JACC Vol. 28, No. 4
October 1996:1012–6

Agreement Between Artificial Neural Networks and Experienced Electrocardiographer on Electrocardiographic Diagnosis of Healed Myocardial Infarction

BO HEDÉN, MD, MATTIAS OHLSSON, PHD, RALF RITTNER, MSC, OLLE PAHLM, MD, PHD, WESLEY K. HAISTY, JR., MD,* CARSTEN PETERSON, PHD, LARS EDENBRANDT, MD, PHD

Lund, Sweden and Winston-Salem, North Carolina

Objectives. **The purpose of this study was to compare the diagnoses of healed myocardial infarction made from the 12-lead electrocardiogram (ECG) by artificial neural networks and an experienced electrocardiographer.**

Background. **Artificial neural networks have proved of value in pattern recognition tasks. Studies of their utility in ECG interpretation have shown performance exceeding that of conventional ECG interpretation programs. The latter present verbal statements, often with an indication of the likelihood for a certain diagnosis, such as "possible left ventricular hypertrophy." A neural network presents its output as a numeric value between 0 and 1; however, these values can be interpreted as Bayesian probabilities.**

Methods. **The study was based on 351 healthy volunteers and 1,313 patients with a history of chest pain who had undergone diagnostic cardiac catheterization. A 12-lead ECG was recorded in each subject. An expert electrocardiographer classified the ECGs in five different groups by estimating the probability of anterior myocardial infarction. Artificial neural networks were trained and tested to diagnose anterior myocardial infarction. The network outputs were divided into five groups by using the output values and four thresholds between 0 and 1.**

Results. **The neural networks diagnosed healed anterior myocardial infarctions at high levels of sensitivity and specificity. The network outputs were transformed to verbal statements, and the agreement between these probability estimates and those of an expert electrocardiographer was high.**

Conclusions. **Artificial neural networks can be of value in automated interpretation of ECGs in the near future.**

(J Am Coll Cardiol 1996;28:1012–6)

Artificial neural networks are computer-based decision tools that have proved of particular value in pattern recognition tasks. Their utility has been tested in processing of the electrocardiogram (ECG) (1–4), and studies concerning detection of myocardial infarction and lead reversal have reported performance exceeding that of conventional rule-based ECG interpretation programs (5,6). The diagnostic performance of the artificial neural networks in those studies makes it of interest to assess the possibility of implementing artificial neural networks in conventional ECG interpretation programs. However, neural networks present numeric output values, whereas conventional ECG interpretation programs present verbal statements. For some diagnoses the latter also present different levels of likelihood, such as "possible left ventricular hypertrophy" or "probable inferior myocardial infarction." This approach is now widely used and accepted by ECG readers.

Statements with probability estimates can also be obtained with artificial neural networks. It has been shown that a neural network output under certain circumstances indicates a Bayesian probability (see Appendix). An artificial neural network classifying ECGs as indicative or not indicative of anterior myocardial infarction has output values between 0 and 1. Values close to 0 should be assigned by the network to normal ECGs, and values close to 1 assigned to ECGs with clear-cut changes consistent with anterior myocardial infarction, such as a QS pattern in leads V_2 to V_4. Intermediate values should be assigned to ECGs with borderline findings (such as poor R wave progression in anterior leads). Therefore it would be appropriate, also from a theoretic point of view, to introduce several thresholds to the network output and, hence, several categories, such as "no," "possible," "probable" and "definite" infarction. The purpose of the present study was to transform numeric artificial neural network outputs into verbal statements and to compare these verbal probability estimates with those of an experienced electrocardiographer. A data base of digitized ECGs was therefore analyzed for the presence or

From the Departments of Clinical Physiology and Theoretical Physics, Lund University, Lund, Sweden, and *Department of Medicine/Cardiology, Bowman Gray School of Medicine, Winston-Salem, North Carolina. This study was supported by grants from the Swedish Medical Research Council (B95-14X-09893-04B), Stockholm; Swedish National Board for Industrial and Technical Development, Stockholm; the Faculty of Medicine at Lund University, Lund, Sweden; the Göran Gustafsson Foundation for Research in National Science and Medicine, Stockholm; and the Swedish Natural Science Research Council, Stockholm.

Manuscript received October 24, 1995; revised manuscript received March 27, 1996, accepted June 3, 1996.

Address for correspondence: Dr. Lars Edenbrandt, Department of Clinical Physiology, University Hospital, S-221 85 Lund, Sweden. E-mail: lars.edenbrandt@klinfys.lu.se.

0735-1097/96/$15.00
PII S0735-1097(96)00269-0

Table 1. Number of Patients in Total Study Group and Subgroups

	Patients (no.)
Total anterior myocardial infarction group	414
Isolated anterior myocardial infarction	272
Combined anterior + inferior myocardial infarction	142
Total control group	1,250
Healthy volunteers	351
Catheterization normal	479
Isolated inferior myocardial infarction	356
Isolated posterior myocardial infarction	64
Total study group	1,664

absence of healed anterior myocardial infarction, and ECG-independent methods were used as a reference standard.

Methods

Study group. A total of 1,664 subjects were included in the study; 351 healthy volunteers and 1,313 patients with a history of chest pain. The healthy volunteers were selected at random from a defined urban population. They were without any known or suspected heart disease, lung disease or any other pathologic condition that might influence the ECG (7). All patients had undergone diagnostic cardiac catheterization at the North Carolina Baptist Hospital, Winston-Salem, North Carolina. Patients with normal coronary arteries, normal findings on contrast left ventriculography, no evidence of valve dysfunction or congenital heart disease, ejection fraction ≥50% and an overall study evaluation of "normal" were classified as "catheterization-normal." Anterior myocardial infarction was defined by presence of ≥75% diameter stenosis of the left main coronary artery, the left anterior descending coronary artery or its major diagonal branches and akinesia or dyskinesia of the anterosuperior wall in the right anterior oblique ventriculogram. Inferior myocardial infarction was defined by presence of ≥75% diameter stenosis of the right coronary artery and akinesia or dyskinesia of the inferior wall in the right anterior oblique ventriculogram. Posterolateral myocardial infarction was defined by the presence of ≥75% diameter stenosis of the left circumflex artery or any of its major branches and akinesia or dyskinesia of the posterolateral wall in the left anterior oblique ventriculogram.

Patients with isolated anterior myocardial infarction and patients with both anterior and inferior myocardial infarction constituted the anterior myocardial infarction group. A control group was composed of the healthy volunteers, patients classified as catheterization-normal and patients with isolated inferior or posterolateral myocardial infarction. Patients with technically deficient ECGs or ECGs showing left bundle branch block were excluded. The number of patients in the different subgroups of the overall study group is presented in Table 1.

ECG analysis. A 12-lead ECG was recorded in each subject by using a computerized electrocardiograph. The frequency range was in accordance with American Heart Association specifications (0.05 to 100 Hz). Noise reduction was made by time-coherent averaging. Averaged complexes were transferred to a computer and stored for further analysis. Measurements of amplitudes and durations of the ECG complexes were performed by using custom software. The following automated measurements from leads V_2, V_3 and V_4 were used as inputs to the artificial neural networks: Q, R and S wave amplitudes, Q and R wave durations as well as three amplitudes within the ST-T segment. The interval between the ST junction and the end of the T wave was divided into six segments of equal duration, and the amplitudes at the end of segments 1, 3 and 5 were used as network inputs.

Electrocardiographer. An experienced electrocardiographer classified each of the electrocardiograms into one of the following five classes: I = definitely no anterior myocardial infarction; II = probably no anterior myocardial infarction; III = possible anterior myocardial infarction; IV = probable anterior myocardial infarction; V = definite anterior myocardial infarction.

The ECGs, showing only leads V_1 to V_6, were presented in random order to the electrocardiographer. No personal data, clinical findings or results from the neural networks were available at the classification procedure.

Artificial neural networks. A multilayered perceptron artificial neural network architecture (8) was used. A more general description of neural networks can be found elsewhere (9). The neural networks consisted of one input layer, one hidden layer and one output layer. The number of neurons in the input layer equals the number of input variables (i.e., 24 measurements from leads V_2 to V_4, as presented above. The hidden layer contained six neurons, and a single output unit encoded the probability of anterior myocardial infarction. Each variable in the training set is normalized such that the mean of all examples is 0 with a unit variance.

The data set was divided into a training set and a test set. The training set was used to adjust the connection weights, whereas the test set was used to assess the performance. To obtain as reliable performance as possible a K-fold cross-validation procedure was used. The data set was randomly divided into K equal parts. Each of the K different parts of the data was used once as a test set, while training was performed on the remaining (K-1) parts. We used threefold cross validation to decide when to terminate learning in order to avoid "overtraining" and eightfold cross validation to train the networks and assess their performances. The results presented are based on 10 independent training/test runs; that is, the eightfold cross-validation procedure was repeated 10 times.

During the training process the connection weights between the neurons were adjusted by using the backpropagation algorithm. A sigmoid transfer function was used. The learning rate (η) had a start value of 0.5. During the training η was decreased geometrically between epochs by using the following equation:

$$\eta = k\eta \quad \text{with} \quad k = 0.998.$$

Table 2. Electrocardiographer and Artificial Neural Network Classifications of 1,664 Electrocardiograms

ECG Class	Artificial Neural Network Class				
	I	II	III	IV	V
I	*986*	99	10	8	1
II	92	*58*	16	11	10
III	16	14	*7*	10	8
IV	8	14	10	*23*	18
V	2	2	12	21	*208*

Numbers in italics indicate agreement between the electrocardiographer (ECG) and artificial neural network classification. See Methods for definition of classes I to V.

The momentum α was set to 0.7. Updating occurred after each 10 patterns. Training was terminated at a training error of 0.245, which was achieved after 18 to 21 epochs. The network weights were initiated with random numbers between −0.025 and 0.025. All calculations were done using the JETNET 3.0 package (10).

The ECGs were classified into five groups by using the network outputs and four different thresholds between 0 and 1. The thresholds were selected so as to give the same number of ECGs in classes I to V as were the result of the classification of the electrocardiographer. Complete agreement between the neural network and the electrocardiographer could be obtained only by using these thresholds.

Statistical methods. The significance of the difference in sensitivities between the artificial neural networks and the electrocardiographer was tested with attention to the fact that the same ECGs were used; that is, a McNemar type statistic was used.

Results

Performance. The electrocardiographer classified 1,291 ECGs as "definitely no anterior myocardial infarction" (n = 1,104) or "probably no anterior myocardial infarction" (n = 187). Of these ECGs, 1,185 were control ECGs, resulting in a specificity of 94.8%. A classification as "possible anterior myocardial infarction" (n = 55), "probable anterior myocardial infarction" (n = 73) or "definite anterior myocardial infarction" (n = 245) was assigned to 373 ECGs. A true positive classification was made in 308 of these cases, resulting in a sensitivity of 74.4%. The sensitivity for the neural network was 81.4% at a specificity of 94.8% and this difference in sensitivity was significant ($p < 0.001$).

Agreement/disagreement. The classifications of the ECGs by the electrocardiographer and the neural network are presented in Table 2. There was agreement in 1,282 ECGs (77.0%), a difference of one class or less in 1,562 ECGs (93.9%) and a difference two classes or less in 1,633 (98.1%). In 31 cases a difference of more than two classes was found. The electrocardiographer was correct in 9 of these ECGs and the network in 22.

The nine ECGs on which the electrocardiographer and the network disagreed by more than two classes, and on which the network was incorrect constitute a particularly interesting group. One of the nine ECGs had serious errors in the data of the measurement program and was therefore not properly presented to the networks. ECGs with errors of this kind may impair the performance of the artificial neural network both when they appear in the training set and the test set. Leads V_2 to V_4 of the remaining eight ECGs and the network outputs (means of 10 different runs) are presented in Figures 1 and 2.

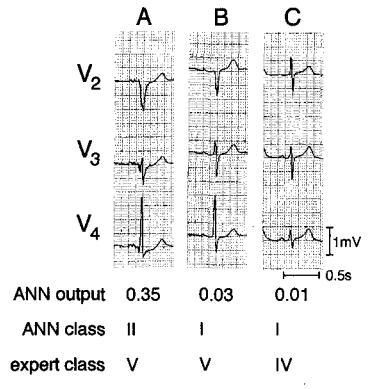
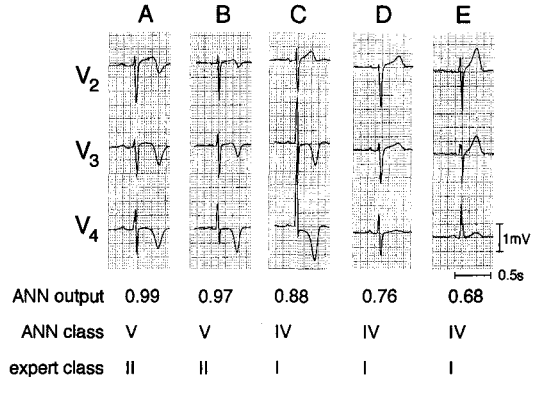


Figure 1. Three ECGs from the anterior myocardial infarction group correctly classified by the electrocardiographer (expert) but incorrectly classified as definitely no anterior myocardial infarction (class I) or probably no anterior myocardial infarction (class II) by the artificial neural network (ANN).

Three ECGs in the anterior myocardial infarction group classified as probable or definite anterior myocardial infarction by the electrocardiographer and as definitely or probably no anterior myocardial infarction by the neural network are presented in Figure 1. All three ECGs have R waves, though with small amplitudes in some leads. They also have normal T waves. Some QRS complexes have abnormal notches. This information is not given to the network but could be used by an ECG expert. The reversed R wave progression found in panel

Figure 2. Five ECGs from the control group correctly classified by the electrocardiographer (expert) but incorrectly classified as probable anterior myocardial infarction (class IV) or definite anterior myocardial infarction (class V) by the artificial neural network (ANN).



C of Figure 1 was not a common finding in the material. Therefore, this pattern might be difficult for the network to learn.

Figure 2 presents five ECGs from the control group that the network falsely classified as definite or probable anterior myocardial infarction. The extremely negative T waves found in three cases were probably important in the network classifications. The ECG in panel E has a decreasing R wave amplitude from lead V_2 to lead V_3. This is not a normal finding and the network classification is therefore not surprising. However, it is not obvious why the network output of the ECG in panel D is as high as 0.76, resulting in a classification of probable anterior myocardial infarction. A network trained and tested using QRS measurements only (without ST amplitudes) obtained a lower output value and hence correctly classified this case. This indicates that the ST amplitudes were important for the high output value of the neural network, which used both QRS and ST measurements as input variables.

Discussion

Main findings. The results of this and an earlier study (5) show that neural networks can be trained to diagnose myocardial infarction from the ECG with greater accuracy than that obtained with use of a conventional interpretation program and an experienced electrocardiographer. This study also showed a high level of agreement between the artificial neural network and the electrocardiographer. When there was obvious disagreement the artificial neural network was correct somewhat more often than the expert, with regard to the reference standard of this study material. Most users of black box methods like artificial neural networks worry that the methods make obvious and severe misclassifications in some cases even though their overall performance is very good. The worst network errors made in the 1,664 ECGs in this study are presented in Figures 1 and 2.

Reasons for misclassification. Why were some ECGs misclassified by the artificial neural network and correctly classified by the electrocardiographer? A relatively small number of input variables was used to train the neural networks in this study. A network fed with many input variables requires many examples in the training set. As a rule of thumb, the number of training examples needed for appropriate training is 10 times the total number of interneuron connections in the neural network. In this study only eight variables from each of three leads were used, but the number of weights was as high as 157. A network of this size could be trained by using a data base of some 1,500 ECGs, as in this study, but much larger networks would probably not be sufficiently trained. In contrast, the electrocardiographer makes his decision based on much more data—in this study the QRS complexes and ST-T segments of six leads. Therefore, it is not surprising that the electrocardiographer outperforms the neural network in a few ECGs with minor configurational deviations, such as notches in the QRS complex.

Another reason for misclassification by the neural networks may have been that the networks in this study were only trained to diagnose anterior myocardial infarction. Therefore, some ECGs with deep inverted T waves but normal QRS configuration, as in Figure 2, are likely to be classified as showing anterior myocardial infarction. However, when all precordial leads are taken into account, left ventricular hypertrophy with strain is a probable diagnosis. However, a neural network could only learn this pattern if a sufficient number of examples of left ventricular hypertrophy were added to the data base.

Clinical implications. One advantage of artificial neural networks over rule-based criteria is the enhanced diagnostic performance. Another advantage is the ability to easily adjust the network outputs in different clinical situations. Neural network outputs can be regarded as Bayesian a posteriori probabilities if the a priori probabilities of the classes in the training data base are the same as the a priori probabilities in the test situation. In this study the a priori probabilities were 0.25 for anterior myocardial infarction and 0.75 for non-anterior myocardial infarction. Consequently, the networks will only provide good Bayesian probabilities if used in environments with these a priori class probabilities. It is also possible to use the network in test situations with different a priori probabilities without retraining (see Appendix). Consider, for example, an ECG with a network output of 0.85, which was interpreted as probable anterior myocardial infarction in this study. If this ECG were analyzed by an artificial neural network from this study but recorded in a screening situation, where the a priori probability of anterior myocardial infarction is 0.05, the output value of the network would be adjusted from 0.85 to 0.47 to represent a true a posteriori probability. If the same ECG were recorded in a third situation with a high a priori probability (0.50), the a posteriori probability would be 0.94. With use of the same thresholds for ECG classification, the resulting statement would be "possible anterior myocardial infarction" in the screening situation and "definite anterior myocardial infarction" in the high a priori probability situation. Also, an experienced electrocardiographer takes into account the clinical situation in which an ECG is recorded and adjusts the interpretation accordingly.

A disadvantage with artificial neural networks is the lack of reasons for a certain diagnosis, which at least in theory can be presented from rule-based criteria. However, these criteria are usually very complex. They are rarely studied in clinical practice and probably not easy for many ECG readers to understand. Nevertheless, they are well accepted by millions of users.

Conclusions. Artificial neural networks can be trained to diagnose healed anterior myocardial infarction at high levels of sensitivity and specificity. The outputs from the neural networks can be transformed to verbal statements, and the agreement between these probability estimates and those of an expert electrocardiographer is high. Reasons for misdiagnosis by the artificial neural network are the limited number of variables of the ECG used as input values and the presence of ECGs with uncommon features. Use of a large number

of examples to train the artificial neural network will lower the risk of misdiagnosis.

Appendix

Recall that (11) a Bayesian probability P(Cix) represents the conditional probability for a class Ci given input x. The Bayes rule tells us that it can be expressed as

$$P(Cix) = (P(xCi)P(Ci))/P(x),$$

where P(xCi) is the conditional probability for producing the input vector x given the class Ci, P(Ci) is the a priori probability of class Ci and P(x) is the input probability distribution. In conventional Bayesian analysis P(xCi) is given by well known parametric distributions (e.g., Gaussian), and the training involves estimating the parameters.

The artificial neural network is a black box method. However, it has been shown (12) that output values from a multilayered perceptron can be interpreted as Bayesian probabilities P(Cix) provided that 1) the training is accurate; 2) the outputs are of 1-of-M-type, which means that the task is coded such that only one output unit should be "on" at a time; and 3) a summed squared error or cross entropy error function is used. In addition, the a priori class probabilities P(Ci) have to be representative of actual use or test conditions. However, it is possible to vary the class probabilities P(Ci) during classification without retraining the network because P(Ci) only occur as a multiplicative term in the expression for P(Ci). One simply divides the outputs by the training class probabilities and multiplies by the correct class probabilities. The benefit from this Bayesian interpretation of the artificial neural network output units is that they can be subjected to higher level decision analysis.

References

1. Baxt WG. Use of an artificial neural network for the diagnosis of myocardial infarction. Ann Intern Med 1991;115:843–8.
2. Bortolan G, Willems JL. Diagnostic ECG classification based on neural networks. J Electrocardiol 1993;26:S75–9.
3. Edenbrandt L, Devine B, Macfarlane PW. Classification of electrocardiographic ST-T segments—human expert versus artificial neural network. Eur Heart J 1993;14:464–8.
4. Clayton RH, Murray A, Campbell RWF. Recognition of ventricular fibrillation using neural networks. Med Biol Eng Comput 1994;32:217–20.
5. Hedén B, Edenbrandt L, Haisty WK Jr, Pahlm O. Artificial neural networks for the electrocardiographic diagnosis of healed myocardial infarction. Am J Cardiol 1994;74:5–8.
6. Hedén B, Ohlsson M, Edenbrandt L, Rittner R, Pahlm O, Peterson C. Artificial neural networks for recognition of electrocardiographic lead reversal. Am J Cardiol 1995;75:929–33.
7. Lundh B. On the normal scalar ECG. A new classification system considering age, sex and heart position. Acta Med Scand 1984; suppl 691.
8. Rumelhart DE, McClelland JL, editors. Parallel Distributed Processing, Vol. 1 and 2. Cambridge (MA): MIT Press, 1986.
9. Hertz J, Krogh A, Palmer RG. Introduction to the Theory of Neural Computation. Redwood City (CA): Addison-Wesley, 1991.
10. Peterson C, Rögnvaldsson T, Lönnblad L. JETNET 3.0. A versatile artificial neural network package. Comp Phys Commun 1994;81:185–220.
11. Duda RO, Hart PE. Pattern classification and scene analysis. New York: Wiley & Sons, 1973.
12. Richard MD, Lippmann RP. Neural network classifiers estimate a posteriori probabilities. Neural Comput 1991;3:461–83.