Multi-Classifier System Configuration using Genetic Algorithms

D. Impedovo, G. Pirlo, D. Barbuzzi
Dipartimento di Informatica
Università degli Studi di Bari
via Orabona 4, 70126, Bari
pirlo@di.uniba.it

Abstract - **Classifier combination is a powerful paradigm to deal with difficult pattern classification problems. As matter of this fact, multi-classifier systems have been widely adopted in many applications for which very high classification performance is necessary. Notwithstanding, multi-classifier system design is still an open problem. In fact, complexity of multi-classifiers systems make the theoretical evaluation of system performance very difficult and, consequently, also the design of a multi-classifier system. This paper presents a new approach for the design of a multi-classifier system. In particular, the problem of feature selection for a multi-classifier system is addressed and a genetic algorithm is proposed for automatic selecting the optimal set of features for each individual classifier of the multi-classifier system. The experimental results, carried out in the field of handwritten digit recognition, demonstrate the effectiveness of the proposed approach.**

Keywords: Multi-classifier System, Digit Recognition, Genetic Algorithms.

I. INTRODUCTION

Classifier combination is an effective strategy to solve difficult classification problems, like those related to on-line and off-line handwriting recognition [1]. On the basis of the kind of decisions combined, methods for classifier combination can be categorized into measurement-level, ranked-level and abstract-level methods [2]: Measurement-level combination methods combine values provided by individual classifiers as a measure of the degree of membership of the input pattern to each class; Ranked-level combination methods combine ranked lists of class labels ordered according to the degree of membership of the input pattern; Abstract-level combination methods simply combine top-class labels.

Whatever combination method is used, the efficacy of the combined classifier depends on the performance of the individual classifiers and on the degree of diversity among them [3,4,5]. For this purpose, research has been devoted so far in order to the the analysis of multi-classifier system behaviour [6, 7, 8, 9, 10, 11, 12, 13] as well as on the techniques to design individual classifiers well-suited for the realization of multi-expert systems [14, 15, 16].

In this field the role of feature selection is crucial. In fact feature selection allows the choice of the best subset of the features available from the data to be considered for pattern classification. The best subset should contains the least number of features ensuring high accuracy, whereas unimportant features should be discarded. In the past, many approaches have been proposed for feature selection. Kira and Rendell described a statistical feature selection algorithm that uses instance based learning to assign a relevance weight to each feature [17]. Koller and Sahami proposed a method for feature selection in which cross-entropy is evaluated to minimize the amount of information lost during feature elimination [18]. Kohavi and John introduced wrappers for feature subset selection [19]. Their approach searches for an optimal feature subset tailored to a particular learning algorithm and a particular training set. John, Kohavi and Pfleger addressed the problem of irrelevant features detection and elimination [20]. Pudil, Novoviˇcov´a and Kittler presented sequential search methods characterized by a dynamically changing number of features included or eliminated at each step [21]. Guyon and Elisseeff proposed variable selection in two alternate ways: (1) with a variable ranking method using a correlation coefficient; (2) with a nested subset selection method performing forward or backward selection [22]. Yang and Honavar used a genetic algorithm for feature subset selection [23]. Raymer et al. presented a hybrid approach that combines a k-nearest-neighbors classifier and a genetic algorithm. In this case they perform a simultaneous optimization of feature weights and selection of key features by including a masking vector on the genetic algorithm chromosome [24]. Also Wan et al. demonstrate that a genetic algorithm, combined with a classification algorithm, can improve classification efficiency, precision and robustness [25].

This paper focuses on the selection of the most profitable feature sets of the individual classifiers, for combination purposes. The problem of feature selection – of a multi-classifier system - is here considered as an optimization problem and the optimal feature sets are determined by a binary-coded genetic algorithm as those sets which maximize the performance of the combined classifier. The experimental results, carried out in the field of handwritten digit recognition, demonstrate the effectiveness of the approach.

The organization of the paper is the following: Section 2 presents the problem of multi-expert system design by

978-0-7695-4774-9/12 $26.00 © 2012 IEEE
DOI 10.1109/ICFHR.2012.237

selection of the best feature set of each individual classifier. The binary coded genetic algorithm for the optimization of the feature sets of each individual classifier of the multi-expert system is discussed in Section 3. Section 4 presents the experimental results, carried out in the field of handwritten digit recognition, that demonstrate the effectiveness of the proposed approach.

II. Multi-expert System Design

Let $C=\{C_1,C_2,\ldots,C_M\}$ be the set of pattern classes and $A=\{A_1,A_2,\ldots,A_K\}$ be the set of K abstract-level classifiers of the multi-classifier system, being F_i the feature set of A_i, for $i=1,2,\ldots,K$. In addition, let CR be the combination rule used for decision combination. When an unknown input pattern x has to be classified, it is first fed to each A_i that provides the response $A_i(x)\in C$. Finally, all responses are combined to obtain the final classification result: $E(x)= E(A_1(x), A_2(x),\ldots,A_K(x), CR)$ [1, 2].

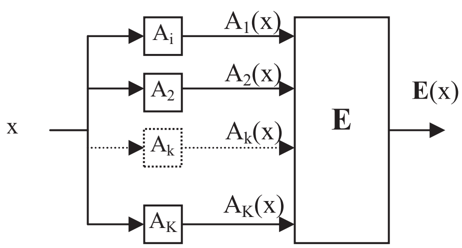

Figure 1. Multi-classifier System

In this paper the problem of multi-expert system design is considered as an optimization problem. More specifically, the optimal features of each individual classifier are selected as those features for which the classification performance of the multi-expert system is maximum. More precisely, the optimization problem related to multi-expert design is formulated as follows:

Select the optimal feature sets $F_1, F_2,\ldots, F_K$, for which the Cost Function associated to the classification result of the multi-expert system is minimum:

$$CF(E(A_1,A_2,\ldots A_k,\ldots,A_K, CR)) \rightarrow \text{Min.} \tag{1}$$

where, in this work, it is assumed that:

- Fk is a subset of $F=\{f_1,f_2,\ldots,f_T\}$, $k=1,2,\ldots,K$;
- $CF(E(A_1(F_1),A_2(F_2),\ldots,A_k(F_k),\ldots,A_K(F_K),CR))$ is the Error Rate of the multi-expert classifier, being F_k the feature set of A_k, $k=1,2,\ldots,K$.

III. A Genetic Algorithm for Multi-expert Optimization

Genetic algorithms are adaptive methods inspired by the processes of natural evolution and widely used to solve optimization problems [26, 27]. In particular, when a genetic algorithm is considered, the initial population – that corresponds to a set of possible solutions - evolves according to the principles of natural selection and survival of the best individuals. In this section, a binary-coded genetic algorithm is presented for the optimization problem of eq. (1). The initial – population

$$Pop=\{\Phi_1, \Phi_2, \ldots,\Phi_i, \ldots ,\Phi_{Npop}\} \tag{2}$$

is created by generating N_{pop} random individuals (N_{pop} even). Each individual is a binary vector

$$\Phi_i= \langle p_1,p_2,\ldots,p_T,\ p_{T+1},p_{T+2},\ldots,p_{2T},\ldots,\ p_{kT+1},p_{kT+2},\ldots,p_{(k+1)T},\ldots,\ p_{(K-1)T+1},p_{(K-1)T+2},\ldots,p_{KT}\rangle \tag{3}$$

where, for each $t=1,2,\ldots,(K+1)\cdot T$, each element p_t is a flag defining the absence / presence of the feature f_i in F_k (the feature set of the classifier A_k), being $t=(k-1)\cdot T+i$ ($k=1,2,\ldots K$; $i=1,2,\ldots,T$). More precisely:

- $p_t=0 \rightarrow f_i$ *is in* F_k
- $p_t=1 \rightarrow f_i$ *is not in* F_k.

The fitness value of an individual $\Phi_i=\langle p_1,p_2,\ldots,p_j,\ldots,p_{KT}\rangle$ is taken as the classification cost

$$CF(E(A_1(F_1),A_2(F_2),\ldots,A_k(F_k),\ldots,A_K(F_K),CR)).$$

From the initial - population, the following genetic operations are used to generate the new populations of individuals [28]:

Individual selection
Crossover
Mutation
Elitist strategy.

1) In the individual selection procedure $N_{pop}/2$ random pairs of individuals are selected, according to a *roulette-wheel* strategy. Figure 2 shows an example of individuals with different fitness values. In this case the probability for an individual Φ_i to be selected is equal to $P_{selection}(\Phi_i)=Fitness(\Phi_i)/Total\ Fitness$.

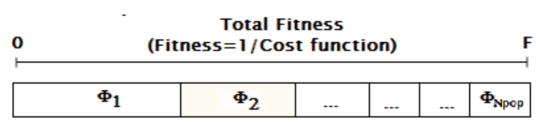

Figure 2. Roulette-wheel strategy

2) The crossover operator uses the one-point strategy. Therefore let

$$\langle p^a_1,p^a_2,\ldots,p^a_{s-1},\ p^a_s,\ldots,p^a_{KT}\rangle \tag{4a}$$

and

$$\langle p^b_1,p^b_2,\ldots,p^b_{s-1},\ p^b_s,\ldots,p^b_{KT}\rangle, \tag{4b}$$

be two parent individuals, the two offspring individuals of the next generation are:

$$\langle \bar{p}^a_1,\bar{p}^a_2,\ldots,\bar{p}^a_{s-1},\ \bar{p}^a_s,\ldots,\bar{p}^a_{KT}\rangle \tag{5a}$$

and

$$\langle \bar{p}^b_1,\bar{p}^b_2,\ldots,\bar{p}^b_{s-1},\ \bar{p}^b_s,\ldots,\bar{p}^b_{KT}\rangle; \tag{5b}$$

where s is a random integer, $1<s<KT$.

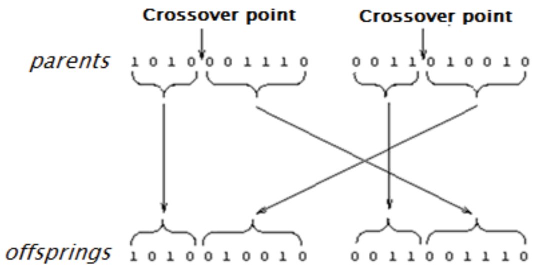

Figure 3. Single point crossover

3) The mutation operator changes each element p_t of an individual Φ_I, according to the mutation probability Mut_prob.

Figure 4. Mutation

4) From the N_{pop} individuals generated by the above operations, the elitist strategy removes one random individual from the current population that is substituted by the individual with the minimum cost in the previous population.

Steps from 1 to 4 are repeated until N^{iter} successive populations of individuals are generated. When the process stops, the optimal set of features is obtained by the best individual of the last-generated population.

IV. Experimental Results

For the experimental test the set of pattern classes $\Omega_1=\{0,1,2,3,4,5,6,7,8,9\}$ was considered, concerning the 10 numeral digits. The data sets of the CEDAR database were used in this case [29]: 18223 patterns for learning (BR directory) and 2128 patterns for the test (BS directory). The following features were extracted from each digit image [30,31,32]:

- Contour Profiles (Figure 5) :
  - $f_1,\ldots f_8$: the locations of maxima and minima in the profiles
  - $f_9,\ldots,f_{24}$: the locations of maximum and minimum peaks in the profiles
  - f_{25} , f_{28}: the locations of maxima and minima in the height
  - f_{29} , f_{32} : the locations of maxima and minima in the width;

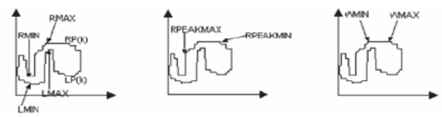

Figure 5. Contour Profile Features

- Geometric Feature (Figure 6):
  - f_{33}: *holes*;
  - f_{34}, f_{35}, f_{36}, f_{37}: *vertical-up, vertical-down, horizontal-right, horizontal-left cavities*;
  - f_{38}, f_{39}, f_{40}, f_{41}: *vertical-up, vertical-down, horizontal-right, horizontal-left end-points*.

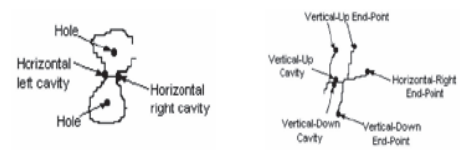

Figure 6. Geometric Feature

- Intersection with lines (Figure 7):
  - $f_{42},\ldots,f_{46}$: intersections with five vertical lines
  - $f_{47},\ldots,f_{51}$: intersections with five horizontal lines
  - $f_{52},\ldots,f_{56}$: intersections with five diagonal lines (+45°)
  - $f_{57},\ldots,f_{61}$: intersections with five diagonal lines (-45°);

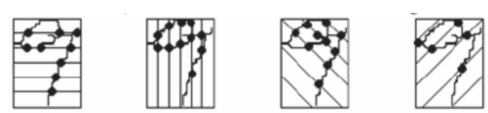

Figure 7. Intersections Feature

- Extrema Points (Figure 8):
  - f_{62}, f_{63}, f_{64}, f_{65}: *Top, bottom, left, right extrema points* ;

Figure 8. Extrema Point Features

- Cross Points (Figure 9):
  - f_{66} : *Cross Points.*

Figure 9. Cross Point Feature

This set of feature was considered to develop a multi-classifier system, in which the decisions of five individual classifiers A_1, A_2, A_3, A_4, A_5 , each one based on a k-NN classification rule (k=3) were combined by a Majority Voting (MV) rule.

Table I. The Optimal Individual (Feature Vectors)

Feature Type	***Feature***	F_1	F_2	F_3	F_4	F_5
Contour Profile	f_1	0	0	0	0	1
	f_2	1	0	0	0	0
	f_3	1	0	0	0	1
	f_4	0	1	1	1	0
	f_5	0	0	0	1	1
	f_6	1	1	1	0	0
	f_7	0	1	0	1	1
	f_8	0	1	1	1	1
	f_9	0	0	1	0	0
	f_{10}	1	1	1	1	1
	f_{11}	1	0	1	1	0
	f_{12}	1	0	0	0	1
	f_{13}	0	1	1	1	1
	f_{14}	0	0	1	0	1
	f_{15}	1	0	0	1	0
	f_{16}	0	1	1	1	1
	f_{17}	0	1	1	1	1
	f_{18}	1	0	0	1	1
	f_{19}	0	0	0	0	0
	f_{20}	1	1	0	1	1
	f_{21}	0	1	1	1	0
	f_{22}	1	0	1	1	0
	f_{23}	0	0	0	0	0
	f_{24}	1	0	1	1	1
	f_{25}	1	1	0	0	1
	f_{26}	0	0	1	1	1
	f_{27}	0	1	0	1	0
	f_{28}	1	0	1	0	0
	f_{29}	1	0	0	0	0
	f_{30}	0	1	0	0	0
	f_{31}	0	1	1	1	1
	f_{32}	0	1	1	1	1
Geometric	f_{33}	0	0	1	0	0
	f_{34}	1	1	1	0	1
	f_{35}	1	1	0	0	1
	f_{36}	1	1	1	1	1
	f_{37}	1	1	1	1	0
	f_{38}	1	1	1	1	0
	f_{39}	1	1	1	1	0
	f_{40}	1	0	1	0	1
	f_{41}	1	0	0	0	0
Intersection	f_{42}	0	1	1	0	0
	f_{43}	1	1	0	1	1
	f_{44}	0	0	1	0	0
	f_{45}	1	0	1	0	0
	f_{46}	1	0	0	1	1
	f_{47}	0	0	1	1	1
	f_{48}	1	1	1	0	1
	f_{49}	1	1	0	1	1
	f_{50}	1	1	1	0	1
	f_{51}	1	0	1	1	1
	f_{52}	0	0	0	0	0
	f_{53}	1	0	1	1	0
	f_{54}	0	0	1	0	1
	f_{55}	0	0	0	1	1
	f_{56}	0	0	1	0	1
	f_{57}	1	0	1	1	1
	f_{58}	1	0	0	0	0
	f_{59}	0	0	0	1	0
	f_{60}	0	0	0	1	0
	f_{61}	1	1	1	1	1
Extrema Points	f_{62}	0	1	0	1	0
	f_{63}	0	1	0	0	0
	f_{64}	0	1	1	0	1
	f_{65}	0	1	0	1	0
Cross Point	f_{66}	0	1	0	0	0

The genetic algorithm must define the optimal feature set F_1, F_2, F_3, F_4, F_5 for A_1, A_2, A_3, A_4, A_5, respectively. For this purpose, the following free-parameter values of the genetic algorithm were pre-estimated: N_{Pop}=10, N^{iter}=40, *Mut_prob*=0.05, *max_displ*=3, *b*=1.0. Figure 10 shows the behaviour of the cost function of the best individual of each generation.

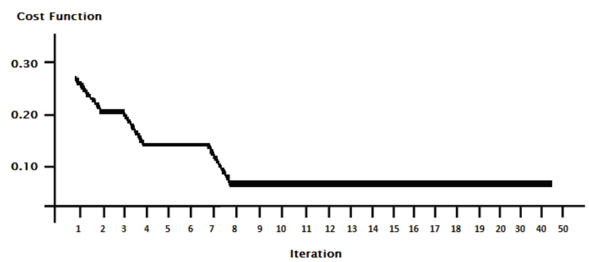

Figure 10. Cost Function vs iterations

In particular, the Genetic Algorithm allows to define the optimal individual, i.e. the best feature set of each classifiers, as reported in Table I.

When the best feature sets are used, the recognition rate of the multi-classifier system is equal to 94%. Conversely, when a random feature set is used , the recognition rate is equal to 72%, on average. Therefore, the GA permits a reduction of the cost function (error rate) from 28% to 6%.

V. Conclusion

In this paper the problem of automated configuration of multi-classifier systems is addressed. A new approach is also proposed, based on a genetic algorithm. In particular the approach allows the definition of the best feature set for each individual classifier, able to optimize the performance of the multi-classifier system. The approach, that has been applied to the field of handwritten digit classification, has demonstrate that genetic algorithms are capable to significantly support automated configuration of multi-classifier systems.

References

[1] J. Kittler, M. Hatef, R.P.W. Duin, J. Matias, "On combining classifiers", *IEEE T-PAMI*, Vol.20, no.3, pp.226-239, 1998.

[2] L.Xu, A.Krzyzak, C.Y-Suen, "Methods of Combining Multiple Classifiers and Their Applications to Handwriting Recognition", IEEE T-SMC - Vol. 22, N. 3, 1992, pp. 418-435.

[3] C.A. Shipp and L.I. Kuncheva, "Relationships between combination methods and measures of diversity in combining classifiers", *Information Fusion*, Vol. 3, No.2, 2002, pp. 135-148.

[4] L. Bovino, G. Dimauro, S.Impedovo, M.G. Lucchese, R. Modugno, G. Pirlo A.Salzo, L. Sarcinella, "On the Combination of *Abstract-Level* Classifiers", *IJDAR,* 2003, Vol. 6, pp. 42-54.

[5] V. Di Lecce, G. Dimauro, A. Guerriero, S. Impedovo, G. Pirlo, A. Salzo, "Classifier Combination: the role of a-priori knowledge", *Proc. IWFHR-7*, Sept. 2000, Amsterdam, pp. 143-152.

[6] L. Lam, "Classifier Combinations: Implementations and Theoretical Issues", *Multiple Classifier Systems*, J.Kittler F.Roli (eds.),LNCS, vol.1857,Springer,2000,pp.77-86.

[7] G. Dimauro, S. Impedovo, G. Pirlo, "Multiple Experts:A New Methodology for the Evaluation of the Combination Processes", Proc. IWFHR-5, Colchester, Uk, 1996, pp.131-136.

[8] G. Dimauro, S.Impedovo, M.G. Lucchese, G. Pirlo, A. Salzo "Discovering Rules for Dynamic Configuration of Multi-Classifier Systems", Document Analysis Systems V, LNCS 2002, Vol. 2423, D. Lopresti et al. (eds.), Springer Verlag Publ., 2002, pp. 157-166.

[9] L. Bovino, G. Dimauro, S.Impedovo, G. Pirlo, A. Salzo, “Increasing the number of classifiers in multi-classifier systems: a complementary-based analysis”, Document Analysis Systems V, LNCS 2002, Vol. 2423, D. Lopresti et al. (eds.), Springer Verlag Publ., 2002, pp. 145-156.

[10] N. Greco, S. Impedovo, R.Modugno, G. Pirlo, "Generation of Ensambles of Synthetic Classifiers for the Evaluation of Combination Methods", IJIT, World Enformatika Society Publ., 2004, Vol.1, N.2, pp. 90-93.

[11] D. Impedovo, G. Pirlo, L. Sarcinella, E. Stasolla, "Artificial Classifier Generation for Multi-Expert System Evaluation”, Proceedings of the 12th Interational Conference on Frontiers in Handwriting Recognition (ICFHR-12), Nov. 16-18, 2010, IEEE Computer Society Press, Kolkata, India, pp. 421-426.

[12] H. Zouari, L. Heutte, Y. Lecourtier, A. Alimi, “ A New Classifier Simulator for Evaluating Parallel Combination Methods”, Proc. ICDAR, Edinburgh, *2003*, pp.26-30.

[13] D. Impedovo, G. Pirlo, “Generating Sets of Classifiers for the Evaluation of Multi-expert Systems”, Proc of 20th Int. Conf on Pattern Recognition (ICPR), 2010. Istanbul - Turkey, August 2010, IEEE Computer Society , pp. 2166-2169.

[14] R. E. Schapire and Y. Singer, “Improved Boosting Algorithms Using Confidence-Rated Predictors”, *Machine Learning*, Vol. 37, No. 3, 1999, pp. 297-336.

[15] R. E. Schapire, Y. Freund, P. Bartlett, and W. S. Lee, ”Boosting the margin: A new explanation for the effectiveness of voting methods”, *The Annals of Statistics*, 1998, Vol. 26, No. 5, pp. 1651–1686.

[16] Y. Freund, R. Iyer, R. E. Schapire, Y. Singer, “ An Efficient Boosting Algorithm for Combining Preferences“, Journal of Machine Learning Research, Vol. 4, 2003, pp. 933-969.

[17] K. Kira, L. A. Rendell, “A practical approach to feature Selection”, Proceedings of the Ninth International Conference on Machine Learning, San Francisco, CA, USA, Morgan Kaufmann Publishers Inc., 1992, pp. 249–256.

[18] D. Koler and M. Sahami, ”Toward optimal feature selection”, Proceedings of the Thirteenth International Conference on Machine Learning, Morgan Kaufmann, 1996, pp. 284–292.

[19] R. Kohavi, G. H. John, “Wrappers for feature subset selection”, Artificial Intelligence, Vol. 97, no. 1–2, 1997, pp. 273–324.

[20] G. H. John, R. Kohavi, K. Pfleger, ”Irrelevant features and the subset selection problem”. Proceedings of the Eleventh International Conference on . Machine Learning, San Francisco, CA: Morgan Kaufmann Publishers, 1994, pp. 121–129.

[21] P. Pudil, J. Novoviˇcov´a, J. Kittler, “Floating search methods in feature selection”, Pattern Recognition Letters, 1994, Vol. 15, No. 11, pp. 1119–1125.

[22] I. Guyon, A. Elisseeff, “An introduction to variable and feature selection”, Journal of Machine Learning Research, Vol. 3, 2003, pp. 1157–1182.

[23] J. Yang and V. Honovar, “Feature subset selection using a genetic algorithm”. IEEE Intelligent Systems, Vol. 13, No. 2, 1998, pp. 44–49.

[24] M. L. Raymer, W. F. Punch, E. D. Goodman, P. C. Sanschagrin and L.A. Kuhn, “Simultaneous Feature Extraction and Selection Using a Masking Genetic Algorithm”;

[25] J. Wan, Z. Chen, Y. Chen and Z. Bai, “GA Based Optimal Feature Extraction Method for Functional Data Classification”.

[26] D. Beasley, D.R.Bull, R.R.Martin, “An Overview of Genetic Algorithms: Part 1, Fundamentals”, University Computing, Vol 15, n. 2, pp. 58-69, 1993 .

[27] D. Beasley, D.R.Bull, R.R.Martin, “An Overview of Genetic Algorithms: Part 2, Research Topics”, University Computing, Vol. 15, n. 2, pp. 170-181,1993.

[28] Z. Michalewicz, *Genetic Algorithms + Data Structure=Evolution Programs*, Springer Verlag, Berlin, Germany, 1996.

[29] J. Hull, “A database for handwritten text recognition research”, *IEEE T_PAMI*, Vol. 16, n. 5, pp. 550–554, 1994.

[30] S. Impedovo, M.G. Lucchese, G. Pirlo, “Optimal Zoning Design by Genetic Algorithms” , *IEEE Transactions on Systems, Man and Cybernetics - Part A: Systems and Humans,* Vol. 36 , Issue: 5, Sept. 2006, pp. 833-846.

[31] G. Dimauro, S. Impedovo, M.G. Lucchese, R. Modugno, G. Pirlo, “Alphanumeric Hand-Prints Classification: Similarity Analysis between Local Decisions”, *International Journal of Information Technology (IJIT)*, World Enformatika Society Publ., 2004, Vol. 1, N. 2, pp. 86-89.

[32] G. Dimauro, S.Impedovo, G. Pirlo, A.Salzo, "Zoning Design for Handwritten Numeral Recognition", in *Lecture Notes in Computer Sciences*, Vol. 1311, Ed. by A. Del Bimbo, Springer Verlag, Berlin, 1997, pp. 592-599.