Monograph:
Downs, J., Harrison, R.F. and Cross, S. (1994) A Neural Network Decision-Support Tool for the Diagnosis of Breast Cancer. Research Report. ACSE Research Report 548 . Department of Automatic Control and Systems Engineering

---

A Neural Network Decision-Support Tool for the Diagnosis of Breast Cancer

Joseph Downs, Robert F Harrison
Department of Automatic Control and Systems Engineering
The University of Sheffield

Simon S Cross
Department of Pathology
University of Sheffield Medical School
The University of Sheffield

Research Report # 548
14 October 1994

Abstract

An application of the ARTMAP neural network to the diagnosis of breast cancer is described. Performance results are given for 10 individual ARTMAP networks and the five most accurate such networks using "pooled" decision making (the so-called *voting strategy*). The results are compared with those of expert and neophyte human pathologists. These show that ARTMAP diagnoses are at least as accurate as those of the expert, and can approach the optimum for the domain. However, human pathologists bias their predictions in order to minimize false positive predictions at the expense of increased false negatives. The same effect is achieved in ARTMAP by pruning category cluster nodes which make positive predictions. Overall diagnostic accuracy is not changed by this pruning. Symbolic rules are then extracted from the pruned networks. The appropriateness of the rules has been verified by a domain expert, confirming that ARTMAP can self-organise a valid mapping between input features and diagnoses in this domain. It is concluded that ARTMAP has strong potential as a decision-support tool for this task, since it learns by example, provides very accurate predictions and possesses explanatory rules for its diagnoses.

Correspondence Address:
Joseph Downs
Department of Automatic Control and Systems Engineering
University of Sheffield
Sheffield, S1 4DU
e-mail:downs@acse.shef.ac.uk

1. Introduction

Neural networks have great potential value as decision-support tools in medical domains. Unlike expert systems, they bypass the difficult and time-consuming knowledge acquisition process by learning complex associations directly from domain examples. This potentially allows them to adapt so as to perform the same task under varying conditions, as occurs for example because of different demographic populations of different regional hospitals.

However, most neural networks suffer from the opaqueness of their associations. This "black box" nature may make clinicians reluctant to utilise a neural network decision-support tool, no matter how great the claims made for its performance. Thus, there is a need to supplement neural networks with symbolic rule extraction capabilities in order to provide explanatory facilities for the network's "reasoning".

In this paper we describe an application of the ARTMAP neural network (Carpenter, Grossberg and Reynolds, 1991) to the diagnosis of breast cancer. ARTMAP has recently been provided with symbolic rule extraction capabilities (Carpenter and Tan, 1993; Tan, 1994) enabling the network to justify its predictions. Thus it seems highly suitable for medical diagnosis tasks.

The structure of the remainder of this paper is as follows. Section 2 provides an outline description of ARTMAP for the benefit of those unfamiliar with this particular neural network model. Section 3 describes the application domain. Section 4 gives performance results for ARTMAP in this domain. Section 5 discusses the findings and provides suggestions for future work.

2 ARTMAP

ARTMAP (Carpenter, Grossberg and Reynolds, 1991) is a self-organizing, supervised learning, neural network model for the classification of binary patterns[1]. It is one of a series of models based upon Adaptive Resonance Theory, or ART, (Carpenter and Grossberg, 1991) an outgrowth of competitive learning which overcomes the stability problems of that paradigm (Grossberg, 1987). This is achieved by utilizing feedback between layers of input and category nodes in addition to the standard feedforward connections of competitive learning. Thus, in ART models, an input pattern is not automatically assigned to the category that is initially maximally activated by the input. It should also be noted that most ART models, including ARTMAP, employ a localist representation for category nodes owing to the so-called "winner-take-all" competitive learning dynamics.

ARTMAP itself consists of three modules, two ART 1 systems (Carpenter and Grossberg, 1987) termed ART_a and ART_b, and a related structure termed the map field. During training, input patterns are presented to ART_a together with their associated teaching stimuli at ART_b. Associations between patterns at ART_a and ART_b are then formed at the map field. During testing, supervisory inputs at ART_b are omitted, and instead the inputs at ART_a are used to recall a previously learned association with an ART_b pattern via the map field.

---

[1] In actuality, our implementation is most closely akin to Simplified Fuzzy ARTMAP (Kasuba, 1993), which can process analogue or binary data. However, with the purely binary data of this application, cf. section 3, the implementation coincides with ARTMAP.

However, ARTMAP does not directly associate inputs at ART_a and ART_b. Rather, such patterns are first self-organized into prototypical category clusters before being associated at the map field. Hence generalized associations are formed[2].

Training in ARTMAP almost always results in multiple category clusters forming at ART_a for each teaching category present at ART_b. Each such ART_a cluster thus represents a significant sub-region of the overall state space covered by a particular teaching category. It can be seen therefore that ARTMAP instantiates a many-to-one mapping between ART_a input patterns and their actual classification.

In addition to the symbolic rule extraction facilities (which will be described in section 4), ARTMAP has a number of other desirable properties for potential use as a decision-support tool in medical domains.

First, ARTMAP has few user-changeable parameters, which allows the model to be tuned to a particular problem without undue effort. Indeed, the only parameter which has a significant effect in this domain is the ART_a baseline vigilance parameter. This determines how close a match is required between an ART_a input pattern and a category cluster prototype before accepting an input as a member of the cluster. This parameter (indirectly) controls the size of the category clusters that will form, since the higher it is set, the closer acceptable matches must be, and the smaller the coverage of the state space each cluster will have.

Second, ARTMAP does not perform optimization of an objective function and is not therefore prone to the problem of local minima as occurs with feedforward networks using backpropagation. Instead, as described before, it self-organizes its own structuring of the data, automatically creating new category nodes for itself as and when they become needed.

Third, the model is able to discriminate rare events from a "sea" of similar cases with different outcomes owing to the feedback mechanism based on top-down matching of learned categories to input patterns. This is again in contrast to feedforward networks using backpropagation where weights are refined by a process which effectively averages together similar cases and thus fails to acknowledge rare events.

Fourth, successful learning in ARTMAP can occur with only one pass through the data set. Furthermore, the model is capable of incorporating new data items at a later time without degradation of performance on previous data, or the necessity of retraining on such past data. (This solution to the so-called *stability-plasticity dilemma* is claimed to be a feature unique among neural networks to the ART models.)

Collectively these features make ARTMAP potentially suitable for on-line learning in non-stationary environments. This accords with our long-term goal of developing a decision-support tool which can adapt itself *in situ* to clinical conditions which can vary from hospital to hospital as well as over time as local clinical practices change. However, our goal here is to establish that ARTMAP can show adequate classification performance with real-world medical data, and to demonstrate the usefulness of its rule-extraction capabilities. Accordingly, off-line learning using a static data set is described in this paper.

---

[2] In practice, domains (such as this one) which perform many-to-one classification do not usually require generalization of the teaching inputs and a simplified ART_b module can be used which simply codes these patterns directly.

3 The Application Domain

The data set consisted of 413 patient records, each comprising ten binary-valued features recorded from human-observation of breast tissue samples, together with the actual outcome for each case (i.e. whether a lesion proved to be malignant or benign.) The distribution of categories within the data was fairly even - 53% of cases were malignant, 47% benign. The features themselves are all claimed to have predictive value for the diagnosis task (Trott, 1991; Koss,1992).

Comprehension of the precise definition of the features requires a level of specialist medical knowledge unlikely to be possessed by most readers. Therefore, the features will be referred to throughout this paper by the following abbreviations: DYS, ICL, 3D, NAKED, FOAMY, NUCLEOLI, PLEOMORPH, SIZE, NECROTIC and APOCRINE. For the benefit of those who are conversant with medical terminology however the full definitions of the features are provided in the Appendix.

As with almost all information gathered from a real-world medical domain, the data set possesses a degree of "noise". Specifically, some feature-states do not always have the same outcome in every case. Analysis of the data set revealed the existence of 12 such states, which collectively account for 188 cases. Assuming that the most frequent outcome should always be chosen when an ambiguous feature-state occurs will result in 17 of these cases being misclassified. This represents approximately 4% of the data-set, and thus optimal performance in the domain is a diagnostic accuracy of 96%.

However, it should further be noted that in medical domains overall diagnostic accuracy is not always the key criterion by which performance is judged. Two other indicators are also used - *sensitivity* and *specificity*. While accuracy is the proportion of all cases correctly diagnosed, sensitivity is the proportion of positive outcomes correctly diagnosed, and similarly specificity is the proportion of negative outcomes correctly diagnosed.

Obviously, high accuracy requires both high sensitivity and high specificity. However, there is also a trade-off between sensitivity and specificity. If all cases are diagnosed as having a positive outcome, 100% sensitivity but 0% specificity will be achieved (and vice versa if all cases are classed as negative outcomes). Although this is an extreme example, it is often the case that diagnoses will be deliberately made that are to some extent biased towards one or the other outcome.

On this particular data-set, assigning malignant cases as "positive" and benign cases as "negative", an expert human pathologist (of Consultant status with 10 years experience in the field) performed with accuracy, 91%, sensitivity, 83%, and specificity, 100%; a "neophyte" (Senior House-Officer with 18 months experience) with accuracy, 71%, sensitivity, 57% and specificity, 98%. These figures clearly demonstrate the point made above, in this instance diagnoses being slanted towards maximizing specificity at the expense of sensitivity. The reason for this is that the pathologist's prime concern is to avoid false positive predictions (i.e. diagnosing benign tumours as malignant), since these may result in unnecessary mastectomy (breast-removal) operations. The resultant increase in false negatives (diagnosing malignant tumours as benign) is tolerated because, if the clinical suspicion of malignancy remains, the surgeon will then take further samples to be sent to the pathologist for additional testing.

4 Results

One hundred records were randomly selected from the data to serve as test items in the evaluation of ARTMAP for the task. The remaining 313 records served as the teaching data. Ten ARTMAP networks were trained with "one-pass" learning, each on a different random ordering of the teaching data. (The order of presentation of data items is known sometimes to have quite large effects on the formation of category clusters in ARTMAP.) During training, the ART_a baseline vigilance parameter was set to 0.9 to ensure narrow category clustering, during testing this was relaxed to 0.6 to ensure that a category prediction (diagnosis) was made for all data items. (High vigilance during testing can lead to items failing to match sufficiently to any existing category clusters.) The subsequent performance of the 10 networks on the test set is shown in table 1 below. The mean performance of the 10 networks was accuracy, 93.9%, sensitivity, 96.0% and specificity, 91.7%.

Table 1: Performance of 10 ARTMAP Networks on 100 Item Test Set

Total ART_a Category Clusters Formed in Training	False Positive Diagnoses	False Negative Diagnoses	Accuracy (%)	Sensitivity (%)	Specificity (%)
60	5	2	93	96.2	89.6
61	4	4	92	92.3	91.7
59	3	1	96	98.1	93.8
58	3	2	95	96.2	93.8
58	5	2	93	96.2	89.6
60	5	1	94	98.1	89.6
61	5	2	93	96.2	89.6
60	3	2	95	96.2	93.8
68	2	4	94	92.3	95.8
65	5	1	94	98.1	89.6

The five most accurate individual networks were then tested collectively, using the so-called *voting strategy* (Carpenter et al., 1992). This works as follows: each individual network makes its prediction for a test item in the normal way, the number of predictions made for each category is then totalled and the one with the highest score (or the most "votes") is the final predicted category outcome. The voting strategy is claimed to offer improved ARTMAP performance. In addition it also provides an indication of the confidence of a particular prediction, since the larger the voting majority the more certain is the prediction.

In this particular domain the voting strategy yielded performance figures of accuracy, 95%, sensitivity, 96.1% and specificity, 93.8%. Although this may seem to be only a slight

improvement in comparison with the individual ARTMAP results, it should be noted that diagnostic accuracy with the voting strategy is almost at the optimum possible for the domain. Furthermore, when unanimous voting decisions were considered alone, performance becomes near-perfect on a large subset of the test cases. Five-nil category decisions accounted for 91% of the test set and showed accuracy, 99%, sensitivity, 100% and specificity, 98% on this subset of the data. Thus the voting strategy can provide a useful partitioning between data items with high and low certainty of outcome.

Symbolic rule extraction (Carpenter and Tan, 1993; Tan, 1994) was then performed upon all 10 of the previously trained ARTMAP networks.The act of rule extraction is a straightforward procedure in ARTMAP compared with that required for feedforward networks since there are no hidden units with implicit meaning. In essence, each category cluster in ART_a represents a symbolic rule whose antecedent is the category prototype weights and whose consequent is the associated ART_b category (denoted via the map field).

However, an ARTMAP network often becomes "over-specified" on the training set, generating many low-utility ART_a category clusters which represent rare but unimportant cases, and subsequently provide poor-quality rules. The problem is particularly acute when a high ART_a baseline vigilance level is used during training (as with this data set). To overcome this difficulty, rule extraction involves a "preprocessing" stage of *category pruning*[3]. This involves the deletion of these low utility nodes. Pruning is guided by the calculation of a *confidence factor* between nought and one for each category cluster, based upon a node's usage (proportion of training set exemplars it encodes) and accuracy (proportion of correct predictions it makes on the test set). All nodes with a confidence factor below a user-set threshold are then pruned. Full details of the process are given in Carpenter and Tan (1993) or Tan (1994).

Severe pruning was performed upon the 10 trained ARTMAP networks, using a threshold confidence level of 0.7. The number of category cluster nodes remaining for each individual network after pruning ranged from 3-9. Thus the networks were reduced to a small number of ART_a category nodes of strong predictive power from which rules could be extracted. Before doing so however, the test data was re-applied to each of the pruned networks to check that pruning had not adversely affected performance. Since pruning necessarily reduced ARTMAP's coverage of the feature space, the baseline ART_a vigilance was this time relaxed further to 0.5. Despite this, some pruned networks were still unable to generate category predictions on all test set items. Full details of the performance of the 10 pruned networks is shown in table 2 overleaf. The mean performance of the 10 networks after pruning was accuracy, 94.1%, sensitivity, 92.3% and specificity, 97.9%.

It can be seen that pruning has virtually no effect upon overall diagnostic accuracy but has led to increased specificity and reduced sensitivity. This matter will be discussed further in the next section. The five most accurate pruned networks (excluding those which did not generate predictions on all test set items) were then tested using the voting strategy. This gave results of accuracy, 95%, sensitivity, 92.3% and specificity, 97.9%, again confirming that the voting strategy allows the optimum accuracy for the domain to be closely approached.

---

[3] With continuously-valued category weights, rule extraction also involves a second preprocessing stage of *quantization*. However, since we are using binary data under so-called fast-learn conditions (Carpenter et al.,1992), category weights are all binary and this stage is omitted.

Table 2: Performance of 10 ARTMAP Networks on 100 Item Test Set after Category Pruning

ART_a Category Clusters after Pruning	False Positive Diagnoses	False Negative Diagnoses	No Diagnosis Possible	Accuracy (%)	Sensitivity (%)	Specificity (%)
5	1	5	0	94	90.4	97.9
6	0	6	0	94	88.5	100.0
7	1	4	0	95	92.3	97.9
3	1	4	1	94	90.4	97.9
9	0	6	0	94	88.5	100.0
4	0	7	0	93	86.5	100.0
4	1	4	1	94	90.4	97.9
4	1	4	1	94	90.4	97.9
6	1	4	0	95	92.3	97.9
8	1	5	0	94	90.4	97.9

Rule extraction from the 10 pruned nets yielded 14 distinct rules, 12 for malignant outcomes and 2 for benign. The full list of rules is shown overleaf, ranked by how many of the 10 pruned networks a rule occurred in.

Before discussing the specific rules discovered, it should be stressed that the symbolic rules shown here are of a somewhat different nature to those found in expert systems. Expert system rules are "hard" - an input must match to each and every feature in a rule's antecedent before the consequent will be asserted. In ARTMAP the rules are "soft", recall that they are derived from prototypical category clusters which are in competition with each other to match to the input data. Exact matching between inputs and categories is not necessary, merely a reasonably close fit. (The degree of inexactitude that is tolerated being determined by the value of the ART_a vigilance parameter.) It is this method of using the "closest" match which allows ARTMAP to maintain very high performance in the domain with an extremely small number of rules, far fewer than would be needed by an expert system requiring exact matches to its rules.

The other important difference is that the rules in ARTMAP have been self-discovered by learning from a set of examples. This is in contrast to expert systems which must be pre-configured with externally provided rules derived from human domain specialists. This knowledge-acquisition process can be difficult and time-consuming (Hayes-Roth, Waterman and Lenat, 1983).

Symbolic Rules Extracted from Pruned ARTMAP Networks

Rule 1 (10 Occurrences)
IF
    NO-SYMPTOMS
THEN
    BENIGN

Rule 2 (8 Occurrences)
IF
    3D=TRUE
    NUCLEOLI=TRUE
    PLEOMORPH=TRUE
    SIZE=TRUE
THEN
    MALIGNANT

Rule 3 (8 Occurrences)
IF
    3D=TRUE
    FOAMY=TRUE
    NUCLEOLI=TRUE
    PLEOMORPH=TRUE
    SIZE=TRUE
THEN
    MALIGNANT

Rule 4 (7 Occurrences)
IF
    FOAMY=TRUE
THEN
    BENIGN

Rule 5 (4 Occurrences)
IF
    ICL=TRUE
    3D=TRUE
    NUCLEOLI=TRUE
    PLEOMORPH=TRUE
    SIZE=TRUE
THEN
    MALIGNANT

Rule 6 (4 Occurrences)
IF
    DYS=TRUE
    NUCLEOLI=TRUE
    PLEOMORPH=TRUE
    SIZE=TRUE
THEN
    MALIGNANT

Rule 7 (3 Occurrences)
IF
    FOAMY=TRUE
    NUCLEOLI=TRUE
    PLEOMORPH=TRUE
    SIZE=TRUE
THEN
    MALIGNANT

Rule 8 (3 Occurrences)
IF
    NUCLEOLI=TRUE
    PLEOMORPH=TRUE
    SIZE=TRUE
THEN
    MALIGNANT

Rule 9 (2 Occurrences)
IF
    3D=TRUE
    FOAMY=TRUE
    NUCLEOLI=TRUE
    PLEOMORPH=TRUE
    SIZE=TRUE
    NECROTIC=TRUE
THEN
    MALIGNANT

Rule 10 (2 Occurrences)
IF
    3D=TRUE
    FOAMY=TRUE
    PLEOMORPH=TRUE
    SIZE=TRUE
    NECROTIC=TRUE
THEN
    MALIGNANT

Rule 11 (2 Occurrences)
IF
    DYS=TRUE
    ICL=TRUE
    NUCLEOLI=TRUE
    PLEOMORPH=TRUE
    SIZE=TRUE
THEN
    MALIGNANT

Rule 12 (1 Occurrence)
IF
    ICL=TRUE
    NUCLEOLI=TRUE
    PLEOMORPH=TRUE
    SIZE=TRUE
THEN
    MALIGNANT

Rule 13 (1 Occurrence)
IF
    FOAMY=TRUE
    NUCLEOLI=TRUE
    PLEOMORPH=TRUE
    SIZE=TRUE
    NECROTIC=TRUE
THEN
    MALIGNANT

Rule 14 (1 Occurrence)
IF
    ICL=TRUE
    3D=TRUE
    PLEOMORPH=TRUE
    SIZE=TRUE
THEN
    MALIGNANT

Turning to the rules per se, it can be seen that an absence of features, or the FOAMY feature present in isolation, leads to a benign diagnosis. PLEOMORPH and SIZE are found in all rules for malignant diagnoses, and NUCLEOLI is additionally present in all but two of these same rules (both of which have low frequency of occurrence). Thus these three features in combination seem to be the strongest indicators of malignancy. Other features are weaker indicators of malignancy, and indeed two input features, NAKED and APOCRINE, are conspicuous by their absence from any of the rules. We would conclude therefore that these two features are the least useful in forming a diagnosis, at least for this particular data set.

An expert human pathologist confirmed the relative importance of the features listed above in making his own diagnoses, with the exception that he places no value on the presence or absence of the FOAMY feature. It should be noted that this feature has a somewhat ambiguous status within the ARTMAP rules. In isolation, it is indicative of a benign diagnosis. However, when it occurs in combination with other features, a malignant diagnosis results.

It should also be noted that there is some disagreement between different domain experts as to the relative importance of the features in making diagnoses. Thus another pathologist states "I think the presence of bipolar naked nuclei and foamy macrophages can be taken as indicative of benignancy. This is not to say however, that when these features are combined with cells showing obvious features of malignancy, malignancy should not be diagnosed." (Dr. P.A. Trott, personal correspondence). This accords with the self-discovered ARTMAP rules for the FOAMY feature.

5 Discussion

Table 3 below summarizes the performance figures for ARTMAP in comparison with human pathologists in this domain. It can be seen that in terms of diagnostic accuracy ARTMAP always performs at least as well as the human expert and much better than the "neophyte". However, the weak spot in the unpruned ARTMAP networks' performance is the lower specificity in comparison to the human pathologists. As pointed out in section 3, it is vital that false positive cases (which lower specificity) are avoided in this domain

.

Table 3: Relative Performance of Human Pathologists and ARTMAP

	Accuracy (%)	Sensitivity (%)	Specificity (%)
Human Expert	91	83	100
Human Neophyte	71	57	98
Unpruned ARTMAP - Individual Mean	94	96	92
Unpruned ARTMAP - Voting Strategy	95	96	94
Pruned ARTMAP - Individual Mean	94	90	99
Pruned ARTMAP - Voting Strategy	95	92	98

The pruning procedure achieves this goal, by increasing specificity at the expense of sensitivity

without changing overall diagnostic accuracy. The reason for this is that the category clusters formed at ART_a predominantly indicate positive (malignant) cases. (On average, 70% of ART_a category nodes in the unpruned networks denoted malignant outcomes.) Pruning therefore mostly deletes nodes with malignant outcomes, and so coverage of these cases in the state space is reduced disproportionately more than for benign cases.

This effect of biasing the trade-off between sensitivity and specificity was achieved naturally in this domain as a side-effect of the rule-extraction process. However, we envisage achieving the same trade-off deliberately in other domains where category clusters are more evenly distributed across outcomes. This will be accomplished by deleting only those nodes of the particular outcome that we wish to have reduced within the domain.

The performance figures for ARTMAP, together with its rule extraction facilities indicate that ARTMAP has strong potential as a decision-support tool. Useful performance results have already been achieved with ARTMAP (and its variants) in other medical domains - the early diagnosis of heart attacks (Harrison, Lim and Kennedy, 1994) and coronary care patient prognosis (Downs et al., 1994) However, in all these cases the networks were generally employed off-line with stationary data. Thus future work needs to test if useful results are still possible under the more stringent on-line, non-stationary, learning conditions that will be found in a genuine hospital environment.

One immediate problem is that on-line learning cannot be used in conjunction with the ARTMAP voting strategy since inputs arrive in a fixed order when operating on-line, whereas the voting strategy requires a randomized ordering of training for each ARTMAP network. A compromise solution may be to utilise a mixture of static and dynamically trained networks. Thus, for example, four networks might be trained off-line on random orderings of an initial data set for a domain. Another network is trained on-line without any artificial termination of learning, and all five networks are then used in the voting strategy. Initially, the dynamically learning network will perform worse than the ones with static training which have received more domain examples. Over time however the dynamically learning network will become more closely attuned to the nuances of the operating environment and, unlike the static networks, should be able to adapt to changed operating conditions. Thus, one envisages gradually increasing the importance of the dynamic network in relation to the static ones by periodically increasing its number of votes, until perhaps the static networks are not consulted at all.

Another interesting possibility might be to encode "text-book" rules for the domain into an ARTMAP network prior to training. Thus the net does not need to start as a *tabula rasa*, but will possess a reasonable initial baseline performance provided by these rules. Experience of actual domain examples may then lead the network to refine or over-write some or all of the rules as it adapts to the distinct operating conditions of its environment.

A third area of future work concerns the verification of the rules discovered by ARTMAP. As noted previously in section 4, there is some disagreement between experts as to the relative importance of the diagnostic features. We intend to survey a number of such experts in an attempt to establish a baseline "average" importance for the features. The possibility also exists that the rules discovered by ARTMAP might serve to provide supporting evidence to resolve disagreements between experts, or even establish useful novel diagnostic criteria.

In conclusion of the current work, however, we consider our findings to be highly promising. It has been demonstrated that ARTMAP achieves near-optimum diagnostic accuracy on a real-world medical data set. Moreover, performance can be altered to increase specificity at the expense of sensitivity without affecting overall accuracy, thus mimicking the diagnostic behaviour of human pathologists. ARTMAP also provides symbolic rule extraction capabilities enabling clinicians to validate that the network has learned an acceptable mapping between input features and diagnoses. Such rules may be used further to provide explanatory facilities for ARTMAP's diagnoses when this neural network model is used as a decision-support tool in medical domains.

Acknowledgement

Thanks to Ah-Hwee Tan of the Department of Cognitive and Neural Systems, Boston University, for clarifying details of the rule extraction procedure.

References

Carpenter, G.A. and Grossberg, S. (1987) A Massively Parallel Architecture for a Self-Organizing Neural Pattern Recognition Machine, *Computer Vision, Graphics and Image Processing*, 37, pp.54-115.
Reprinted in Carpenter and Grossberg (1991), 316-382.

Carpenter, G.A. and Grossberg, S., eds (1991) *Pattern Recognition by Self-Organizing Neural Networks*.
Cambridge, MA: MIT Press.

Carpenter, G.A., Grossberg, S., Markuzon, N., Reynolds, J.H. and Rosen, D.B. (1992) Fuzzy ARTMAP: A Neural Network Architecture for Incremental Supervised Learning of Analog Multidimensional Maps, *IEEE Transactions on Neural Networks*, 3(5), 698-712.

Carpenter, G.A., Grossberg, S. and Reynolds, J.H. (1991) ARTMAP: Supervised Real-Time Learning and Classification of Nonstationary Data by a Self-Organizing Neural Network, *Neural Networks*, 4(5), 565-588.

Carpenter, G.A. and Tan, A.H. (1993) Rule Extraction, Fuzzy ARTMAP, and Medical Databases, *Proceedings of the World Congress on Neural Networks*, Volume I, 501-506.

Downs, J., Harrison, R.F., Kennedy, R.L. and Woods, K. (1994) The Use of Fuzzy ARTMAP to Identify Low Risk Patients Hospitalized with Acute Chest Pain, Research Report, Dept. Of Automatic Control and Systems Engineering, University of Sheffield, submitted to ICANNGA-95.

Grossberg, S. (1987) Competitive Learning: From Interactive Activation to Adaptive Resonance, *Cognitive Science*, 11(1), 23-63.

Harrison, R.F., Lim, C.P. and Kennedy, R.L. (1994) Autonomously Learning Neural Networks for Clinical Decision Support, in E.C. Ifeachor and K.G. Rosen, eds *Proceedings of the International Conference on Neural Networks and Expert Systems in Medicine and Healthcare* (NNESMED-94), 15-22.

Hayes-Roth, F., Waterman, D.A. and Lenat, D.B. (1983) *Building Expert Systems.* London: Addison-Wesley.

Kasuba, T. (1993) Simplified Fuzzy ARTMAP, *AI Expert*, 8(11), 18-25.

Koss, L.G. (1992) *Diagnostic Cytology and its Histopathologic Basis* (4th Edition), 1293-1315.
Philadelphia: J.B. Lippincott.

Tan, A.H. (1994) Rule Learning and Extraction with Self-Organizing Neural Networks, in M. Mozer, P. Smolensky, D. Touretzky, J. Elman and A. Weigend, eds *Proceedings of the 1993 Connectionist Models Summer School*, 192-199.
Hillsdale, NJ: Lawrence Erlbaum Associates.

Trott, P.A. (1991) Aspiration Cytodiagnosis of the Breast, *Diagnostic Oncology*, 1, 79-87.

Appendix - Medical Definition of Features

DYS: True if majority of epithelial cells are dyhesive, false if majority of epithelial cells are in cohesive groups.

ICL: True if intracytoplasmic lumina are present, false if absent.

3D: True if some clusters of epithelial cells are not flat (more than two nuclei thick) and this is not due to artefactual folding, false if all clusters of epithelial cells are flat.

NAKED: True if bipolar "naked" nuclei in background, false if absent.

FOAMY: True if "foamy" macrophages present in background, false if absent.

NUCLEOLI: True if more than three easily visible nucleoli in some epithelial cells, false if three or fewer easily visible nucleoli in epithelial cells.

PLEOMORPH: True if some epithelial cell nuclei with diameters twice that of other epithelial cell nuclei, false if no epithelial cell nuclei twice the diameter of other epithelial cell nuclei.

SIZE: True if some epithelial cells with nuclear diameters at least twice that of lymphocyte nuclei, false if all epithelial cell nuclei with nuclear diameters less than twice that of lymphocyte nuclei.

NECROTIC: True if necrotic epithelial cells present, false if absent.

APOCRINE: True if apocrine change present in all epithelial cells, false if not present in all epithelial cells.