Multiple Classifier Combination for Character Recognition: Revisiting the Majority Voting System and Its Variations

A.F.R. Rahman[1], H. Alam[1], and M.C. Fairhurst[2]

[1] Document Analysis and Recognition Team (DART)
BCL Technologies Inc.
Santa Clara, CA 95050, USA, Tel: +1 408 557 5279, Fax: +1 408 249 4046
fuad@bcltechnologies.com
[2] Department of Electronics
University of Kent, Kent, CT2 7NT, UK
Tel: +44 1227 823389, Fax: +44 1227 456084
M.C.Fairhurst@ukc.ac.uk

Abstract. In recent years, strategies based on combination of multiple classifiers have created great interest in the character recognition research community. A huge number of complex and sophisticated decision combination strategies have been explored by researchers. However, it has been realized recently that the comparatively simple Majority Voting System and its variations can achieve very robust and often comparable, if not better, performance than many of these complex systems. In this paper, a review of various Majority Voting Systems and their variations are discussed, and a comparative study of some of these methods is presented for a typical character recognition task.

Keywords: Multiple classifier combination, majority voting, character recognition.

1 Introduction

Combination of multiple classifiers is now accepted as a very important method in achieving robustness and accuracy in many recognition tasks, especially in character recognition problems (Ho *et al.*[1], Huang *et al.*[2], Rahman and Fairhur-st [3,4,5], Lam *et al.*[6], Kittler *et al.*[7]) etc.). Despite huge advances, the recognition of handwritten characters is still an unsolved problem. With the advent of new computer technologies and the explosion in Internet-based on-line data manipulation applications, automatic conversion of written and spoken information into computer readable forms is becoming increasingly important. Against this backdrop, researchers are exploring various ways of combining decisions from multiple classifiers as a viable way of delivering very accurate and robust performance over a wide range of applications. This has resulted in the design of

many decision combination algorithms. In recent years, attention has been refocused on using more simple techniques, e.g. Majority Voting Systems, precisely because of their simplicity and very high level of accuracy and robustness which can be achieved in appropriate circumstances (Lam and Suen[8,9]. In this paper, a review of various multiple classifier approaches exploiting the principle of majority voting is presented. In addition, a comparative study of some of these methods on a typical handwritten character recognition task is also reported.

2   Majority Voting and Its Variations

Majority Voting Systems have quite a number of variations in terms of application and methodology, although the underlying principle is the same. A basic issue to be addressed here concerns two potentially conflicting strategies for achieving decision combination, which may be broadly identified as:

- Should the decision agreed by the majority of the experts (in some fashion) be accepted without giving due credit to the competence of each expert? or,
- should the decision delivered by the most competent expert be accepted, without giving any importance to the majority consensus?

This ultimately amounts to a choice between selecting either the "consensus decision" or the "decision delivered by the most competent expert", yet both strategies could contribute overall to a successful decision, and ideally require careful integration within the decision making process. Moreover, by their very nature, these two considerations are often contradictory and hence most multiple expert decision combination algorithms emphasize either one or other of the strategies. This section presents some of the principal techniques based on Majority Voting System.

2.1   Simple Majority Voting

If there are n independent experts having the same probability of being correct, and each of these experts produces a unique decision regarding the identity of the unknown sample, then the sample is assigned to the class for which there is a consensus, i.e. when at least k of the experts agree, where k can be defined as:

$$k = \begin{cases} \frac{n}{2} + 1 \text{ if n is even,} \\ \frac{n+1}{2} \quad \text{if n is odd.} \end{cases} \tag{1}$$

Assuming each expert makes a decision on an individual basis, without being influenced by any other expert in the decision-making process, the probabilities of various different final decisions, when $x + y$ experts are trying to a reach a decision, are given by the different terms of the expansion of $(P_c + P_e)^{x+y}$, where, P_c is the probability of each expert making a correct decision, P_e is the probability of each expert making a wrong decision, with $P_c + P_e = 1$. Bernoulli[10] is credited with first realizing this group decision distribution. The

probability that x experts would arrive at the correct decision is $\frac{(x+y)!}{x!y!}(P_c)^x(P_e)^y$ and the probability that they arrive at the wrong decision is $\frac{(x+y)!}{x!y!}(P_c)^y(P_e)^x$. So in general, the precondition of correctness (Condorcet[11]) of the combined decision for $x > y$ can be conveniently expressed as:

$$\kappa = \frac{(P_c)^x(P_e)^y}{(P_c)^x(P_e)^y + (P_c)^y(P_e)^x} \tag{2}$$

Reordering Eq 2 and assuming the fraction of the experts arriving at the correct decision to be fixed, (e.g. x and y to be constant), it is possible to show that,

$$\frac{\delta\kappa}{\delta P_c} = \kappa^2(x-y)\frac{(P_e)^{x-y-1}}{(P_c)^{x-y-1}}(P_c + P_e) \tag{3}$$

Since $(x - y - 1 \geq 0)$, $\frac{\delta\kappa}{\delta P_c}$ is always positive. Thus when x and y are given, as P_c increases κ increases continuously from zero to unity. This demonstrates that the success of the Majority Voting Scheme (like most decision combination schemes) directly depends on the reliability of the decision confidences delivered by the participating experts. It is also clear that as the confidences of the delivered decisions increase, the quality of the combined decision increases.

Recently, it has been demonstrated that although majority vote is by far the simplest of the variety of strategies used to combine multiple experts, properly applied it can also be very effective. Suen *et al.*[12] presented a method for decision combination incorporating different types of classifiers based on a straightforward voting scheme. A detailed study of the working of the majority voting scheme has been presented by Lam and Suen[8]. Ng and Singh[13], have discussed the applicability of majority voting techniques and have proposed a support function to be used in the combination of votes. Researchers have also used various types of classifiers in these majority voting schemes. Stajniak *et al.*[14] presented a system having three voting nonlinear classifiers: two of them based on the multilayer perceptron (MLP), and one using the moments method. Belaid and Anigbogu[15] reported a character recognition system using six classifiers built around first and second order hidden Markov models (HMM) as well as nearest neighbor considerations. Parker[16] has reported voting methods for multiple autonomous agents. Ji and Ma[17] have reported a learning method to combine weak classifiers, where weak classifiers are linear classifiers (perceptrons) which can do little better than making random guesses. The authors have demonstrated, both theoretically and experimentally, that if the weak classifiers are properly chosen, their combinations can achieve a good generalization performance with polynomial space- and time-complexity.

2.2   Weighted Majority Voting

A simple enhancement to the simple majority systems can be made if the decisions of each classifier are multiplied by a weight to reflect the individual confidences of these decisions. In this case, Weighting Factors, ω_k, expressing

the comparative competence of the cooperating experts, are expressed as a list of fractions, with $1 \leq k \leq n$, $\sum_{k=1}^{n} \omega_k = 1$, n being the number of participating experts. The higher the competence, the higher is the value of ω. So if the decision by the kth expert to assign the unknown to the ith class is denoted by d_{ik} with $1 \leq i \leq m$, m being the number of classes, then the final combined decision d_i^{com} supporting assignment to the ith class takes the form of: $d_i^{com} = \sum_{k=1,2,...,n} \omega_k * d_{ik}$. The final decision d^{com} is therefore: $d^{com} = \max_{i=1,2,...,m} d_i^{com}$.

Lam and Suen[9] have studied the performance of combination methods including a Bayesian formulation and a weighted majority vote with weights obtained through a genetic algorithm. Alpaydin[18] has employed a weighted majority voting scheme by adopting a Bayesian framework where 'weights' in voting may be interpreted as plausibilities of the participating classifiers. Additional discussion of the weighted majority voting technique may be found in Kittler et al.[19] and Lam et al.[6].

2.3   Class-Wise Weighted Majority Voting Scheme

The Class Confidence Index, $\beta_{i,j}$, $1 \leq i \leq n$, $1 \leq j \leq m$, where n is the number of decisions and m is the number of classes under consideration, denotes the ranking of the different decisions on a class by class basis. The higher the class recognition rate, the higher the ranking. These class-wise confidences can then be conveniently converted to suitable weight values ω_k^i, expressing the comparative competence of the decisions on a class-wise basis, are expressed as a list of fractions, with, $1 \leq k \leq n$, $\sum_{k=1}^{n} \omega_k^i = 1$, n again being the number of the decisions being combined. So if the kth decision to assign the unknown pattern to the ith class is denoted by d_{ik} with $1 \leq i \leq m$, m being the number of classes, then the final combined decision d_i^{com} supporting assignment to the ith class takes the form of: $d_i^{com} = \sum_{k=1,2,...,n} \omega_k^i * d_{ik}$. The final decision d^{com} is therefore: $d^{com} = \max_{i=1,2,...,m} d_i^{com}$.

2.4   Restricted Majority Voting

Sometimes it is important to shift the emphasis of decision combination in selecting the best appropriate classifier from an array of classifiers. The Overall Confidence values, γ_k, expressing the comparative competence of the classifiers, are expressed as a ranking list, with $1 \leq k \leq n$, and n being the number of experts. The higher the recognition rate, the higher the ranking. It is straightforward to convert this ranking to a set of weighting factors. These weighting factors, ω_k, expressing the comparative competence of the cooperating classifiers, are expressed as list of fractions, with $1 \leq k \leq n$, $\sum_{k=1}^{n} \omega_k = 1$, n again being the number of classifiers. The best representative expert for a particular problem domain can then be noted by finding the maximum weighting factor ω_m, so that, $\omega_m = max(\omega_k)$. The final decision by this method is then expressed by, $d_{Best} = d_m$.

Gang *et al.*[20] have described such a modularized neuroclassifier. This classifier combines four modularized classifiers using MLP modules. The same idea of identifying the best classifier in a voting scheme can be further extended in a more generalized framework (Rahman and Fairhurst[21]). Defining γ_k, $1 \leq k \leq n$, as Overall Confidence Indices, representing the ranking of the experts ($\gamma_k = 1, 2, ..., n$), β_{ij}, ($1 \leq i \leq n, 1 \leq j \leq m$, where m is the number of classes under consideration), as Class Confidence Indices, denoting the ranking of the different experts ($\beta_{ij} = 1, 2, ..., n$) on a class by class basis, and finally α_{ijk}, as Sample Confidence Indices, denoting the confidence value assigned by the ith expert to the kth sample of the jth class in the test set, this generalized decision combination framework selects the best final combined decision by the following logical structure: The decision associated with the highest sample confidence index is accepted as the final decision, provided the sample confidence index is greater than or equal to a threshold value from the nearest confidence value assigned by a competing expert, so that, $|\alpha_{uwt} - \alpha_{vwt}| \geq \psi_c$, where u and v denote the top two competing experts, w is the class under consideration, t is the sample in question and ψ_c is the threshold. In circumstances where this criteria is not met, a decision combination is attempted based on $|\beta_{sw} - \beta_{tw}| \geq \theta_c$, where s and t denote the top two competing experts, w is the class under consideration by that particular expert and $\theta_c > 0$. The motivation behind using this criterion is to give due preference to the strengths of particular experts in recognizing a particular class of characters. In the unlikely event of a failure to draw a final decision at this stage, the criterion $|\gamma_s - \gamma_w| \geq \phi_c$ can be applied, where s and w are the top two competing experts and $\phi_c > 0$. If no decision can be achieved at this stage, this denotes a classifier redundancy and an arbitrary decision of accepting the decision of the first expert in the array is taken. The characters which are rejected by one or more of the experts are channelled to the re-evaluating expert. If a sample is rejected, then the expert rejecting the character is excluded from the decision making process. Hence, if a sample is rejected by u experts out of a total of n experts, then the decision combination problem is reduced to a simpler problem of combining n-u decisions according to the decision-making algorithm of the decision fusion expert. In the extreme case, where the test sample is rejected by all the experts, no decision can be taken and the sample is finally rejected (Rahman and Fairhurst[22]).

2.5   Class-Wise Best Decision Selection

The Class Confidence Index, $\beta_{i,j}$, $1 \leq i \leq n$, $1 \leq j \leq m$, where n is the number of decisions and m is the number of classes under consideration, denotes the ranking of the different decisions $\beta_{i,j} = 1, 2, ..., n$ on a class by class basis. The higher the class recognition rate, the higher the ranking. These class-wise confidences can then be conveniently converted to suitable weight values ω_k^i, expressing the comparative competence of the decisions being combined on a class-wise basis, are expressed as a list of fractions, with $1 \leq k \leq n$, $\sum_{k=1}^n \omega_k^i = 1$, n, n, again being the number of decisions. In this case, the best decision is selected based on the class confidence indices rather than on the overall confidence indices.

Therefore, the best representative decision for a particular problem domain for a particular class can then be noted by finding the maximum weighting factor ω_m^i , so that $\omega_m^i = \max \omega_k^i$. The final decision is then expressed by $d_{Best} = d_m^i$.

2.6   Enhanced Majority Voting

Simple and weighted majority voting are very robust provided adequate number of classifiers are available to reach the correct consensus. There are various ways in which qualitative enhancement can be added to this framework. Rovatti *et al.*[23] have discussed a form of cooperation between the k-nearest neighbors (NN) approach to classification and their neural like property of adaptation. A tunable, high level k-nearest neighbor decision rule is defined that comprehends most previous generalizations of the common majority rule. ENCORE (Enhanced Consensus in Recognition) has been proposed by Fairhurst and Rahman[24]. This approach implements a decision consensus approach, but the quality of the consensus is evaluated in terms of the past track record of the consenting experts before it is accepted. The logical structure of this approach can be summarized as follows:

– Find the class $\omega(X)$ having the consensus support of the experts for any arbitrary pattern X.
– Find the corresponding sample confidence indices α_{ijk}, which denote the confidence of the ith expert in identifying the kth sample coming from the jth class.
– Apply the following rule:
  • The decision associated with the highest sample confidence value α is accepted as the final decision, provided the highest sample confidence value is not separated from the next highest confidence value assigned by a cooperating expert by an amount greater than (or equal to) a threshold, so that,
$$\alpha_{uwt} \sim \alpha_{vwt} \leq \acute{\psi_c} \tag{4}$$
  where u and v are the top two cooperating experts, w is the class under consideration, t is the sample in question and $\acute{\psi_c}$ is the threshold.
  • In circumstances where the criteria of Eqn.(4) are not met, it is deemed that the top two cooperating experts show an unacceptable confidence gap and alternative solutions should be sought. In this case, the second group of consensus decisions are considered and the same process of applying Eqn.(4) to the corresponding sample confidence values is attempted. This process is repeated until all the consensus groups are exhausted.
  When this process has been exhausted, the consensus decision approach is abandoned and the decision combination approach changes. Now the top two decisions having the maximum sample confidence index are examined in terms of Eqn.(4). In circumstances where the criteria of Eqn.(4) are not met decision combination is attempted based on the criteria of Eqn.(5) so that,
$$\beta_{sw} \sim \beta_{tw} \leq \acute{\theta_c} \tag{5}$$

where β is the class confidence index, s and t are the top two competing experts, w is the class under consideration by that particular expert and $\acute{\theta}_c$ is the threshold value. The motivation behind using this criterion is to give due preference to the strengths of particular experts in recognizing a particular class of patterns.

- In the event of a failure to reach a final decision at this stage, the criterion of Eqn.(6) can be applied.

$$\gamma_s \sim \gamma_w \leq \acute{\phi}_c \tag{6}$$

where γ is the overall confidence index, s and w are the top two competing experts and $\acute{\phi}_c$ is the corresponding threshold. If no decision can be arrived at even at this stage, either an arbitrary decision of accepting the decision of the first expert in the array is taken, or the pattern is rejected.

2.7   Ranked Majority Voting

It is entirely possible to include additional information derived from participating classifiers in reaching the final consensus. Instead of only using the final class labels to which a sample is estimated to belong, it is also possible to produce a ranked list of suggested decisions covering multiple classes. These ranked lists can then be manipulated to reach a final consensus. A very interesting approach to majority voting has been put forward by Ho *et al.*[1]. In this case, decisions by the classifiers have been represented as rankings of classes so that they are comparable across different types of classifiers and different instances of a problem. The rankings are combined by methods that either reduce or re-rank a given set of classes. An intersection method and a union method have been proposed for class set reduction. Three methods based on the highest rank, the Borda count, and logistic regression are proposed for class set re-ranking. Elsewhere, Ho *et al.*[25] emphasis the re-ranking of the ranked outputs delivered by the cooperating classifiers. In [26], Ho *et al.* described a concise and focused version of the ideas presented in [1], again emphasizing the substantial improvements achievable from these multiple expert systems. Duong[27] has discussed the problem of combination of forecasts employing a ranking and subset selection approach. The ranking and subset selection approach is suggested as a statistical procedure for ranking alternative forecasts. This simple method is shown to compare favorably with those based on other optimality criteria when applied to some real data.

2.8   Committee Methods

Mazurov *et al.*[28] have discussed theorems concerning the existence of p-commit-tee for an arbitrary finite system of sets and for the finite systems of half-spaces. The existence theorem for a discriminating committee consisting of affine functions which were used in the solution of the problem of pattern recognition has been presented in detail. Yu *et al.*[29] report a multiple expert

decision combination method based on stacked generalization and committee methods. Kimura and Shridhar[30] have combined two algorithms for unconstrained handwritten numeral recognition. The first of their algorithms employs a modified quadratic discriminant function utilizing direction sensitive spatial features of the numeral image. The other algorithm utilisers features derived from the profile of the character in a structural configuration to recognized the numerals.

2.9 Regression

Ho *et al.*[31] described in detail the regression approach to the combination of decisions by multiple character recognition algorithms. Ho[32] reported an investigation on different strategies to coordinate and combine different alternative classifiers that can adapt to certain conditions in the input which concern both accuracy and speed. The design of such strategies is based on a detailed analysis of the classifiers' performances on test data using a parameterized defect model.

Table 1. Performance of the individual classifiers

Expert	Accepted	Recognized	Error	Rejected
FWS	97.35	78.76	18.59	2.65
MPC	97.62	85.78	11.84	2.38
BWS	95.50	72.31	23.19	4.50
MLP	95.13	82.31	12.82	4.87

Table 2. Performance of various majority voting systems

Combination Method	Accepted	Recognized	Error	Rejected
Simple Majority Voting	96.59	90.59	6.00	3.41
Weighted Majority Voting	96.85	90.64	6.21	3.15
Class-wise Weighted Majority Voting	96.86	90.70	6.16	3.14
Restricted Majority Voting (Top Choice)	95.68	88.97	6.71	4.32
Class-wise Best Decision Selection	96.76	89.64	6.79	3.24
Restricted Majority Voting (Generalized)	96.54	90.63	5.91	3.46
Enhanced Majority Voting (ENCORE)	97.14	90.91	6.23	2.86
Ranked Majority Voting (Borda Count)	96.99	90.77	6.22	3.01
Committee Methods	95.98	89.63	6.35	4.02
Regression Methods	97.68	90.85	6.83	2.32

3   A Sample Problem Domain

In order to compare some of these widely differing multiple classifier decision combination methods implementing forms of majority consensus, a sample problem domain of handwritten character recognition has been chosen. The source of the handwritten characters (numerals) was the database compiled by the

U.S. National Institute of Standards and Technology (NIST) [33], which contains samples of numerals 0 to 9. Four experts were chosen to be combined in the framework of the various chosen methods, which included a Binary Weighted Scheme (BWS), which employs a technique based on *n-tuple* sampling or memory network processing (Rahman and Fairhurst[34]), a Frequency Weighted Scheme (FWS), which calculates the relative frequencies of the sampled features indicating the probability distribution of the group of points or *n-tuples* (Rahman and Fairhurst[5]), a Multi-layer Perceptron Network (MLP), the familiar multilayer perceptron neural network structure, employing the standard error backpropagation algorithm (Rahman and Fairhurst[35]) and a Moment-based Pattern Classifiers (MPC), which is a maximum likelihood classifier, employing the nth order mathematical moments derived from the binarized patterns (Rahman and Fairhurst[36]).

4   Performance

Before reporting the performance of the proposed structure, it is important to assess the performance of the chosen experts individually on the selected problem. Table 1 presents the optimum performance achieved on the chosen database. Table 2 presents the results of decision combination using various decision combination schemes based on majority voting. It is clearly seen that some of these methods offer very high levels of top choice recognition (e.g. ENCORE, Regression Methods), and some offer very low levels of error rates (Restricted Majority Voting: Generalized Framework). Depending on what is required from the decision combination system, different methods can be selected in different problem domains.

From the short discussion presented so far, it is clear that the performances of these various approaches are directly related to their design emphasis. In some of these approaches, emphasis is given to assess how a consensus can be reached given the often conflicting opinions of these classifiers. Simple and Weighted Majority Methods are examples of this approach. On the other hand, other methods try to assess the 'appropriateness' or the 'suitability' of a classifier over other classifiers and favor its opinion over others. So it is seen that there are two ways of looking at this problem. One is to assume that the classifiers are cooperating with each other in reaching a final decision, the other is to assume that the classifiers are competing with each other to win an argument. Either of these approaches can be the 'most appropriate' solution to a particular problem, given that the best solution depends on the localized nature of the problem, and in this respect there is no global winner. However, a combination of these two approaches has been found to offer some advantages (ENCORE, for example, can be treated as a combination of Simple Majority and Restricted Majority techniques). Combination of decision combination methods is not a new concept (Paik *et al.*[37], Anzai *et al.*[38] etc.), but the use of this concept is relatively unexplored and preliminary results are very encouraging (Rahman and Fairhurst[24]).

A final note about the applicability of these majority voting techniques on document analysis applications. In addition to the obvious applications to character level recognition used an example in this paper, these techniques have been used in word level recognition, cursive recognition, signature recognition, PDA applications, automatic segmentation, automatic check processing, layout analysis, mail sorting, archiving, indexing and many more similar applications.

5   Conclusion

The main objective of the paper is to revisit the issue of Majority Voting in classifier design, to illustrate huge number of variations that are possible on the theme of majority voting, and to demonstrate their applicability in typical task domains. A discussion of the various multiple expert decision combination strategies implementing some form of majority voting in the context of character recognition has been presented. The review includes short descriptions of various methods and how they are related in terms of the underlying design philosophy. A comparative study of these methods on a typical handwritten character recognition problem is also included to show how, even with this apparently narrowly-defined framework, different strategic strands can be productively identified.

References

1. T. K. Ho, J. J. Hull, and S. N. Srihari. Decision combination in multiple classifier systems. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 16(1):66–75, January 1994.
2. Y. S. Huang, K. Liu, and C. Y. Suen. The combination of multiple classifiers by a neural-network approach. *Int. Jour. of Pattern Recognition and Artificial Intelligence*, 9(3):579–597, 1995.
3. A. F. R. Rahman and M. C. Fairhurst. Serial combination of multiple experts: A unified evaluation. *Pattern Analysis and Applications*, 2:292–311, 1999.
4. A. F. R. Rahman and M. C. Fairhurst. Enhancing multiple expert decision combination strategies through exploitation of a priori information sources. *IEE Proc. on Vision, Image and Signal Processing*, 146(1):1–10, 1999.
5. A. F. R. Rahman and M. C. Fairhurst. Machine-printed character recognition revisited: Re-application of recent advances in handwritten character recognition research. *Special Issue on Document Image Processing and Multimedia Environments, Image & Vision Computing*, 16(12-13):819–842, 1998.
6. L. Lam, Y. S. Huang, and C. Y. Suen. *Combination of multiple classifier decisions for optical character recognition in Handbook of Character Recognition and Document Image Analysis*, pages 79–101. World Scientific Publishing Company, 1997. H. Bunke and P. S. P. Wang(Eds.).

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

8. L. Lam and C. Y. Suen. Application of majority voting to pattern recognition: An analysis of its behavior and performance. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 27(5):553–568, 1997.

9. L. Lam and C. Y. Suen. A theoretical-analysis of the application of majority voting to pattern-recognition. In *Proc. 12th IAPR Int. Conf. on Pattern Recognition, Conf. B: Pattern Recognition and Neural Networks*, volume 2, pages 418–420, Jerusalem, Israel, 1994.

10. I. Todhunter. *A History of the Mathematical Theory of Probability from the Time of Pascal to that of Laplace.* Macmillan, Cambridge, UK, 1865.

11. N. C. de Condorcet. *Essai sur l'Application de l'Analyze à la Probabilité des Décisions Rendues à la Pluralité des Voix.* Imprimérie Royale, Paris, France, 1785.

12. C. Y. Suen, C. Nadal, T. A. Mai, R. Legault, and L. Lam. Recognition of totally unconstrained handwritten numerals based on the concept of multiple experts. In *Proc. IWFHR*, pages 131–143, Montréal, Canada, 1990.

13. G. S. Ng and H. Singh. Democracy in pattern classifications: Combinations of votes from various pattern classifiers. *Artificial Intelligence in Engineering*, 12(3):189–204, 1998.

14. A. Stajniak, J. Szostakowski, and S. Skoneczny. Mixed neural-traditional classifier for character recognition. In *Proc. Int. Conf. on Imaging Sciences and Display Technologies: International Society for Optical Engineering (SPIE)*, volume 2949, pages 102–110, Berlin, Germany, 1997.

15. A. Belaid and J. C. Anigbogu. Use of many classifiers for multifont text recognition. *Traitement du Signal*, 11(1):57–75, 1994.

16. J. R. Parker. Voting methods for multiple autonomous agents. In *Proc. 3rd Australian and New Zealand Conf. on Intelligent Information Systems*, pages 128–133, Perth, WA, Australia, 1995.

17. C. Ji and S. Ma. Combination of weak classfiers. *IEEE Trans. on Neural Networks*, 8(1):32–42, 1997.

18. E. Alpaydin. Improved classification accuracy by training multiple models and taking a vote. In *Sixth Italian Workshop. Neural Nets Wirn Vietri-93*, pages 180–185, 1994.

19. J. Kittler, A. Hojjatoleslami, and T. Windeatt. Weighting factors in multiple expert fusion. In *Proc. British Machine Vision Conference*, pages 41–50, 1997.

20. S. R. Gang, T. K. Woo, and I. C. Sung. Recognition of printed and handwritten numerals using multiple features and modularized neural networks. *Journal of the Korean Institute of Telematics*, 32B(10):101–1111, 1995.

21. A. F. R. Rahman and M. C. Fairhurst. Exploiting second order information to design a novel multiple expert decision combination platform for pattern classification. *Electronics Letters*, 33(6):476–477, 1997.

22. A. F. R. Rahman and M. C. Fairhurst. A New Multiple Expert Framework for Decision Fusion. In *Proc. 9th Int. Graphonomics Society Conference (IGS'99)*, pages 161–166, Singapore, 1999.

23. R. Rovatti, R. Ragazzoni, Z. M. Kovacs, and R. Guerrieri. Voting rules for k-nearest neighbors classifiers. *Neural Computation*, 7(3):594–605, 1995.

24. M. C. Fairhurst and A. F. R. Rahman. Enhancing Consensus in Multiple Expert Decision Fusion. *IEE Proc. on Vision, Image and Signal Processing*, 147(1):39–46, 2000.

25. T. K. Ho, J. J. Hull, and S. N. Srihari. *Combination of Decisions by Multiple Classifiers in Structured Document Image Analysis*, pages 188–202. S-V, 1992. H. S. Baird, H. Bunke and K. Yamamoto(Eds.).

26. T. K. Ho, J. J. Hull, and S. N. Srihari. On multiple classifier systems for pattern recognition. In *Proc. 11th ICPR*, pages 84–87, The Hague, Netherlands, 1992.

27. Q. P. Duong. The combination of forecasts: A ranking and subset selection approach. *Mathematical and Computer Modelling*, 12(9):1131–1143, 1989.

28. V. D. Mazurov, A. I. Krivonogov, and V. L. Kazantsev. Solving of optimisation and identification problems by the committee methods. *Pattern Recognition*, 20(4):371–378, 1987.

29. Y. H. Hu, J. M. Park, and T. Knoblock. Committee pattern classifiers. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, pages 3389–3392, Munich, Germany, 1997.

30. F. Kimura and M. Shridhar. Handwritten numeral recognition based on multiple algorithms. *Pattern Recognition*, 24(10):969–983, 1991.

31. T. K. Ho, J. J. Hull, and S. N. Srihari. A regression approach to combination of decisions by multiple character recognition algorithms. In *Machine Vision Applications in Character Recognition and Industrial Inspection, Proc. SPIE 1661*, pages 137–145, 1992.

32. T. K. Ho. Adaptive coordination of multiple classifiers. In *Proc. of Int. Workshop on Document Analysis Systems*, pages 338–351, Malvern, Pennsylvalia, USA, 1996.

33. NIST Special Databases 1-3, 6-8, 19, 20, National Institute of Standards and Technology, Gaithersburg, MD 20899, USA.

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

35. A. F. R. Rahman and M. C. Fairhurst. Towards the design of a generic multiple expert image classification system capable of automatic self-configuration. In *Proc. Int. Conf. on Quality Control by Artificial Vision (QCAV'99)*, pages 259–264, Québec, Canada, 1999.

36. M. C. Fairhurst and A. F. R. Rahman. A generalised approach to the recognition of structurally similar handwritten characters. *IEE Proc. on Vision, Image and Signal Processing*, 144(1):15–22, 1997.

37. J. Paik, S. Jung, and Y. Lee. Multiple combined recognition system for automatic processing of credit card slip applications. In *Proceedings of the Second International Conference on Document Analysis and Recognition (Cat. No.93TH0578-5)*, pages 520–523, 1993.

38. Y. Anzai, H. Mori, M. Ito, and Y. Hayashi. A serial-parallel integrated information-processing model for complex human problem solving. In *Cognitive Engineering in the Design of Human-Computer Interaction and Expert Systems. Proceedings of the Second International Conference on Human-Computer Interaction. Vol.II*, pages 175–182, 1987.