Discovering Rules for Dynamic Configuration of Multi-classifier Systems

G. Dimauro[1], S. Impedovo[1], M.G. Lucchese[2], G. Pirlo[1], and A. Salzo[2]

[1] Dipartimento di Informatica, Università degli Studi di Bari
Via Orabona 4, 70126 Bari-Italy
impedovo@di.uniba.it
[2] Consorzio Interuniversitario Nazionale per l'Informatica (CINI),
Via G.Petroni 15/F.1, 70124 Bari– Italy

Abstract. This paper addresses the problem of dynamic configuration of multi-classifier systems. For this purpose, the performance of combination methods for abstract-level classifiers is predicted, under different working conditions, and sets of rules are discovered and used for dynamic configuration of multi-classifier systems. The experimental tests have been carried out in the field of hand-written numeral recognition. The result demonstrates the validity of the proposed approach.

1 Introduction

The combination of classifiers is a diffuse strategy to design high-performance classification systems. In fact it is common experience that the complexity of the classification problem and pattern variability do not allow the development of classifiers as good as required for many practical applications [1].

Up to now, several methods have been proposed to combine different classifiers [2,3,4]. This notwithstanding, the problem of multi-classifier system design is still open [5]. For instance, in the field of hand-written numeral recognition, several decades of research activity have lead to develop thousands of different algorithms that can be combined in a multi-classifier system. A lot of different features have been considered (based on mathematical transforms, structural decomposition, geometrical and topological characteristics, etc.), and many different classification strategies (based on pattern-matching, structural-analysis, etc.) have been used. Similarly, many combination methods are also available. Hence, the prediction of the performance of a multi-classifier system is difficult, since it depends on both the classifiers and the method considered [6]. Moreover, in many cases, the working conditions can change dynamically. For instance, this is the case of neural network classifiers whose characteristics can change significantly, depending on the learning conditions.

D. Lopresti, J. Hu, and R. Kashi (Eds.): DAS 2002, LNCS 2423, pp. 157–166, 2002.

Another case of change of the working conditions can be due to modification in the characteristics of the input data (different types of writing styles, or input data from different sources, which require different pre-processing algorithms, etc.). Therefore, the development of advanced strategies is required for the design of multi-classifier systems able to change dynamically their configurations, depending on the modifications of the working conditions.

This paper presents a first attempt to solve the problem. In particular, the most effective combination method of a multi-classifier system is dynamically selected, on the basis of the estimation of the degree of complementarity among the individual classifiers. For this purpose, a simulation procedure is used to determine systematically different working conditions in which the performance of various combination methods for abstract-level classifiers is evaluated a-priori. This information is used to determine sets of rules which are used, during the run-time, for the dynamic selection of the optimal combination method. Two combination methods for abstract-level classifiers are used for the experimental tests. The Dempster-Shafer (DS) method and the Behavioural Knowledge Space (BKS) method. In Section 2 the combination methods are briefly described. Section 3 describes the methodology for the evaluation of combination methods. Section 4 presents the new approach for dynamic selection of combination method. Section 5 reports the experimental results that have been obtained in the field of hand-written numeral recognition, using the data from the CEDAR database. The conclusion of this work is presented in Section 6.

2 Analysis of Combination Methods

In a multi-classifier parallel system, the input pattern x_t is fed to K individual classifiers in parallel. Each classifier A_i provides its response $A_i(x_t)$, i=1,2.,..K. The responses obtained by all the classifiers are then combined to obtain the final results $E(x_t)$ according to a suitable combination strategy $E(A_1(x_t), A_2(x_t),...,A_K(x_t)) \rightarrow E(x_t)$[1].

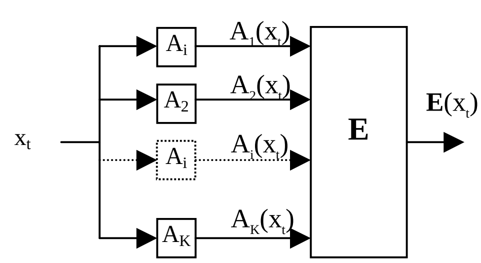


Fig. 1. A Multi-classifier Parallel System

The effectiveness of a multi-classifier system depends not only on the performance of the individual classifiers but also on the collective behavior of the entire set of classifier as well as on the capability of the combination method to integrate the results. On the basis of this consideration, a methodology for the evaluation of a combination process has been recently proposed [6]. It uses a Similarity Index which provides a measure of the degree of complementarity among the decisions of a set of classifiers. Precisely, let A_1 , A_2 be two classifiers and $A_1(x_t)$ and $A_2(x_t)$ respectively the top-candidates provided for the pattern x_t, for x_t belonging to a database $T=\{x_1,x_2,x_3,\dots,x_N\}$ and let be

$$E\,(A_1\,(x_t),\,A_2\,(x_t)\,) = \begin{cases} 1 & \text{if } A_1\,(x_t) = A_2(x_t) \\ 0 & \text{otherwise} \end{cases} \quad . \tag{1}$$

The Similarity Index between A_1 , A_2 is defined as:

$$\rho_{A_1,A_2} = \frac{1}{\mathrm{Card(T)}} \sum_{x_t \in T} C(A_1(x_t),\, A_2(x_t)) \quad . \tag{2}$$

Figure 2a shows the decisions of two classifiers A_1 and A_2 , for 10 input patterns belonging to the set of the ten digits. The recognition rate of both classifiers is 70%. Moreover it is easy to verify that A_1 and A_2 always provide the same response. Thus, we have ρ=1 (fig. 2a). Also the recognition rate of both classifiers in Figure 2b is 70%. However, in this case ρ=0.4.

Pattern n.	A_1	A_2
x_1∈ '4'	4	4
x_2∈ '0'	8	8
x_3∈ '8'	8	8
x_4∈ '2'	2	2
x_5∈ '0'	0	0
x_6∈ '1'	7	7
x_7∈ '9'	9	9
x_8∈ '3'	3	3
x_9∈ '8'	0	0
x_{10}∈ '7'	7	7

(a) ρ=1

Pattern n.	A_1	A_2
x_1∈ '4'	2	4
x_2∈ '0'	8	0
x_3∈ '8'	8	8
x_4∈ '2'	2	2
x_5∈ '0'	0	8
x_6∈ '1'	1	7
x_7∈ '9'	9	9
x_8∈ '3'	8	3
x_9∈ '8'	8	0
x_{10}∈ '7'	7	7

(b) ρ=0.4

Fig. 2. Variability range for the Similarity Index

In general, for a set of K classifiers $A =\{A_i \mid i=1,2,\dots,K\}$, the Similarity Index is [6]:

$$\rho_A = \frac{\sum_{\substack{i,\ j=1,\dots K \\ i<j}} \rho_{A_i,A_j}}{\binom{K}{2}} \tag{3}$$

The performance of a method E for classifier combination is then considered as a function $P_E(K,\underline{R}, \rho)$, where K is the number of individual classifiers that are combined; $\underline{R}=(R_1,R_2,\ldots,R_K)$ is the vector of the recognition rates of the classifiers (for the sake of simplicity in this paper we suppose that all classifiers have the same recognition rate - i.e. $R=R_i$, $i=1,2,\ldots,K$); ρ is the Similarity Index of the set of classifiers. Hence, in order to evaluate the performance of a classifier combination method in different working conditions, several sets of classifiers are simulated and grouped into different categories each one characterized by: the number of classifiers (K), the recognition rate (R), the similarity index (ρ). Hence, each value $P_E(K,R, \rho)$ is obtained as the average performance of the method E, when the sets of the category (K,R, ρ) are considered [6].

3 Dynamic Selection of the Combination Method

The analysis on the performance of combination methods, carried out by simulated data, can be useful to select dynamically the combination method for a multi-classifier system. For this purpose let us consider the advanced multi-classifier system in figure 3. It works as a traditional multi-classifier system (see fig. 1) even if it contains a control module for the run-time monitoring of the Similarity Index ρ of the set of classifiers. Depending on the value of ρ (computed by the analysis of the agreements among the outputs of the classifiers) and taking into consideration the performance of the combination methods $E_{BKS}(K,R,\rho)$ and $E_{DS}(K,R,\rho)$, the control module selects dynamically the most profitable method.

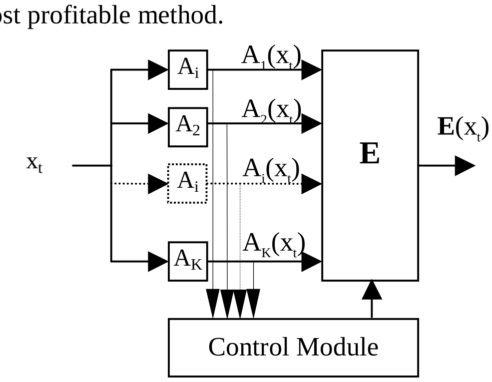


Fig. 3. Dynamic Selection of the Combination Method

4 The Combination Methods

In this paper two combination methods for abstract-level classifiers are considered: the Dempster-Shafer (DS) [5] and the Behaviour Knowledge Space (BKS) [4].

(a) Dempster-Shafer Method (DS)
The Dempster-Shafer method combines different classifiers using their recognition and substitution rates as a priori knowledge [4]. For a given input pattern x, all classifiers having the same output are collected into a group E_k, k=1,...,K', (where K' is the number of different outputs). $E_1, E_2, \ldots, E_{k'}$ are then considered as new classifiers and for each of them the recognition and substitution rates are estimated. Successively, $E_1, E_2, \ldots, E_{k'}$ are combined in order to calculate the belief of the correct output $Bel(A_j)$ and the belief of a misrecognition output $Bel(\neg A_j)$ [4]. The result of the combined classifier E is defined by the following decision rule:

$$E(x) = \begin{cases} j, & \text{if} \quad Bel(A_j) = \max\{Bel(A_i) \mid Bel(\neg A_i) \le \alpha,\ \mathrm{i} = 0,1,...,9\} \\ \text{Reject,} & \text{otherwise} \end{cases} \tag{4}$$

where α is a suitable threshold value.

(b) Behaviour-Knowledge Space Method (BKS)
The Behaviour Knowledge Space method is based on two processing phases: the "learning" phase and the "operation" phase [3].

- In the "learning" phase the set of learning pattern is fed to K classifiers. The result is used to fill a discrete K-dimensional space in which each dimension corresponds to the decision of a specific classifier. So, the K-tuple of decisions provided by the K classifiers defines a unit in the space generally called "Focal Unit". When a "Focal Unit" is addressed by the vector of recognition responses, the index corresponding to the class of the input pattern is incremented. This index counts the number of times in which a pattern belonging to that class generates the specific K-tuple of decisions.
- In the "operation" phase, the K-tuple of decisions provided by the classifiers is used to address a "Focal Unit". From the analysis of the data in the "Focal Unit" the result of the combined classifier E is obtained by the following rule:

$$\mathrm{E(x)} = \begin{cases} \mathrm{j}_{\mathrm{FU(x)}} & \text{if} \ \ \mathrm{T}_{\mathrm{FU(x)}} > 0 \ \text{ and } \ \dfrac{\mathrm{Card}(\mathrm{j}_{\mathrm{FU(x)}})}{\mathrm{T}_{\mathrm{FU(x)}}} > \mathrm{a} \\ \text{Reject} & \text{otherwise} \end{cases} \tag{5}$$

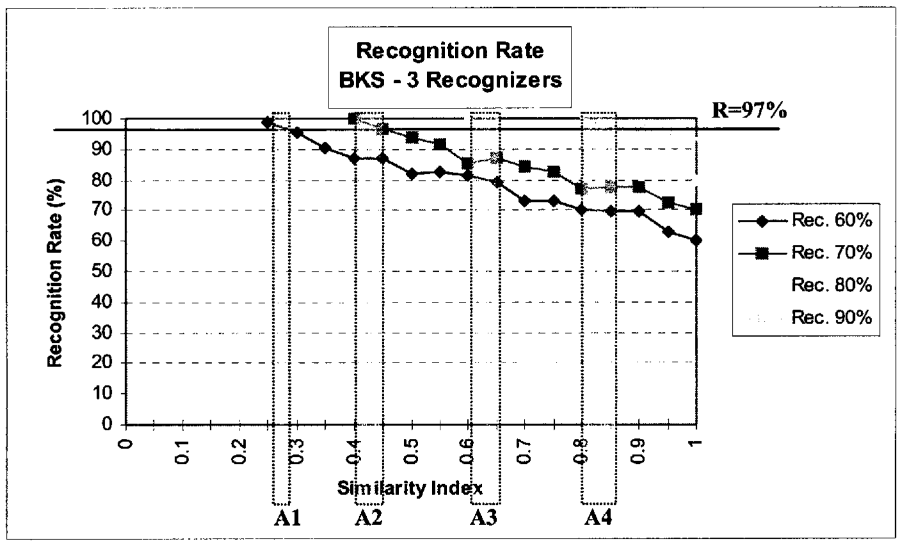
where:

- ♦ *FU(x)* is the Focal Unit selected by the input pattern *x*;
- ♦ $j_{FU(x)}$ is the best representative class;
- ♦ $T_{FU(x)}$ is the total number of samples in the Focal Unit;
- ♦ a is a suitable threshold value.

5 Experimental Results

The performance of each method has been evaluated by considering sets of 3 and 4 classifiers each one with an individual recognition rate of 60%, 70%, 80% and 90%.

For instance, figures 4 and 5 show the recognition rate of the BKS and DS, respectively, as a function of the Similarity Index of the set of individual classifiers [8].



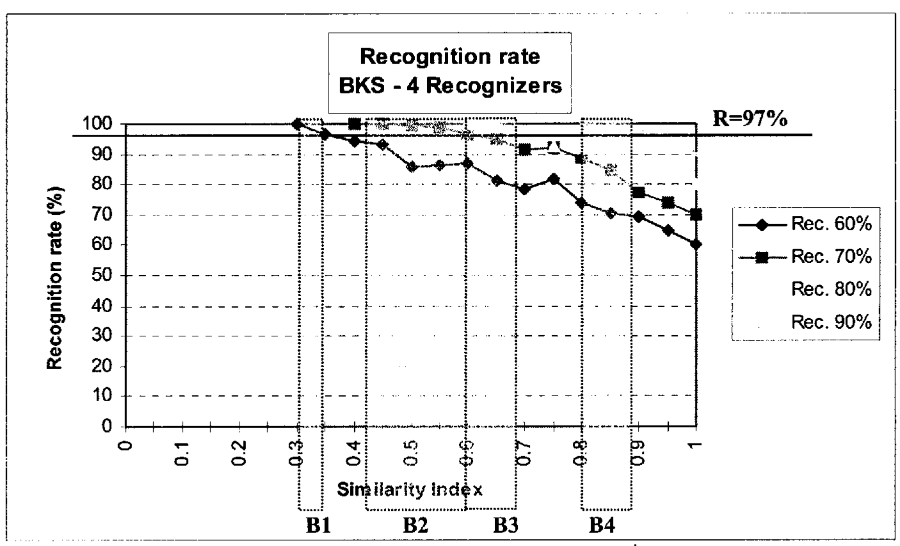


Fig. 4. Recognition Rate of the BKS method.

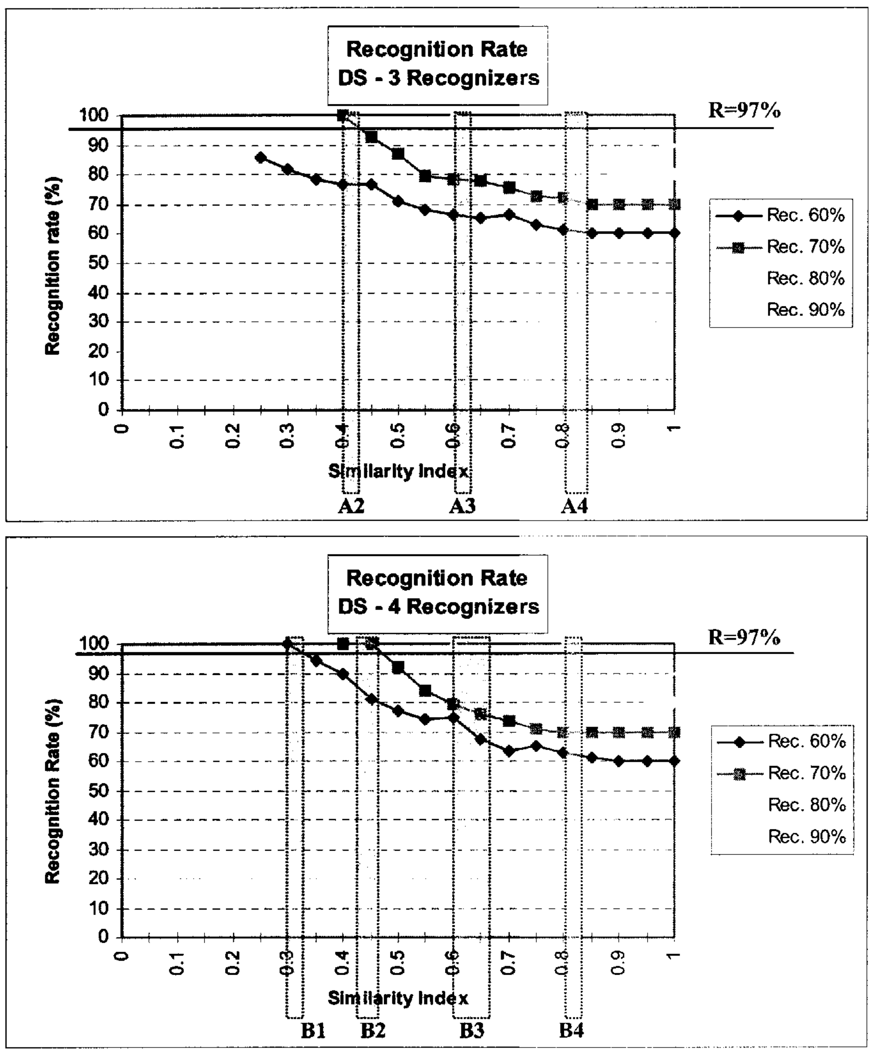


Fig. 5. Recognition Rate of the DS method.

These results first allow the determination of the expected performance of the combination methods. For example, if the target requirement is the recognition rate greater than 97%, we have that, when BKS is used, one of the following conditions must be satisfied (see the regions delimited in figure 4):

BKS- 3 classifiers:

A1) recognition rate of 60%, similarity index $\rho \in [0.26,0.28]$;
A2) recognition rate of 70%. similarity index $\rho \in [0.43,0.45]$;
A3) recognition rate of 80%. similarity index $\rho \in [.60,0.65]$;
A4) recognition rate of 90%. similarity index $\rho \in [0.80,0.86]$;

BKS- 4 classifiers:

B1) recognition rate of 60%. similarity index $\rho \in [0.30,0.35]$;
B2) recognition rate of 70%. similarity index $\rho \in [0.43,0.60]$;
B3) recognition rate of 80%. similarity index $\rho \in [0.60,0.68]$;
B4) recognition rate of 90%. similarity index $\rho \in [0.80,0.88]$;

In a similar way, DS provides a recognition rate greater than 97% only if one of the following conditions is satisfied (see the grey regions in figure 5):

DS- 3 classifiers:

A2) recognition rate of 70%. similarity index $\rho \in [0.43,0.42]$;
A3) recognition rate of 80%. similarity index $\rho \in [0.60,0.62]$;
A4) recognition rate of 90%. similarity index $\rho \in [0.80,0.83]$;

DS- 4 classifiers:

B1) recognition rate of 60%. similarity index $\rho \in [0.30,0.32]$;
B2) recognition rate of 70%. similarity index $\rho \in [0.43,0.46]$;
B3) recognition rate of 80%. similarity index $\rho \in [0.60,0.67]$;
B4) recognition rate of 90%. similarity index $\rho \in [0.80,0.83]$;

To confirm this result, obtained by simulated data, we consider the set of classifiers for hand-written numerals reported in Table 1 and used to recognise the courtesy amount in a system for Italian bank-check processing [9]. The classifiers have been initially trained by data from the CEDAR database (BR directory). On average, the recognition rate of the classifiers (at zero rejection) is equal to 89,8%.

Table 1. The individual classifiers

Classifiers	Recognition
C1: Regions	91.3%
C2: Contour Slope	90.2%
C3: Enhanced Loci	89.5%
C4: Histogram	88.2%

When combined by BKS and DS, the classification performance is reported in Table 2 (tested on the CEDAR database – BS directory). Table 2 confirms the previous results obtained by simulated data. For instance, if the recognition rate must be greater than 97%, we have: if BKS is used (Fig. 4), the set of classifiers 1,2 and 3 (condition A4) and also 5 (condition B4) must be selected; when the DS is used (Fig.5), the set of classifiers 1 and 2 (condition A3) must be selected.

This analysis also provides information useful for the dynamic selection of the combination method. In our tests, the classifiers of Table 1 have been trained with additional sets of hand-written numerals. Three learning levels are considered: at the

Table 2. Performance of the combination methods

Set	ρ	Classifiers	BKS Recognition	BKS Reliability	DS Recognition	DS Reliability
1	0.82	C1-C2-C3	98.0% (A4)	99.2%	97.8% (A3)	98.7%
2	0.83	C1-C3-C4	97.1% (A4)	98.3%	97.2% (A3)	98.4%
3	0.85	C2-C3-C4	97.3% (A4)	98.4%	95.6%	97.1%
4	0.87	C1-C2-C4	95.6%	96.8%	95.1%	96.9%
5	0.87	C1-C2-C3-C4	97.5% (B4)	99.0%	96.9%	98.7%

first level (L1) a first set of 1000 patterns is provided to each individual classifier for learning, at the second level (L2) another set of 1000 patterns is added to the first set, at the third level (L3) an additional set of 1000 patterns is added to the first two sets. Table 3 reports the characteristics of the set of experts at the three learning levels and the performance of their combination by BKS and DS. The dynamic selection procedure, based on the a-priori information obtained by simulated data (Figs. 4,5), allows the selection of the best combination method (indicated with a grey background in Table 3), depending on the degree of correlation among the four classifiers. At the first level (ρ=0,84) the DS achieves the best result (R=95,3%). At the second and third learning levels (ρ=0,85 and ρ=0,87 respectively), BKS is the best (R=94,4% and R=93.3%, respectively).

Table 3. Dynamic Selection of Combination Method

Learning Level	ρ	Average Recognition Rate	Recognition Rate DS	Recognition Rate BKS
L1	0,84	89,5	95,3	94,5
L2	0,85	89,7	93,7	94,4
L3	0,87	89,8	92,5	93,3

6 Conclusion

This paper presents a new approach for the dynamic configuration of multi-classifier systems. Rules are discovered by the a-priori analysis of the combination methods and used, during the run time, to select dynamically the most effective combination method, depending on the degree of complementarity among the individual classifiers. The experimental results are promising and lead to continue researches in this direction.

Acknowledgements. This paper has been supported by the Italian Ministry "Ministero dell'Istruzione, dell'Università e della Ricerca ", MIUR under law 488, Project "Rete Puglia" CINI-BA - D.M. n.600, 11-11-1999.

References

1. Ley Xu, Adam Krzyzak, Ching Y-Suen, “Methods of Combining Multiple Classifiers and Their Applications to Handwriting Recognition” , IEEE Transaction on Systems, Man and Cybernetics- Vol. 22, N. 3, 1992, pp. 418–435.
2. T. Matsui, I. Yamashita, T. Wakahara, “Results of Second IPTP Character Recognition Competition and Studies on Multi-Expert Handwritten Numeral Recognition”, Proc. of IWFHR-4, 1994, pp.338–346.
3. Huang, C.Y. Suen, “An Optimal Method of Combining Multiple Classifiers for Unconstrained Handwritten Numeral Recognition”, Proc. of IWFHR-3, Buffalo, NY, 1993, pp. 11–20.
4. Y. Lu, F. Yamaoka, “Integration of Handwritten Digit Recognition results using Evidential Reasoning”, Proc. of IWFHR-4, 1994, pp. 456–463.
5. J. Kittler, M. Hatef, R.P.W. Duin, J. Matias, "On combining classifiers", *IEEE Trans. on Pattern Analysis Machine Intelligence*, Vol.20, no.3, pp.226–239, 1998.
6. G.Dimauro, S.Impedovo, G.Pirlo, “Multiple Experts:A New Methodology for the Evaluation of the Combination Processes”, IWFHR-5, Colchester,Uk,1996,pp.131–136.
7. V. Di Lecce, G.Dimauro, A. Guerriero, S.Impedovo, G.Pirlo, A. Salzo, “Classifier Combination: the role of a-priori knowledge”, *Proc. of IWFHR-7*, Sept. 2000, Amsterdam, The Netherlands, pp. 143–152.
8. S. Impedovo and A. Salzo, "A new methodology for expert combination in multi-expert system designing". In Lecture Notes in Computer Science, (vol. 1857), J. Kittler and F.Roli Eds., MCS2000, Cagliari, Italy, pp.230–239.
9. G. Dimauro, S. Impedovo, G. Pirlo, A. Salzo, “Automatic Bankchecks Processing: A New Engineered System”, *International Journal of Pattern Recognition and Artificial Intelligence*, Vol.11, N.4, World Scientific Publ., Singapore, 1997, pp.1–38.