
Monograph:
Harrison, R.F. and Kennedy, R. Lee (1990) The Multilayer Perceptron as an Aid to the Early Diagnosis of Myocardial Infarction. Research Report. Acse Report 395 . Dept of Automatic Control and System Engineering. University of Sheffield

---

The Multi-layer Perceptron as an Aid to the Early Diagnosis of Myocardial Infarction

Robert F Harrison and Stephen J Marshall*
Department of Control Engineering
and
R Lee Kennedy
Clinical Sciences Centre
University of Sheffield
Sheffield S1 3JD
England

Tel: (0742) 768555
Fax: (0742) 731729

4 June 1990

Research Report No 395

Abstract

The establishment of a decision aid for the early diagnosis of myocardial infarction is described. The system uses a multi-layer Perceptron structure and is trained in the usual way. It is shown that the performance of the network can exceed that of the admitting clinicians, a panel of senior physicians in a large teaching hospital and a protocol derived using conventional statistical methods, over a wide range of performance measures. In particular, the network demonstrates the highly specific behaviour necessary when making the decision whether or not to administer thrombolytic therapy—a potentially life-saving decision which must be taken in the very early stages, long before confirmatory laboratory test results are available. The network is compact and has been implemented on a portable computer. In operation it responds very quickly, giving its diagnosis and recommendations (taking account of clinical opinion) in a fraction of the time taken to input the patient's symptoms.

Keywords: clinical decision support, myocardial infarction, chest pain, neural networks, connectionism, computer aided diagnosis

Category: Applications

*Supported by the Science and Engineering Research Council, UK

1 Introduction

The diagnosis and management of suspected acute myocardial infarction (AMI), or heart attack, is a complex decision making process. It is crucial that diagnosis be accurate and speedy in order that appropriate action be taken at the earliest possible moment. In particular, the use of thrombolytic agents to dissolve clots in the coronary blood vessels can significantly decrease the mortality and complications related to necrosis of cardiac muscle, but only if administered within a short time of the onset of AMI (typically four to six hours and certainly fewer than 24 hours). If used appropriately, thrombolytic therapy can reduce mortality by up to 50% in those patients who safely reach hospital [1]. It has been calculated that appropriate use of thrombolytic agents could prevent at least 50 deaths per 1000 patients treated. However, because of the dangers associated with inappropriate administration, a working party of the British Heart Foundation has recently recommended that the use of thrombolytic agents be confined to hospital and to those patients with *definite* evidence of infarction [2].

It is well known that many—perhaps 50% or more—patients admitted to coronary care units (CCU's) with suspected AMI are diagnosed incorrectly at presentation [3]—[5]. Apart from the electrocardiograph (ECG), which often does not show diagnostic changes, there is no convenient and reliable test which can be used at presentation to help the clinician formulate a diagnosis.

It is not surprising therefore that there is much interest in the development of computer-based algorithms to aid in the clinical diagnosis of AMI. The approaches most frequently used are either data-driven or rule-based. Data-driven systems are derived from the statistical analysis of large patient cohorts [6]—[8] while expert systems [9] and [10] attempt to model the diagnostic process by explicitly programming the complex and often obscure set of rules involved. In practice both types of system have proved difficult to establish. Neither type is adaptive and none of the above systems has found wide usage in the diagnosis of AMI.

Of course, a neurocomputational approach to this problem is also data driven in that a system, such as the multi-layer Perceptron (MLP) used here, learns from exposure to its data environment (patient records). However, as we shall see, a remarkably high level of performance can be achieved using relatively small data sets. In addition, no *a priori* assumptions need be made about the probabilistic structure of the data set. It is interesting to note that in contrast to the work presented here, a similar study [11] also using an MLP, achieves an unacceptably low level of accuracy. We are unable to ascertain the reason for this.

We believe that neurocomputing, although in its infancy, has an important rôle in medical informatics, even in the short term. Some theoretical aspects of its use in medical diagnosis have been addressed, *eg* [12] and [13] but, to date, few formal clinically-based studies of its use have been undertaken.

In this study we establish an MLP for the early diagnosis of AMI. The system is compared

directly with the protocol of Goldman and colleagues [7] [14] which uses nine clinical and electrocardiographic features and was derived using conventional statistics from 5000 patient records, the admitting clinicians' unaided diagnoses and the opinions of a panel of senior physicians.

1.1 Diagnosis and management of AMI

It is often difficult in the first stages of AMI to be certain about the diagnosis because symptoms are not always typical and may arise from a variety of causes including pulmonary, gastro-intestinal and musculo-skeletal disorders. Even in patients with ischaemic heart disease it is often not possible to be certain initially whether the patient is suffering from AMI or from an attack of reversible ischaemia (angina). The diagnosis of AMI is made when two out of three standard criteria are satisfied: (1) a compatible clinical history, (2) changes in serial cardiac enzymes and (3) evolving ECG changes. Items (2) and (3) rely, unfortunately, on data which must be collected serially over a period of 24—48 hours. Therefore, the decision of whether or not to admit a patient either to hospital, for observation, or to a CCU must be made before test results are available. Clearly, if they have to wait for these serial test results, those patients with AMI will be diagnosed too late for the administration of thrombolytic therapy.

Moreover, the use of thrombolytic agents is not without risk, particularly in conditions which may be confused with AMI such as peptic ulcer disease and dissecting aortic aneurysm, and where an increased risk of haemorrhage is likely to be especially dangerous. They increase the likelihood of haemorrhage and the most widely used agent—streptokinase—can be associated with severe allergic reactions especially if the patient has been previously exposed to it. Inappropriate use during, say, an angina attack could therefore preclude the patient from receiving therapy if he or she subsequently presents with AMI.

In addition to the risk that inappropriate administration of thrombolytic therapy poses to patient health there are also major resource implications due to the lack of an early diagnosis. The efficient use of hospital beds or vastly more expensive CCU beds and their attendant equipment, consumables and staff is a major issue in modern health care. Furthermore, the cost of thrombolytic agents is high—those developed using recombinant DNA technology and having the lowest associated risk of allergy cost more than £500 per dose.

The need for a rapid and accurate diagnosis of AMI is clear. A demonstration that such a goal is attainable using neural networks is presented below.

2 Methods

2.1 Data collection and coding

The study included 300 consecutive emergency referrals, with a complaint of chest pain, to the Medical Department of the Northern General Hospital, Sheffield, England. Information from all patients was recorded on a standard proforma which consisted of 78 items of demographic, clinical and electrocardiographic data. In addition, the admitting clinician was asked to estimate the likelihood of the patient having suffered AMI. This was recorded as a percentage. Each proforma was completed before the serial measurements of cardiac enzyme and ECG were available. Cardiac enzymes (creatine, phosphokinase and lactate dehydrogenase) were measured each day for three days following admission and 12-lead ECG recordings were taken over the same period and during any episode of pain. When the serial test results were available myocardial infarction was diagnosed according to the two out of three criterion of 1.1 thus supplying the "teacher" for the network.

We took 38 features from each patient record and coded them as a 53 dimensional, bipolar vector (each element taking the value -1, 0 or 1 where 0 indicates missing data) and the target vector (diagnosis) was coded as 1 for AMI, 0 otherwise. Continuous valued variables such as age and duration of pain were coded as bipolar vectors eliminating logically redundant elements as described by Widrow *et al* [15].

2.2 Network architecture and training

The network architecture used here is the feed-forward structure known as the multi-layer Perceptron or MLP. One intermediate layer of "neurons" is used throughout, and this is fully interconnected to the input and output layers. Neither intra-layer connections nor direct coupling of input and output layers is allowed. The effect of the size of the hidden layer is investigated and guidance is obtained using the algorithm proposed in [16]

The network was trained in the usual way, using gradient descent to adjust the weights to minimize the ensemble mean-square prediction error, over the training set, see *eg* [17]. Both a "momentum" term and a weight decay term (which derives from including the sum of the squares of the weights in the cost function) are allowed in the weight update equation. The effects of learning rate, inertia and weight-decay parameters on speed of learning and accuracy of representation were studied.

2.2.1 Implementation

The network was programmed in "C" and implemented on a Neurocomputing AstraCard neural network co-processor residing in a 25 MHz 80386 AT compatible machine with 80387 maths co-processor. The speed of the host machine is irrelevant as far as the Astracard is

concerned but we include its specification for comparison. The peak performance for the AstraCard is claimed to be 33 Mflops. Although we have not been able to check this directly our experiments have shown that it runs approximately fifteen times more quickly than the 80386/80387 host machine alone. This means that, for instance, comprehensive experiments on the effects of different numbers of hidden units, using jack-knifing to improve statistical accuracy, can be run over the course of an afternoon rather than the course of a week.

We have also implemented a demonstrator system using only the feedforward network (no training) with a menu driven data entry system, on a portable machine. This system is intended for use at the bedside, for evaluation purposes. After data entry the response time of the system is less than one second.

2.3 Performance indicators

The predictive power of the trained MLP was assessed by calculating a number of performance indicators. These indicators use the numbers of true-positive (TP), true-negative (TN), false-positive (FP) and false-negative (FN) diagnoses made by the network for the test set of (previously unseen) data. They are defined as follows.

- Diagnostic accuracy: ratio of number of correct diagnoses to total number of cases $(= \frac{TP + TN}{TP + TN + FP + FN})$
- Sensitivity: ratio of number of correct *positive* diagnoses to the total number of patients *with* the disease $(= \frac{TP}{TP + FN})$
- Specificity: ratio of number of correct *negative* diagnoses to the total number of patients *without* the disease $(= \frac{TN}{TN + FP})$

Sensitivity therefore compares the number of people who are diagnosed as being ill with those who actually are ill, and specificity, the converse. It is easy to construct a highly sensitive system since, if all patients were diagnosed as having MI, sensitivity would be 100% but specificity would be 0%. Similarly, if none of the patients was diagnosed as having MI then specificity would be 100% and sensitivity 0%. A good predictive system is therefore one which is both highly specific *and* highly sensitive, that is, it would indicate *those and only those* patients who require treatment. Clearly, high sensitivity at the expense of specificity or *vice versa* is not tolerable in the present situation.

2.4 Incorporation of clinical opinion

The objective of this work is not to develop an alternative to the clinician, but rather an aid which can make optimum use of all available data at presentation *including* the clinician's opinion. For instance, we might make the combined network and clinician system more

specific by insisting that both diagnoses should be positive if AMI is to be diagnosed (the logical AND operation). Such a protocol might be valid when trying to decide whether or not to administer a thrombolytic agent. A less critical decision, that of whether or not to admit the patient to CCU, say, might benefit from the combination of network and clinician using the logical OR operation. This says if either diagnosis is positive then AMI is present and increases sensitivity.

A third option arises which admits a "grey" area of clinical opinion, representing probabilities of AMI lying between, say, 10% and 90%. Table 1 indicates a possible scoring system for such a fuzzy logical operation. Using this weighted combination of clinical opinion and output from the neural network, MI is diagnosed if a combined score of 3 or more is obtained. Notice how responsibility ultimately devolves to the clinician.

Clinician		Network	
Prob MI (%)	Score	Prob MI (%)	Score
0—10	0	0—10	0
11—50	1	11—50	1
51—90	2	51—90	2
91—100	3	91—100	2

Table 1: Scoring system for fuzzy logical operation

It is not suggested that this protocol is in any way optimum. It was chosen arbitrarily to demonstrate how a workable system might be achieved. Clearly, a good deal of effort is required to define the way in which machine and clinician should best co-operate.

Whilst we realize that the output from the MLP, using our particular choice of weight update rule, can only be regarded as a true probability under some restrictive assumptions, we choose to use the term probability to indicate the network's degree of certainty. Clearly, optimization of the likelihood function or the cross-entropy function would make the probabilistic interpretation of the network's output more rigourous.

3 Computational experiments

A number of experiments were conducted to ascertain the effect of a number of adjustable parameters on the diagnostic capability of the MLP.

3.1 The effect of hidden units

The network was set-up in a 54-N-1 configuration* and was allowed to train for a fixed number of epochs using the first 90 patterns of the training set for N in the range 0 to 30. The resulting systems were tested using the remaining 210 patterns and the performance indicators calculated. It was found that little was to be gained by including more than three hidden units and that the case of 0 hidden units — corresponding to logistic regression — caused a degradation in all of the performance indicators of approximately three percent. This indicates that the 54-N-1 configuration with $N > 3$ is able correctly to diagnose about three patients in every hundred more than a straightforward logistic regression. The advantage of using hidden units becomes more apparent when the network's output is combined with clinical judgement. The heuristic algorithm proposed in [16] suggests that the current problem requires 18 hidden units and henceforth we have used this number.

3.2 The effect of parameters on the learning algorithm

3.2.1 Learning rate and inertia

The general effects of learning rate and inertia are well documented and have been borne out here. We found that a learning rate of 0.1 gave reasonably fast convergence without undue problems of oscillation in the mean-square prediction error. There was no apparent advantage in using non-zero inertia other than to speed-up convergence, indicating the existence of an acceptable minimum having desirable "bowl-like" properties. In order to keep the number of parameters to a minimum we have henceforth set inertia to zero.

3.2.2 Weight decay

The reason for introducing a weight decay term into the weight update equation is to limit the size of the weights, $w_{i,j}$, (in a mean-square sense). A term $\rho \sum_{i,j} w_{i,j}^2$ with $\rho > 0$, is added to the standard cost function resulting in the addition of the term $-\rho w_{i,j}$ to the right-hand side of the weight change equation. For $\rho < 1$ this results in a decrease in the absolute value of the weights at each step. In the long run, in addition to minimizing the mean-square value of the weights, weight decay also serves to highlight the relative importance of individual units. For example, those units whose output is fed forward through a set of weights having a "small" ℓ_1-norm (sum of absolute values) may be deemed to be relatively unimportant in predicting the final outcome. The converse is also true. Since our primary objective is to minimize the ensemble mean-square error, E, it is important not to let the minimization of the weights dominate the optimization and so ρ must be very much less than 1. Weight decay can also be helpful if the network becomes "paralysed" during learning due to very large net inputs to units causing very small weight changes.

*the notation A-B-C indicates A units in the input layer, B in the hidden layer and C in the output layer.

Figure 1(a) indicates the rôle played by each unit in the diagnosis by showing the ℓ_1-norm of the weights from each input unit to those in the hidden layer for $\rho = 0.0$ and $\rho = 0.04$. Figure 1(b) shows the ℓ_1-norm of the weights from the hidden units to the output unit. The network was allowed to train for 2000 epochs. By removing units 3, 4 and 18 from the *hidden* layer the network's performance remained unimpaired but removal of a number of the *input* units with small weight norms resulted in a significant decrease in performance. However, it is clear that units such as 33 and 51 play little part in the diagnostic procedure whilst units 13, 20, 25, 29, 42, and 48—50 have a much larger effect. Indeed, these units correspond to clinical features which we would expect to have just such an effect. For instance, units 48—50 indicate diagnostic ECG changes and have a strong excitatory effect whereas unit 25 corresponds to the patient's description of the pain as being "sharp"This symptom is not usually indicative of MI and indeed its weights have a strong inhibitory effect. We are not able to draw any firm conclusions from our investigations of the values of the weights other than to say that they support our intuition.

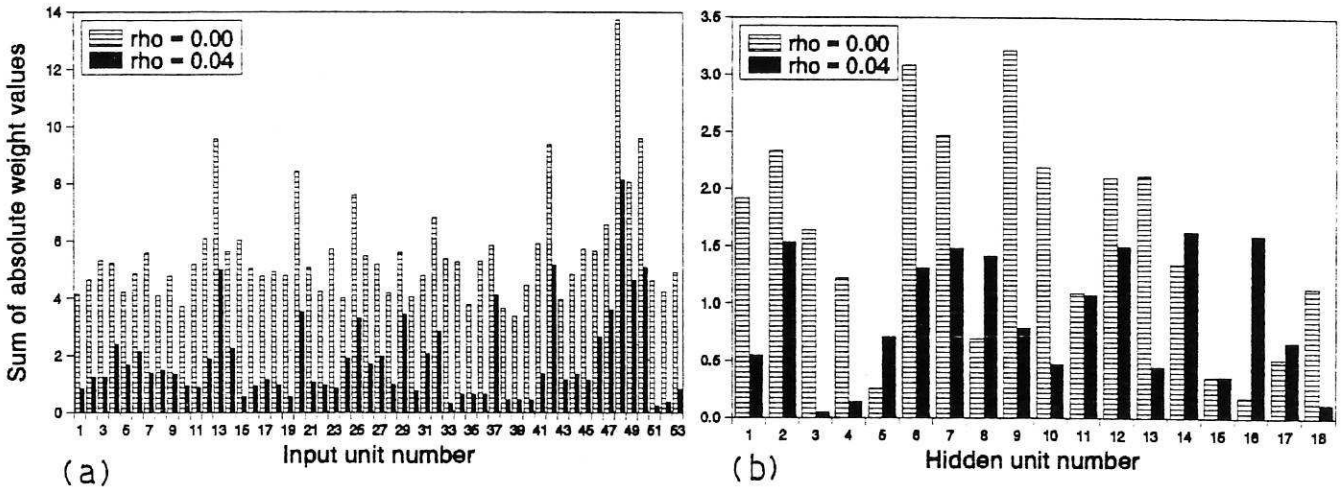

Figure 1: Reduction of weights due to weight decay. (a) Input layer to hidden layer. (b) Hidden layer to output layer.

Interestingly, the introduction of weight decay with $\rho = 0.04$ had no effect on the predictive performance of the network.

3.3 The effect of training time

The question of how long should training be allowed to continue is of great practical importance. The usual measure of how well a network is performing during training is the ensemble mean-square prediction error, E. We have found that this does indeed give a reliable indication of how well the network might perform. Figure 2(a) shows how our performance indicators vary as a function of the number of training epochs. Also shown is the variation in E/E_{peak}. Even when E/E_{peak} is as high as 10% the performance of the network is very good, with all indicators better than 80%. However, our objective is to produce a diagnostic aid, to be used in concert with clinical opinion. Figure 2(b) demonstrates that the same measures, incorporating clinical judgment using the scoring system of Table 1, do

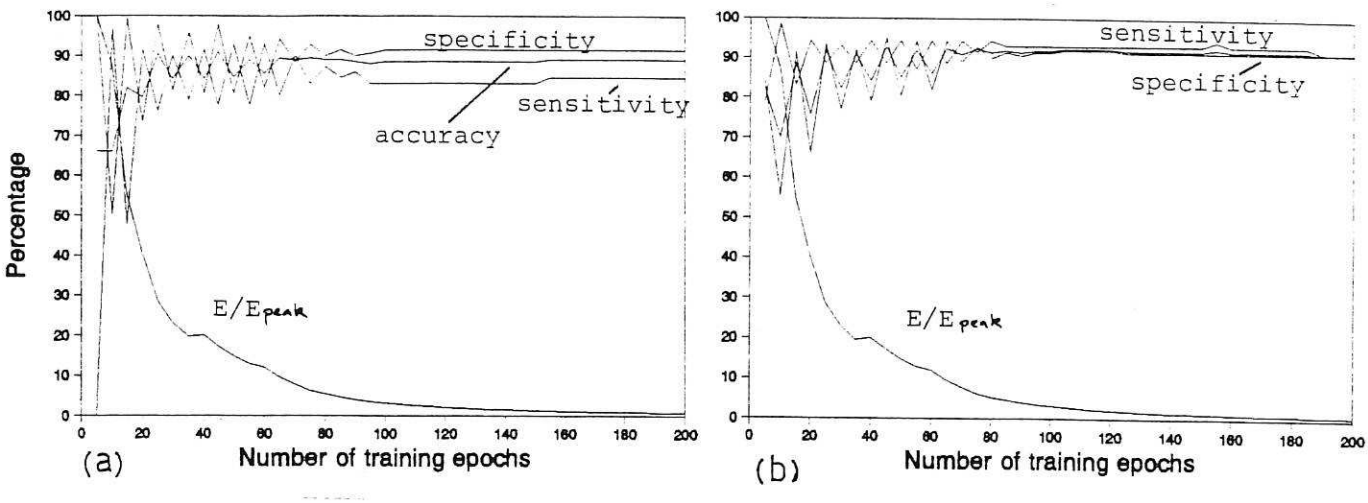

Figure 2: Variation of accuracy, sensitivity, specificity and E with number of training epochs. (a) Network alone. (b) Network and clinician using Table 1.

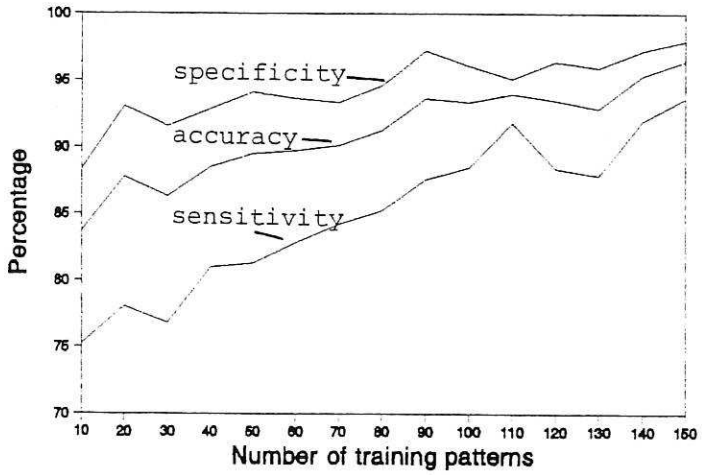

Figure 3: Variation of accuracy, sensitivity, specificity with size of training set.

not settle to their best values until E/E_{peak} has dropped below 5%. Little is to be gained from allowing training to continue beyond this point.

It is interesting to note that if the symptoms are coded as binary rather than bipolar vectors the time taken to converge to an acceptable level more than doubles. This is because, for a binary coding the absence of a symptom (coded as zero) does not disturb the network whereas, for bipolar coding, absence of a symptom is coded as -1 which does cause a change in the network's activity. Bipolar coding has the added advantage that missing items (coded as zero) do not disturb the network or introduce bias during training, and hence do not interfere with the diagnostic process.

3.4 The effect of the size of the training set

A technique akin to "jack-knifing" was used to gauge the effect of the size of training set. Here the network was trained on $\mathcal{N}$ patterns selected at random from the ensemble and tested on 150 others, also selected at random but not including the training set. Each of these train / test cycles was carried out 20 times and the resulting performance indicators were calculated by averaging over the 20 sets of results. The number of patterns, $\mathcal{N}$, included in each experiment was allowed to vary from 10 to 150. Figure 3 clearly demonstrates the way in which the network's predictive capability improves with increasing training sample size.

4 Discussion

At first sight the problem at hand appears to be one in which learning is relatively easy. Indeed, a simple linear regression can achieve an accuracy of approximately 85% indicating, perhaps, that the boundary between the AMI and non-AMI classes is not very complex and that E has a "nice" structure in weight space. However, this simplicity obscures some important points. In a small sample such as the cohort of 300 gathered here, there are many incidences of patients presenting with typical symptoms. These are easy to diagnose. Those presenting with atypical symptoms are, however, far fewer and their symptom patterns tend to be highly dissimilar. Therefore the effect of these statistical outliers is very pronounced. It is reasonable to assume then that a small number of hidden units can adequately represent the boundary between the *typical* AMI and non-AMI classes. However, in the present situation we are mainly interested in being able to predict the outliers — clinicians have little need of aids to diagnose typical patients. It is for this reason that the additional effort expended in including hidden units to improve performance by a few percent is worthwhile.

We have found that the use of hidden units has an increased advantage over logistic regression when the network output is combined with clinical judgement. This is because logistic regression tends to give outputs which are closer to zero or to one than the 53-18-1 configuration. The ability of the this configuration to be less "certain" about difficult cases can then be exploited by the fuzzy logic of Table 1.

4.1 Comparison of AMI predictions

Table 2 indicates the performance of the neural network in comparison with the admitting clinicians, the protocol derived by Goldman *et al* [7] [14] and logistic regression[†]. The network's architecture was 53-18-1 with a learning rate of 0.1, no momentum term and no weight decay term. It was trained on the first 90 cases for 155 epochs.

	Sensitivity	Specificity	Accuracy
Admitting clinician	78.9	84.2	82.4
Goldman's protocol	84.5	58.3	67.1
MLP (53-0-1)	85.9	89.9	88.6
MLP (53-18-1)	87.3	90.7	89.5
MLP (53-0-1) + Clin.	90.1	89.2	89.5
MLP (53-18-1) + Clin.	94.4	91.4	92.4

Table 2: Comparative performance of neural network

The protocol of Goldman and colleagues achieves greater sensitivity than the unaided clinician but only at the expense of specificity, which is very poor, and overall diagnostic accuracy. In contrast, logistic regression performs well in all three measures. The introduction of 18 hidden units into the network gives the 53-18-1 architecture only a slight advantage over logistic regression. However, the true advantage of hidden units can be seen in the combined results. Here the introduction of hidden units results in a significant improvement in performance relative to logistic regression and would have allowed the majority of atypical cases to be correctly diagnosed at presentation.

4.2 Comparison with a panel of experts

A panel of three senior clinicians was asked to give an opinion on 30 case histories (14 AMI, 10 angina and 6 other diagnoses). The cases included six which were relatively straightforward and 24 in which there had been a disparity between the admitting clinician and the ultimate diagnosis.

The neural network diagnosed 12 of the 14 cases of AMI and was "suspicious" of the other two, compared with 11 cases correctly diagnosed by the panel. Of the ten angina cases, two were incorrectly diagnosed as AMI by the neural network compared with five by the panel. Only one of the six other diagnoses was said to be AMI by the panel while all were correctly excluded by the network.

The panel of three clinicians would have administered thrombolytic therapy to 15 of the 30

[†]Equivalent to a 53-0-1 architecture

patients. Four of these did not have AMI while three patients with AMI would not have been treated. With the combination of the neural network and the opinion of the admitting clinician, thrombolytic treatment would have been given to all 14 with AMI and, at worst, only two (compared to six with the panel) would have received inappropriate thrombolytic therapy.

5 Conclusions

For a computer-based diagnostic aid to be of value it must be statistically valid, diagnostically accurate and its use must enhance the clinician's diagnostic performance. In addition it must be readily usable in the clinical setting. In this initial study we have not rigourously addressed the question of statistical validity but we have developed a system whose performance can exceed that of experienced physicians and is also superior to that of a conventionally derived system which has been widely evaluated. Our system is readily usable on a portable computer, could be developed as a hand held instrument and gives an instant prediction of the likelihood that a patient has sustained AMI.

We have also indicated how our system could be used to enhance the clinician's judgement. However, the true performance of the system can only be assessed by a formal, clinical trial.

References

[1] ISIS-2 (Second International Study of Infarct Survival) Collaborative Group (1988) Randomized trial of intravenous streptokinase, oral aspirin, both or neither, among 17,187 cases of suspected acute myocardial infarction *Lancet* **ii** 349-360.

[2] Rawles, J and Julian, D (1989) Rôle of the general practitioner in managing patients with myocardial infarction—impact of thrombolytic treatment (Brit Med J) **299** 555-557.

[3] Schroeder, J, Lamb, I and Harrison, D (1977) Patients admitted to coronary care unit for chest pain: high-risk subgroup for subsequent cardiovascular death *Am J Cardiol* **39** 829-832.

[4] Schroeder, J, Lamb, I and Hu, M (1978) The prehospital course of patients with chest pain: analysis of the prodromal, symptomatic, decision-making, transportation and emergency room periods *Am J Med* **65** 742-748.

[5] Fuchs, R and Scheidt, S (1981) Improved criteria for admission to cardiac care units *J Am Med Assoc* **426** 2037-2041.

[6] Boissel, J and Vanarie, R (1977) Système d'aide à la phase aiguë de l'infarctus du myocarde *Proc Int Symposium on Medical Informatics, Toulouse* 571-583.

[7] Goldman, L *et al* (1982) A computer-derived protocol to aid the diagnosis of emergency room patients with acute chest pain *New England J Med* **307** 588-596.

[8] Pozen, M *et al* (1984) A predictive instrument to improve coronary care unit admission practices in acute ischaemic heart disease: a prospective multi-centre clinical trial *New England J Med* **310** 1273-1278.

[9] Fattu, J, Blomberg, D and Patrick, E (1987) CONSULT learning system applied to early diagnosis of chest pain *Proc* 11th *Annual Symposium on Computer Applications in Medical Care* 71-77.

[10] Emerson, P *et al* (1988) The development of ACORN, an expert system enabling nurses to make admission decisions about patients with chest pain in an accident and emergency department *Medical Informatics in Clinical Medicine, Nottingham* 37-40.

[11] Hart, A and Wyatt, J (1989) Connectionist models in medicine: an investigation of their potential *Proc European Conf on AI in medicine, London* 115-124.

[12] Tonkonogy, J and Armstrong, J (1988) Diagnostic algorithms and clinical diagnostic thinking *Proc Symposium on Engineering of Computer-based Medical Systems, Minneapolis* 71-74.

[13] Peng, Y and Reggia, J (1989) A connectionist model for diagnostic problem solving *IEEE Trans on Systems, Man and Cybernetics* **19** 285-298.

[14] Goldman, L *et al* (1988) A computer protocol to predict myocardial infarction in emergency department patients with chest pain *New England J Med* **318** 797-803.

[15] Widrow, B, Gupta, N and Maitra, S (1973) Punish/reward: learning with a critic in adaptive threshold systems *IEEE Trans on Systems, Man and Cybernetics* **SMC 3** 455-465.

[16] Gutierrez, M, Wang, J and Lipschik, G (1989) Estimating hidden units for two-layer Perceptrons *Proc IEE 1st Int Conf on Artificial Neural Networks, London* 120-124.

[17] Rumelhart, D, Hinton, G and Williams, R (1986) Learning representations by back-propagating errors *Nature* **323** 533-536.

WHITING, B AND KELMAN, A.W, "The modelling of drug response," *Clinical Science*, vol. 59, pp. 311-317.

YAMAZAKI, T AND MAMDANI, E.H, (1982). "On the performance of a rule-based self-organising controller," *IEEE Conference on application of adaptive and multivariable control, Hull*, pp. 50-55.

YING, H, SHEPPARD, L, AND TUCKER, D, (1988). "Expert-system-based fuzzy control of arterial pressure by drug infusion," *Med Prog Though Tech*, vol. 13, pp. 203-215.

ZADEH, L.A, (1965). "Fuzzy sets," *Information and Control*, vol. 8, pp. 28-44.

ZADEH, L.A, (1973). "Outline of a new approach to the analysis of complex systems and decision processes," *IEEE Trans on Systems Man and Cyb*, vol. SMC-3, pp. 338-353.