Monograph:

Downs, J., Harrison, R.F. and Kennedy, R. Lee. (1994) A Prototype Neural Network Decision-Support Tool for the Early Diagnosis of Acute Myocardial Infarction. UNSPECIFIED. ACSE Research Report 552 . Department of Automatic Control and Systems Engineering

A Prototype Neural Network Decision-Support Tool for the Early Diagnosis of Acute Myocardial Infarction

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

R. Lee Kennedy
Department of Medicine
The University of Edinburgh

Research Report #552
22 November 1994

Abstract

An application of the ARTMAP neural network model to the early diagnosis of acute myocardial infarction is described. Performance results are given for 10 individual ARTMAP networks, and for combinations of the networks using "pooled" decision making (the so-called *voting strategy*). Category nodes are pruned from the trained networks in different ways so as to improve accuracy, sensitivity and specificity respectively. The differently pruned networks are employed in a novel "cascaded" variation of the voting strategy. This allows a partitioning of the test data into predictions with a high and a lower certainty of being correct, providing the diagnosing clinician with an indication of the reliability of an individual prediction. Additionally, symbolic rule extraction is performed upon the networks, allowing a domain expert to verify that the networks have learned autonomously a valid set of predictive rules for the domain.

Correspondence Address

R.F. Harrison,
Department of Automatic Control and Systems Engineering,
The University of Sheffield,
PO Box 600, Mappin Street,
Sheffield, S1 4DU,
United Kingdom.

Telephone: +44 (0)114 2825139
Facsimile: +44 (0)114 2780409
E-mail: r.f.harrison@sheffield.ac.uk

1 Introduction

The early identification of patients with acute ischaemic heart disease remains one of the greatest challenges in emergency medicine. The ECG only shows diagnostic changes in about half of acute myocardial infarction (AMI) patients at presentation. (Stark and Vacek, 1987; Adams, Trent and Rawles, 1993) None of the available biochemical tests becomes positive until at least three hours after symptoms begin, making such measurements of limited use for the early triage of patients with suspected AMI (Adams, Abendschein and Jaffe, 1993). The early diagnosis of AMI, therefore, relies on an analysis of clinical features along with ECG data. A variety of statistical and computer-based algorithms has been developed to assist with the analysis of these factors (for review see Kennedy, Harrison and Marshall, 1993). Although none of these has yet found widespread usage in clinical practice, this remains an important area of research not only because of its clear potential to improve triage practices for the commonest of all medical problems, but also because of the light it may shed on techniques for the development of decision aids for use in other areas of medicine.

This paper describes the application of the ARTMAP neural network model to this task. This powerful, but little-used, model has a number of advantages for medical domains, outlined in section 2 (see also Harrison, Lim and Kennedy, 1994). Section 3 describes the provenance of the patient data used in this study, as well as the training and testing procedures applied to the ARTMAP model with this data. Section 4 gives performance results for different ARTMAP configurations. Section 5 describes and evaluates symbolic rules for the diagnosis of AMI that were extracted from the ARTMAP networks. Section 6 concludes with a discussion of the strengths and weaknesses of the approach, and suggests areas for future work.

2 ARTMAP

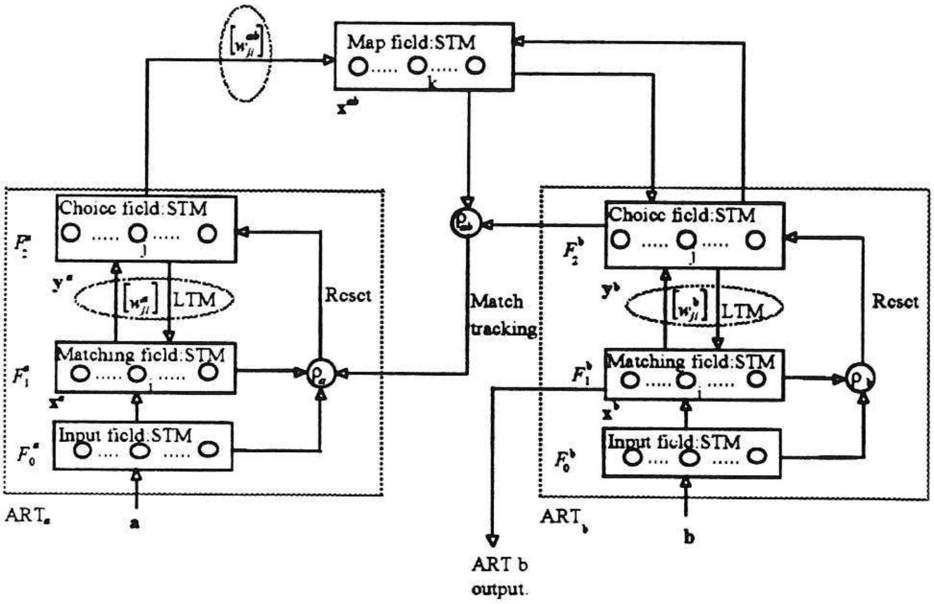
ARTMAP (Carpenter, Grossberg and Reynolds, 1991) is a self-organizing, supervised learning, neural network model for the classification of binary patterns[1]. It is one of a series of models based upon Adaptive Resonance Theory, or ART, (Carpenter and Grossberg, 1991) an outgrowth of competitive learning 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 (see figure 1). 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. ARTMAP does not directly associate inputs at ART_a and ART_b. Instead, such patterns are first self-organized into prototypical category clusters before being associated at 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 (see section 3) the implementation coincides with ARTMAP.

Hence generalized associations are formed[2].

Figure1: ARTMAP



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.

ARTMAP has a number of desirable properties for potential use as a decision-support tool in medical domains. First, it has few user-changeable parameters, which allows the model to be tuned to a particular problem without undue effort. The single most important parameter is that controlling the *vigilance* of the ART_a module. 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. Generally, higher vigilance provides better classification performance, although this must be balanced against the potential proliferation of category clusters, providing poor data compression and leading the net to become little more than an "look-up table" (Marriot and Harrison, In Press). Additionally, with small training sets and/or "wide" input vectors with many features, high vigilance can lead to incomplete coverage of the state space by the network.

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.

---

[2] In practice, domains (such as AMI diagnosis) 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.

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 hence fails to acknowledge rare events. ARTMAP is thus suitable for domains were the distribution of data items is highly skewed between different categories. (See Downs et al., In Press, for a particularly marked example of this phenomena.)

Fourth, successful learning in ARTMAP can occur with only one pass through the data set (termed single-epoch training). 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 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.)

For the purposes of this paper, three further features of ARTMAP are of particular note, the *voting strategy*, *symbolic rule extraction* and *category pruning*. These are described in detail next.

2.1 Voting Strategy

The formation of category clusters in ARTMAP is effected by the order of presentation of input data items (Carpenter et al., 1992). Thus the same data presented in a different order to separate ARTMAP networks can lead to the formation of quite different clusters within the two nets. This subsequently leads to different categorisations of test data, and thus different performance scores. This effect is particularly marked with small training sets and/or "wide" input vectors, where the input items may not be fully representative of the domain, and with single-epoch training.

This effect can be compensated for by the use of the ARTMAP voting strategy (Carpenter et al., 1992). This works as follows: a number of ARTMAP networks are trained on different orderings of the training data. During testing, 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 can provide improved ARTMAP performance in comparison with the individual networks. 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.

2.2 Symbolic Rule Extraction

Most neural networks suffer from the opaqueness of their learned associations (Towell and Shavlik, 1993). In medical domains, this "black box" nature may make clinicians reluctant to utilise a neural network application, 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". ARTMAP has recently been endowed with such capabilities (Carpenter and Tan, 1993; Tan, 1994). 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).

2.3 Category Pruning

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. 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* (CF) between nought and one for each category cluster, based equally upon a node's usage (proportion of training set exemplars it encodes) and accuracy (proportion of correct predictions it makes on a separate prediction 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).

The pruning process can provide significant reductions in the size of a network. In addition, it also has the very useful side-effect that a pruned network's performance is usually superior to the original, unpruned net on both the prediction set and on entirely novel test data.

In the original formulation of the pruning process, a uniform CF threshold is used to select nodes for deletion, irrespective of their category class. In this application, we have generalised the pruning process to allow separate CF thresholds for nodes belonging to different category classes. This allows us to vary the proportion of the state-space covered by different categories. This is useful for medical domains since it allows an ARTMAP network to be pruned so as to trade sensitivity for specificity and vice versa.

3 Patients and Methods

3.1 Patients and Clinical Data

The data used in this study were derived from consecutive patients attending the Accident and Emergency Department of the Royal Infirmary, Edinburgh, Scotland, with non-traumatic chest pain as the major symptom. The relevant clinical and ECG data (see below) were entered onto a purpose-designed proforma at, or soon after, the patient's presentation. The study included both patients who were admitted and those who were discharged. 970 patients were recruited during the study period (September to December 1993). The final diagnosis for these patients was assigned independently by a Consultant Physician, a Research Nurse and a Cardiology Registrar. This diagnosis made use of follow-up ECGs, cardiac enzyme studies and other investigations as well as clinical history obtained from review of the patient's notes. Patients discharged from Accident and Emergency were contacted directly regarding further symptoms and, where necessary, their General Practitioners were also contacted and the notes of any further hospital follow-up reviewed. The final diagnosis in the 970 patients was Q wave AMI in 146 cases, non-Q wave AMI in 45, unstable angina in 69, stable angina in 271 and other diagnoses in 439 cases. The patients were 583 men and 387 women with a mean age of 58.2 years (range 14 – 92). Unstable angina was defined as either more than two episodes of pain lasting more than 10 minutes in a 24-hour period or more than three episodes in a 48 hour period, or as angina which was associated with the development of new ECG changes of

[3] With continuously-valued category weights, rule extraction also involves a second preprocessing stage of *quantization* (see Carpenter and Tan, 1993). However, we prefer to use binary data under so-called fast-learn conditions (Carpenter et al.,1992) which yields purely binary category weights and subsequently provides rules of greater clarity. Quantization is therefore omitted in this application.

ischaemia (either at diagnosis or in the subsequent three days).

The input data items for the ARTMAP model were all derived from data available at the time of the patient's presentation. In all, 35 items were used, coded as 37 binary inputs. The full list of the inputs is given in Appendix 1, together with their feature names, used for symbolic rule extraction from the networks. For the purposes of this application, the final diagnoses were collapsed into two classes termed "AMI" (Q wave AMI and non-Q wave AMI) and "not-AMI" (all other diagnoses). AMI cases were assigned as positive diagnoses, not-AMI cases as negative diagnoses. Informed consent was obtained from all patients participating in the study which was approved by the local Medical Ethics Committee.

3.2 Method

The 970 patient records were divided into three data sets; 150 randomly selected records formed the *prediction set*, a further 150 randomly chosen records formed the *test set*, and the remaining 670 comprised the *training data*. The prediction set consisted of 28 cases of AMI and 122 not-AMI; the test set of 30 AMI and 120 not-AMI.

The training data was randomly ordered in ten different ways, and each ordering applied to a different ARTMAP network using single-epoch training. The ART_a base-line vigilance was set to a medium level (0.6) for training, all other parameters were set to their standard values (see Kasuba, 1993). The performance of the ten trained ARTMAP networks was then measured on both the prediction and test sets. During this testing phase the ART_a baseline vigilance was relaxed slightly (to 0.5) in order to ensure that all test items were matched to an existing category cluster (i.e. forced choice prediction).

The performance of the trained networks on the prediction set alone was then used to calculate accuracy scores for the category nodes in each network, as a prerequisite of the category pruning process described in section 2.3.

The "standard" form of category pruning (Carpenter and Tan, 1993) was performed on the original networks, such that all nodes with a CF below 0.5 were deleted from the networks in order to improve predictive accuracy. Performance of the resultant pruned networks was then measured on the prediction and test sets. Vigilance was further relaxed to 0.4 for testing these (and all other) pruned networks, again to ensure forced choice prediction.

The original networks were then pruned using different CF thresholds for the AMI and not-AMI nodes in order to produce pruned networks which maximized *sensitivity*. CF thresholds of 0.2 for AMI nodes and 0.95 for not-AMI nodes were employed, the criterion for setting the CF thresholds being to produce a mean sensitivity greater than 95% on the prediction set for the 10 pruned networks. Performance of the resultant nets was recorded for both the prediction and test sets. A similar procedure was then conducted to produce 10 networks which maximized *specificity*. CF thresholds of 0.7 AMI and 0.5 not-AMI were sufficient to yield a mean specificity greater than 95% on the prediction set.

The final pruning procedure was to produce 10 networks with approximately equal sensitivity and specificity (ESAS), the criterion for setting the CF thresholds being a performance on the prediction set where sensitivity and specificity were within 5% of each other. The performance of the pruned networks was again recorded on both the prediction and test sets.

Performance results using the voting strategy were then obtained for the unpruned networks and all classes of pruned network. Three voters were used with all networks types, except the ESAS class, where five voters were used. Voters for the unpruned, uniformly pruned, and ESAS network classes were selected on the basis of the networks with the highest accuracy on the prediction set. Selection criteria for the set of sensitive networks was maximum specificity, while maintaining a minimum sensitivity of 95% on the prediction set. The converse criteria were used for the set of specific networks.

Lastly, a novel "cascaded" variant of the voting strategy was employed utilizing 3 sensitive nets, 2 specific nets and 5 ESAS nets (see figure 2). This operated as follows: data items were first applied to the sensitive voting nets. If these yielded a unanimous (3-0) verdict that the category prediction was not-AMI, this was taken as the final category prediction. If not, the input was presented to the specific voting nets. If these yielded a unanimous (2-0) verdict of AMI, this was taken as the final prediction. Otherwise the final prediction of the category class of the test item was obtained by majority verdict from the ESAS nets.

4 Results

The mean performance on the prediction and test sets for all classes of ARTMAP networks is shown in table 1. (Performance figures for each individual net are given in Appendix 2.) As a baseline for comparisons, the Casualty Doctors showed an accuracy, sensitivity and specificity of 83.0%, 81.3% and 83.5% respectively over the entire data set.

Average accuracy for the unpruned networks can be seen to be only slightly below this baseline. However this is largely an artefact of the unequal prior probabilities of the category distributions—specificity accounts for the majority of accuracy, and although the networks' sensitivity is much poorer than the humans', this is compensated for by the superior specificity.

As expected, the uniformly pruned networks show an across-the-board increase in accuracy over the unpruned nets, with a 2.7% increase on the test set, and a 7.3% increase on the prediction set. (The greater increase in performance on the prediction set is explained by the fact that pruning utilized the accuracy scores for this data, and the networks are consequently optimized for this data.) However, the increase in accuracy is largely because of an overall improvement in specificity rather than sensitivity, which actually drops on the test set.

Figures for the sensitive nets show that almost all AMI cases can be diagnosed by the network, while approximately 36% of the not-AMI cases are detected. Conversely, with the sensitive nets, almost all not-AMI cases are trapped while approximately 40% of the AMI cases are detected.

The performance of the ESAS class networks is most directly comparable with that of the Casualty Doctors, since they are not unduly biased towards specificity or sensitivity. It can be seen that the mean individual accuracy of such networks is approximately 7% worse than the human diagnoses.

When the voting strategy is employed the accuracy of all network types except the specific nets is improved, as shown in table 2. Furthermore, unlike pruning, performance improvements owing to the voting strategy are almost always because of both increased sensitivity and specificity.

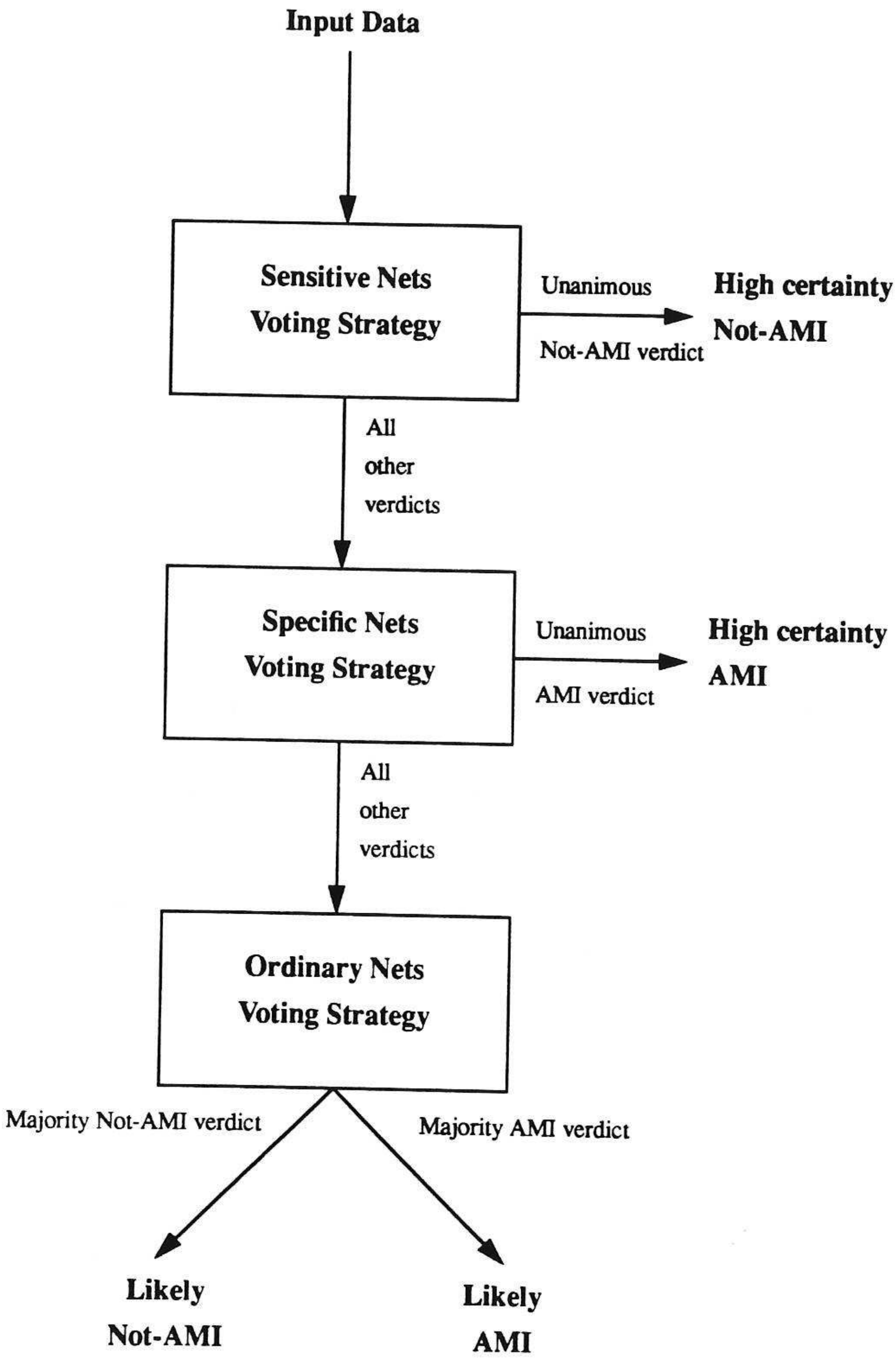


Figure 2: ARTMAP Configuration for AMI Diagnosis

Accuracy for the ESAS nets is now much closer to that of the Casualty Doctors and sensitivity is slightly better. Accuracy for the unpruned and uniformly pruned networks is now higher than the human diagnoses, particularly with the latter network class. However, this is again because of the networks' very high specificity while their sensitivity remains relatively poor.

Table 1: Mean Performance of 10 Differently Pruned Networks

	Prediction Set			Test Set		
Network Type	Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)
Unpruned	80.9	51.8	87.5	80.9	59.0	86.3
Uniform Pruning	88.2	60.7	94.5	83.6	52.0	91.5
Pruning for Sensitivity	50.0	96.4	39.3	47.3	94.3	35.5
Pruning for Specificity	86.9	41.8	97.2	84.7	39.7	96.0
Pruning for Equal Sensitivity and Specificity	76.6	76.1	76.7	75.6	80.0	74.5

Table 2: Voting Strategy Performance of Differently Pruned Networks

	Prediction Set			Test Set		
Network Type	Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)
Unpruned	86.0	64.3	91.0	83.3	56.7	90.0
Uniform Pruning	92.0	78.6	95.1	88.0	56.7	95.8
Pruning for Sensitivity	55.3	96.4	45.9	51.3	96.7	40.0
Pruning for Specificity	88.7	46.4	98.4	84.7	33.3	97.5
Pruning for Equal Sensitivity and Specificity	82.0	82.1	82.0	81.3	83.3	80.8

Use of the voting strategy with the sensitive networks on the test set results in increased coverage of the not-AMI cases while trapping even more AMI cases than previously. However, the converse is not true for the specific nets, where a gain in not-AMI coverage is offset by poorer coverage of the AMI cases in comparison to the individual network means.

The best overall network performance was achieved by the cascaded voting strategy, shown in table 3 below.

Table 3: Performance of the Cascaded Voting Strategy

	Prediction Set			Test Set		
	Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)
High Certainty Voters	100.0	100.0	100.0	96.3	88.9	97.8
Lower Certainty Voters	71.0	73.7	70.3	72.9	81.0	70.7
Overall Performance	82.0	82.1	82.0	82.7	86.7	81.7

The cascade's overall performance can be seen to be almost identical to that of the Casualty Doctors. Moreover, the cascade provides a partitioning of input items into those with a high and a lower certainty of a correct diagnosis. Unanimous not-AMI decisions by the highly specific networks (i.e. the first stage of the cascade) are almost certain to be correct, similarly unanimous AMI decisions by the highly sensitive networks (the second stage of the cascade) are also almost certain to be correct. The ESAS class voters then provide lower certainty predictions for the remaining data items at the bottom of the cascade. High-certainty predictions accounted for 38% of items in the prediction set and 36% of items in the test set.

Perfect performance by the high-certainty voters on the test set was prevented by the occurrence of one false positive case and one false negative case. At least one of these data items is highly atypical, and will be discussed further in section 6.

5 Symbolic Rule Extraction

As mentioned previously in section 2.2, the ability to extract symbolic rules from neural networks is an important enhancement to their use as decision-support tools in medical domains. Such symbolic rules provide two advantages which, taken collectively, should help to overcome reluctance to utilize a neural network decision-support tool.

First, a domain expert can examine the complete rule set in order to validate that the network has acquired an appropriate mapping of input features to category classes.

Second, the symbolic rules provide explanatory facilities for the network's predictions during on-line operation. In the case of ARTMAP this corresponds to displaying the equivalent rule for the ART_a cluster node that was activated to provide a category decision. (In the case of the voting strategy, a number of such rules, one per voting network, would be displayed.) The diagnosing clinician is then able to decide whether or not to concur with the network's prediction, based upon how valid they believe that rule to be.

In this domain, a single network still averaged 49 cluster nodes remaining after uniform CF pruning (see table 6 in Appendix 2). Space limitations therefore preclude the display of a typical complete rule set in this paper. Instead, we have opted to provide a list of all rules for diagnosing AMI from nodes with a CF greater than 0.8 from the 10 original networks. In order to pass such a high threshold a node must encode a large proportion of the training exemplars and possess high predictive accuracy. Hence, these nodes are the "pick of the crop" in the sense of being the most useful to their originating networks for the purpose of diagnosing AMI. In all 18 such nodes occurred, their equivalent rules are shown in table 4.

Examination of the rules as a whole allows the following picture of a typical AMI case to be constructed: The patient is likely to be a smoker, aged over 45 (and most likely over 65), exhibiting central chest pain which possibly radiates to the left arm. The pain itself is likely to be described as "tight" or "heavy". Other physical symptoms may include sweating and nausea. ECG readings are very likely to show ST segment or T wave changes suggestive of ischaemia, and perhaps also new ST segment elevation and/or new pathological Q waves.

This picture closely corresponds to a "text-book" example of AMI, although it has been discovered by ARTMAP through self-organisation of the input data without any pre-specified knowledge of the domain. Thus the ARTMAP decision-support tool encodes rules which provide valid classifications for the domain, while bypassing the difficult and time-consuming knowledge-acquisition process found with rule-based expert systems (Hayes-Roth, Waterman and Lenat, 1983).

ARTMAP's symbolic rules also differ from those of expert systems as regards the way they are matched to input features. 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 suffices. (The degree of inexactitude that is tolerated being determined by the value of the ART_a vigilance parameter.) This provides greater coverage of the state space for the domain using fewer rules.

A drawback of the approach is that the rules are "correlational" rather than causal, since ARTMAP possesses no underlying theory of the domain but simply associates conjunctions of input features with category classes. Of course, this problem is not specific to ARTMAP but occurs with neural networks generally. However, this matter is probably not of great importance since useful diagnostic performance can often be achieved from correlational features without recourse to any "deep" knowledge of the domain.

6 Discussion

We consider the prototype decision-support tool that has been described here to be potentially valuable in assisting the early diagnosis of AMI. Furthermore, the general architecture should be of utility in other medical domains.

The ARTMAP application employs two novel techniques. First, the generalization of the category pruning method to allow for different threshold confidence factors for nodes of differing category classes. Second, the employment of a cascaded voting strategy employing differently pruned networks.The use of different CF thresholds for category pruning allows networks to be created which trade either sensitivity for specificity or vice versa. This should be particularly useful in domains where the costs of misdiagnosis of one class are much greater than for another, since it allows biasing of network performance so as to avoid the high-cost

Table 4: Symbolic Rules for AMI Diagnosis Extracted from ARTMAP Networks

IF RETRO = TRUE THEN AMI	IF RETRO = TRUE SWEAT = TRUE STTWAVE = TRUE THEN AMI	F AGE=45-65 = TRUE RETRO = TRUE STELEV = TRUE THEN AMI
IF AGE>65 = TRUE RETRO = TRUE SWEAT = TRUE THEN AMI	IF SMOKES = TRUE RETRO = TRUE STTWAVE = TRUE THEN AMI	IF RETRO = TRUE NEWQ = TRUE STTWAVE = TRUE THEN AMI
IF AGE>65 = TRUE RETRO = TRUE STTWAVE = TRUE THEN AMI	IF AGE>65 = TRUE RETRO = TRUE ALLTIGHT = TRUE SWEAT = TRUE THEN AMI	IF AGE>65 = TRUE RETRO = TRUE LARM = TRUE STTWAVE = TRUE THEN AMI
IF RETRO = TRUE LARM = TRUE SWEAT = TRUE STTWAVE = TRUE THEN AMI	IF SMOKES = TRUE RETRO = TRUE ALLTIGHT = TRUE STTWAVE = TRUE THEN AMI	IF AGE=45-65 = TRUE RETRO = TRUE NEWQ = TRUE STTWAVE = TRUE THEN AMI
IF AGE>65 = TRUE RETRO = TRUE SWEAT = TRUE LIKEMI = TRUE THEN AMI	IF AGE=45-65 = TRUE SMOKES = TRUE RETRO = TRUE STTWAVE = TRUE THEN AMI	IF AGE=45-65 = TRUE SMOKES = TRUE SWEAT = TRUE NAUSEA = TRUE STTWAVE = TRUE THEN AMI
IF SMOKES = TRUE RETRO = TRUE LARM = TRUE NAUSEA = TRUE STELEV = TRUE THEN AMI	IF AGE>65 = TRUE RETRO = TRUE ALLTIGHT = TRUE SWEAT = TRUE NAUSEA = TRUE STTWAVE = TRUE THEN AMI	IF SMOKES = TRUE RETRO = TRUE ALLTIGHT = TRUE SWEAT = TRUE NAUSEA = TRUE STTWAVE = TRUE THEN AMI

misdiagnoses. (One example of such a domain is the diagnosis of breast cancer by a pathologist from fine needle aspirate samples, see Downs, Harrison and Cross, 1994). It also allows for the correction of any biases in classification arising with the initial training of ARTMAP. Such an approach may be compared with weighting "risk" in a Bayesian classification system.

The benefits of the cascaded voting strategy are twofold. First, overall classification performance is improved. Second, it allows those input data items to be identified which ARTMAP has a very high likelihood of diagnosing correctly.

We shall now consider these claims in the context of the system implemented for AMI diagnosis.

It is certainly the case that the variable CF thresholds allowed the construction of both highly sensitive and highly specific nets (see tables 7 and 8 in Appendix 2). Moreover, the construction of the ESAS nets (table 9 in Appendix 2) demonstrates the use of the technique to correct an initial bias in the network performance—in this case high specificity at the cost of poor sensitivity owing to disparate prior category probabilities.

The drawback of the technique is that it can be difficult to select the best CF thresholds that are needed to cause the desired changes in network performance. This problem was particularly acute for the construction of the ESAS class nets, where each individual network required a different CF threshold setting to achieve the desired performance. In the present implementation the CF thresholds were "hand-set" by the system's designer using a rather laborious trial-and-error process. A useful area for future work therefore would be to automate this process.

Considering the cascaded voting strategy, the overall performance results for this method were the best achieved by any of the ARTMAP configurations tested (see tables 1, 2 and 3). (However, performance of the ESAS voting nets was identical on the prediction set and only slightly weaker on the test set). This performance was very similar to that of the Casualty Doctors and we have some confidence that further improvements could be achieved with an enlarged data set. In particular, we believe the prediction and test sets were probably too small, particularly given the unequal distribution of category classes with relatively few AMI cases. The small number of AMI cases in the prediction set is the cause of most concern, since optimum benefit from category pruning is achieved only if the prediction set is truly representative of the overall domain. Otherwise, pruning will optimize a net's performance on the prediction set, but not will not generalize well to novel test data. It is intended to test this using more data that will shortly become available from the same site.

The cascaded voting strategy was also intended to partition input data so to identify those cases on which the ARTMAP system makes near-perfect predictions. In the AMI domain such cases accounted for over one-third of the test items although perfect performance was prevented by the occurrence of one false positive and one false negative prediction. Examination of the input features for these cases is revealing however.

The false positive case had the following features (see Appendix 1 for definitions): AGE=45-65, SMOKES, FAM_IHD, RETRO, JAW, S_O_BREATH, NAUSEA, STELEV, NEWQ, STTWAVE. This exhibits almost all of the "classic" features of AMI (see section 5), the latter three features being regarded as particularly strong AMI indicators. The false negative case had the following features: AGE<45, SMOKES, LCHEST, LARM, ALLSHARP, OLD_Q, OLD_ST. This displays none of the classic features of AMI, although the existence of OLD_Q should mean a human clinician probably would not entirely discount the possibility of AMI. We conclude therefore that these cases are highly idiosyncratic, particularly the false positive, and would cause most human experts to make the wrong diagnosis. Thus the general ability of the cascaded voting strategy to identify cases with high-certainty of a correct diagnosis is not greatly undermined.

In summary therefore, we have described the application of the ARTMAP neural network model to the diagnosis of acute myocardial infarction, and introduced two new enhancements to this model - the use of variable threshold category pruning, and a cascaded voting strategy. The strengths and weaknesses of these new techniques, and the ARTMAP model in general, have been discussed. We conclude that the model is of potential value in both the diagnosis of AMI and in medical domains generally.

Acknowledgement

Thanks to Shaun Marriott for providing the diagram of ARTMAP.

This research was supported by the Science and Engineering Research Council (SERC) of the UK, grant number GR/J/43233.

References

Adams, J.E., Abendschein, D.R. and Jaffe, A.S. (1993) Biochemical Markers of Myocardial Injury. Is MB Creatine Kinase the Choice for the 1990s?, *Circulation*, 88, 750–63.

Adams, J., Trent, R. and Rawles, J. (1993) Earliest Electrocardiographic Evidence of Myocardial Infarction: Implications for Thrombolytic Treatment, *British Medical Journal*, 307, 409–13.

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. and Cross, S.S. (1994) A Neural Network Decision Support Tool for the Diagnosis of Breast Cancer, Research Report 548, Department of Automatic Control and Systems Engineering, University of Sheffield.

Downs, J., Harrison, R.F., Kennedy, R.L. and Woods, K. (In Press) The Use of Fuzzy ARTMAP to Identify Low Risk Patients Hospitalized with Acute Chest Pain, to appear in *Proceedings of the 1995 International Conference on Artificial Neural Networks and Genetic Algorithms* (ICANNGA–95), Alès, France.

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), Portsmouth, UK, 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.

Kennedy, R.L., Harrison, R.F. and Marshall, S.J. (1993) Do We Need Computer-Based Decision Support for the Diagnosis of Acute Chest Pain?, *Journal of the Royal Society of Medicine*, 86, 31–4.

Marriott, S. and Harrison, R.F. (In Press) A Modified Fuzzy ARTMAP Architecture for the Approximation of Noisy Mappings, to appear in *Neural Networks.*

Stark, M.E. and Vacek, J.L. (1987) The Initial Electrocardiogram During Admission for Myocardial Infarction. Use as a Predictor of Clinical Course and Facility Utilization, *Archives of Internal Medicine*, 147, 843–6.

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.

Towell, G. and Shavlik, J.W. (1993) Extracting Refined Rules from Knowledge-Based Neural Networks, *Machine Learning*, 13(1), 71–101.

Appendix 1: Binary Inputs to ARTMAP

AGE<45: Age less than 45 years old.

AGE=45-65: Age between 45 and 65 years old.

AGE>65: Age greater than 65 years old.

SMOKES: Smokes.

EX_SMOKER: Ex-smoker.

FAM_IHD: Family history of ischaemic heart disease.

DIABETES: Diabetes mellitus.

HYPERTENSE: Hypertension.

HYPERLIPID: Hyperlipidaemia.

RETRO: Central chest pain.

LCHEST: Pain in left side of chest.

RCHEST: Pain in right side of chest.

BACK: Pain radiates to back.

LARM: Pain radiates to left arm.

JAW: Pain radiates to neck or jaw.

RARM: Pain radiates to right arm.

BREATHING: Pain is worse on inspiration.

POSTURE: Pain related to posture.

TENDER_CW: Chest wall tenderness

ALLTIGHT: Pain described as tight, heavy, gripping or crushing

ALLSHARP: Pain described as sharp or stabbing.

SWEAT: Sweating

S_O_BREATH: Short of breath.

NAUSEA: Nausea

VOMIT: Vomiting

SYNCOPE: Syncope

EPIS: Episodic pain.

LIKEMI: Worse than usual angina/similar to previous AMI.

LVF: Fine crackles suggestive of pulmonary oedema.

ADDED_HS: Added heart sounds.

HYPOPERF: Signs of hypoperfusion

STELEV: New ST segment elevation.

NEWQ: New pathological Q waves.

STTWAVE: ST segment or T wave changes suggestive of ischaemia.

BBB: Bundle branch block.

OLD_Q: Old ECG features of myocardial infarction.

OLD_ST: ECG signs of ischaemia known to be old.

Appendix 2: Performance Results for Different ARTMAP Configurations

Table 5: Performance of 10 Unpruned ARTMAP Networks

		Prediction Set			Test Set		
Network Number	Total ART_a Category Clusters	Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)
1	122	81.3	60.7	86.1	80.0	63.3	84.2
2	119	78.0	42.9	86.1	80.0	60.0	85.0
3	122	79.3	46.4	86.9	84.0	70.0	87.5
4	129	79.3	46.4	86.9	76.7	50.0	83.3
5	131	80.0	46.4	87.7	81.3	63.3	85.8
6	129	82.7	53.6	89.3	76.7	53.3	82.5
7	120	78.7	50.0	85.2	78.7	56.7	84.2
8	118	82.7	53.6	89.3	84.7	56.7	91.7
9	118	86.0	75.0	88.5	83.3	66.7	87.5
10	124	80.7	42.9	89.3	83.3	50.0	91.7
Mean	123	80.9	51.8	87.5	80.9	59.0	86.3

Table 6: Performance of 10 ARTMAP Networks with Uniform Pruning (Min. CF = 0.5)

		Prediction Set			Test Set		
Network Number	Total ART_a Category Clusters	Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)
1	55	88.7	75.0	91.8	80.0	53.3	86.7
2	44	88.0	50.0	96.7	82.7	53.3	90.0
3	47	88.7	53.6	96.7	84.0	40.0	95.0
4	54	88.7	57.1	95.9	82.0	46.7	90.8
5	45	87.3	53.6	95.1	84.7	56.7	91.7
6	53	88.0	67.9	92.6	83.3	56.7	90.0
7	40	88.7	60.7	95.1	79.3	43.3	88.3
8	51	86.7	57.1	93.4	86.7	50.0	95.8
9	58	89.3	78.6	91.8	84.7	60.0	90.8
10	42	88.0	53.6	95.9	88.7	60.0	95.8
Mean	49	88.2	60.7	94.5	83.6	52.0	91.5

Table 7: Performance of 10 ARTMAP Networks Pruned to Maximise Sensitivity (Min. CFs = 0.2 MI, 0.95 not-MI)

		Prediction Set			Test Set		
Network Number	Total ART_a Category Clusters	Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)
1	24	53.3	92.9	44.3	48.0	93.3	36.7
2	25	60.7	92.9	53.3	55.3	93.3	45.8
3	21	56.0	96.4	46.7	52.0	93.3	41.7
4	15	52.0	96.4	41.8	48.7	93.3	37.5
5	21	46.0	100.0	33.6	42.7	96.7	29.2
6	29	44.0	100.0	31.1	40.7	93.3	27.5
7	22	52.7	96.4	42.6	53.3	96.7	42.5
8	21	43.3	100.0	30.3	40.7	93.3	27.5
9	13	39.3	100.0	25.4	40.7	96.7	26.7
10	23	52.7	89.3	44.3	50.7	93.3	40.0
Mean	21	50.0	96.4	39.3	47.3	94.3	35.5

Table 8: Performance of 10 ARTMAP Networks Pruned to Maximise Specificity (Min. CFs = 0.7 MI, 0.5 not-MI)

		Prediction Set			Test Set		
Network Number	Total ART_a Category Clusters	Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)
1	47	88.0	53.6	95.9	82.7	43.3	92.5
2	38	84.7	25.0	98.4	84.0	33.3	96.7
3	43	88.7	46.4	98.4	86.0	40.0	97.5
4	49	86.0	39.3	96.7	82.0	40.0	92.5
5	41	87.3	32.1	100.0	85.3	36.7	97.5
6	51	89.3	67.9	94.3	85.3	53.3	93.3
7	37	87.3	42.9	97.5	84.0	33.3	96.7
8	47	87.3	46.4	96.7	86.7	40.0	98.3
9	49	83.3	17.9	98.4	83.3	20.0	99.2
10	40	86.7	46.4	95.9	88.0	56.7	95.8
Mean	44	86.9	41 8	97.2	84.7	39.7	96.0

Table 9: Performance of 10 ARTMAP Networks Pruned for Approximately Equal Sensitivity and Specificity on Prediction Set

		Prediction Set			Test Set		
Network Number	Total ART_a Category Clusters	Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)
1	25	80.7	78.6	81.1	77.3	76.7	77.5
2	24	74.7	78.6	73.8	74.7	83.3	72.5
3	15	73.3	71.4	73.8	79.3	90.0	76.7
4	21	77.3	75.0	77.9	79.3	76.7	80.0
5	18	72.0	71.4	72.1	70.0	76.7	68.3
6	19	73.3	71.4	73.8	73.3	76.7	72.5
7	14	82.7	82.1	82.8	75.3	83.3	73.3
8	17	74.7	71.4	75.4	71.3	73.3	70.8
9	26	84.0	85.7	83.6	78.7	83.3	77.5
10	17	73.3	75.0	73.0	76.7	80.0	75.8
Mean	20	76.6	76.1	76.7	75.6	80.0	74.5