Application of Dempster-Shafer theory in condition monitoring applications: A case study

Chinmay R. Parikh, Michael J. Pont[1] and N. Barrie Jones

Control & Instrumentation Research Group
Department of Engineering
University of Leicester
University Road
LEICESTER LE1 7RH
UK.

Abstract

This paper is concerned with the use of Dempster-Shafer theory in 'fusion' classifiers. We argue that the use of predictive accuracy for basic probability assignments can improve the overall system performance when compared to 'traditional' mass assignment techniques. We demonstrate the effectiveness of this approach in a case study involving the detection of static thermostatic valve faults in a diesel engine cooling system.

Keywords: Dempster-Shafer theory, pattern classification, diesel engine.

Pre-print of:

Parikh C.R., M.J. Pont and N.B. Jones (2001) "Application of Dempster-Shafer theory in condition monitoring systems: A case study", *Pattern Recognition Letters*, **22** (6-7): 777-785.

---

[1] To whom correspondence should be addressed. E-mail: M.Pont@leicester.ac.uk

1 Introduction

Many recent papers have examined the gains that may be obtained through the use of "fusion classifiers": that is, by combining the outputs from two or more primary classifiers in an effort to improve the overall classification performance. Fusion classifiers have recently been applied in many areas including handwriting recognition (Xu, *et al.*, 1992; Lu, 1996; Alimoglu and Alpaydin, 1997; Rahman and Fairhurst, 1998; Huang et al, 1995; Ho *et al.*, 1994), optical character recognition (Mao and Mohiuddin, 1997; Huang *et al.*, 1997), speech recognition (Chen *et al.*, 1997; Yu *et al.*, 1997; Bowles, 1992), biometrics (Chen, 1998; Frenstermacher, 1997), medical diagnosis (Kitler *et al.*, 1997), sales forecasting (Tchaban *et al.*1998), earth quake evaluation (Giacinto et al, 1997) and industrial diagnostics applications (Parikh, *et al.*, 1998; Li *et al.*, 1999). In almost all cases, the 'fusion' classifier has demonstrated higher levels of performance than any of the 'primary' classifiers (but see Li and Jain, 1998).

Various techniques have been used to create fusion classifiers depending upon the form of information from the primary classifiers (Xu *et al.*, 1992). One such technique, Dempster-Shafer evidence theory (Shafer, 1976), is considered in detail in this paper: this approach has previously been applied in a number of studies (e.g. Bowles and Damper, 1989; Xu *et al.*, 1992; Chen *et al.*, 1997; Rogova, 1994; Lu, 1996).

Dempster-Shafer theory is a powerful method of combining accumulative evidence or for changing prior opinions in the light of new evidence (Shafer, 1976). Briefly, the theory may be summarised as follows. For a finite set of mutually exclusive and exhaustive propositions Θ, sometimes referred to as a Frame of Discernment (FOD), a power set 2^{Θ} is the set of all the subsets of Θ including itself and a null set, ϕ. Each subset is called a focal element. Based on the evidence, a numeric value between [0,1] is assigned to each focal element. The value 0 indicates no belief in a proposition, the value 1 indicates total belief, and any values between these two limits indicate partial beliefs. A portion of belief committed to one focal element is also committed to any other implied focal element(s) and cannot be further subdivided among the subsets.

Dempster-Shafer theory allows the mass or *basic probability assignment (bpa)* to individual prepositions and also to any sub-sets of the power set provided that the sum of all basic probability masses is equal to one (Equation 1).

$$\sum_{A\subset\Theta} m(A) = 1 \qquad \text{and} \qquad m(f) = 0 \qquad \text{Equation 1}$$

If the probability number for only a partial set of hypotheses is known then the remaining complementary probability number is assigned to the FOD, $m(\Theta)$, thus allowing the representation of ignorance.

The measure of total belief committed to A can be obtained by computing the belief function for A which adds the mass of all the proper subsets of A (Equation 2).

$$Bel(A) = \sum_{B\subset A} m(B) \qquad \text{Equation 2}$$

The belief function represents the lower limit of the probability and *plausibility* function (Equation 3) provides the upper limit of the probability.

$$pl(A) = 1 - Bel(\neg A) = \sum_{B\cap A\neq f} m(B) \qquad \text{Equation 3}$$

The difference between the two functions (Equation 2 and Equation 3) represents the ignorance.

Evidence can be combined by computing the orthogonal sum using Dempster's rule of combination[2] where Evidence A and Evidence B are used for computing a new belief function for a focal element C (Equation 4).

$$m(C) = \frac{\sum_{A\cap B=C} m(A)\times m(B)}{1 - \sum_{A\cap B=f} m(A)\times m(B)} \qquad \text{Equation 4}$$

The issue of *basic probability assignment (bpa)* is crucial to the success of the technique (Rogova, 1994). A number of different ways have been employed to interpret outputs from a primary classifier as a *bpa*, viz.,

---

[2] This is the normalised Dempster's rule and there have been various arguments in favour and against the normalised masses (Zadeh, 1986; Parsons, 1994; Guan and Bell, 1991).

(i) training the classifier to estimate probability (Richard and Lippmann, 1991, Wan, 1990, Giacinto *et al.*, 1997),

(ii) using transformation methods to compute probability masses (e.g. Chen *et al.*, 1997, Huang *et al.*, 1995) and

(iii) utilising the information available from confusion matrices (Xu *et al.*, 1992; Chen *et al.*, 1997 and Rogova, 1994).

A key problem with approach (i) when working with Condition Monitoring and Fault Diagnosis (CMFD) applications - the focus of this paper - is that data used during training of the various primary classifiers must be statistically valid. In many CMFD systems, data representing fault conditions can only be obtained with great difficulty: as a result, training classes may vary greatly in size. Also, it is not always possible to quantify *a priori* the likelihood of each fault condition. These two factors present major problems for conventional statistically-based techniques.

An additional problem, associated with both approaches (i) and (ii), is that each classifier will usually generate outputs on a different scale, and it is rarely easy to identify equivalent confidence levels for any two (or more) sources (Lu, 1996, Ho *et al.*, 1994).

Approach (iii) appears more promising. This exploits the information available from the confusion matrices of different classification techniques for belief assignment. Each primary classifier supports one output class and rejects the rest. When using such binary outputs, transformation of output scores or parity of outputs among different classification techniques is not an issue.

While some researchers have used transformation methods for computing beliefs (Rogova, 1994; Lu, 1996; Bowles and Damper, 1989; Giacinto *et al.*, 1997), it is not always possible to justify the validity of transformed scores or posterior probability outputs to be used in combining different classifiers for non-linear systems without having a complete understanding of the system and relationship between the different classifiers.

To address these problems we propose in this paper the use of predictive rates, a measure that is common in ‘receiver operating characteristic’ (ROC) analysis (e.g. Swets and Picket, 1982, Bradley, 1997), for computing beliefs for individual decisions from the classifiers. We argue that as the predictive rate accounts for misclassifications it is logically valid, and a more appropriate measure for mass assignment then 'recognition-substitution-rejection' (RSR) rates. In this paper, we apply both the RSR and predictive rate methods to a diesel engine thermostatic valve diagnostic problem, and we demonstrate that the predictive rate approach results in substantially improved performance.

The paper is organised as follows. In Section 2 we briefly describe the diesel engine cooling system case study. In Section 3, we use the real engine data on three primary classifiers, viz., multi-layer Perceptron (MLP), radial basis function (RBF) neural networks and k-nearest neighbour (kNN). A brief introduction to the use of Dempster-Shafer theory in classifier combination using RSR rates is presented in Section 4. Following the discussion of the RSR results, the alternative predictive rate approach is presented in Section 5. Finally the paper is concluded in Section 6.

2 The case study

In this paper, we conduct three experiments using different classifier systems. In each case, the classifiers are applied to the condition monitoring and fault diagnosis of the cooling system for a four-litre, four-cylinder, turbo-charged diesel engine, used in a diesel (electricity) generator system, usually referred to as a 'diesel genset'.

The cooling system has a key role in this type of application. This system has the task of removing the heat generated during the combustion process, so as to avoid engine seizure or - in an extreme - the melting of engine components. In water-cooled engines, like that considered here, combustion heat is dissipated to water circulated by a pump around the engine cylinder through a thermostatic valve and radiator (or heat exchanger). For efficient operation the thermostatic valve is set to open at a preset temperature which then allows the hot coolant (water) to pass through the radiator where secondary coolant (usually air draft in mobile applications) cools it down prior to re-circulation.

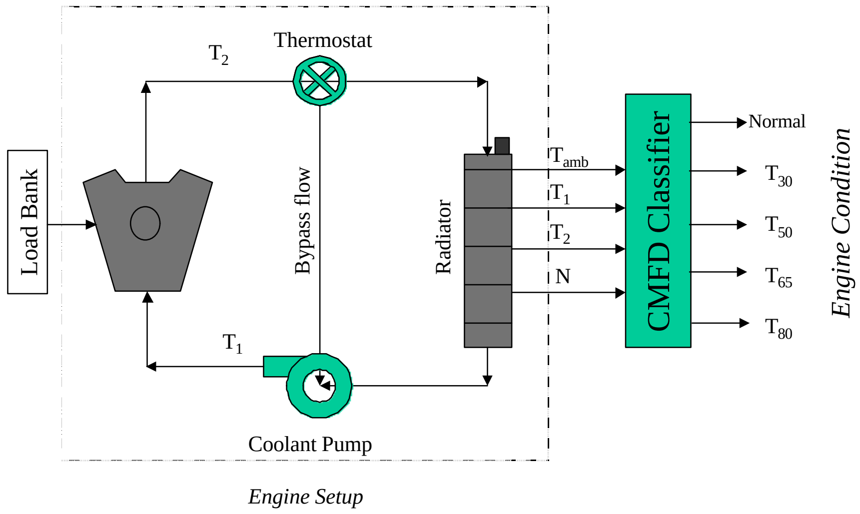

Figure 1: Schematic diagram for diesel engine cooling system classifier *(see text for details)*

Various faults may occur in this form of cooling system. In this study, we are concerned only with faults in the thermostatic valve: specifically, the faults considered consist of a thermostatic valve damaged so that the maximum coolant flow rate is reduced to 30, 50, 65 and 80 % (approximately) of the normal (100%) value.

The flow reductions were introduced with the engine stationary, before the system was brought up to a speed of approximately 1500 rpm. A 60 kW load was used in all cases. In the steady-state condition, engine speed (Z), ambient temperature (T_{amb}), radiator inlet (T_2) and engine inlet (T_3) temperatures were recorded. As the coolant system is a slow dynamic system, data were sampled at 2Hz.

For each of the flow restrictions tests were run to collect 200 samples for training the classifiers, 200 samples for validating[3] the trained classifiers, and a further 800 samples for

---

[3] The validation data set is used to measure the true error rate for computing the predictive rate of various primary classifiers for *basic probability assignment* in Dempster-Shafer theory. See Section 3.

testing the classifiers. All the data were filtered with a low-pass Butterworth filter and then normalised to a 0 to 1 scale after sampling.

3 Using the primary classifiers - Experiment A

The multi-layer Perceptron (MLP) neural network, Radial Basis Function (RBF) neural network and non parametric k-nearest neighbour (kNN) classifiers are very successful techniques in pattern recognition, and are employed for the primary classification task in our CMFD application. Further details of these classifiers are available from numerous reference books (e.g. Webb, 1999).

The MLP had one hidden layer with 25 neurons. The network was trained to 0.05% epoch error using the pattern-by-pattern training mode with the momentum constant 0.25 and a variable learning rate.

The RBF network was trained with maximum 25 Gaussian basis functions in the hidden layer using the orthogonal least-square algorithm in MATLAB. The spread constant of 4 with 0.05% as the training error goal was used for training the classifier.

The kNN classifier was tested with different values of k. However, all the values gave very similar performance results. For the minimal computational load, a value of k equal to 1 was chosen in this implementation.

3.1 Results

The performance of the classifiers on the independent test data set is shown in *confusion matrix* form in Table 1. Table 2 shows the performance of the classifier on the validation data set.

Input Simulated	**Classifier**	**Output**				
		N	T_{30}	T_{50}	T_{65}	T_{80}
N	MLP	***99***				1
	RBF	***85.25***				14.75
	kNN	***69.25***		30.75		
T_{30}	MLP		***100***			
	RBF		***60***			40
	kNN		***100***			
T_{50}	MLP	2.37		***82.5***		15.38
	RBF			***100***		
	kNN			***100***		
T_{65}	MLP				***76.88***	23.12
	RBF				***99.50***	0.50
	kNN				***100***	
T_{80}	MLP			6.62	17.25	***76.13***
	RBF					***100***
	kNN			5.88	25.12	***69***

Table 1 Performance of Primary classifiers on test data (***italics*** figures indicate recognition rate)

Input Simulated	**Classifier**	**Output**				
		N	T_{30}	T_{50}	T_{65}	T_{80}
N	MLP	***99***				1
	RBF	***86***				14
	kNN	***69***		31		
T_{30}	MLP		***100***			
	RBF		***59.75***			40.25
	kNN		***100***			
T_{50}	MLP	2025		***81***		15.75
	RBF			***100***		
	kNN			***100***		
T_{65}	MLP				***76.75***	23.25
	RBF				***99.50***	0.50
	kNN				***100***	
T_{80}	MLP			6.5	17.25	***76.25***
	RBF					***100***
	kNN			5.75	25.25	***69***

Table 2 Performance of primary classifiers on validation data (***italics*** figures indicate recognition rate)

3.2 Discussion

From the above results, it is clear that none of the techniques is able to classify all the engine conditions at a high level of accuracy. While the MLP performs well for normal and T_{30} conditions, the RBF classifier does well on T_{50}, T_{65}, and T_{80} engine conditions and the kNN classifier on T_{30}, T_{50} and T_{65} conditions.

4 Using Dempster-Shafer - Experiment B

In Experiment A, it was demonstrated that none of the individual (primary) classifiers was able to classify the unseen fault data at a high level of accuracy. In Experiment B, Dempster-Shafer theory was applied to combine the outputs from these primary classifiers in an attempt to improve the overall performance in the classification task.

The Frame of Discernment (Θ) for a classification task consists of m mutually exclusive and exhaustive propositions. For an input pattern using winner-take-all (WTA) approach, each primary classifier would assign a true label (or 1) to proposition i, $i \in \Theta$ and remaining classes would be labelled as false (or 0). Hence for every classification there are two focal elements, i and $\neg i = \Theta - \{i\}$, either confirming or denying a single proposition for mass assignment in the Dempster-Shafer theory. For each classifier k, the confusion matrix would provide a recognition rate $e_{r[i]}^{(k)}$, substitution rate $e_{s[i]}^{(k)}$, and rejection rate (1-$e_{r[i]}^{(k)}$-$e_{s[i]}^{(k)}$) for each of the classes i which is used as *bpa*, $m(i)$, $m(\neg i)$ and $m(\Theta)$ respectively[4].

In order to obtain beliefs $bel(i)$ and $bel(\neg i)$ for $\forall i \in m$ based on all the evidence from the three classifiers, Dempster's rule of combination (Equation 4) can be used to combine two classifiers. This mass can then be combined with the output from the remaining classifier. For each of the proposition beliefs, plausibility and contradiction information may be used for making a final decision. While 'winner-take-all' would be the simplest implementation for making decisions, other variations could include setting a threshold on ignorance, contradiction and so forth.

Here, Dempster-Shafer was implemented as follows. The masses of K evidence are combined recursively, $m = m_1 \oplus m_2 \ldots \oplus m_K$, using Dempster's rule of combination (Equation 4). This is computationally very expensive, particularly when the number of classes and number of evidence increase. As suggested by earlier research (Xu, *et al.* 1992) evidence may be

---

[4] Recognition rate, $e_{r[i]}$ for a class '*i*' is the ratio of the number of patterns classified as the class '*i*' to total number of patterns presented to the classifier belonging to the class '*i*'. Substitution rate, $e_{s[i]}$ for a class '*i*' is the ratio of the number of patterns classified as classes other than '*i*' to total number of patterns presented to the classifier belonging to the class '*i*'.

combined in two stages. First, combine all the evidence suggesting the same proposition. When there are K classifiers and P classes, the first stage combination reduces the number of combined groups to K_1 [min(P, K)]. In the second stage all the groups which suggest different classes are combined. Since, after the first stage, the number of focal elements for all the groups will be different, the second stage is computationally expensive because the number of focal elements increases after each combination. Therefore, if computational power is limited, the efficient implementation suggested by Xu, *et al.* (1992) may be useful. However, in our three-evidence implementation, we are able to use normal combination rules for the Stage Two.

For a sample from the test data, based on the recognition, substitution and rejection rates (RSR) (see Table 2) the *bpa* used for $m(i)$ and $m(\neg i)$ is shown in Table 3. Since the rejection rate is zero for all the classes, and therefore $m(\Theta) = 0$, this value is not shown in the table.

	N		**T_{30}**		**T_{50}**		**T_{65}**		**T_{80}**	
	$m(N)$	**$m(\neg N)$**	**$m(T_{30})$**	**$m(\neg T_{30})$**	**$m(T_{50})$**	**$m(\neg T_{50})$**	**$m(T_{65})$**	**$m(\neg T_{65})$**	**$m(T_{80})$**	**$m(\neg T_{80})$**
MLP	0.99	0.01	1.00	0	0.81	0.19	0.7675	0.2325	0.7625	0.2375
RBF	0.86	0.14	0.5975	0.4025	1.00	0	0.9950	0.0050	1.00	0
kNN	0.69	0.31	1.00	0	1.00	0	1.00	0	0.69	0.31

Table 3 *basic probability assignment (bpa)* based on recognition, substitution and rejection rates.

4.1 Results

The five simulated engine conditions form the Frame of Discernment, Θ, and Table 3 is used for calculating the *bpa* for the test data sample and for computing beliefs. The result of combining the three classifiers is shown in Table 4.

Simulated	**Condition classified**					
	N	T_{30}	T_{50}	T_{65}	T_{80}	Unclassified
N	***69.25***		16			14.75
T_{30}		***60***				40
T_{50}			***100***			
T_{65}				***99.50***		0.50
T_{80}					***69***	31

Table 4 Performance of Dempster-Shafer combination on test data using conventional approach.

The average recognition rate for the combined classifier is 79.55%.

4.2 Discussion

Results for N, T_{30} and T_{80} engine conditions after combination is similar to the lowest of that obtained by one of the primary classifiers. The higher percentage of unclassified samples has resulted in lower recognition rate for all the conditions with the exception of T_{50} and T_{60}.

In this Dempster-Shafer implementation, a sample would remain unclassified if, after the combination, two or more prepositions have equal maximum belief: this leads to poor performance. In our three-classifier case, it is not difficult to understand how two propositions could have similar beliefs when using the conventional RSR approach. A particular classifier classifies a particular proposition with 100% recognition rate and therefore, any samples classified to this class label would have *bpa* 1. However, the same class label is also possible due to misclassification from samples belonging to other propositions and the mass assigned to the misclassified samples will also be 1. This may lead to contradiction during combination, if there is another primary classifier that is able to classify the 'misclassified' sample with the same level of accuracy (100% in this example). This means that the latter classifier assigns mass 1, with correct class label and the former classifier has assigned mass 1 to the misclassified label. Thus this combination means that each primary classifier puts complete faith in a different (contradictory) proposition resulting, after combination, in an unclassified condition.

5 Using modified Dempster-Shafer - Experiment C

Experiment B demonstrated the failure of Dempster-Shafer combination approach using recognition, substitution and rejection rates. In Experiment C, the aim was to evaluate an alternative form of Dempster-Shafer combination using predictive rates.

Before considering this implementation, we first clarify the meaning of the term *predictive rate*. For m class classification problem, an input sample belonging to class i ($i \in m$), is classified to one of the j ($j \in m+1$) classes including the rejection class (m+1). The *Predictive rate* of a classifier k for an output class j ($j \in m$), $e_{p[j]}^{(k)}$ is the ratio of number of input patterns classified correctly ($e_{ij}|\ i = j;\ j \in m$) to the total number of patterns classified as class j ($\sum_{n=1}^{m} e_{nj}$), for input patterns belonging to all i classes is presented to the classifier.

Note that predictive rates are also referred to as True/False Positives (TP/FP) and True/False Negatives (TN/FN) in the relative operating characteristic (ROC) analysis, and as hits-correct rejection, false alarms and misses in signal detection (Bradley, 1997; Swets and Pickett, 1982).

5.1 Combining classifiers using Dempster-Shafer and predictive rates

The proposed approach uses predictive rate $e_{p[i]}^{(k)}$ information for each of the classes *i* from the confusion matrix for every classifier *k,* instead of recognition, substitution and rejection rates. When a particular classifier classifies the result $j^{(k)} \in m$, we consider that for all the instances when the class is classified as *j*, the likelihood that *j* is the actual class is $e_{p[j]}^{(k)}$ and that *j* is not the correct class is $(1-e_{p[j]}^{(k)})$; instances when the classifier wrongly classified the input sample as *j*. Therefore, the predictive rate is used for *basic probability assignment or mass m*(*j*) and the disbelief is assigned to $m(\neg j)$. When the output is the class ($m+1$), i.e. the input sample is rejected by the classifier, all the propositions are equally likely to occur and hence full support is allocated to Θ with m(Θ) = 1. In order to obtain beliefs *bel*(*j*) and $bel(\neg j)$ for $\forall j \in m$ based on all the evidence, a two stage combination approach is implemented, as in Experiment B.

Overall, the key advantage obtained through the use of predictive rates is that, unlike RSR rates (e.g. Xu et al, 1992), this modified approach takes misclassification of patterns into account.

5.2 Results

For a test data sample, the *bpa* for the five simulated engine conditions on the basis of the proposed predictive rate approach using Table 2 is shown in the Table 5.

	N		**T_{30}**		**T_{50}**		**T_{65}**		**T_{80}**	
	m(N)	**m(¬N)**	**$m(T_{30})$**	**$m(\neg T_{30})$**	**$m(T_{50})$**	**$m(\neg T_{50})$**	**$m(T_{65})$**	**$m(\neg T_{65})$**	**$m(T_{80})$**	**$m(\neg T_{80})$**
MLP	0.9766	0.0234	1.00	0	0.9255	0.0745	0.8167	0.1833	0.6584	0.3416
RBF	1.00	0	1.00	0	1.00	0	1.00	0	0.6440	0.3560
kNN	1.00	0	1.00	0	0.7393	0.2607	0.7992	0.2008	1.00	0

Table 5: *Basic probability assignment (bpa)* based on predictive rates. See text for details.

The result of combining the three classifiers is shown in Table 6.

Simulated	**Condition classified**				
	N	**T_{30}**	**T_{50}**	**T_{65}**	**T_{80}**
N	**99**				**1**
T_{30}		**100**			
T_{50}			**100**		
T_{65}				**100**	
T_{80}			6.63	18.50	**74.87**

Table 6 Performance of Dempster-Shafer combination on test data using proposed approach

Average recognition rate for the proposed technique has increased to 94.77% .

5.3 Discussion

From the results of two different Dempster-Shafer implementations, it is clear that the predictive rate method is performing better than the conventional one in this study. Successful classification of all samples after combination has led to improved performance compared with the conventional approach, primarily because the use of the *bpa* based on predictive rates has avoided the contradictory mass assignment problem.

6 Conclusions

This paper has described a novel implementation of the Dempster-Shafer theory for use in condition monitoring and fault diagnosis applications. This new approach, based on predictive rates, is intuitively sensible. In the case study described in this paper, we have demonstrated that the use of such predictive rates was much more effective than the traditional Dempster-Shafer implementation.

Acknowledgements

This work was supported in part by EPSRC grant GR/L42018. We thank Perkins Engines Company Limited, Peterborough, and JREI grant GR/M30777 for providing the Gen-Set and instrumentation used in this study. Chinmay Parikh is supported in part by an Overseas Research Studentship from the CVCP and Perkins Engines Company Limited. We thank John Twiddle (University of Leicester) for his help in collecting the data for this study.

References

Alimoglu, F., Alpaydin, E., 1997, Combining multiple representations and classifiers for pen-based handwritten digit recognition, Proceedings of the International Conference on Document Analysis and Recognition, ICDAR, 2, 637-640.

Bowles, R. L., 1992, Combination of evidence in speech recognition, MPhil Thesis, University of Southhampton, England.

Bowles, R.L., Damper, R.I., 1989, Application of Dempster-Shafer theory of evidence to isolated word speech recognition, Research Journal, Department of Electronics and Computer Science, University of Southampton, England.

Bradley, A. P., 1997, The use of the area under the ROC curve in the evaluation of machine learning algorithms, Pattern Recognition, 30(7), 1145-1159.

Chen, K., Wang, L., Chi, H.S., 1997, Methods of combining multiple classifiers with different features and their applications to text-independent speaker identification, International Journal of Pattern Recognition and Artificial Intelligence, 11(3), 417-445.

Chen, K., Wang, L., Chi, H.S., 1998, A connectionist method for pattern classification with diverse features, Pattern Recognition Letters, 19(7), 545-558.

Fenstermacher, L.H., 1997, Techniques for higher confidence target ID, Proceedings of SPIE - The International Society for Optical Engineering, 3077, 356-366.

Giacinto, G., Paolucci, R., Roli, F., 1997, Application of neural networks and statistical pattern recognition algorithms to earthquake risk evaluation, Pattern Recognition Letters, 18(11-13), 1353-1362.

Guan, J.W., Bell, D.A., 1991, Evidence Theory and its Applications, Amsterdam:North-Holland.

Ho, T.K., Hull, J.J., Srihari, S.N., 1994, Decision combination in multiple classifier systems, IEEE Transactions on Pattern Analysis and Machine Intelligence, 16(1), 66-75.

Huang, K., Wu, J., Yan, H., 1997, Off-line writer verification utilizing multiple neural networks, Optical Engineering, 36(11), 3127-3133.

Huang, Y.S., Liu, K., Suen, C.Y., 1995, The Combination of multiple classifiers by a neural network approach, International Journal of Pattern Recognition and Artificial Intelligence, 9(3), 579-597.

Kittler, J., Hojjatoleslami, A., Windeatt, T., 1997, Strategies for combining classifiers employing shared and distinct pattern representations, Pattern Recognition Letters, 18(11-13), 1373-1377.

Kittler, J., Hatef, M., Duin, R.P.W., Matas, J., 1998, On combining classifiers, IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(3), 226-239.

Kittler, J., Hojjatoleslami, A., Windeatt, T., 1997, Strategies for combining classifiers employing shared and distinct pattern representations, Pattern Recognition Letters, 18(11-13), 1373-1377.

Li, Y.H., Jain, A.K., 1998, Classification of text documents, Computer Journal, 41(8), 537-546.

Li, Y.H., Pont, M.J., Parikh, C.R., Jones, N.B., 1999, Using a combination of RBFN, MLP and kNN classifiers for engine misfire detection, In: John, R., Birkenhead, R. (eds.), Advances in Soft Computing, Physica-Verlag, Heidelberg, 22-27.

Lu, Y., 1996, Knowledge integration in a multiple classifier system, Applied Intelligence, 6(2), 75-86.

Mao, J.C., Mohiuddin, K.M., 1997, Improving OCR performance using character degradation models and boosting algorithm, Pattern Recognition Letters, 18(11-13), 1415-1419.

Parikh, C.R., Pont, M.J., Li, Y.H., Jones, N.B., Twiddle, J.A., 1998, Towards a flexible application framework for data fusion using real-time design patterns, 6th European Congress on Intelligent Techniques & Soft Computing, Aachen, Germany, September, 1131-1135.

Parsons, S., 1994, Some qualitative approaches to applying the Dempster-Shafer theory, Information and Decision Technologies, 19(4), 321-337.

Rahman, A.F.R., Fairhurst, M.C., 1998, An evaluation of multi-expert configurations for the recognition of handwritten numerals, Pattern Recognition, 31(9), 1255-1273.

Richard, M., Lippmann, R.P., 1991, Neural network classifiers estimate Bayesian a posteriori probabilities, Neural Computation, 3(4), 461-483.

Rogova, G., 1994, Combining the results of several neural network classifiers, Neural Networks, 7(5), 777-781.

Shafer, G., 1976, A mathematical theory of evidence, Princeton University Press.

Swets, J.A., Pickett, R.M., 1982, Evaluation of Diagnostic Systems - Methods from Signal Detection Theory, Academic Press.

Tchaban, T., Griffin, J.P., Taylor M.J., 1998, A Comparison between single and combined backpropagation neural networks in the prediction of turnover, Engineering Applications of Artificial Intelligence, 11, 41-47.

Wan, E.A., 1990, Neural network classification: A Bayesian interpretation, IEEE Transactions on Neural Networks, 1(4), 303-305.

Webb, A., 1999, Statistical Pattern Recognition, Arnold Publishers, ISBN 0340741643.

Xu, L., Krzyzak, A., Suen, C.Y., 1992, Methods of combining multiple classifiers and their applications to handwriting recognition, IEEE Transactions on Systems Man and Cybernetics, 22(3), 418-435.

Yu, K., Jiang, X., Bunke, H., 1997, Lip reading: A classifier combination approach, Pattern Recognition Letters, 18 (11-13), 1421-1426.

Zadeh, L.A., 1986, A simple view of the Dempster-Shafer theory of evidence and its implication for the rule of combination, AI Magazine, 7, 85-90.