Comparison of Fusion Algorithms Based on Logistic Model of Correlated Classifiers

S. Y. Sohn
Department of Computer Science
& Industrial Systems Engineering,
Yonsei University, 134 Shinchon-dong,
Seoul 120-749
Republic of Korea
sohns@yonsei.ac.kr

Y. S. Kim
Department of Computer Science
& Industrial Systems Engineering,
Yonsei University, 134 Shinchon-dong,
Seoul 120-749
Republic of Korea
yoonseong@yonsei.ac.kr

Abstract **-** *This study compares the classification ability of various fusion algorithms (average, majority vote, median, max/min) when individual classifiers are potentially correlated. A logistic transformation of multivariate normal distribution (MVN) is used to generate the posterior probability estimates, assuring that the probability exists between 0 and 1. With varying parameters of MVN and number of classifiers, we measure the relative performance of the fusion algorithms to that of single classifier. Our results can be utilized for the selection of the most effective fusion method for given situation.*

Keywords**:** Classifier fusion method, Posterior probability estimates, Correlation, Logistic model, Multivariate normal distribution.

1 Introduction

The research on classifier combination has recently been considered with great interest [1,2-9]. It is because combining the results of classifiers is expected to provide greater performance improvement than that of a single classifier [10]. In order to combine classifier outputs, average, majority vote, maximum, minimum and median approaches are frequently used. Performances of these fusion methods vary depending on the diversity in the classifier team.

The distribution of posterior probability estimates were assumed as both truncated normal and uniform distributions in [11]. It compared the performance of the product, sum, maximum, minimum, median and majority vote methods to that of individual classifier, with different mean, standard deviation, and the number of classifiers thorough Monte Carlo simulation.

The classification ability of several fusion methods were analytically derived in [1] based on the assumption that the distribution of posterior probability estimates follows either truncated normal or uniform distribution. However, the truncation was not really considered in derivation and the author justified this choice.

Combining multiple probability distributions representing potentially dependent expert opinion was reviewed. However, it was not done in the context of posterior probability of classifiers which varies within 0 to 1 [7].

As reviewed, most of the former studies did not consider dependent classifiers in the context of the distribution of the posterior probability. Independence is a very restrictive and admittedly unrealistic assumption.

In this study, we compare the performance of the several classifier fusion methods by considering the situation where individual classifiers are potentially correlated. Multivariate normal (MVN) distribution is used to generate the correlated structure and subsequently logistic transformation is applied to the generated MVN to create the posterior probabilities of individual classifiers assuring that such probabilities exist between 0 and 1.

With the generated probabilities we measure the improved performance of the fusion methods in comparison of the single classifier based on the designed experiment. Factors used are the parameters of MVN (mean, variance, and correlation), the number of classifiers and types of fusion methods. With these five factors and corresponding levels for each factor, we perform Monte Carlo simulation. We expect that as a result of our study one can select the best fusion method for given situation.

The organization of this paper is as follows. In section 2, we conduct a simulation based on the designed experiment in order to compare the performance of the several classification fusion methods. The study results are explained in section 3. Finally, in section 4, summary is given along with further study areas.

2 Monte Carlo Simulation

We consider the problem of binary classification where two classes are labeled as $\{w_0, w_1\}$. Let P_i be the posterior probability $P(w_1|x)$ given by classifier i for an input x. Then, the probabilities of L number of classifiers judging a case as class w_1 are $P_1, P_2, \cdots, P_i, \cdots, P_L$. When they are applied to the fusion method of classifiers S, the probability of selecting class w_1, denoted $\hat{P}_1$, is

$$\hat{P}_1 = S(P_1, P_2, \cdots, P_i, \cdots, P_L) \quad (1)$$

The fusion method (S) we consider includes average, median, majority vote, maximum and minimum methods. As for the average method, if the average of P_i s of L classifiers is greater than 0.5, then class w_1 is selected, and otherwise class w_0 is selected. Similarly, for the median method, if the median of P_i s of L classifiers is greater than 0.5, then class w_1 is selected, and otherwise class w_0 is selected. The minimum method is as follows. First, we select the smallest P_i value and the smallest $1-P_i$ value. Second, the two minimums are compared and a class with the larger value is selected. If the two minimums are the same, then the second smallest values are compared. The maximum method is the same as the minimum except that the biggest values are compared. In the binary classification, the median and majority vote methods are considered the same. The minimum and the maximum methods are also considered the same [1].

Based on the definition of the fusion methods, an error occurs if $\hat{P}_1$ is less than 0.5. However, for the maximum method, an error will occur if

$$\max_i \{P_i\} < \max_i \{1 - P_i\}. \quad (2)$$

For the minimum fusion method, an error will occur if

$$\min_i \{P_i\} < \min_i \{1 - P_i\}. \quad (3)$$

Including [1], most studies assumed that $P_1, P_2, \cdots, P_i, \cdots, P_L$ are independent of each other. In our study we consider they are correlated. So we first generate an L × 1 vector $\boldsymbol{\varepsilon} = [\varepsilon_1, \varepsilon_2, ..., \varepsilon_L]^T$ which follows a multivariate normal distribution representing the pattern of correlated posterior probabilities:

$$\boldsymbol{\varepsilon} \sim MVN(\boldsymbol{\mu}, \sigma^2 \Omega) \quad (4)$$

where $\boldsymbol{\mu} = [\mu, \mu, ..., \mu]^T$ is an L × 1 mean vector and σ^2 is the variance and Ω is an L × L correlation matrix.

Next, in order to guarantee correlated P_i being between 0 and 1, we apply a logistic model based on the multivariate normal distributed ε_i as follows:

$$P_i = \frac{\exp(\varepsilon_i)}{1 + \exp(\varepsilon_i)}. \quad (5)$$

In this way, $\boldsymbol{\mu}$ can be related to the mode of posterior probabilities of individual classifiers and σ^2 controls the dispersion of individual P_i's. Ω indirectly decides the degree of correlation between any set of two classifiers.

In Kuncheva (2002), error rate of each classifier fusion method is analytically derived from the probability density functions (pdf). However, one cannot obtain the pdf for multivariate normal distribution which is transformed by logistic model. Thus we need simulation to compare the performance of the fusion methods such as average, max/min, and median.

In order to compare the performance of the fusion methods in various situations, we conduct a factorial design using five factors. Factors used are the μ, σ^2, Ω, the number of classifiers (L) and the fusion methods (S). The levels of each factor are as shown in Table 1.

The mean of the ε_i s is chosen as 0.4, 0.8 and 1.5 because transformed probability values by logistic model are about 0.6, 0.7 and 0.8 respectively. They would be the mode of transformed distribution. The variance of the ε_i s is chosen as 0.4, 1 and 7. When the variance is larger than 7, most of the probability is densely concentrated near 0 and 1. When the variance is less than 0.4, the probabilities are densely aggregated around the mean of probabilities. The levels of correlation of classifiers are chosen as 'high positive,' 'weak positive' and 'high negative.' In case of 'high positive,' all the correlation coefficients in correlation matrix (Ω) are 0.9 while the correlation coefficients for 'weak positive' are 0.1. In case of 'high negative,' all correlation coefficients can not be negative because the determinant of the correlation matrix (Ω) must be positive. Thus, we simulate that a half of the classifiers are correlated with each other with the degree of 0.9 and another half of the classifiers are correlated with the degree of -0.9 each other.

A response variable for the factorial design is the improvement of each fusion method's performance compared to that of individual classifier. The performance is calculated as the ratio of the number of cases classified as class w_1 among a thousand of samples.

Table 1 : 3^5 Factorial Design

Factors	Level 1	Level 2	Level 3
Mean (Probability)	0.4 (0.6)	0.8 (0.7)	1.5 (0.8)
Variance	0.4	1	7
Number of classifiers	3	9	49
Correlation of classifiers	High positive (0.9)	Weak positive (0.1)	High negative (-0.9 or 0.9)
Classifier combination method	Average	Max/Min	Median

3 Results

To determine which factors are significant on the changes of the improvement, we conducted an analysis of variance (ANOVA). For this analysis each treatment of the 3^5 factorial design was replicated 100 times each. According to the result of ANOVA, all the interaction effects were significant.

In order to detect which fusion method is significantly different from the others, multiple comparisons are necessary. For that purpose, we performed Duncan test. [12] Because there were too many combinations to conduct Duncan test, we divided whole dataset into three subgroups based on the 'the number of classifiers' factor. For three levels of 'the number of classifiers' we performed the Duncan test. Table 2 shows the best fusion method for each situation. 'X' in Table 2 indicates that the performances of three fusion methods are not significantly different.

Table 2-1: The best performance fusion method in each case, when the correlation is high negative

When the number of classifiers is								
(i) 3			(ii) 9			(iii) 49		
Variance	Mean	Method	Variance	Mean	Method	Variance	Mean	Method
0.4	0.4 (0.6)	Average	0.4	0.4 (0.6)	Average & Min/Max	0.4	0.4 (0.6)	Average & Min/Max
0.4	0.8 (0.7)	Average & Min/Max	0.4	0.8 (0.7)	Average & Min/Max	0.4	0.8 (0.7)	Average & Min/Max
0.4	1.5 (0.8)	X	0.4	1.5 (0.8)	Average & Min/Max	0.4	1.5 (0.8)	X
1	0.4 (0.6)	Average	1	0.4 (0.6)	Min/Max	1	0.4 (0.6)	Average & Min/Max
1	0.8 (0.7)	Average & Min/Max	1	0.8 (0.7)	Average & Min/Max	1	0.8 (0.7)	Average & Min/Max
1	1.5 (0.8)	Average & Min/Max	1	1.5 (0.8)	Average & Min/Max	1	1.5 (0.8)	Average & Min/Max
7	0.4 (0.6)	Min/Max	7	0.4 (0.6)	Min/Max	7	0.4 (0.6)	Average
7	0.8 (0.7)	Min/Max	7	0.8 (0.7)	Min/Max	7	0.8 (0.7)	Min/Max
7	1.5 (0.8)	Min/Max	7	1.5 (0.8)	Min/Max	7	1.5 (0.8)	Min/Max

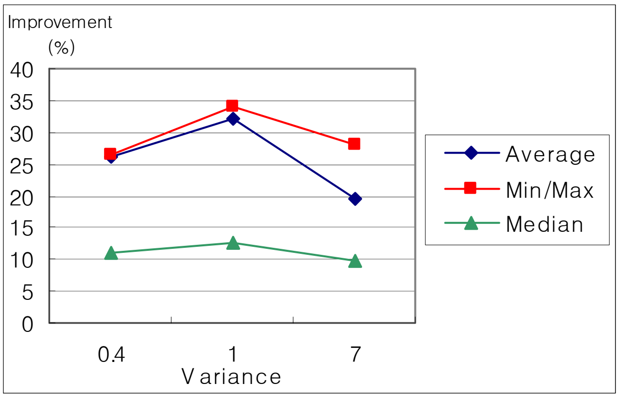
In order to detect the pattern of the performance of the fusion methods in a better manner, we use Figures 1, 2 and 3 for the nine classifier case. For different number of classifiers (e.g., three and forty nine), different patterns were found but we present and analyze the patterns in case of nine classifiers as an example. Fig. 1 depicts the changes in the relative performance improvement for varying levels of variance for given means ($\mu = 0.4, 0.8, \textit{and } 1.5$) when individual classifiers have high negative correlation.

Table 2-2: The best performance fusion method in each case, when the correlation is weak positive

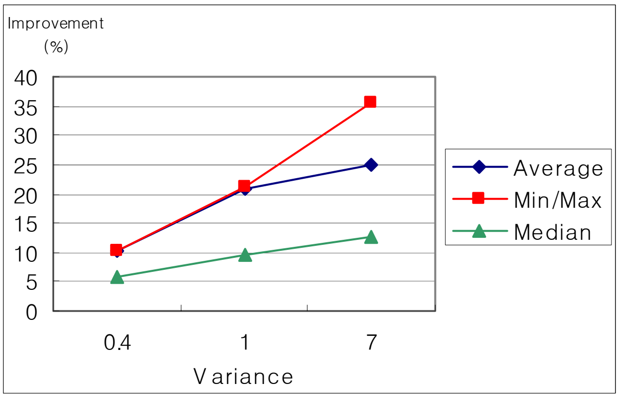
When the number of classifiers is								
(i) 3			(ii) 9			(iii) 49		
Variance	Mean	Method	Variance	Mean	Method	Variance	Mean	Method
0.4	0.4 (0.6)	Average	0.4	0.4 (0.6)	Average	0.4	0.4 (0.6)	Average
0.4	0.8 (0.7)	Average	0.4	0.8 (0.7)	Average	0.4	0.8 (0.7)	X
0.4	1.5 (0.8)	Average & Min/Max	0.4	1.5 (0.8)	X	0.4	1.5 (0.8)	X
1	0.4 (0.6)	Average	1	0.4 (0.6)	Average	1	0.4 (0.6)	Average
1	0.8 (0.7)	Average	1	0.8 (0.7)	Average	1	0.8 (0.7)	Average
1	1.5 (0.8)	Average	1	1.5 (0.8)	X	1	1.5 (0.8)	X
7	0.4 (0.6)	Average & Min/Max	7	0.4 (0.6)	Average	7	0.4 (0.6)	Average
7	0.8 (0.7)	Min/Max	7	0.8 (0.7)	Average	7	0.8 (0.7)	Average
7	1.5 (0.8)	Average & Min/Max	7	1.5 (0.8)	Average	7	1.5 (0.8)	Average

Table 2-3: The best performance fusion method in each case, when the correlation is high positive

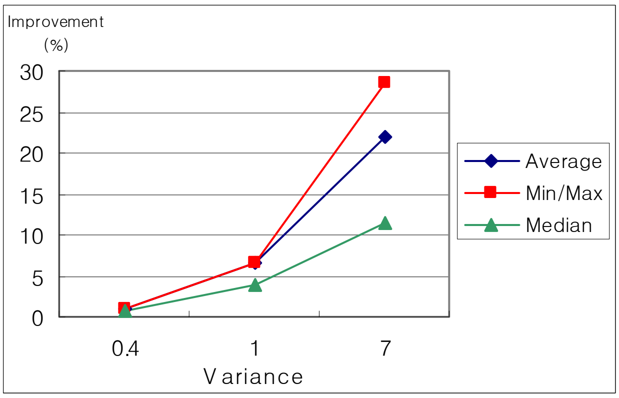
When the number of classifiers is								
(i) 3			(ii) 9			(iii) 49		
Variance	Mean	Method	Variance	Mean	Method	Variance	Mean	Method
0.4	0.4 (0.6)	X	0.4	0.4 (0.6)	Average & Min/Max	0.4	0.4 (0.6)	Average
0.4	0.8 (0.7)	Average & Min/Max	0.4	0.8 (0.7)	Average	0.4	0.8 (0.7)	Average & Median
0.4	1.5 (0.8)	X	0.4	1.5 (0.8)	X	0.4	1.5 (0.8)	X
1	0.4 (0.6)	Min/Max	1	0.4 (0.6)	X	1	0.4 (0.6)	Average & Median
1	0.8 (0.7)	Min/Max	1	0.8 (0.7)	Average & Median	1	0.8 (0.7)	X
1	1.5 (0.8)	Min/Max	1	1.5 (0.8)	Average & Median	1	1.5 (0.8)	Average & Median
7	0.4 (0.6)	Average	7	0.4 (0.6)	Average & Median	7	0.4 (0.6)	X
7	0.8 (0.7)	Min/Max	7	0.8 (0.7)	Average	7	0.8 (0.7)	Median
7	1.5 (0.8)	Average & Min/Max	7	1.5 (0.8)	Average	7	1.5 (0.8)	X



(a) When the mean is 0.4 (0.6)

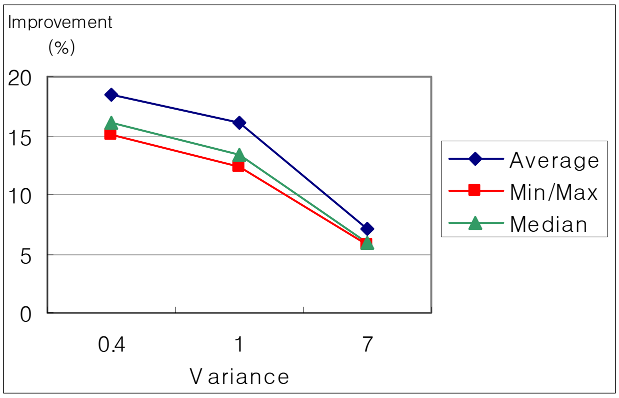


(b) When the mean is 0.8 (0.7)

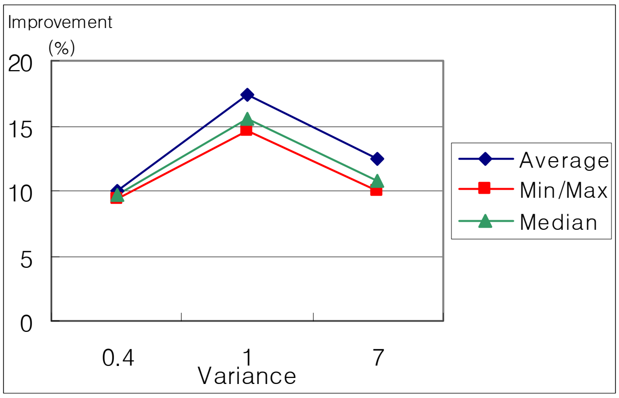


(c) When the mean is 1.5 (0.8)

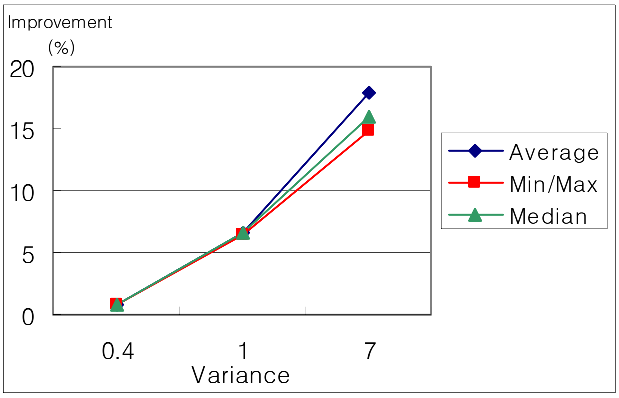
Fig. 1. The change of performance improvement of fusion methods according to the change of the variance when the correlation is high negative and the number or classifiers is 9.



(a) When the mean is 0.4 (0.6)

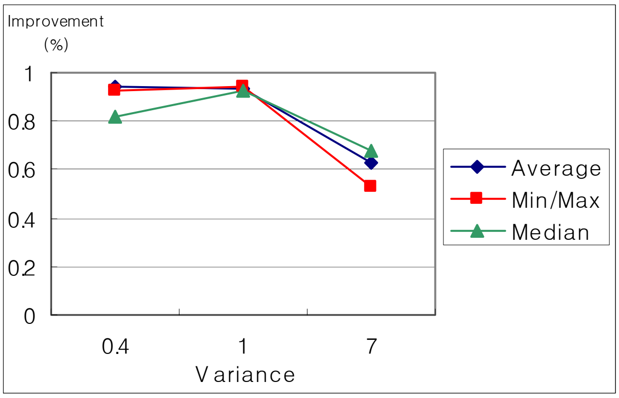


(b) When the mean is 0.8 (0.7)

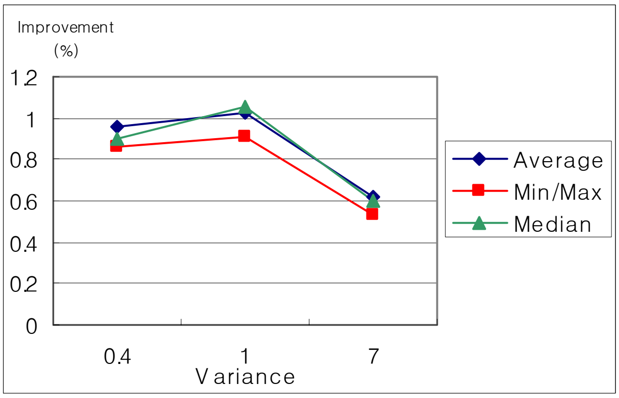


(c) When the mean is 1.5 (0.8)

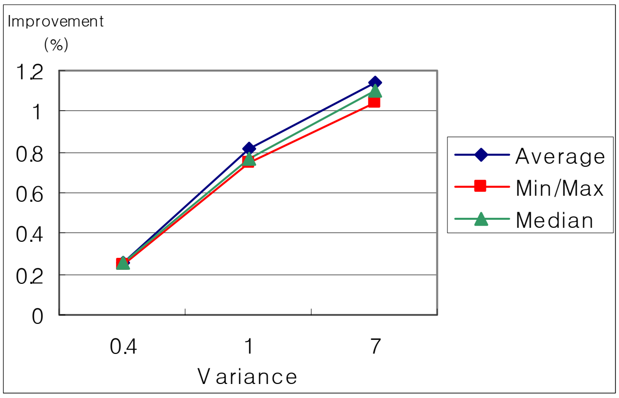
Fig. 2. The change of performance improvement of fusion methods according to the change of the variance when the correlation is weak positive and the number of classifiers is 9.



(a) When the mean is 0.4 (0.6)



(b) When the mean is 0.8 (0.7)



(c) When the mean is 1.5 (0.8)

Fig. 3. The change of performance improvement of fusion methods according to the change of the variance when the correlation is high positive and the number of classifiers is 9.

In general, fusion methods are better than an individual classifier. Improvement level is highest when classifiers are negatively correlated, while much improvement is not achieved when they are positively correlated. When classifiers are negatively correlated, median method is the worst approach; Min/max is recommended over the average method when the variance is large. Also, when the number of classifiers is nine, it is interesting to note that the performances of fusion methods increase as the variance increases when the mean of posterior probabilities is 0.8 as shown in Figures 1(c), 2(c) and 3(c). However, when the mean is not large, medium level of variance usually provides better performance (see Figures 1(a), 3(a) and (b)).

When the number of classifiers is three, like the case of 9 classifiers, median is generally the worst method in all situations while average and min/max are almost same in the improvement of the performance. In addition, the performances of fusion methods increase as the variance increases with the large mode of probabilities (0.8).

However, when the number of classifiers is 49, min/max method is the worst in almost all situations except when the classifiers are negatively correlated. In case of negative correlation, min/max is the best like other cases of 3 and 9 classifiers and the performances of fusion methods increase as the variance increase. However, medium level of the variance provides better performances when the classifiers are high-positively correlated.

Among our results of ANOVA and Duncan test, the followings turn out to be the same as those of Kuncheva (2002) assuming that it corresponds to our weak positive correlation case:

1. When the individual classifiers are independent (weak positive correlated), the average method generally outperforms the min/max and the median/vote method.
2. The more classifiers we have in the team, the lower the error.
3. As the mean of the probabilities increases and the variance of the probabilities decreases, the error rate of the fusion methods decreases.

Meanwhile, the results which could not be obtained by the study of Kuncheva (2002) can be summarized as follows:

1. When individual classifiers have high negative correlation, the average and the min/max outperform the median. The min/max outperforms the average especially as the variance of the posterior probabilities increases.
2. When the correlation is weak positive or high negative, the extent of the improvement of fusion methods against single classifier increases as the number of classifiers increases.
3. When the correlation of the probabilities is negative, there is a great deal of achievement compared to the positively relates case.

For the practical use of our study the classification probabilities of classifiers should be sampled many times. From the distribution of the classification probabilities which a classifier generates, we can infer the mean (μ) and the variance (σ^2). Also, the correlation of the classifiers (Ω) can be inferred from the distribution of the probabilities which a team of classifiers generates.

4. Conclusion

The classification ability of the classifier fusion methods was compared and analyzed. In order to solve the problem of clipped normal distribution and the limitation of independence assumption regarding individual posterior probabilities, we applied logistic model for multivariate normal distribution. Like the study of Kuncheva (2002) when the correct classification probabilities are independent, the average method mainly outperforms min/max and median vote. However, the performance of the fusion methods except the oracle is not significantly different when the correlation of the probabilities is high positive. Also, the classification accuracy of the min/max method was better than average and median vote method when the probabilities are negatively correlated and especially when the variance of the probabilities is high.

In addition to these patterns, we obtained the best performance fusion method in each case which is represented as the combination of four factors: mean, variance, correlation of the probabilities and the number of the classifiers. These results can be utilized to select the most effective fusion method in a given situation.

References

[1] L.I. Kuncheva. A theoretical study on six classifier fusion strategies. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(2), 281-286, 2002.

[2] T. Ho, J. Hull, and S. Srihari. Decision combination in multiple classifier systems. *IEEE Trans. Pattern Anal. Mach. Intell.*, 16 , 66-75, 1994.

[3] J. Kittler, M. Hatef, R.P.W. Duin, and J. Matas. On combining classifiers. *IEEE Trans. Pattern Anal. Mach. Intell.*, 20(3), 226-239, 1998.

[4] K. Chen, and H. Chi. A method of combining multiple probabilistic classifiers through soft competition on different feature sets. *Neural Comput.*, 20, 227-252, 1998.

[5] K. Al-Ghoneim, and B.V.K.V. Kumar. Unified decision combination framework. *Pattern Recognition Letters*, 31(12), 2077-2089, 1998.

[6] I. Bloch. Information combination operators for data fusion: a comparative review with classification. *IEEE Trans. Systems Man Cybernet*, 26 (1), 52-67, 1996

[7] R.A. Jacobs. Methods of combining experts' probability assessments, *Neural Comput.*, 7, 865-888, 1995.

[8] L. Lam, and C.Y. Suen. Optimal combinations of pattern classifiers. *Pattern Recognition Letters*, 16, 945-954, 1995.

[9] H. Altincay, and M. Demirekler. An information theoretic framework for weight stimation in the combination of probabilistic classifiers for speaker identification. *Speech Commun.*, 30(4), 255-272, 2000.

[10] M. Demirekler, and H. Altincay. Plurality voting-based multiple classifier systems: statistically independent with respect to dependent classifier sets, *Pattern Recognition Letters*, 35,2365-2379, 2000.

[11] F.M. Alkoot, and J. Kittler. Experimental evaluation of expert fusion strategies. *Pattern Recognition Letters*, 20, 1361-1369, 1999.

[12] D.C. Montgomery. *Design and Analysis of Experiments*. Wiley, NJ, 2000.