Ludmila I. Kuncheva
School of Computer Science
Bangor University, UK
l.i.kuncheva@bangor.ac.uk

*Teaching and Research Practices in Pattern Recognition (Personal Views and Experiences)**

* A disclaimer: This document is a reflection based upon 25 years of research and over a decade of teaching Pattern Recognition at Bangor University, UK. The paper represents my personal views and should not be generalised lightly. More importantly, it should not be taken as the view of my School of Computer Science or that of Bangor University.

1. Introduction

1.1 Pattern Recognition

Pattern recognition is about labelling objects into classes. To predict the class label of an object, its feature values are fed into an algorithm, called a 'classifier'. Originating within the realm of statistics, dating back to the famous Fisher's linear discriminant [12] and Fix and Hodges' nearest neighbour classifier [13], the pattern recognition theme evolved through several parallel streams. Pattern Recognition grew strong on the image processing front, Machine Learning claimed new territories on the theoretical side of classification, both intertwined with neurocomputing (neural networks) and data mining. Figure 1 shows diagrammatically areas related to pattern recognition (the arrangement is arbitrary). With the intricate interconnections and overlap between these areas, an important task is to unify the terminology and facilitate cross-disciplinary exploitation of the individual advances. The variety of challenges posed by pattern recognition applications serve as a catalyst for new developments. For example, spam e-mail classification requires that the classifier keeps on learning during its operation in order to respond to the ever changing spamming pattern. This fuels the research into concept drift (changing environments). On the other hand, the main difficulty in microarray data analysis is the large number of genes (features) describing the objects (tissue samples) to be classified. Methods for feature selection and extraction, radically different from the mainstream methods, have to be developed for this type of applications. Network security systems operate on data characterised by unbalanced classes: class 'abnormal behaviour' is a fraction of the size of class 'normal behaviour'. This calls for cost-sensitive classification methods. Pattern recognition underpins all such developments, hence the need for teaching the subject's ontology.

1.2 Classifier ensembles

Combination of classifiers (commonly called 'classifier ensembles') received much attention in the past decade [22] although original works suggesting such combinations can be traced back to the 70s and early 80s [1, 9, 27]. The idea is to use "collective intelligence" instead of a single "expert". Collective decision making (voting, for example) has been in human society since the dawn of civilisation; *Vox populi, vox dei*. We list below some of the reasons for replacing a single classifier with an ensemble

Figure 1. Areas related to pattern recognition

1. Ensembles are, on average, more accurate than single classifiers.
2. Using an ensemble eliminates the risk of a poor choice of a single classifier.
3. Some problems are too complex as a whole but can be decomposed into easier sub-problems; an ensemble of simple classifiers can be used to solve a complex classification problem.
4. The data or other resources may be distributed and training of a single classifier may be infeasible.

Figure 2 (a) shows a digram of a classifier ensemble. The outputs of the component classifiers are combined and a final class label for the input data point x is produced. A scatterplot of a data set consisting of two classes is given in Figure 2 (b). The decision boundaries of 5 neural network classifiers (multi-layer perceptron) are overlaid with thin lines. The ensemble boundary is shown using a thick line. The ensemble decision is more accurate than that of any of the component classifiers; its boundary is the smoothest and also the closest to the optimal boundary guaranteeing minimum error.

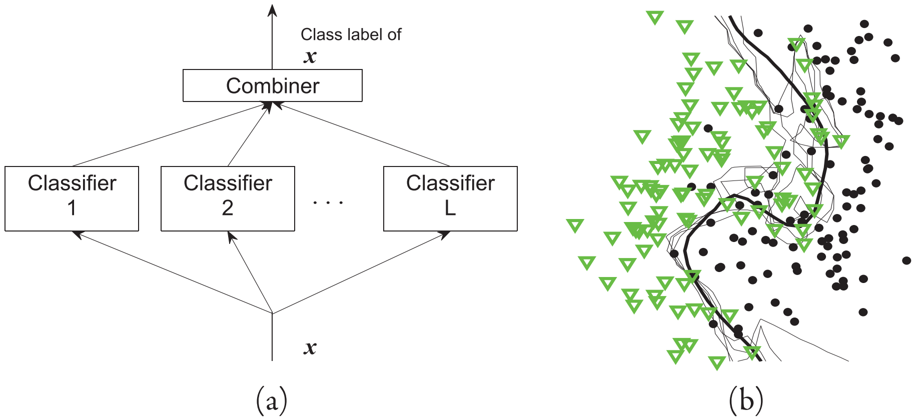

Figure 2. (a) Classifier combination (ensemble structure); (b) Scatterplot of 2 classes and classification boundaries of 5 individual classifiers (thin lines) and the ensemble (thick line)

An introduction to the area of classifier ensembles would not be complete without listing some of the numerous "aliases" of classifier ensembles in past literature (Table 1). The series of Workshops on *Multiple Classifier Systems* (MCS) [16, 26, 29–32, 35]1[1] played a pivotal role in unifying the terminology, stopping us from rediscovering each other's results, and encouraging us to work together towards building our "Tower of Babel". The credit for establishing the MCS Workshops goes to Professor Fabio Roli (Cagliary, Italy) and Professor Josef Kittler (Surrey, UK).

1 http://www.diee.unica.it/mcs/

Table 1. Classifier ensemble "aliases" in the literature

1	combination of multiple classifiers [21, 23, 28, 38, 39];
2	classifier fusion [6, 8, 14, 15, 20];
3	mixture of experts [17–19, 25];
4	committees of neural networks [5, 10];
5	consensus aggregation [3, 4, 24];
6	voting pool of classifiers [2];
7	dynamic classifier selection [38];
8	composite classifier systems [9];
9	classifier ensembles [10, 11, 33];
10	bagging, boosting, arcing, wagging [33];
11	modular systems [33];
12	collective recognition [1, 27]
13	stacked generalization [37];
14	meta-learning [36];
15	divide-and-conquer classifiers [7];
16	pandemonium system of reflective agents [34];

2. Teaching Pattern Recognition

2.1 The Computer Science Degree at Bangor

Enthusiastic as I was when I joined the School of Mathematics in Bangor in 1997, I was planning to raise students' understanding of mathematics to the level where I can teach advanced pattern recognition topics such as classifier ensembles. Gradually my enthusiasm was cooled down. I watched with sadness and a feeling of doom the shrinking trickle of students applying to study mathematics at Bangor followed by the demise of the School of Mathematics and shutting down the mathematics degree.

At the same time the Computer Science (CS) degree was shaping up and attracting students with various academic backgrounds and abilities. While mathematics is a "well channelled" degree with long tradition and structure almost carved in stone, CS is far from it. We had to design a degree course that would pass the accreditation of the relevant professional bodies and at the same time suit our particular student population. Teaching plans were constantly changing, modules were being introduced, discontinued or modified; the competition with the other UK universities was fierce. Just when our young CS degree was producing its first graduates it was hit by the "bursting of the dot-com bubble" (2001-2002) heralding a rapid withdrawal from computing degrees all over the UK.[2] A low tide of applications for CS affected Bangor University too but not as sverely as other leading universities. The degree is still recovering across UK, with applications levelling off. In 2008/9 we have enrolled 85 students - the largest intake ever for computer science degrees at Bangor.

With all the turbulence, the CS degree at Bangor University was ranked 4th in the Guardian league table for 2008[3] after Edinburgh, Oxford and York and before Southampton and Cambridge. The table was produced from a student satisfaction survey across the UK. We scored low on our (quite liberal)entry requirements but scored 10/10 on employability after graduation. I believe our success hinges on three main features of our CS degree programme

1. Software Hut. This is a module taught in the second year. The students are split into groups of 3 or 4, and each team is assigned a mini-project. The team has to elect a leader, make a plan, manage and execute the project, and finally give an oral presentation. Controversial as team work is, students get invaluable experience of "the good, the bad and the ugly". More importantly, IT Wales[4] has an office in Bangor which helps to create and foster links with local industries. Thus the mini-projects that students work on are often simplified versions of real-life problems suggested by the industrial collabora-

2 http://www.ucas.com/about_us/stat_services/stats_online/
3 http://education.guardian.co.uk/universityguide2008
4 http://www.itwales.com/

tors. The real benefit for both parties is that Software Hut gets students and managers thinking on the same wavelength. Of course, getting to know each other in person is a big factor in a potential future employment.

2. Individual projects. The 3rd (final) year project accounts for 1/4 of the mark for the third year. Failing the project (scoring less than 40%) automatically forfeits the degree. Each student has an individually assigned project, usually rlated to the student's strengths and interests. Project topics in pattern recognition and data analysis that I have supervised include cricket score prediction, analysis of data about children behaviours, diagnosing scrapie in sheep, identifying fossils in microscope images, designing a maths-millionnaire game and many more. The students enjoy working on the individual projects, so much so that sometimes they tend to neglect the rest of their third year modules.

3. Teaching our research topics. A great part of our CS degree is built around the research interests of the staff. Thus we offer pattern recognition along with medical data visualisation, virtual environments, computer graphics, artificial agents, and so on.

2.2 Subjects to stay and subjects to go

The School of Computer Science has consulted with external experts on teaching this subject in higher education. This process has identified subjects which are likely to become less relevant to CS and others that are likely to surface, move more centre-stage or stay in the curricula in a new disguise (Table 2).

The experts gave their predictions based upon projections about the employment market of CS graduates as well as on the socio-cultural expectations of potential CS students. In other words, we have to make the degree attractive, feasible and useful at the same time. The key to success is to keep the delicate balance between universal and unique, academic and vocational, high standard and accessible; and make the degree coherent at the same time. Not easy...

Table 2. Subjects from CS curriculum whose importance is predicted to decline or rise (listed in no particular order)

On the decline	On the rise
maths	real-time and embedded systems
hardware	natural language processing
design UML	Internet-based development
formal methods	security (cryptography)
data structures	distributed systems
software engineering	IT and business
programming languages	management
(advanced programming)	robotics
analysis and specification	AI

2.3 Maths or no maths, that is the question

The dilemma. In the UK, students come to university after two final years in school where they study 3 subjects of their choice known as A-levels (some students take 4 or even 5 subjects). Mathematics suffered a substantial drop of number of students taking it at A-level in 2000 and has slowly been rising since then. Striving to increase our undergraduate intake we were quick to drop mathematics from our entry requirements. The maths that is taught in the first year resembles a recipe book of algorithms; we test students' memory rather than their understanding. The problems appear later when a certain level of mathematics is expected from the students. Pattern Recognition and Neural Networks is a third year module which is supposed to introduce the fundamentals of classifier design. Basic knowledge of probability and statistics is vital for understanding the material. Usually the class would have mixed abilities including extreme cases such as students not being able to sum up simple fractions as well as students who would comfortably grasp the concept of multi-dimensional probability density functions. What makes the problem particularly acute is that there is no textbook on pattern recognition that is so unassuming when it comes to maths. Maybe it is time I wrote a textbook titled “Pattern Recognition for Dummies”, “Idiot’s Guide to Pattern Recognition” or “Pattern Recognition without tears”?

The “polite mayor” script. This is a true story from the time before we dropped Probability and Statistics from the curriculum. One of the problems on the exam, related to continuous-valued random variables and probability density functions, was phrased approximately like this: “In the opening ceremony for a new monument, the Mayor of the town cuts the ribbon into two pieces. Assuming that there is an equal chance the ribbon to be cut at any particular place, what is the probability (in theory!) that the two pieces have exactly the same length?” Of course the answer is zero, and I would have accepted an answer “negligibly small” too. The answer in one of the scripts (meant to be a genuine answer!) was something like this:

> *“The probability is very small. The Mayor is a polite man; he would not stand right in the middle because that way he would obscure the view for the crowd that had gathered for the opening. So he would cut the ribbon somewhere on the side.”*

Needless to say, from the 2 marks this question was worth, I gave the student 1 mark for originality. Sadly, this quote is indicative of the level of understanding and appreciation of the subject among first year CS students. The dislike of maths, rather the mental block, is partly due to lack of mathematical abilities and partly due to attitude - “Oh, that is maths; I don’t understand maths! And I have come to study computer science, not maths!” Figure 3 illustrates the mutual love and understanding between me and the Pattern Recognition class. This is a photo I took after a lecture showing one of the desks in the classroom. Engraved on the desk was “Lecturer destruct button. Please press”. Students may be stubborn in their aversion to maths but they have met their match - I am not giving up just yet.

Figure 3. Student's graffiti engraved on a desk in the classroom.

3. Research on Pattern Recognition and Machine Learning

3.1 Breeding PhDs

As in every branch of science, capable PhD students are our best asset and hope for the future. Bangor has been fortunate to attract and retain exceptional students in the past; mostly ones who live in Wales and prefer to stay in the area. One problem, possibly cultural, is UK students' reluctance to travel abroad and share their research achievements with the rest of Europe and the world. The practices I have seen in other countries regarding PhD development are quite different. I have had visiting PhD students working on classifier ensembles coming from Spain, Russia, Lithuania and Italy and staying from 3 to 6 months. Exchange programmes and other types of funding are available to UK postgraduates but they rarely choose to use these opportunities.

Having served as an external examiner for PhD theses from The Netherlands, Italy and Spain, as well as the UK, it strikes me that the length of a PhD course is quite variable across Europe, and so are the standards, requirements and examination procedures. In some countries the defence of the thesis is a mere formality after up to 10 years full-time productive research on the candidate's part. Since the PhD degree is recognised world-wide, more effort should be put in trying to equalise standards, if not across Europe, at least within the UK.

Publishing is an essential part of PhD training. A good practice would be to push PhD students to publish as early as possible during their course. Admittedly, at the early stages, the student would not have gathered sufficient material for publication of their own research. Nonetheless, the supervisor may involve the student as an apprentice in the supervisor's ongoing publishing activities. This always pays off because the student learns the art of writing and can later approach writing their PhD thesis with skill and confidence. At the later stages of the PhD, capable students should be encouraged to publish a single-author paper, of course vetted by the supervisor. This goal is attainable for pattern recognition PhDs, helped by the fact that the area is very active and there are many old and new publication fora.

3.2 Time, staff and money

Organisation of own research at a higher education institution is a complex task. The three 'elephants': *time*, *staff* and *money*, are hardly ever in place.

I would rank time as the most important, most deficient and most unforgiving of the three. Administrative, teaching and pastoral duties progressively take over. The little bouts left in-between other activities are hardly of any use, as research requires larger uninterrupted periods of time. Things were quite different at the Bulgarian Academy of Sciences where I used to work until 1997: time was solely assigned for research; teaching, student supervision or heavy administrative duties were optional. Time is unforgiving in the sense that once you "step off the train" having been distracted by other duties, you may need months of unproductive catching up in order "get back on". What is the solution? Well, there isn't one! While staff and money can be accumulated, time is a limited resource. We can try planning, compromising and toning down our ambition.

Having a viable research group is paramount for producing world-class research. Two or more full-time members of staff in one research group is a luxury that only rich universities can afford. As everywhere else, research groups consists of a leader and a team of PhD students and temporary researchers. Building a team can be done by securing contract funding for post-doc researchers as well as winning PhD grants. The model, which I have adopted in Bangor follows that of Professor Jim Bezdek, University of Pensacola, Florida, where I visited as a post-doc in 1996-1997. The group around the leading academic is mostly composed of visiting researchers, and further contacts are maintained through "e-collaboration". Thus the "group" in Bangor has had collaborations with various researchers and groups across the world. Our European links are shown on the map in Figure 4 (Spain, Italy, The Netherlands, Russia and Lithuania). In contrast again, at the Bulgarian Academy of Sciences, having a group of full-time senior-level staff with similar expertise was not a problem. A group would include one or two staff at professorial level, 3-4 at senior-research level and 3-4 younger researchers (PhD students or post-doc research assistants). The group's research outcome used to be measured collectively; collaborations within the group were dynamically formed. Occasionally we had foreign visitors too.

So what was wrong with the Bulgarian Academy of Sciences then? Mostly the funding, or rather the lack thereof. Salaries were inadequate, literature was inaccessible, travelling to conferences and inviting colleagues over were both infeasible, and the incentives to boost the groups' work rather than one's individual career were dubious.

In the UK, limited funding is distributed to the universities to maintain a minimum research standard. Funds need to be sought externally for PhD fees and subsistence, post-doc salaries, visiting collaborators, research and teaching help, equipment, as well as for attending conferences. Such external research funding in the UK is administered through research councils. Pattern recognition (CS in general) falls mainly in the remit of EPSRC (Engineering and Physical Sciences Research Council)[5]. Difficulties in getting an EPSRC research grant are notorious. In general, across all UK research councils and disciplines, the top 20 universities secure nearly 2/3 of research funds while others are left with nothing[6]:. In CS, the average success rate is approximately 1/10 projects. Less experienced academics or

5 http://www.epsrc.ac.uk/
6 Times Higher Education supplement (THE), 21st August 2008

Figure 4. Research collaborations on Classifier Ensembles between the Bangor group (Wales, UK) and groups in other European countries (at least one common publication)

novices to the grant bidding system in the UK have much less chance to prepare a fundable application compared to colleagues who have already had successful proposals. Recall the Ecclesiastes "*To the place where the rivers flow, there they flow again*". In an attempt to break the vicious circle, a scheme called "First Grant" is in place, which gives a better chance of success for new academics.

In spite of the difficulties in project funding and the acknowledged increase of time restrictions in academia[7], UK is still a major player in terms of research output (across all disciplines)[8]:

- UK produces the largest number of research papers (8.63%) after USA, followed closely by China. The top 5 countries are: USA, UK, China, Germany, Japan.
- UK is second on the number of citations (12%) after USA (45.5%). The top 5 countries are USA, UK, Germany, Japan, France.

7 THE, 31st July 2008
8 THE, 7th August 2008

– UK is ranked as the top country for "value for money" producing more papers and citations than any other country per unit of investment in research and development (USA is 17th and China is 16th).

CS is a part of this research summary, and we would like to believe that it follows suit. Pattern recognition is a truly international subject that benefits from collaboration across many countries. Due to geographic proximity of the countries, Europe is in an excellent position to foster such collaborations through domestic or pan-European funding programmes. Some European grant schemes have high administrative overload (paperwork and regular inspection meetings) and are less favoured by UK researches than domestic programmes. On the other hand, I have had very positive experiences with research grants supporting international collaboration from The Royal Society and the INTAS project.

4. Conclusions: What should we do?

Teaching. The moral of the story is: we should not be longing for the past but embrace the new realities and rethink our teaching in the new light. For example, we should pick out elements of mathematics that are relevant to pattern recognition, and teach them at a level and pace that students could take. At the same time, we should learn to exploit the abilities of a modern student to multi-task, to filter massive amounts of information scattered across different media, and sometimes invent ingenious solutions within limited resources.

Research. Second, research in pattern recognition and machine learning is much more profiled now than it was ten or fifteen years ago. Subject-specific research directions in pattern recognition have branched out so as to respond better to the challenges of the particular field. For example, the problem of imbalanced classes is a challenge in network security; "mind-reading" (neuroimaging) is faced with handling humongous data sets; biometric verification has to deal with large number of classes with only few examples in each class; spam e-mail recognition must respond to an ever changing classification problem, and so on. General-purpose pattern recognition research will still be needed but the leaps ahead can be expected from within subject-specific branches.

References

1. Y. L. Barabash. *Collective Statistical Decisions in Recognition*. Radio i Sviaz', Moscow, 1983. (In Russian).

2. R. Battiti and A.M. Colla. Democracy in neural nets: Voting schemes for classification. *Neural Networks*, 7:691–707, 1994.

3. J.A. Benediktsson, J.R. Sveinsson, J.I. Ingimundarson, H. Sigurdsson, and O. K. Ersoy. Multistage classifiers optimized by neural networks and genetic algorithms. *Nonlinear Analysis, theory, Methods & Applications*, 30(3):1323–1334, 1997.

4. J. A. Benediktsson and P. H. Swain. Consensus theoretic classification methods. *IEEE Transactions on Systems, Man, and Cybernetics*, 22:688–704, 1992.

5. C.M. Bishop. *Neural Networks for Pattern Recognition*. Clarendon Press, Oxford, 1995.

6. I. Bloch. Information combination operators for data fusion: a comparative review with classification. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 26:52–67, 1996.

7. C. C. Chiang and H. C. Fu. A divide-and-conquer methodology for modular supervised neural network design. *In IEEE International Conference on Neural Networks*, pages 119–124, Orlando, Florida, 1994.

8. S-B. Cho and J.H. Kim. Combining multiple neural networks by fuzzy integral for robust classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 25(2):380–384, 1995.

9. B. V. Dasarathy and B. V. Sheela. A composite classifier system design: concepts and methodology. *Proceedings of IEEE*, 67:708–713, 1978.

10. H. Drucker, C. Cortes, L. D. Jackel, Y. LeCun, and V. Vapnik. Boosting and other ensemble methods. *Neural Computation*, 6:1289–1301, 1994.

11. E. Filippi, M. Costa, and E. Pasero. Multy-layer perceptron ensembles for increased performance and faulttolerance in pattern recognition tasks. In *IEEE International Conference on Neural Networks*, pages 2901– 2906, Orlando, Florida, 1994.

12. R.A. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7:179–188, 1936.

13. E. Fix and J. L. Hodges. Discriminatory analysis : Non parametric discrimination : Small sample performance. Technical Report Project 21 - 49 - 004 (11), Randolph Field,Texas, 1952.

14. P. D. Gader, M. A. Mohamed, and J. M. Keller. Fusion of handwritten word classifiers. *Pattern Recognition Letters*, 17:577–584, 1996.

15. M. Grabisch and F. Dispot. A comparison of some for fuzzy classification on real data. In *2nd International Conference on Fuzzy Logic and Neural Networks*, pages 659–662, Iizuka, Japan, 1992.

16. M. Haindl, J. Kittler, and F. Roli, editors. *Proc 7th Int. Workshop on Multiple Classifier Systems (MCS 2005), Lecture Notes in Computer Science LNCS* **4472**. Springer Verlag, Prague, Czech Republic, 2007.

17. R. A. Jacobs. Methods for combining experts' probability assessments. *Neural Computation*, 7:867–888, 1995.

18. R. A. Jacobs, M. I. Jordan, S. J. Nowlan, and G.E. Hinton. Adaptive mixtures of local experts. *Neural Computation*, 3:79–87, 1991.

19. M. I. Jordan and L. Xu. Convergence results for the EM approach to mixtures of experts architectures. *Neural Networks*, 8:1409–1431, 1995.

20. J.M. Keller, P. Gader, H. Tahani, J-H. Chiang, and M. Mohamed. Advances in fuzzy integration for pattern recognition. *Fuzzy Sets and Systems*, 65:273–283, 1994.

21. 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.

22. L. I. Kuncheva. *Combining Pattern Classifiers. Methods and Algorithms*. John Wiley and Sons, N.Y., 2004.

23. L. Lam and C.Y. Suen. Optimal combination of pattern classifiers. *Pattern Recognition Letters*, 16:945–954, 1995.

24. K. C. Ng and B. Abramson. Consensus diagnosis: A simulation study. *IEEE Transactions on Systems, Man, and Cybernetics*, 22:916–928, 1992.

25. S. J. Nowlan and G. E. Hinton. Evaluation of adaptive mixtures of competing experts. In R.P. Lippmann, J.E. Moody, and D.S. Touretzky, editors, *Advances in Neural Information Processing Systems 3*, pages 774–780, 1991.

26. N. Oza and R. Polikar, editors. *Proc 6th Int. Workshop on Multiple Classifier Systems (MCS 2005), Lecture Notes in Computer Science LNCS* **3541**. Springer Verlag, California, USA, 2005.

27. L. A. Rastrigin and R. H. Erenstein. *Method of Collective Recognition*. Energoizdat, Moscow, 1981. (In Russian).

28. G. Rogova. Combining the results of several neural network classifiers. *Neural Networks*, 7:777–781, 1994.

29. F. Roli and J. Kittler, editors. *Proc. 1st International Workshop on Multiple Classifier Systems (MCS 2000), Lecture Notes in Computer Science LNCS* **1857**. Springer-Verlag, Cambridge, UK, 2001.

30. F. Roli and J. Kittler, editors. *Proc. 2nd International Workshop on Multiple Classifier Systems (MCS 2001), Lecture Notes in Computer Science LNCS* **2096**. Springer-Verlag, Cambridge, UK, 2001.

31. F. Roli and J. Kittler, editors. *Proc 3rd Int. Workshop on Multiple Classifier Systems (MCS 2002), Lecture Notes in Computer Science LNCS* **2364**. Springer Verlag, Cagliari, Italy, 2002.

32. F. Roli, J. Kittler, and T. Windeatt, editors. *Proc 5th Int. Workshop on Multiple Classifier Systems (MCS 2004), Lecture Notes in Computer Science LNCS* **3077**. Springer Verlag, Cagliari, Italy, 2004.

33. A. J. C. Sharkey, editor. *Combining Artificial Neural Nets. Ensemble and Modular Multi-Net Systems*. Springer-Verlag, London, 1999.

34. F. Smieja. The pandemonium system of reflective agents. *IEEE Transactions on Neural Networks*, 7:97–106, 1996.

35. T. Windeatt and F. Roli, editors. *Proc 4rd Int. Workshop on Multiple Classifier Systems (MCS 2003), Lecture Notes in Computer Science LNCS* **2709**. Springer Verlag, Guildford, UK, 2003.

36. I. H. Witten and E. Frank. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, 2nd edition, 2005.

37. D. H. Wolpert. Stacked generalization. *Neural Networks*, 5(2):241–260, 1992.

38. K.Woods, W. P. Kegelmeyer, and K. Bowyer. Combination of multiple classifiers using local accuracy estimates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:405–410, 1997.

39. L. Xu, A. Krzyzak, and C. Y. Suen. Methods of combining multiple classifiers and their application to handwriting recognition. *IEEE Transactions on Systems, Man, and Cybernetics*, 22:418–435, 1992.