Usefulness of Serial Electrocardiograms for Diagnosis of Acute Myocardial Infarction

Mattias Ohlsson, PhD, Hans Öhlin, MD, PhD, Susanna Maria Wallerstedt, MD, PhD, and Lars Edenbrandt, MD, PhD

The purpose of this study was to determine whether the automated detection of acute myocardial infarction (AMI) by utilizing artificial neural networks was improved by using a previous electrocardiogram (ECG) in addition to the current ECG. A total of 4,691 ECGs were recorded from patients admitted to an emergency department due to suspected AMI. Of these, 902 ECGs, in which diagnoses of AMI were later confirmed, formed the study group, whereas the remaining 3,789 ECGs comprised the control group. For each ECG recorded, a previous ECG of the same patient was selected from the clinical electrocardiographic database. Artificial neural networks were then programmed to detect AMI based on either the current ECG only or on the combination of the previous and the current ECGs. On this basis, 3 assessors—a neural network, an experienced cardiologist, and an intern—separately classified the ECGs of the test group, with and without access to the previous ECG. The detection performance, as measured by the area under the receiver operating characteristic curve, showed an increase for all assessors with access to previous ECGs. The neural network improved from 0.85 to 0.88 (p = 0.02), the cardiologist from 0.79 to 0.81 (p = 0.36), and the intern from 0.71 to 0.78 (p < 0.001). Thus, the performance of a neural network, detecting AMI in an ECG, is improved when a previous ECG is used as an additional input. ©2001 by Excerpta Medica, Inc.

METHODS

Study population: This retrospective study was based on ECGs recorded in patients who presented to an emergency department of a university hospital from January 1990 to June 1997. Each ECG was classified as: (1) if the recorded ECG originated from a patient who was eventually discharged from the coronary care unit with a diagnosis of AMI, this ECG was defined as belonging to the AMI category; and (2) if the recorded ECG originated from a patient eventually discharged with a diagnosis other than AMI, this ECG was defined as belonging to the “control category.”

For each ECG recorded, a previous ECG of the same patient was selected from the clinical electrocardiographic database. ECGs of the AMI category and their predecessors joined the study group and the corresponding pairs of the control category were assigned to the control group.

Several patients contributed >1 pair of ECGs to the study; for example, a patient presenting to the emergency department on 3 different occasions could contribute 2 pairs of ECGs. If AMI was diagnosed only on the last of these 3 occasions, the pair consisting of the second and last ECGs was included as part of the AMI group, whereas the pair consisting of the first and second ECGs was included in the control group. Only ECGs with severe technical deficiencies and pacemaker ECGs were excluded.

During the study period, AMI was diagnosed according to the following criteria, in which ≥2 of the following 3 criteria had to be fulfilled: characteristic
chest pain lasting >20 minutes, elevated creatine kinase levels, or characteristic serial electrocardiographic changes. Creatine kinase-MB values >0.23 µkat/L together with a typical increase and decrease were used as diagnostics for AMI. Electrocardiographic evidence of AMI included new Q waves in ≥2 adjacent leads and/or persistent T inversions in >2 adjacent leads after a newly developed ST elevation in those leads. A senior cardiologist at the Department of Cardiology confirmed each discharge diagnosis.

The material was divided into a training group used for the development of the artificial neural network and a test group for its evaluation. A total of 1,000 cases (200 AMI cases and 800 control cases) were randomly selected for the test group with the remaining cases comprising the training group. After exclusion of ECGs with severe technical deficiencies and pacemaker ECGs, the test group comprised 988 cases (199 AMI cases and 789 control cases) and the training group 3,703 cases (703 AMI and 3,000 control cases). The final AMI group consisted of 902 pairs of ECGs (546 men and 356 women; mean age 70 ± 11 years). The final control group consisted of 3,789 pairs of ECGs (1,954 men and 1,835 women; mean age 70 ± 15 years).

**Electrocardiography:** The 12-lead ECGs were recorded by the use of computerized electrocardiographs (Siemens-Elema AB, Solna, Sweden), and the following 11 measurements taken from each of the 12 leads were selected for further analysis: QRS duration; QRS area; Q, R, and S amplitudes; and 6 ST-T measurements (ST-J amplitude, ST slope, ST amplitude 2/8, ST amplitude 3/8, positive T amplitude, and negative T amplitude). The ST amplitude 2/8 and ST amplitude 3/8 were obtained by dividing the interval between the ST-J point and the end of the T wave into 8 parts of equal duration. The amplitudes at the end of the second and the third intervals were denoted ST amplitude 2/8 and ST amplitude 3/8. A total of 132 measurements from each 12-lead ECG were used. For each measurement of the current ECG, a measurement describing the difference between the corresponding measurements of the current and the previous ECGs was also calculated.

Each case consisted of 2 ECGs, the current and the previous ECG, and from these, 2 different data sets were generated. The first consisted of the 132 measurements of the current ECGs, and the second set of 264 comprised the 132 measurements of the first data set and an additional 132 measurements that described the difference between the current and the previous ECG.

Because there is a high degree of correlation between the measurements from the 12-lead ECGs, the 2 data sets comprising 132 and 264 measurements for each case were reduced to smaller sets of more “effective” variables. This reduction was accomplished using principal component analysis applied to the measurements and grouped into the following 6 sets of measurements: QRS area, QRS duration, QRS amplitudes, ST amplitudes, ST slope, and positive and/or negative T amplitudes. Thus, for example, the principal component analysis reduced the ST slope variables from 12 (1 from each lead) to 2. The first data set was reduced to 16 variables (QRS area 1 variable, QRS duration 1, QRS amplitudes 2, ST amplitudes 5, ST slope 2, and positive and/or negative T amplitudes 5) and the second to 32, which were then used as inputs to the neural networks.

**Artificial neural networks:** A general introduction to the subject of artificial neural networks has been written by Cross et al. In the present study, networks utilizing standard feed-forward, multilayer, perceptron architecture were used. The networks consisted of 1 input layer, 1 hidden layer, and 1 output layer with 16 (first data set) or 32 (second data set) nodes in the input layer, 1 for each of the input variables. The hidden layer contained 10 nodes for the first data set and 15 nodes for the second data set. The output layer consisted of 1 node that encoded whether the patient had AMI or not. A summed-square error function was used together with a Langevin extension of the backpropagation updating rule. Langevin updating consists of adding a random Gaussian component to the weight updates, which has the effect of speeding up the minimization procedure. To avoid over-training, a weight elimination regularization term was used. The regularization parameter $\alpha$ was set using a fivefold cross-validation scheme on the training group, i.e., an average validation error was calculated as the mean of the 5 individual validation errors and the $\alpha$ that corresponded to the smallest average validation error was selected as the optimal $\alpha$. Finally, a committee of 50 networks was trained on the full training group using the obtained regularization parameter $\alpha$. The classification performance of the test group was calculated using the average output of the committee. All calculations were undertaken using the JETNET 3.0 package.

**Physicians’ interpretation:** Two physicians assessed and classified the test ECGs with and without access to the previous ECGs. One of the physicians, a cardiologist and head of the coronary care unit, had >25 years of experience reading ECGs, whereas the other physician, an intern, had only approximately 4 months. All ECGs in the test group were presented in random order to both physicians separately on 2 different occasions. On the first occasion, only the current ECGs were shown, whereas on the second both the current and the previous ECGs were presented. The physicians classified the ECGs (both the single and the pair) into 1 of the following 4 classes: D1: definite AMI; D2: probable AMI; D3: probably no AMI, and D4: definitely no AMI.

Patient data and clinical findings were not made available during both assessments; the only extra information provided during the second assessment was the time difference (in years, months, and days) between the current and the previous ECG.

**Statistical analysis:** The 4-grade scale used by the physicians for classification of the test ECGs made it possible to calculate 3 sensitivity–specificity pairs (D1 vs D2 + D3 + D4; D1 + D2 vs D3 + D4; D1 + D2 + D3 vs D4) for each of the 4 classification schemes.
sessions (2 physicians with and without previous ECG). The 3 pairs were used to construct a receiver operating characteristic (ROC) curve, the area under the curve being used as a measure of the physicians’ performance.

The neural network output values for the test ECGs were in the range from 0 to 1, with a threshold in this interval being used, such that above it all values were regarded as consistent with AMI. For each threshold, a sensitivity–specificity pair was calculated and by varying this threshold between 0 and 1, an ROC curve was obtained; the area under the ROC curve represented the network’s performance.

For each of the physicians and the network, 2 ROC curves were constructed, 1 based on the current ECG only and the other on both the current and the previous ECG. The difference in performance between the 2 classifications was measured as the difference in area under the ROC curves. The significance of this area difference was calculated using a permutation test. The test is performed by repeatedly and randomly permuting the 988 test cases in the 2 lists. For each permutation, the difference of the 2 resulting areas was calculated (test statistic). The evidence against the null hypothesis, of no difference between the 2 original ROC areas, was given by the fraction of area differences of the test statistic larger than the actual difference.

The number of true positive and true negative classifications made by the physicians with the previous ECGs was compared with the corresponding numbers without previous ECGs. The significance of a difference was tested by noting that the same ECGs were used, that is, a McNemar type of statistics was employed.

RESULTS

The neural network performance, measured by the area under the ROC curve, was significantly higher when current and previous ECGs were used as inputs (0.88) rather than when only the current ECG was used as an input (0.85) (p = 0.02); the resulting ROC curves are shown in Figure 1.

The cardiologist did not improve significantly when a previous ECG was also consulted (Figure 1); the areas under the ROC curves were 0.81 with and 0.79 without the previous ECG (p = 0.36).

The intern improved most when able to compare the current ECG with the previous one (Figure 1). The area under the ROC curve based on the analysis of only the current ECG was 0.71, with the corresponding value for the interpretation based on serial analysis as 0.78 (p < 0.001). Although the intern’s performance improved with serial ECG analysis, it still was not as good as the cardiologist’s or the neural network’s analyses.

The intern’s improvement was due to correctly detecting 113 of the 199 AMI cases as “definite” or “probable AMI” when using the previous ECGs. The corresponding value for the analysis without the previous ECGs was 100 (p = 0.059).

In contrast, the cardiologist detected fewer AMI cases using the previous ECG (156) than when no previous ECG was available (166) (p = 0.049). This result in the AMI group was counterbalanced by his
improved ability to correctly classify control cases as “definite no” or “probably no AMI” when using the previous ECGs. He made true negative classifications in 584 control cases with, and in 499 cases without, the previous ECG (p <0.001).

DISCUSSION
Main findings: The results of this study show that access to a previous ECG from the same patient was of value to the artificial neural network detecting AMI from the 12-lead ECG; the improvement was a small but significant increase in area under the ROC curves. The cardiologist’s performance with access to a previous ECG showed a small but not statistically significant increase in ROC area, whereas serial electrocardiographic analysis improved the performance of the intern to a larger extent. Thus, it is not evident that all electrocardiographic readers benefit from serial electrocardiographic analysis. For the trained electrocardiographic interpreter, almost all useful information seems to be obtained from a single electrocardiographic recording at the acute event, whereas a substantial gain from serial electrocardiographic analysis appears to be limited to inexperienced electrocardiographic readers. This conclusion may only be valid for the diagnosis of AMI on the ECG and it is still possible that access to a previous ECG may improve other types of electrocardiographic diagnosis, such as arrhythmias, identification of unstable coronary syndromes, and so forth.

Clinical implications: The findings discussed corroborate earlier reports from our group that neural networks are able to diagnose AMI in the 12-lead ECG at least as well as an experienced physician.