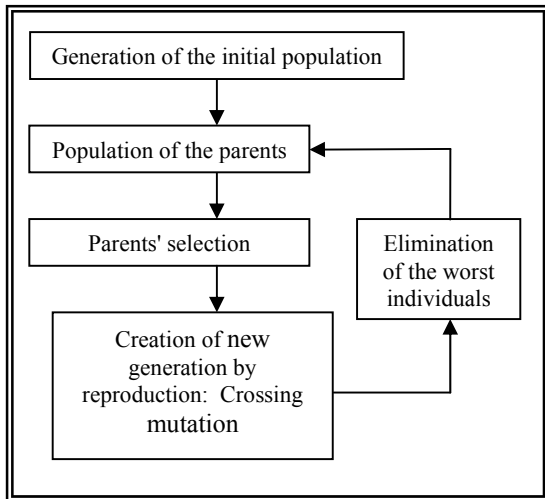


Fig. 4 Cycle of natural evaluation

A. Coding

The solution vector, i.e. the chromosome can be represented by the figure (5). Each set of classifiers is represented by a bit string of size equal to the number of available classifiers (the chromosome is sizeable 10).

Value 1 of a box *i* means that the *i*^{ème} classifier will be taken into account in combination, and value 0 means that the *i*^{ème} classifier will not be taken into account in combination.

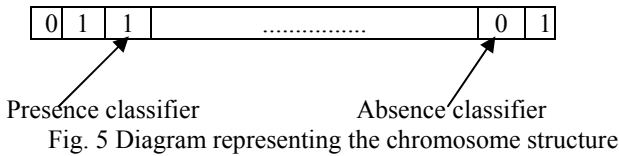


Fig. 5 Diagram representing the chromosome structure

B. Creation of the Initial Population

We chose to create randomly an initial population of N binary chromosomes, by generating N vectors sizeable 10 (two PMC with three layers, two PMC with four layers, two PMC with five layers, four k-nearest neighbors classifier).

C. Evaluation of the Solutions

Each chromosome has an adaptation function *F_i*, measuring its adaptation which in our context is calculated by summing of the exactitude rates and rejection rate. So, the chromosome more adapted will be the one has the maximum value of fitness or the one has the fitness equal to 100.

$$F_i = A_i + \lambda R_i \quad \lambda \in [0,1] \quad (5)$$

Where *A_i* is the exactitude rate of *i*^{ème} set combined by the Bayes method.

R_i is the rejection rate of *i*^{ème} set combined by the Bayes method.

The parameter λ is used to take into account the rejection rate.

D. Reproduction

In this phase, we create in each iteration, a new population, by applying the following genetic operators:

a. Selection

We will use a proportional selection which is based on the principle of “each one has a chance to be selected, according to his effectiveness”. Thus, one will give the chance to any individual to be selected according to his adaptation.

The probability of selection of each individual “*p_i*” is equal to:

$$P_i = F_i / \sum_{j=1}^N F_j \quad (6)$$

Where *N* is a number of individuals in the population.

F_i is the fitness of individual *i*.

The number *O_i* of children generated by one parent is equal to.

$$O_i = N \cdot P_i = F_i / \sum_{n} F_i \quad (7)$$

b. Crossing

We will apply that of the binary mask. Consisting in generating randomly a string bit which have the same length that of the chromosomes, known as the mask. If the box of this mask is to 0, then each two boxes of the same index of the two children E1 and E2 will receive the values of the corresponding boxes of the parents P1 and P2 respectively. And they will receive the values of P2, P1 respectively, if the value of the mask is to 1. This choice we allow to obtain an important variety.

c. Mutation

After a successive application of the crossing, the chromosomes will become almost identical. In order to invoke more variety, and thus to exploit all the space of research, we will apply a standard mutation by making invert one or more bits in chromosomes of the children.

E. Replacement of the New Population

We will proceed to replace the unlike solution. For that, we classify all the solutions (parents and children) according to their adaptations and we keep in our population only the N first individuals, in other words, the greater values of fitness; thus, obtaining a new population made up of the N better individuals. If the number of generations is reached, then the best solution is extracted, and we jump to the next stage; else, a novel iteration of reproduction is made.

F. Stop Condition

Our stop condition will be to find fitness equal to 100 or to reach a maximum number of iterations fixed with the parameters of the algorithm. If one of these two conditions is satisfied, then we keep the individual of greater value of adaptation function as a final solution, and the genetic algorithm ends. else we pass to the following stage.

G. Parameters of a Genetic Algorithm

- Population size.
- Rate of replacement.
- Mechanism of selection.
- Rate of crossing.

- Rate of mutation.
- Number of generations.

VI. COMBINATION

Each classifier generates a result which will be combined with the others to give the final response. Several systems [Giacinto, 1998] use the *vote method* as combination method. It consists in retaining the privileged class by the majority of the classifiers as being the final response of the system. This method ignores the importance which can have a classifier compared to the others; which decreases the confidence rate of the system [Azizi, 2002]. On this observation we preferred used the Bayes method which consists in determining the C_j class for which posterior probability $P(C_j/e(1)=C_1, \dots, e(k)=C_k)$ is maximum, i.e.:

$$e(x) = \begin{cases} C_j & \text{if } P(C_j / e(1) = C_1, \dots, e(k) = C_k) = \\ & \text{MAX}_{i=1}^m P(C_i / e(1) = C_1, \dots, e(k) = C_k) \\ \text{Reject} & \text{else} \end{cases}$$

To estimate the posterior probability, several studies suppose that the decisions of the classifiers must be independent [Zouari, 2002]. Under this assumption and by applying the formula of inversion of Bayes, this probability is written:

$$P(C_j / e(1) = C_1, \dots, e(k) = C_k) =$$

$$P(C_j) \frac{\prod_{i=1}^k P(C_j / e(i) = C_i)}{P(C_{ji})} \quad (8)$$

$P(C_j / e(k) = C_k)$ and $P(C_j)$ are estimated from the confusion matrix obtained for the classifier K.

VII. HELPFUL HINTS

The combination of classifiers is an effective tool to have a great performance without increasing complexity of the existing techniques of classification. It is suitable, for applications requiring a great exactitude of classification. We wanted through this article, to contribute to the resolution of the problem of hepatic pathologies recognition which is until now in phase of research. The system suggested is a method of automatic design of a multi classifiers system based combination. The principal idea is to begin with a large set of classifiers from which we seek all the possible combinations then we combine the results of each subset by the Bayes method. The results of the combination will be then evaluated and the subset providing the best result will be selected.

The system is in the implementation phase. A base of scanographic images will be used for the validation and the

test. The experimental results are in hand, we hope to be able to communicate them within the reasonable times.

REFERENCES

- [1] [Azizi, 2002] Azizi NR., Sari T., Souici L., Sellami Mr., “*Un système multi-classifieurs neuronaux à combinaison floue pour la reconnaissance de l'écriture arabe manuscrite*”. MCSEAI'2002, seventh magrèbine conference in Data processing, Annaba, Algérie (2002).
- [2] [Djebbar, 2004] Djebbar A., Grainia S., Merouani H F., “*Un système de reconnaissance des images scanographiques appliqué au diagnostic des pathologies hépatiques*”. Jetim 2004, Algerian-French studies Days in medical imaging, pp. 131-136, Blida (2004).
- [3] [Gasmı, 2004] Gasmı I, Bouzid A, Souici L, Merouani H., “*Un système multi-classifieurs pour la reconnaissance des montants littéraires arabes*”. Majecstic 04, Manifestation of the young researchers stic, Acts of the congress, Calais, France (2004).
- [4] [Giacinto, 1998] Giacinto, G.: “*Design of multiple classifiers systems*”. Doctorate Thesis, University of deglistudi Disalerno, Italy (1998).
- [5] [Laurent, 1999] Laurent M., “*Combinaison de classifieurs pour la reconnaissance de formulaires structurés*”. DESS/DEA, University of Rouen (1999).
- [6] [Madhvanath, 2001] Madhvanath S., Govindaraju V., ‘*the role of Holistic paradigms in handwritten word recognition*’. IEEE transition on pattern analysis and machine intelligence, Vol 23, N:02, pp: 149-164, (2001).
- [7] [Montoliu, 1995] Montoliu, L.: “*Architecture multi-agents et réseaux connexionnistes : Application à la lecture de chèques manuscrits*”. Doctorate Thesis, Laboratory LIX, polytechnic School, Palaiseau, France (1995).
- [8] [Plamondon, 2000] Plamondon R., Srihari S. N., “*On-Line and off-line handwriting recognition: A comprehensive survey*”. IEEE on pattern analysis and machine intelligence, Vol 22, N:01, pp: 63-78, (2000).
- [9] [Roli, 2001] Roli, F., Giacinto, G., Vernazza, G.: ‘*Methods for designing multiple classifier systems*’. MCS'2001, 2nd International Workshop on multiple classifier systems, pp. 78-87, cambridge, UK, (2001).
- [10] [Ruta, 2000] Ruta, D., Gabrys, B.: “*An overview of classifier fusion methods*”. Computing and information systems., Vol. 7, pp. 1-10, (2000).
- [11] [Spalanzani, 1999] Spalanzani A.: “*Algorithmes évolutionnaires pour l'étude de la robustesse des systèmes de reconnaissance automatique de la parole*”. Doctorate Thesis, Joseph Fourier University - Grenoble I (Data-processing Specialist Systems and Communication) October 1999.
- [12] [Towell, 1994] Towell G. G., Shavlic J. W. “*Knowledge-based artificial neural networks*”. Artificial Intelligence, Vol 70, pp: 119-165, (1994).
- [13] [XU, 1992] Xu, I., Krzyzak, A., Suen, C. Y.: “*Methods of combining multiple classifiers and their application to handwriting recognition*”. IEEE transactions on systems, man and cybernetics., vol. 22, N03, pp. 418-435, (1992).
- [14] [Zouari, 2002] Zouari, H., Heutte, L., Lecourtier, Y., Alimi, A.: “*Un panorama des méthodes de combinaison de classifieurs en reconnaissance de formes*”. RFIA2002, 11th French-speaking congress of AFRIF-AFIA of pattern recognition and Artificial Intelligence., pp. 499-508, Angers (2002).