Communicated by Richard Lippmann

# Use of an Artificial Neural Network for Data Analysis in Clinical Decision-Making: The Diagnosis of Acute Coronary Occlusion

**William G. Baxt**
*Department of Medicine, University of California, San Diego Medical Center, San Diego, CA 92103 USA*

**A nonlinear artificial neural network trained by backpropagation was applied to the diagnosis of acute myocardial infarction (coronary occlusion) in patients presenting to the emergency department with acute anterior chest pain. Three-hundred and fifty-six patients were retrospectively studied, of which 236 did not have acute myocardial infarction and 120 did have infarction. The network was trained on a randomly chosen set of half of the patients who had not sustained acute myocardial infarction and half of the patients who had sustained infarction. It was then tested on a set consisting of the remaining patients to which it had not been exposed. The network correctly identified 92% of the patients with acute myocardial infarction and 96% of the patients without infarction. When all patients with the electrocardiographic evidence of infarction were removed from the cohort, the network correctly identified 80% of the patients with infarction. This is substantially better than the performance reported for either physicians or any other analytical approach.**

## 1 Introduction

Decision-making under uncertainty is often fraught with great difficulty when the data on which the decision is based are imprecise and poorly linked to predicted outcome (Holloway 1979). Clinical diagnosis is an example of such a setting (Moskowitz et al. 1988) because multiple, often unrelated, disease states can present with similar or identical historical, symptomalogic, and clinical data. In addition, singular disease states do not always present with the same historical, symptomalogic, and clinical data. As a result, physician accuracy in diagnosing many of these diseases is often disappointing. A number of approaches have been developed to analyze data collected during patient evaluation to improve on diagnostic accuracy, but none of these approaches has been able to improve significantly on the performance of well-trained physicians (Reggia and Tuhrim 1985; Szolovits et al. 1988). The question still remains as to whether there is any means by which the data available in the clinical setting can be analyzed to yield information that can be utilized to improve diagnostic accuracy.

Acute myocardial infarction is an example of a disease process that has been difficult to diagnose accurately. A considerable number of methodologies have been developed in attempts to improve on the diagnostic accuracy of physicians in identifying the presence of acute myocardial infarction (Pozen et al. 1977, 1980, 1984; Goldman et al. 1982, 1988a; Patrick et al. 1976, 1977; Lee et al. 1985, 1987a,b; Tierney et al. 1985). Stepwise discriminate analysis (Pozen et al. 1977), logistic regression (Pozen et al. 1980), recursive partition analysis (Goldman et al. 1982), and pattern recognition (Patrick et al. 1976, 1977) have been utilized. The best of these approaches has performed with the same detection rate (sensitivity) (88%) and slightly better false alarm rate (1.0-specificity) (26% vs. 29%) than physicians (Goldman et al. 1988a). The following reports on the use of artificial neural network techniques (Widrow and Hoff 1960; Rumelhart et al. 1986; McClelland and Rumelhart 1988; Weigend et al. 1990; Mulsant and Servan-Schreiber 1988; Hudson et al. 1988; Smith et al. 1988; De Roach 1989; Saito and Nakano 1988; Marconi et al. 1989) to determine if the data collected during the routine evaluation of patients for acute myocardial infarction contain previously inapparent information that can be used to improve on the diagnostic accuracy of predicting the presence of acute myocardial infarction.

## 2 Methods

The nonlinear artificial neural network was a multilayer perceptron trained with backpropagation by use of the McClelland and Rumelhart simulator (McClelland and Rumelhart 1988). Figure 1 depicts the topology of the network utilized.

The network was trained by dividing the available data into a training set and a test set. Training took place by choosing input patterns from the training set and allowing activation to flow from the inputs through the hidden units to the output unit. The value of the output unit activation was then compared to the documented diagnosis for each pattern. The difference (error) between the actual activation of the output unit and the correct value was then utilized by the backpropagation algorithm (Rumelhart et al. 1986; McClelland and Rumelhart 1988) to modify all weights of the network so that future outputs approximate the correct diagnosis.

Because most patients presenting to the emergency department with anterior chest pain are not suffering from acute myocardial infarction (Goldman et al. 1988a), a subset of patients with a much greater probability of having sustained infarction were chosen for this study. To this end, only patients admitted to the coronary care unit were studied. In this

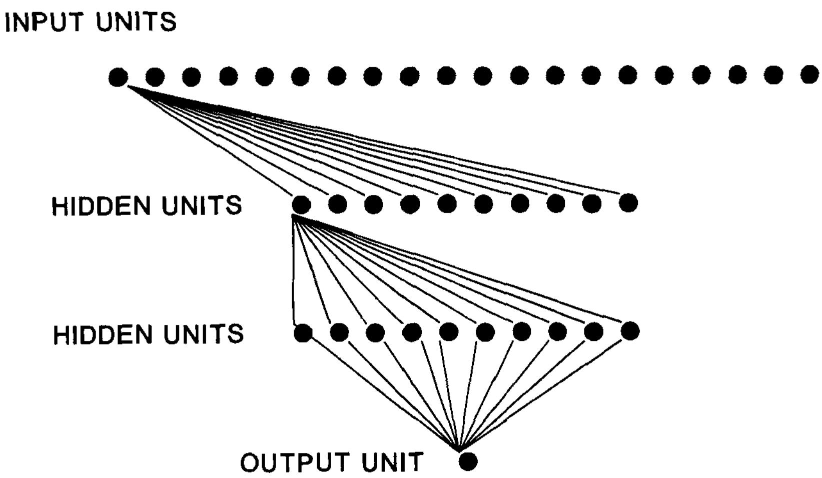


Figure 1: $20 \times 10 \times 10 \times 1$ nonlinear artificial neural network. The network has 20 input units, two layers of 10 hidden units each, and one output unit. Only the connections from one input and one hidden unit from each layer are shown. The network simulator program was run on a 80386 microcomputer with an 80387 math coprocessor running at 20 mHz. Epsilon was set at 0.05. Alpha was set at 0.9. Initial weights were random. Training times ranged between 8 and 48 hr.

way, the network was presented with the potentially most challenging pattern sets to differentiate. A retrospective chart review was performed on 356 patients who were admitted through the emergency department to the coronary care unit to rule out the presence of infarction. Forty-one variables reported to be predictive of the presence of acute myocardial infarction (Pozen et al. 1977, 1980, 1984; Goldman et al. 1982, 1988a; Patrick et al. 1976, 1977; Lee et al. 1985, 1987a,b,; Tierney et al. 1985) (depicted in Table 1) were collected on all patients from the emergency department record. The manner in which the presence or absence of infarction was determined was also documented from the inpatient record. The presence of infarction was confirmed as reported elsewhere (Goldman et al. 1988a).

The input patterns were generated by a specially written program that coded most of the clinical input variables in a binary manner such that 1 equalled the presence of a finding and 0 the absence of a finding. Patient age, blood pressure, pulse, and pain intensity were coded as analog values between 0.0 and 1.0. The target value for the output was coded as 0 for the subsequently confirmed absence of acute myocardial infarction and 1 for the confirmed presence of infarction.

| History | Past History | Examination | Electrocardiogram findings |
|---|---|---|---|
| Age* | Past AMI* | Systolic BP | 2 mm ST elevation* |
| Sex* | Angina* | Diastolic BP | 1 mm ST elevation* |
| Location of pain* | Congestive heart failure | Pulse | ST depression* |
| Intensity of pain | Diabetes* | Jugular venous distension* | T wave inversion* |
| Duration of pain | Hypertension* | Rales* | Peaked T wave |
| Radiation of pain | Family history AMI | Third heart sound | Premature ventricular contractions |
| Pain pleuritic | High cholesterol | Fourth heart sound | Heart block |
| Similar to past AMI | Coffee | Edema | Intraventricular conduction defect |
| Response to pressure | Cigarettes | | Significant ischemic change* |
| Response to nitroglycerin* | | | |
| Nausea and vomiting* | | | |
| Diaphoresis* | | | |
| Syncope* | | | |
| Shortness of breath* | | | |
| Palpitations* | | | |

Table 1: Input Variables. Variables marked with "*" utilized in final pattern sets.

To find a predictive set of input variables, different input pattern formats utilizing different numbers and combinations of the input variables were tested on networks that had as little as 5 to as many as 41 input units. To find a more optimal network architecture, different numbers of hidden units arranged in different numbers of layers were tested. A network with 20 inputs and 2 layers of 10 hidden units each, as depicted in Figure 1, utilizing the 20 clinical input variables noted in Table 1, was chosen on the basis of this analysis. Learning was followed by totaling the sum square (TSS) error over the pattern set (Rumelhart et al. 1986). Input patterns were presented to the network and learning epochs run until the TSS ceased decreasing. The final weights derived from a training session were then saved for use in testing.

Testing of a network was accomplished by using the weights derived in the training set and presenting the network with patterns to which it had *not* been exposed. Performance was scored as correct if the activation of the output unit was equal to or greater than 0.8 when the target was 1 or when the activation of the output unit was equal to or less than 0.2 when the target was 0. The output unit activation in this study was always between 0 and 0.2 and 0.8 and 1.0. Detection rate (sensitivity) was defined as the number of patients in a test population correctly diagnosed as having a disease divided by the total number in the test set with the disease. False alarm rate (specificity) was defined as the number of patients in a test population correctly diagnosed as not having a disease divided by the total number in the test set without the disease.

## 3 Results

The network was trained utilizing a randomly chosen subset of patterns derived from the initial group of 356 patients. Half of the patients who had not sustained acute myocardial infarctions and half of the patients who had sustained infarctions were selected. The subset consisted of 118 patients who were diagnosed as not having sustained an infarction and 60 patients who were diagnosed as having sustained an infarction. The final TSS reached was 0.044 error per pattern. The network was then tested on the remaining 178 patients (118 noninfarction, 60 infarction) to which it had not been exposed. The network correctly diagnosed 55 of the 60 patients with infarction and 113 of the 118 patients without infarction.

This process was repeated utilizing the second pattern set for network training and the first set for testing. The initial TSS achieved during training was 0.02 error per pattern. The network correctly diagnosed 56 of the 60 patients with acute myocardial infarction and 113 of the 118 patients without infarction on the test set. The summed results of the two test sets are depicted in Table 2. The network performed with a detection rate of 92% and a false alarm rate of 96%.

| | Noninfarction | Infarction |
|---|---|---|
| Correct | 226 | 111 |
| Incorrect | −10 | −9 |

Table 2: All Patients. Detection rate (sensitivity), 92%; false alarm rate (1.0-specificity), 4%.

| | Noninfarction | Infarction |
|---|---|---|
| Correct | 47 | 44 |
| Incorrect | −4 | −7 |

Table 3: Infarction Patients without ST Elevation. Detection rate (sensitivity), 86%; false alarm rate (1.0-specificity), 8%.

A significant number of patients who present to the emergency department who have sustained acute myocardial infarction have clear-cut electrocardiographic evidence of infarction. The real diagnostic challenge arises in those patients who have sustained infarction, but do not have clear-cut evidence of infarction on their initial electrocardiogram. When such patients were omitted from one study that attempted to improve on physician diagnostic performance, detection fell significantly (Goldman et al. 1988b).

To determine if this approach could effectively identify new information under the most challenging circumstances, the network was further trained and tested on those patients without clear-cut electrocardiographic evidence of acute myocardial infarction. Fifty-two percent of the patients with a documented infarction had acute ST segment elevation on their initial electrocardiogram and none of the patients without infarction had this finding. To study the effect of eliminating such patients, the network was trained on a pattern set derived from half of the patients who sustained infarctions who did not have ST elevation on their initial electrocardiogram along with an equal number of randomly selected patients who had not sustained infarctions. The network was then tested on the second half of the patients who had sustained infarctions who did not have ST elevation on their initial electrocardiogram along with a randomly chosen equal number of patients from the group that had not sustained infarctions. As above, the process was then reversed utilizing the second pattern set for training and the first set for testing. The results are summarized in Table 3. The network performed with a detection rate of 86% and a false alarm rate of 92%, indicating that network performance was not dependent on the presence of ST elevation on the initial electrocardiogram.

Eighty-three percent of the patients with a documented acute myocardial infarction had either acute ST elevation or new ischemic change on their initial electrocardiogram and none of the patients without infarction had this finding. To further study the effect of clear-cut electrocardiographic markers, a set of patients whose initial electrocardiogram showed neither ST elevation nor new ischemic change were identified.

There were 20 such patients. These patients were combined with 20 randomly chosen patients who had not sustained infarction. Because of the small sample size, a leave-one-out strategy was used to test network performance. Input patterns for training were derived from 19 of the 20 patients who had sustained infarction who had neither ST elevation nor significant ischemic change on their initial electrocardiogram along with 20 randomly chosen patients who had not sustained infarction. Twenty such sets of training data were constructed by removing a different infarction patient in each set. The network was trained on each of these pattern sets and tested on the one infarction patient that had been removed. The network correctly identified 16 of the 20 patients with infarction (detection rate 80%), further indicating that network performance was not predominantly dependent on electrocardiographic markers of infarction.

## 4 Discussion

These data reveal that the artificial neural network had a detection rate of 92% and a false alarm rate of 4%, whereas the best previously reported performance had a detection rate of 88% and a false alarm rate of 26%.

Although these results are encouraging, future studies will need to address some of the questions that were not fully answered in this study. The proof that the nonlinear artificial neural network identified and utilized new information rests on the improvement of diagnostic accuracy derived from comparisons to studies reported in the literature. The results reported here are, thus, compared to studies based on a different set of data. Valid comparisons between methodologies must use the same data sets. In addition, the physician performance on the data set studied herein was not determined and may have been better than that described in the literature. Absolute conclusions about comparative performance will need to be derived from the prospective study of this question. Further, the studies to which these data were compared evaluated all patients presenting to the emergency department with nontraumatic chest pain. This study analyzed data collected only from patients who were admitted to the coronary care unit. The consequence of this will require study.

The good performance afforded by the network deserves comment. Previously utilized statistical strategies have been based on one of three approaches: (1) tree structure rule-based interrelationships, (2) linear pattern matching, or (3) statistical probability calculations (Szolovits et al. 1988). All of these methods are heavily dependent on the consistency of input data for proper performance. One of the striking aspects about the presentation of most disease states is the lack of consistency in their presentation. This emanates from both vague and imprecise clinical histories as well as marked variations in the symptom clusters and clinical findings with which identical disease processes can present. Decision modalities that are highly dependent on consistency of input to arrive at correct diagnostic closure will perform poorly in this setting. All of the other approaches to clinical decision-making alluded to above are based on a highly structured set of rules or statistical probability prediction that are dependent on the accuracy of input data. One possible reason for the good performance of the artificial neural network is that the nonlinear statistical analysis of the data performed tolerates a considerable amount of imprecise and incomplete input data. These networks appear to be able to cope with the subtle variations in the way disease processes present without making categorical decisions solely driven by these variations. The networks appear to be able to discover implicit higher order conditional dependencies in patterns that are not apparent on face value and to utilize these dependencies to derive generalized rules that are resistant to most minor input perturbations. Specifically, the network can shift from one set of input variables to another and still make accurate prediction based on the actual data at hand. It is this resistance to the perversion of accurate generalizations that enables the network to function more accurately in the clinical environment.

The network used input data that are routinely available to and utilized by physicians screening patients for the presence of acute myocardial infarction. The network simply discovered relationships in these data that are evidently not immediately apparent to physicians and was able to use these to come to a more accurate diagnostic closure. Because these relationships can be made explicit by studying the network weighting of input data, physicians could potentially utilize this information to make a more accurate diagnosis. The actual possibility of this will depend on the complexity of the interrelationships defined by the network. It has been demonstrated that when networks with more than one hidden layer are required to achieve optimal training, the solutions are often distributed over multiple units and are difficult to identify (Weigend et al. 1990). However, if these relationships can be elucidated, the use of the network may be unnecessary.

These observations must be validated by extending this study to a larger number of patients, followed by the prospective testing of the relationships identified by the network. If these results hold up to such scrutiny, the improvement in predictive accuracy could have a substantial impact on the reduction of health care costs. Furthermore, these techniques may be able to be extended to other clinical settings.

## Acknowledgments

I thank Dr. David Zipser for his help with this study and Kathleen James for her help in the preparation of this manuscript.

## References

De Roach, J. N. 1989. Neural networks — An artificial intelligence approach to the analysis of clinical data. *Austral. Phys. Engineer. Sci. Med.* **12**, 100–106.

Goldman, L., Weinberg, M., Weisberg, M., Olshen, R., Cook, E. F., Sargent, R. K., Lamas, G. A., Dennis, C., Wilson, C., Deckelbaum, L., Fineberg, H., Stiratelli, R., and the Medical House Staffs at Yale–New Haven Hospital and Brigham and Women's Hospital. 1982. A computer-derived protocol to aid in the diagnosis of emergency room patients with acute chest pain. *N. Engl. J. Med.* **307**, 588–596.

Goldman, L., Cook, E. F., Brand, D. A., Lee, T. H., Rouan, G. W., Weisberg, M. C., Acampora, D., Stasiulewicz, C., Walshon, J., Terranova, G., Gottlieb, L., Kobernick, M., Goldstein-Wayne, B., Copen, D., Daley, K., Brandt, A. A., Jones, D., Mellors, J., and Jakubowski, R. 1988a. A computer protocol to predict myocardial infarction in emergency department patients with chest pain. *N. Engl. J. Med.* **318**, 797–803.

Goldman, L., Cook, E. F., Brand, D. A., Lee, T. H., and Rouan, G. W. 1988b. Letter to the editor. *N. Engl. J. Med.* **319**, 792.

Holloway, C. A. 1979. Behavioral assumptions and limitations of decision analysis. In *Decision Making Under Uncertainty: Models and Choices*, C. A. Holloway, ed., pp. 436–455. Prentice-Hall, Englewood Cliffs, NJ.

Hudson, D. L., Cohen, M. E., Anderson, M. F. 1988. Determination of testing efficacy in carcinoma of the lung using a neural network model. *Symp. Comput. Applic. Med. Care Proc.* **12**, 251–255.

Lee, T. H., Cook, E. F., Weisberg, M., Sargent, R. K., Wilson, C., and Goldman, L. 1985. Acute chest pain in the emergency ward: Identification and examination of low-risk patients. *Arch. Intern. Med.* **145**, 65–69.

Lee, T. H., Rouan, G. W., Weisberg, M. C., Brand, D. A., Cook, F., Acampora, D., Goldman, L., and the Chest Pain Study Group; Boston, MA; New Haven, Danbury, and Milford, CT; and Cincinnati, OH. 1987a. Sensitivity of routine clinical criteria for diagnosing myocardial infarction within 24 hours of hospitalization. *Ann. Intern. Med.* **106**, 181–186.

Lee, T. H., Rouan, G. W., Weisberg, M. C., Brand, D. A., Acampora, D., Stasiulewicz, C., Walshon, J., Terranova, G., Gottlieb, L., Goldstein-Wayne, B., Copen, D., Daley, K., Brandt, A. A., Mellors J., Jakubowski, R., Cook, E. F., and Goldman, L. 1987b. Clinical characteristics and natural history of patients with acute myocardial infarction sent home from the emergency room. *Am. J. Cardiol.* **60**, 219–224.

Marconi, L., Scalia, F., Ridella, S., Arrigo, P., Mansi, C., and Mela, G. S. 1989. An application of back propagation to medical diagnosis. *Symp. Comput. Applic. Med. Care Proc.*, in press.

McClelland, J. L., and Rumelhart, D. E., eds. 1988. Training hidden units. In *Explorations in Parallel Distributed Processing*, pp. 121–160. MIT Press, Cambridge, MA.

Moskowitz, A. J., Kuipers, B. J., and Kassirer, J. P. 1988. Dealing with uncertainty, risks, and trade-offs in clinical decisions. A cognitive science approach. *Ann. Intern. Med.* **108**, 435–449.

Mulsant, G. H., and Servan-Schreiber, E. 1988. A connectionist approach to the diagnosis of dementia. *Symp. Comput. Applic. Med. Care Proc.* **12**, 245–250.

Patrick, E. A., Margolin, G., Sanghvi, V., and Uthurusamy, R. 1976. Pattern recognition applied to early diagnosis of heart attacks. In *Proceedings of the IEEE 1976 Systems, Man, and Cybernetics Conference*, Washington, D.C., November 1–3, pp. 403–406.

Patrick, E. A., Margolin, G., Sanghvi, V., and Uthurusamy, R. 1977. Pattern recognition applied to early diagnosis of heart attacks. In *Proceedings of the 1977 International Medical Information Processing Conference (MEDINFO)*, Toronto, August 9–12, pp. 203–207.

Pozen, M. W., Stechmiller, J. K., and Voigt, G. C. 1977. Prognostic efficacy of early clinical categorization of myocardial infarction patients. *Circulation* **56**, 816–819.

Pozen, M. W., D'Agostino, R. B., Mitchell, J. B., Rosenfeld, D. M., Guglielmino, J. M., Schwartz, M. L., Teebagy, N., Valentine, J. M., and Hood, W. B. 1980. The usefulness of a predictive instrument to reduce inappropriate admissions to the coronary care unit. *Ann. Intern. Med.* **92**, 238–242.

Pozen, M. W., D'Agostino, R. B., Selker, H. P., Sytkowski, P. A., Hood, W. B., Jr. 1984. A predictive instrument to improve coronary-care-unit admission practices in acute ischemic heart disease: A prospective multicenter clinical trial. *N. Engl. J. Med.* **310**, 1273–1278.

Reggia, J. A., and Tuhrim, S., eds. 1985. *Computer Assisted Medical Decision Making. Computers in Medicine Series*, Vol. 2. Springer-Verlag, New York.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. 1986. Learning internal representations by error propagation. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, D. E. Rumelhart and J. L. McClelland, eds., pp. 318–364. MIT Press, Cambridge, MA.

Saito, K., and Nakano, R. 1988. Medical diagnostic expert system based on PDP model. In *Proceedings of the International Joint Conference on Neural Networks*, **I**, 255–262.

Smith, J. W., Everhart, J. E., Dickson, W. C., Knowler, W. C., and Johannes, R. S. 1988. Using the ADAP learning algorithm to forecast the onset of diabetes mellitus. *Symp. Comput. Applic. Med. Care Proc.* **12**, 261–265.

Szolovits, P., Patil, R. S., and Schwartz, W. B. 1988. Artificial intelligence in medical diagnosis. *Ann. Intern. Med.* **108**, 80–87.

Tierney, M. W., Roth, B. J., Psaty, B., McHenry, R., Fitzgerald, J., Stump, D. L., Anderson, F. K., Ryder, K. W., McDonald, C. J., and Smith, D. M. 1985. Predictors of myocardial infarction in emergency room patients. *Crit. Care Med.* **13**, 526–531.

Weigend, A. S., Huberman, B. A., and Rumelhart, D. E. 1990. Predicting the future: A connectionist approach. PDP Research Group Technical Report. *Int. J. Neural Syst.*, submitted.

Widrow, G., and Hoff, M. E. 1960. Adaptive Switching Circuits Institute of Radio Engineering Western Electronic Show and Convention. Convention Record, Part 4, pp. 96–104.

Received 29 August 90; accepted 26 September 90.