Journal of the American College of Cardiology 1996;28:1012-1016
LU TP 95-9

Agreement between Artificial Neural Networks and Human Expert for the Electrocardiographic Diagnosis of Healed Myocardial Infarction

Hedén; Artificial Neural Network and Human Expert

Bo Hedén, MD[1], Mattias Ohlsson, PhD[2], Ralf Rittner, MSc[1], Olle Pahlm, MD, PhD[1], Wesley K. Haisty Jr., MD[3], Carsten Peterson, PhD[2], Lars Edenbrandt , MD, PhD[1]

Departments of [1]Clinical Physiology and [2]Theoretical Physics, Lund University, Lund, Sweden and [3]Department of Medicine/Cardiology, Bowman Gray School of Medicine, Winston-Salem, North Carolina.

Address for correspondence
Lars Edenbrandt
Department of Clinical Physiology
University Hospital

S-221 85 Lund
Sweden

Telephone +46 46 17 33 40
Fax +46 46 15 17 69
E-mail
lars.edenbrandt@klinfys.lu.se

This study was supported by grants from the Swedish Medical Research Council (B95-14X-09893-04B), Swedish National Board for Industrial and Technical Development, the Faculty of Medicine at Lund University, Sweden, the Göran Gustafsson Foundation for Research in National Science and Medicine and the Swedish Natural Science Research Council.

Abstract

Objectives: The purpose of the present study was to compare artificial neural networks and an experienced electrocardiographer diagnosing healed myocardial infarction from the 12-lead electrocardiogram.

Background: Artificial neural networks have proved to be of value in pattern recognition tasks. Studies of their utility in electrocardiographic interpretation have showed performance exceeding that of conventional electrocardiographic interpretation programs. The latter present verbal statements, often with an indication of the likelihood for a certain diagnosis e.g. "possible left ventricular hypertrophy". A neural network presents its output as a numerical value between 0 and 1. The network output values could, however, be interpreted as Bayesian probabilities.

Methods: The study was based on 351 healthy volunteers and 1313 patients who had undergone diagnostic cardiac catheterization. A 12-lead electrocardiogram was recorded on each subject. An expert electrocardiographer classified the electrocardiograms 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 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 was 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 electrocardiograms in the close future.

Key words: Artificial Intelligence, Diagnosis, Computer-Assisted, Electrocardiography, Heart disease

Probability

Condensed Abstract

An artificial neural network that is used as a classifier can, under certain circumstances, indicate a Bayesian probability for a correct classification. The purpose of this study was to compare probability estimates of myocardial infarction from a neural network to those of an experienced electrocardiographer. The study was based on 1664 electrocardiograms from 351 healthy volunteers and 1313 patients with a history of chest pain. The results show that a feed-forward neural network can be trained to diagnose healed anterior myocardial infarction at high levels of specificity and sensitivity, in good agreement with an expert electrocardiographer.

Artificial neural networks are computer based decision tools that have proved to be of particular value in pattern recognition tasks. Their utility has been tested in processing of the electrocardiogram (1-4) and studies concerning detection of myocardial infarction and lead reversal have reported performance exceeding that of conventional rule-based electrocardiographic 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 electrocardiographic interpretation programs. However, neural networks present numerical output values, while conventional electrocardiographic interpretation programs present verbal statements. For some diagnoses the latter also present different levels of likelihood, for example "possible left ventricular hypertrophy" and "probable inferior myocardial infarction". This approach is now widely used and accepted by electrocardiogram readers.

Statements with probability estimates can also be accomplished 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 electrocardiograms as indicative of anterior myocardial infarction or not, outputs values between 0 and 1. Values close to 0 should be assigned by the network to normal electrocardiograms, and values close to 1 to electrocardiograms with clear-cut changes consistent with anterior myocardial infarction, such as a QS pattern in V_2-V_4. Intermediate values should be assigned to electrocardiograms with borderline findings (such as poor R wave progression in anterior leads). Therefore it would be appropriate, also from a theoretical point of view, to introduce several thresholds on the network output, and hence several categories, such as "no infarction", "possible infarction", "probable infarction", and "definite infarction". The purpose of the present study was to transform numerical artificial neural network outputs into verbal statements and to compare these verbal probability estimates to those of an experienced

electrocardiographer. A database of digitized electrocardiograms was therefore analyzed regarding the presence of healed anterior myocardial infarction or not, and electrocardiographic independent methods were used as gold standard.

Methods

Study Population. A total of 1664 subjects were included in the study, 351 healthy volunteers and 1313 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 pathological condition that may influence the electrocardiogram (7). All patients had undergone diagnostic cardiac catheterization at the North Carolina Baptist Hospital, Winston-Salem, North Carolina. Patients with normal coronary arteries, normal contrast left ventriculogram, no evidence of valve dysfunction or congenital heart disease, ejection fraction _ 50 %, and an overall study evaluation of "normal" were classified as "cath normal". Anterior myocardial infarction was defined by presence of _ 75 % diameter stenosis of the left main coronary artery, the left anterior descending artery or its major diagonals, and akinesia or dyskinesia of the anterior-superior wall on 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 on 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 on the left anterior oblique ventriculogram.

Patients with single anterior myocardial infarction and patients with both anterior and inferior myocardial infarction comprised the anterior myocardial infarction group. A control group was composed of the healthy volunteers, cath normals and patients with single inferior or posterolateral myocardial infarction. Patients with technically deficient

electrocardiograms or electrocardiograms with left bundle branch block were excluded. The sizes of the different groups in the study population are presented in Table 1.

Electrocardiographic analysis. A 12-lead electrocardiogram was recorded on each subject using a computerized electrocardiograph. The frequency range was in accordance with AHA specifications (0.05-100 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 electrocardiographic complexes were performed 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 amplitudes, Q and R durations as well as three amplitudes within the ST-T segment. The interval between 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 electrocardiograms, only showing the leads V_1-V_6, were presented in random order to the electrocardiographer. Neither personal data, clinical findings nor the result from the neural networks were available at the classification procedure.

Artificial neural networks. A multilayer 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 1 input layer, 1 hidden layer and 1

output layer. The number of neurons in the input layer equals the number of input variables, i.e. 24 measurements from the leads V_2-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 zero 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. In order to get 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 3-fold cross validation to decide when to terminate learning in order to avoid "overtraining" and 8-fold cross validation to train the networks and assess their performances. The results presented are based on ten independent training/test runs, i.e. the 8-fold cross validation procedure was repeated ten times.

During the training process the connection weights between the neurons were adjusted using the back-propagation 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 using the following equation:

$$\eta = k\,\eta \text{ with} \qquad k = 0.998$$

The momentum α was set to 0.7. Updating occurred after each ten patterns. The 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 electrocardiograms were classified into five groups using the network outputs and four different thresholds between 0 and 1. The thresholds were selected so as to give the same number of electrocardiograms in the classes I to V as was the case for the

electrocardiographer. A complete agreement between the neural network and the electrocardiographer could only be obtained using these thresholds.

Statistical methods. The significance of the difference in performance between the artificial neural networks and the electrocardiographer was tested using the McNemars test with correction for continuity.

Results

The electrocardiographer classified 1291 electrocardiograms as "definitely no anterior myocardial infarction" (1104) or "probably no anterior myocardial infarction" (187). Of these electrocardiograms 1185 were control electrocardiograms resulting in a specificity of 94.8%. A classification as "possible anterior myocardial infarction" (55), "probable anterior myocardial infarction" (73) or "definite anterior myocardial infarction" (245) was assigned to 373 electrocardiograms. 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 ($p<0.001$).

The classifications of the electrocardiograms by the electrocardiographer and the neural network are presented in Table 2. There was agreement in 1282 electrocardiograms (77.0 %), a difference _ 1 class in 1562 electrocardiograms (93.9 %) and a difference _ 2 classes in 1633 (98.1 %). In 31 cases a difference of more than 2 classes was found. The electrocardiographer was correct in nine of these electrocardiograms and the network was correct in 22 of them.

The nine electrocardiograms where the electrocardiographer and the network disagreed by more than two classes, and where the network was incorrect constitutes a particularly interesting group. One of the nine electrocardiograms had serious errors in the data of the measurement program, and was therefore not properly presented to the

networks. It should be noted that erroneous electrocardiograms like this may deteriorate the performance of the artificial neural network both when they appear in the training set and test set. The leads V_2-V_4 of the remaining eight electrocardiograms and the network outputs (means of 10 different runs) are presented in Figures 1-2.

Three cases 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 electrocardiograms 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 it could be used by an electrocardiographic expert. The reversed R-wave progression found in panel C was not a common finding in the material. Therefore, this is a pattern which might be difficult for the network to learn.

Figure 2 presents five cases from the control group which the network falsely classified as definite or probable anterior myocardial infarction. The markedly negative T waves found in three cases were probably important for the network classifications. The electrocardiogram in panel E has a decreasing R wave amplitude from V_2 to 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 electrocardiogram in panel D is as high as 0.76 resulting in a classification as probable anterior myocardial infarction. A network trained without ST-amplitudes correctly classified this case indicating that these variables were important for the high output of the complete network.

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 electrocardiogram with greater accuracy than a conventional interpretation program and an experienced

electrocardiographer. It has also been shown in this study that there is a high level of agreement between the artificial neural network and an electrocardiographer. When there is obvious disagreement the artificial neural network was correct somewhat more often than the expert, with regard to the gold standard of this study material. Most users of black box methods like artificial neural networks are worried that they make obvious and severe misclassifications in some cases even though the overall performance of the method is very good. The worst cases of the 1664 electrocardiograms in this study are presented in Figures 1 and 2. In our opinion, none of these electrocardiograms is a severe misclassification.

Reasons for misclassification. Why are some electrocardiograms misclassified by the artificial neural network and correctly classified by the electrocardiographer? A relatively small number of input variables is 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, ten training examples are needed for each connection weight to be trained. 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 using a database of some 1500 electrocardiograms 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 electrocardiograms with minor morphology deviations, such as notches in the QRS complex.

Another reason for misclassification of the neural networks could be the fact that the networks in this study were only trained to diagnose anterior myocardial infarction. Therefore, some cases with deep inverted T waves, but normal QRS configuration as in Figure 2, are likely to be classified as anterior myocardial infarction. However, taking all precordial leads into account left ventricular hypertrophy with strain is a probable

diagnosis. A neural network, though, could only learn this pattern if a sufficient number of examples of left ventricular hypertrophy is added to the database.

Clinical implications. One advantage with artificial neural networks compared to rule-based criteria is the enhanced diagnostic performance. Another advantage with neural networks is the possibility to easily adjust the network outputs when used in different clinical situations. It has been shown that neural network outputs can be regarded as Bayesian a posteriori probabilities, if the a priori probabilities of the classes in the training database are the same as the a priori probabilities in the test situation. In this study the a priori probabilities were; anterior myocardial infarction 0.25 and non-anterior myocardial infarction 0.75. 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 electrocardiogram with a network output of 0.85, which was interpreted as probable anterior myocardial infarction in this study. If this electrocardiogram 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 electrocardiogram were recorded in a third situation where the a priori probability is high, (0.50), the a posteriori probability would be 0.94. Using the same thresholds for classification of electrocardiograms the resulting statement in the screening situation would be "possible anterior myocardial infarction" and in the high a priori probability situation "definite anterior myocardial infarction". Also an experienced electrocardiographer takes into account the clinical situation in which an electrocardiogram 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 to understand for many electrocardiogram readers. 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 electrocardiogram used as input values, and electrocardiograms with uncommon features. A large number of examples used 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. 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 and P(Ci) is the a priori probability of class Ci. 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 that output values from a multilayer perceptron can be interpreted as Bayesian probabilities (12) P(Cix) provided that: First, the training is accurate. Second, 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. Third, a summed squared error or cross entropy error function is used.

Finally, the a priori class probabilities P(Ci) have to be representatives of actual use or test conditions. It is however possible to vary the class probabilities P(Ci) during classification without retraining the network, since 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;115:843-848.

2 Bortolan G, Willems JL. Diagnostic ECG classification based on neural networks. J Electrocardiol. 1993;26:S75-S79.

3 Edenbrandt L, Devine B, Macfarlane PW. Classification of electrocardiographic ST-T segments - human expert versus artificial neural network Eur Heart J. 1993;14:464-468.

4 Clayton RH, Murray A, Campbell RWF. Recognition of ventricular fibrillation using neural networks. Med Biol Eng Comput. 1994;32:217-220.

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;75:929-933.

7 Lundh B. On the normal scalar ECG. A new classification system considering age, sex and heart position. Acta Med Scand. 1984, suppl No 691.

8 Rumelhart DE, McClelland JL, eds. Parallel distributed processing. Volumes 1 & 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 Comm. 1994;81:185-220.

11 Duda RO, Hart PE. Pattern classification and scene analysis. John Wiley and Sons, New York 1973.

12 Richard MD, Lippmann RP. Neural network classifiers estimate a posteriori probabilities. Neural Comp. 1991;3:461-483.

Figure Legends

Figure 1 Three cases of the anterior myocardial infarction group correctly classified by the electrocardiographer, but falsely classified as class I and II by the neural network. (ANN=artificial neural network)

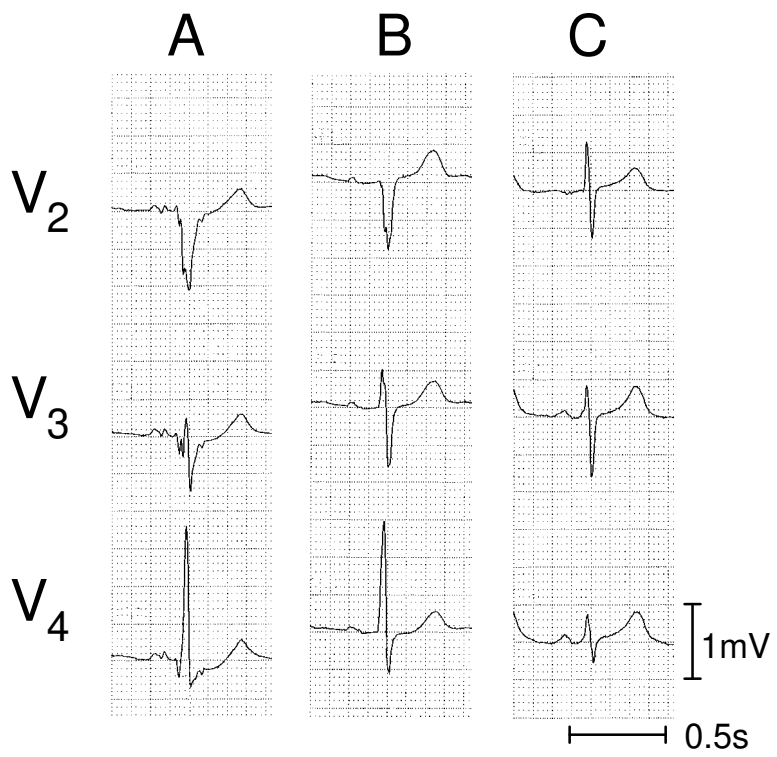
Figure 2 Five examples of disagreement between the neural network and the electrocardiographer where the latter was correct. The electrocardiograms are all from the control group. (ANN=artificial neural network)

Table 1. Study population.

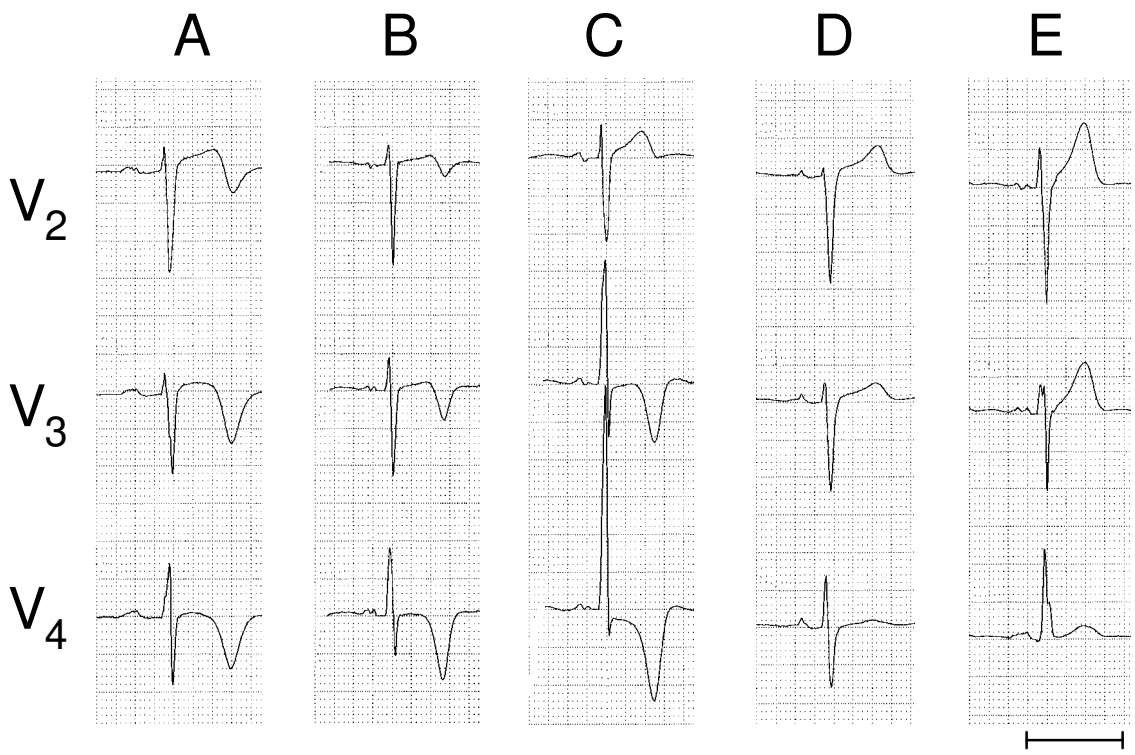
	Number	
Isolated anterior myocardial infarction	272	
Combined anterior + inferior myocardial infarction	142	
Total anterior myocardial infarction group		**414**
Healthy volunteers	351	
Catheterized normal	479	
Isolated inferior myocardial infarction	356	
Isolated posterior myocardial infarction	64	
Total control group		**1250**
Grand total		**1664**

Table 2. Classification table comparing electrocardiographer and artificial neural networks (n=1,664).

		Artificial ı				
	Class	**I**	**II**	**III**	**IV**	**V**
Electrocardio-grapher	**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**



	A	B	C
ANN output	0.35	0.03	0.01
ANN class	II	I	I
expert class	V	V	IV



	A	B	C	D	E
ANN output	0.99	0.97	0.88	0.76	0.68
ANN class	V	V	IV	IV	IV
expert class	II	II	I	I	I