SELECTION OF MULTIPLE CLASSIFIERS WITH NO PRIOR LIMIT TO THE NUMBER OF CLASSIFIERS BY MINIMIZING THE CONDITIONAL ENTROPY

Hee-Joong Kang
Department of Computer Engineering, Hansung University, 389 Samseon-dong 3-ga, Seongbuk-gu, Seoul, Korea

Keywords: Multiple classifier system, Multiple classifiers, Conditional entropy, Mutual information, Selection method.

Abstract: In addition to a study on how to combine multiple classifiers in multiple classifier systems, recently a study on how to select multiple classifiers from a classifier pool has been investigated, because the performance of multiple classifier systems depends on the selected classifiers as well as a combination method. Previous studies on the selection of multiple classifiers select a classifier set based on the assumption that the number of selected classifiers is fixed in advance, or based on the clustering followed the diversity criteria of classifiers in the classifier overproduce and choose paradigm. In this paper, by minimizing the conditional entropy which is the upper bound of Bayes error rate, a new selection method is considered and devised with no prior limit to the number of classifiers, as illustrated in examples.

1 INTRODUCTION

There have been a lot of studies to solve the pattern recognition problems with a multiple classifier system composed of multiple classifiers. The studies on a multiple classifier system have proceeded in both directions of how to combine multiple classifiers(Ho, 2002; Kang and Lee, 1999; Kittler et al., 1998; Saerens and Fouss, 2004; Woods et al., 1997) and how to select multiple classifiers from a classifier pool(Ho, 2002; Kang, 2005; Roli and Giacinto, 2002), because the performance of a multiple classifier system depends on the selected classifiers as well as a combination method. It is well known that a set of classifiers showing only high recognition rates is not superior to other set of classifiers in most cases(Roli and Giacinto, 2002; Woods et al., 1997), furthermore it was also reported that fewer classifiers provided superior results to more classifiers(Woods et al., 1997). Thus, it is well recognized that the selection of component classifiers in a multiple classifier system is one of very difficult research issues(Ho, 2002; Kang, 2005; Roli and Giacinto, 2002; Woods et al., 1997).

Previous studies on the selection of multiple classifiers select a classifier set based on the assumption that the number of selected classifiers is fixed in advance(Kang, 2005), or based on the clustering followed the diversity criteria of classifiers in the overproduce and choose paradigm without such assumption(Roli and Giacinto, 2002). The selected classifier set is used in a multiple classifier system and then multiple classifiers in the set are combined for fusing their classification results.

In this paper, in order to select multiple classifiers in a systematic way, by minimizing the conditional entropy which is the upper bound of Bayes error rate, a new selection method is considered and devised with no prior limit to the number of classifiers, as illustrated in examples. The proposed selection method is briefly compared to the previous selection methods.

2 RELATED WORKS

A preliminary study to select multiple classifiers was conducted on the assumption that the number of selected classifiers is fixed in advance. Under such an assumption, classifiers are ordered according to recognition rates or reliability rates, and then the classifiers are sequentially selected up to the fixed number from the best one. And also, measure of closeness and conditional entropy based on information theory were applied to select a promising classifier set among classifier sets composed of the fixed number of classifiers(Kang, 2005).

Based on the overproduce and choose paradigm

Kang H. (2009).
SELECTION OF MULTIPLE CLASSIFIERS WITH NO PRIOR LIMIT TO THE NUMBER OF CLASSIFIERS BY MINIMIZING THE CONDITIONAL ENTROPY.
In *Proceedings of the Fourth International Conference on Computer Vision Theory and Applications*, pages 356-361
DOI: 10.5220/0001799103560361

proposed by Partridge and Yates, a number of classifiers are created, and then promising classifiers are selected according to criteria(Roli and Giacinto, 2002). In their selection, a heuristic function is used to select the promising classifiers, or the best classifier is respectively selected according to its category of classifiers, or the clustering followed the diversity criteria of classifiers is applied to empirically decide the number of classifiers complementary to each other. Then, the selected classifiers are consisted of a classifier set in a multiple classifier system. By using the clustering, a prior limit to the number of classifiers is avoided and the number is decided. But, a classifier set by the clustering can not guarantee the best performance over others, so the number of selected classifiers still remains an unresolved issue.

Various diversity criteria of classifiers are GD (within-set generalization diversity) proposed by Partridge and Yates, Q statistic proposed by Kuncheva et al., CD (compound diversity) proposed by Roli and Giacinto, and mutual information between classifiers proposed by Kang(Kang, 2005; Roli and Giacinto, 2002). And additional diversity criteria of classifier sets used in the clustering to decide the number of classifiers are GDB (between-set generalization diversity) proposed by Partridge and Yates, and diversity function proposed by Roli and Giacinto(Roli and Giacinto, 2002).

Q statistic criteria are expressed as follows.

$$Q_{i,j} = \frac{N^{11}N^{00} - N^{01}N^{10}}{N^{11}N^{00} + N^{01}N^{10}} \quad (1)$$

$$Q_{i,j,k} = \frac{\sum_{a,b} Q_{a,b}}{{}_3C_2} \quad (2)$$

A diversity between two classifiers is calculated by the Eq. 1, and a diversity among three classifiers is calculated by the Eq. 2 which divides the sum of Q statistic $Q_{a,b}$ for a pair of classifiers by the combination for a pair of classifiers. Here, N^{11} is the number of data elements that both two classifiers correctly classify, and N^{00} is the number of data elements wrongly classified by both classifiers, and N^{10} or N^{01} is the number of data elements that either of classifiers correctly classifies.

CD criterion is based on the compound error probability for two classifiers E_i and E_j and it is expressed as follows when they are respectively the member of each classifier set A and B.

$$CD(E_i, E_j) = 1 - prob(E_i\ fails, E_j\ fails) \quad (3)$$

$$diversity(A, B) = \max_{E_i \in A, E_j \in B} \{CD(E_i, E_j)\} \quad (4)$$

3 SELECTION OF CLASSIFIERS BASED ON THE MINIMIZATION OF CONDITIONAL ENTROPY

In a Bayesian approach to deal with pattern recognition problems using a multiple classifier system, the upper bound of Bayes error rate P_e is defined like Eq. 5 by conditional entropy $H(L|E)$ with a label class L and a K classifiers group E. By minimizing such conditional entropy, the upper bound of Bayes error rate can be lowered and improvement on the performance of a multiple classifier system can be expected. From the definition of conditional entropy, class-decision (C-D) mutual information $U(L;E)$ is derived like Eq. 6. Lowering the upper bound of Bayes error rate means maximizing the C-D mutual information $U(L;E)$.

C-D mutual information is also used to optimally approximate high order probability distribution with a product of low order probability distributions based on the first-order or the second-order or the dth-order dependency, when there are a high order probability distribution $P(E,L)$ composed of a class label L and classifiers E, and a high order probability distribution $P(E)$ composed of only classifiers. Below expressions show the derivation of finding the optimal approximate probability distributions P_a based on the dth-order dependency from the C-D mutual information.

$$P_e \leq \frac{1}{2}H(L|E) = \frac{1}{2}[H(L) - U(L;E)] \quad (5)$$

$$\begin{aligned} U(L;E) &= \sum_e \sum_l P(e,l) \log \frac{P(e|l)}{P(e)} \\ &= \sum_e \sum_l P(e,l) \log \frac{\prod_{j=1}^{K} P(E_{n_j}|E_{n_{id(j)}}, \cdots, E_{n_{i1(j)}}, l)}{P(l)} \\ &\quad - \sum_e P(e) \log \prod_{j=1}^{K} P(E_{n_j}|E_{n_{id(j)}}, \cdots, E_{n_{i1(j)}}) \\ &= H(L) + \sum_{j=1}^{K} \Delta D(E_{n_j}; E_{n_{id(j)}}, \cdots, E_{n_{i1(j)}}, L) \end{aligned} \quad (6)$$

$$P_a(E_1, \cdots, E_K, L) = \prod_{j=1}^{K} P(E_{n_j}|E_{n_{id(j)}}, \cdots, E_{n_{i1(j)}}, L), \quad where (0 \leq id(j), \cdots, i1(j) < j) \quad (7)$$

$$P_a(E_1, \cdots, E_K) = \prod_{j=1}^{K} P(E_{n_j}|E_{n_{id(j)}}, \cdots, E_{n_{i1(j)}}), \quad where (0 \leq id(j), \cdots, i1(j) < j) \quad (8)$$

$$H(L) = -\sum_l P(l) \log P(l) \quad (9)$$

$$\Delta D(E_{n_j}; E_{n_{id(j)}}, \cdots, E_{n_{i1(j)}}, L) = D(E_{n_j}; E_{n_{id(j)}}, \cdots, E_{n_{i1(j)}}, L) - D(E_{n_j}; E_{n_{id(j)}}, \cdots, E_{n_{i1(j)}}) \quad (10)$$

$$D(E_{n_j};E_{n_{id(j)}},\cdots,E_{n_{i1(j)}},L)= \sum_e\sum_l P(e,l)\log\frac{P(E_{n_j}|E_{n_{id(j)}},\cdots,E_{n_{i1(j)}},l)}{P(E_{n_j})} \tag{11}$$

$$D(E_{n_j};E_{n_{id(j)}},\cdots,E_{n_{i1(j)}})= \sum_e P(e)\log\frac{P(E_{n_j}|E_{n_{id(j)}},\cdots,E_{n_{i1(j)}})}{P(E_{n_j})} \tag{12}$$

$$P(E_{n_j}|E_0,E_0,L)\equiv P(E_{n_j},L) \tag{13}$$

$$P(E_{n_j}|E_0,E_{n_{i\cdot(j)}},L)\equiv P(E_{n_j}|E_{n_{i\cdot(j)}},L) \tag{14}$$

$$P(E_{n_j}|E_0,E_{n_{i\cdot(j)}})\equiv P(E_{n_j},E_{n_{i\cdot(j)}}) \tag{15}$$

From Eqs. 5 and 6, for a given set of K classifiers, maximizing the C-D mutual information $U()$ leads to the decision of unknown permutation used in the optimal approximate probability distributions by maximizing the total sum of mutual information $\Delta D()$ like Eq. 10 by the dth-order dependency. Eqs. 13 to 15 are defined for confirming the property of probability and E_0 is a null term. After deciding the unknown permutation, an optimal product of low order distributions can be found.

A selection method based on the conditional entropy assumes that a classifier set having high C-D mutual information is better than other classifier sets, because the higher the C-D mutual information is the lower the upper bound of Bayes error rate is. So, a classifier set with the highest C-D mutual information is selected as the classifier set of a multiple classifier system. For example, for a classifier set composed of one classifier E_1, C-D mutual information is expressed as follows.

$$\begin{aligned} U(L;E_1) &= \sum_e\sum_l P(e_1,l)\log\frac{P(e_1|l)}{P(e_1)} \\ &= \sum_e\sum_l P(e_1,l)\log\frac{P(e_1,l)}{P(l)} - \sum_e P(e_1)\log P(e_1) \\ &= H(L)+H(E_1)+\sum_e\sum_l P(e_1,l)\log P(e_1,l) \end{aligned} \tag{16}$$

And for a classifier set composed of two classifiers E_1,E_2, C-D mutual information is expressed as follows.

$$\begin{aligned} U(L;E_1,E_2) &= \sum_e\sum_l P(e_1,e_2,l)\log\frac{P(e_1,e_2|l)}{P(e_1,e_2)} \\ &= \sum_e\sum_l P(e_1,e_2,l)\log\frac{P(e_1,e_2,l)}{P(l)} \\ &\quad -\sum_e P(e_1,e_2)\log P(e_1,e_2) \\ &= H(L)+H(E_1,E_2) \\ &\quad +\sum_e\sum_l P(e_1,e_2,l)\log P(e_1,e_2,l) \end{aligned} \tag{17}$$

As mentioned above, for a classifier set composed of K classifiers $E_1,\cdots,E_K$, C-D mutual information is expressed as follows without probability approximation.

$$\begin{aligned} U(L;E_1,\cdots,E_K) &= \sum_e\sum_l P(e_1,\cdots,e_K,l)\log\frac{P(e_1,\cdots,e_K|l)}{P(e_1,\cdots,e_K)} \\ &= \sum_e\sum_l P(e_1,\cdots,e_K,l)\log\frac{P(e_1,\cdots,e_K,l)}{P(l)} \\ &\quad -\sum_e P(e_1,\cdots,e_K)\log P(e_1,\cdots,e_K) \\ &= H(L)+H(E_1,\cdots,E_K) \\ &\quad +\sum_e\sum_l P(e_1,\cdots,e_K,l)\log P(e_1,\cdots,e_K,l) \end{aligned} \tag{18}$$

Therefore, possible classifier sets can be built by increasing the number of classifiers to be added from one, and then for a classifier set, when C-D mutual information $U()$ can be computed, if there is no meaningful increment on the C-D mutual information $U()$, then adding a classifier to the classifier set is useless because there is no lowering at the upper bound of Bayes error rate. That is, no more classifiers addible is considered in building a classifier set. A classifier set having the maximum C-D mutual information is selected as the classifier set of a multiple classifier system.

4 EXAMPLE OF MULTIPLE CLASSIFIER SYSTEMS AND ANALYSIS OF SELECTION METHODS

In this section, four examples are shown to illustrate why multiple classifier systems are useful and how a conditional entropy-based selection method affects the performance of multiple classifier systems. Five label classes are {A, B, C, D, E}, and then let us suppose that training data and test data are the same as the label classes. And also, it is supposed that five imaginary classifiers are {Ea, Eb, Ec, Ed, Ee}. Combination methods to combine classifiers are voting method and conditional independence assumption-based Bayesian (CIAB) method commonly used(Kang, 2005; Roli and Giacinto, 2002).

The first example, EX-1, supposes that five imaginary classifiers showing 20% recognition rate show recognition results like Table 1 for the five test data. In this example, classifier sets composed of one to five classifiers are considered. Under these circumstances, C-D mutual information computed by using Eqs. 16 to 18 are shown in Table 2. Additionally, Q statistic is shown in column 'Q' and CD diversity is shown in column 'CD'. A plausible imaginary highest recognition rate is denoted by 'pl' in Table 3. The reason why a classifier set composed of five classifiers can have 100% pl rate assumes that a correct classifier is rightly selected to decide its decision for each data. For a classifier set composed of two classifiers, three possible recognition rates in voting method is that the recognition rates depend on how to deal with tie votes. This is the same situation as CIAB method.

Table 1: Recognition result of EX-1 example.

data	classifier				
	Ea	Eb	Ec	Ed	Ee
A	A	B	C	D	E
B	A	B	C	D	E
C	A	B	C	D	E
D	A	B	C	D	E
E	A	B	C	D	E

Table 2: Diversity result of a classifier set of the EX-1 example.

no. of classifiers	U(L;E)	Q	CD
1	0	-	-
2	0	-1	0.4
3	0	-1	0.4
4	0	-1	0.4
5	0	-1	0.4

Table 3: Recognition performance (%) of a classifier set of the EX-1 example.

no. of classifiers	pl	voting	CIAB
1	20	20	0,20
2	40	0,20,40	0,20
3	60	0	0,20
4	80	0	0,20
5	100	0	0,20

From the Table 2, $U(L;E)$ is computed to 0 (e.g., $\frac{1}{5} * \log \frac{1}{5} * 5 - \log \frac{1}{5} = 0$) regardless of the number of classifiers in a classifier set by using the real probability distributions. And, Q statistic is -1 and CD is 0.4 in all cases except that the number of classifier is one. Q and CD can not be computed when the number of classifier is one, so '-' means unavailable. For a classifier set composed of more than three classifiers, voting method shows 0% and CIAB method shows at most 20% recognition rate. Although the above example is simple and extreme, there is no positive performance by the well representative combination methods, voting and CIAB. If there is an oracle to select an appropriate classifier for a given input, then it is possible to show 100% recognition rate as shown in column 'pl'.

The second example, EX-2, supposes that five imaginary classifiers showing 40% recognition rate show recognition results like Table 4 for the five test data. Classifier sets composed of one to five classifiers are considered. Under these circumstances, diversity calculations are shown in Table 5 and performance results are in Table 6. The reason why a classifier set composed of four classifiers has 100% pl rate assumes that 60% recognition rate is basic for the four classifiers and for remaining rate a correct classifier is rightly selected to decide its decision.

Table 4: Recognition result of EX-2 example.

data	classifier				
	Ea	Eb	Ec	Ed	Ee
A	A	B	C	D	A
B	B	B	C	D	E
C	A	C	C	D	E
D	A	B	D	D	E
E	A	B	C	E	E

Table 5: Diversity result of a classifier set of the EX-2 example.

no. of classifiers	U(L;E)	Q	CD
1	0.5004	-	-
2	0.9503	$\frac{1}{3}$	0.6
3	1.3322	$\frac{1}{3}$	0.6
4	1.6094	$\frac{1}{3}$	0.6
5	1.6094	$\frac{1}{3}$	0.6

Table 6: Recognition performance (%) of a classifier set of the EX-2 example.

no. of classifiers	pl	voting	CIAB
1	40	40	40
2	60	20,40,60	40,60,80,100
3	80	40	60,80,100
4	100	60	100
5	100	100	100

From the Table 5, for a classifier set of one classifier, $U(L;E)$ is approximately computed to 0.5004 (e.g., $-(\frac{4}{5} * \log \frac{4}{5} + \frac{1}{5} * \log \frac{1}{5}) \approx 0.5004$), for a classifier set of two classifiers $U(L;E)$ is approximately computed to 0.9503 (e.g., $-(\frac{3}{5} * \log \frac{3}{5} + \frac{1}{5} * \log \frac{1}{5} * 2) \approx 0.9503$), for a classifier set of three classifiers $U(L;E)$ is approximately computed to 1.3322 (e.g., $-(\frac{2}{5} * \log \frac{2}{5} + \frac{1}{5} * \log \frac{1}{5} * 3) \approx 1.3322$), and for a classifier set of four classifiers $U(L;E)$ is approximately computed to 1.6094 (e.g., $-(\frac{1}{5} * \log \frac{1}{5} * 5) \approx 1.6094$), by using the real probability distributions from the data. And, Q statistic is $\frac{1}{3}$ and CD is 0.6 in all cases except that the number of classifier is one. For a classifier set composed of more than four classifiers, while voting method shows 60%, CIAB method shows 100% recognition rate. Voting method is perfect with all five classifiers, but CIAB method is better than voting method because the CIAB method is perfect with at least four classifiers. And, it is recognized that C-D mutual information tends to show as similar as the performance of CIAB method, but Q and CD did not show any meaningful indication on the selection of classifiers. Therefore, when CIAB method is used as the combination method of a multiple classifier system, the number of classifiers can be decided with reference to C-D mutual information of them, and the number of selected classifiers is four.

The third example, EX-3, supposes that five imaginary classifiers showing 60% recognition rate show recognition results like Table 7 for the five test data. Classifier sets composed of one to five classifiers are considered. Under these circumstances, diversity calculations are shown in Table 8 and performance results are in Table 9. The reason why a classifier set composed of three classifiers has 100% pl rate assumes that 60% recognition rate is basic for the three classifiers and for remaining rate a correct classifier is rightly selected to decide its decision.

Table 7: Recognition result of EX-3 example.

data	classifier				
	Ea	Eb	Ec	Ed	Ee
A	A	B	C	A	A
B	B	B	C	D	B
C	C	C	C	D	E
D	A	D	D	D	E
E	A	B	E	E	E

Table 8: Diversity result of a classifier set of the EX-3 example.

no. of classifiers	U(L;E)	Q	CD
1	0.9503	-	-
2	1.3322	$\frac{1}{3}$	0.8
3	1.6094	$\frac{1}{3}$	0.8
4	1.6094	$\frac{1}{3}$	0.6
5	1.6094	$\frac{1}{3}$	0.6

Table 9: Recognition performance (%) of a classifier set of the EX-3 example.

no. of classifiers	pl	voting	CIAB
1	60	60	40,60
2	80	40,60,80	60,80,100
3	100	60	100
4	100	100	100
5	100	100	100

From the Table 8, for a classifier set of one classifier, $U(L;E)$ is approximately computed to 0.9503 (e.g., $-(\frac{3}{5}*\log\frac{3}{5}+\frac{1}{5}*\log\frac{1}{5}*2)\approx 0.9503$), for a classifier set of two classifiers $U(L;E)$ is approximately computed to 1.3322 (e.g., $-(\frac{2}{5}*\log\frac{2}{5}+\frac{1}{5}*\log\frac{1}{5}*3)\approx 1.3322$), and for a classifier set of three classifiers $U(L;E)$ is approximately computed to 1.6094 (e.g., $-(\frac{1}{5}*\log\frac{1}{5}*5)\approx 1.6094$), by using the real probability distributions from the data. And, Q statistic is $\frac{1}{3}$ and CD is 0.8 in all cases except that the number of classifier is one. For a classifier set composed of more than three classifiers, while voting method shows 60%, CIAB method shows 100% recognition rate. Voting method is perfect with at least four classifiers, but CIAB method is better than voting method because the CIAB method is perfect with at least three classifiers. And, it is recognized that C-D mutual information tends to show as similar as the performance of CIAB method, but Q and CD did not show any meaningful indication on the selection of classifiers. Therefore, when CIAB method is used as the combination method of a multiple classifier system, the number of classifiers can be decided with reference to C-D mutual information of them, and the number of selected classifiers is three.

The fourth example, EX-4, supposes that five imaginary classifiers showing 80% recognition rate show recognition results like Table 10 for the five test data. Classifier sets composed of one to five classifiers are considered. Under these circumstances, diversity calculations are shown in Table 11 and performance results are in Table 12. The reason why a classifier set composed of two classifiers has 100% pl rate assumes that 60% recognition rate is basic for the two classifiers and for remaining rate a correct classifier is rightly selected to decide its decision.

Table 10: Recognition result of EX-4 example.

data	classifier				
	Ea	Eb	Ec	Ed	Ee
A	A	B	A	A	A
B	B	B	C	B	B
C	C	C	C	D	C
D	D	D	D	D	E
E	A	E	E	E	E

From the Table 11, for a classifier set of one classifier, $U(L;E)$ is approximately computed to 1.3322 (e.g., $-(\frac{2}{5}*\log\frac{2}{5}+\frac{1}{5}*\log\frac{1}{5}*3)\approx 1.3322$), and for a classifier set of two classifiers $U(L;E)$ is approximately computed to 1.6094 (e.g., $-(\frac{1}{5}*\log\frac{1}{5}*5)\approx$ 1.6094), by using the real probability distributions from the data. And, Q statistic is -1 and CD is 1 in all cases except that the number of classifier is one. For a classifier set composed of more than two clas-

Table 11: Diversity result of a classifier set of the EX-4 example.

no. of classifiers	U(L;E)	Q	CD
1	1.3322	-	-
2	1.6094	-1	1
3	1.6094	-1	1
4	1.6094	-1	1
5	1.6094	-1	1

Table 12: Recognition performance (%) of a classifier set of the EX-4 example.

no. of classifiers	pl	voting	CIAB
1	80	80	60,80
2	100	60,80,100	100
3	100	100	100
4	100	100	100
5	100	100	100

sifiers, while voting method shows three rates and at most 100%, CIAB method shows 100% recognition rate. Voting method is perfect with at least three classifiers, but CIAB method is better than voting method because the CIAB method is perfect with at least two classifiers. And, it is recognized that C-D mutual information tends to show as similar as the performance of CIAB method, but Q and CD did not show any meaningful indication on the selection of classifiers. Therefore, when CIAB method is used as the combination method of a multiple classifier system, the number of classifiers can be decided with reference to C-D mutual information of them, and the number of selected classifiers is two.

5 DISCUSSION

The minimization of conditional entropy was previously applied to approximate the high order probability distribution with the product of low order probability distributions, but in this paper it is tried to decide the number of classifiers in a classifier set with no prior limit and its promising usefulness was shown in simple and obvious examples, by compared with Q statistic and CD diversity. And also, in a classifier set by the minimization of control entropy, CIAB method is better than voting method. Even although high order probability distribution was directly used in computing the C-D mutual information without approximation in the simple examples, consideration on the approximation of high order probability distributions will be needed because it is often hard to directly obtain actual high order probability distributions. As one of further works, the selection method proposed in this paper will deeply be analyzed and compared by previously proposed selection methods such as Q statistic and CD diversity in other literature with a variety of examples and real pattern recognition problems.

REFERENCES

Ho, T. K. (2002). Chapter 7 Multiple Classifier Combination: Lessons and Next Steps. In Bunke, H. and Kandel, A., editors, *Hybrid Methods in Pattern Recognition*, pages 171–198. World Scientific.

Kang, H.-J. (2005). Selection of Classifiers using Information-Theoretic Criteria. *Lecture Notes in Computer Science*, 3686:478–487.

Kang, H.-J. and Lee, S.-W. (1999). Combining Classifiers based on Minimization of a Bayes Error Rate. In *Proc. of the 5th ICDAR*, pages 398–401.

Kittler, J., Hatef, M., Duin, R. P. W., and Matas, J. (1998). On Combining Classifiers. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20(3):226–239.

Roli, F. and Giacinto, G. (2002). Chapter 8 Design of Multiple Classifier Systems. In Bunke, H. and Kandel, A., editors, *Hybrid Methods in Pattern Recognition*, pages 199–226. World Scientific.

Saerens, M. and Fouss, F. (2004). Yet Another Method for Combining Classifiers Outputs: A Maximum Entropy Approach. *Lecture Notes in Computer Science*, 3077:82–91.

Woods, K., Kegelmeyer Jr., W. P., and Bowyer, K. (1997). Combinition of Multiple Classifiers Using Local Accuracy Estimates. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(4):405–410.