Artificial Intelligence in Medicine (2004) 32, 127–136

ELSEVIER

ARTIFICIAL INTELLIGENCE IN MEDICINE

http://www.intl.elsevierhealth.com/journals/aiim

Detecting acute myocardial infarction in the 12-lead ECG using Hermite expansions and neural networks

Henrik Haraldsson[a], Lars Edenbrandt[b], Mattias Ohlsson[a,*]

[a]*Department of Theoretical Physics, Lund University, Sölvegatan 14A, SE-223 62 Lund, Sweden*
[b]*Department of Clinical Physiology, University Hospital Malmö, SE-205 02 Malmö, Sweden*

Received 16 May 2003; received in revised form 29 December 2003; accepted 8 January 2004

KEYWORDS
Artificial neural networks; Acute myocardial infarction; Computer-assisted diagnosis; Case-based analysis; Hermite expansion; Mean field annealing

Summary We use artificial neural networks (ANNs) to detect signs of acute myocardial infarction (AMI) in ECGs. The 12-lead ECG is decomposed into Hermite basis functions, and the resulting coefficients are used as inputs to the ANNs. Furthermore, we present a case-based method that qualitatively explains the operation of the ANNs, by showing regions of each ECG critical for ANN response. Key ingredients in this method are: (i) a cost function used to find local ECG perturbations leading to the largest possible change in ANN output and (ii) a minimization scheme for this cost function using mean field annealing. Our approach was tested on 2238 ECGs recorded at an emergency department. The obtained ROC areas for ANNs trained with the Hermite representation and standard ECG measurements were 83.4 and 84.3% ($P = 0.4$), respectively. We believe that the proposed method has potential as a decision support system that can provide good advice for diagnosis, as well as providing the physician with insight into the reason underlying the advice.

1. Introduction

Early diagnosis of acute myocardial infarction (AMI) is of vital importance for patients attending the emergency department with chest pain, as there are large benefits for immediate treatment of AMI patients. With appropriate therapy the size of an infarct can be reduced which helps in preserving long-term cardiac function. On the contrary, without proper treatment the result may be severe cardiac damage that significantly reduces the prognosis for the patient. The 12-lead ECG, together with patient history and biochemical markers, are usually used at the emergency department to diagnose AMI. This diagnosis can be difficult, and for short durations of symptoms the biochemical markers may not show any signs at all. For an early diagnosis of AMI one may therefore have to rely on the 12-lead ECG together with patient history. The 12-lead ECG has the advantage of always being available, but the interpretation can be difficult. Computer-based ECG interpretations for the detection of AMI are therefore of importance as they can improve the early diagnosis of AMI.

In healthcare many problems regarding decision making are often difficult and involve complex pattern recognition tasks. This requires expertise, usually in terms of more experienced colleagues, but it could also be in terms of computer-based decision support systems. Artificial neural network

* Corresponding author. Tel.: +46-222-77-82; fax: +46-222-96-86.
E-mail address: mattias@thep.lu.se (M. Ohlsson).

0933-3657/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.artmed.2004.01.003

(ANN) is a technology that has turned out to be useful for many biomedical problems occurring in healthcare [1]. There are several studies where ANNs have been used as a tool for diagnosing AMI. Among the first ones were Baxt [2] and Harrison et al. [3] that used patient history together with ECG data and clinical findings as inputs to the ANNs. A drawback with the ANN of these studies is that some of the input data may be unavailable at the time of initial patient presentation. Later studies [4,5] have developed ANNs with the aim of making a decision support tool that can be used at the emergency department. Using only 12-lead ECG as inputs to the ANNs, Hedén et al. [6] were able to detect AMI with a performance better than an experienced cardiologist (10.5% sensitivity increase at a specificity level of 86.3%, $P < 0.00001$). Another study [7] showed a small, but statistically significant, increase in performance with serial ECG analysis when diagnosing AMI, again only using the 12-lead ECG as input to the ANN. The advantage of using only the 12-lead ECG is the immediate availability and the possibility of an automated interpretation using computer software included with the ECG recorder. An additional advantage will certainly be obtained if the ECG interpretation method can explain the reasoning behind its findings, thereby increasing the support for the physician diagnosing the patient.

Case-based sensitivity analysis of ANNs can be defined as a way of finding the most influential inputs for a given test case and a trained ANN. This differs from the objective of overall sensitivity analysis which aims at finding the most important inputs for a given problem and its corresponding data set. A method following the latter approach was developed by Baxt et al. [8] in connection with the diagnosis of AMI. Their method used a bootstrap technique to estimate input variable effects when diagnosing AMI, which resulted in an impact factor for each input variable. These results may differ from a case-based analysis that aims at finding important inputs for a given test case. The advantage of a case-based approach is that it is possible to indirectly explain how the ANN derived its output. A related approach was developed in [9] that aimed at explaining the reasoning behind the ANN, on a case-based level, by finding similar cases. Similarity was measured using the hidden activation of the ANN.

In this paper we have used Bayesian ANNs as a tool for detecting AMI patients using the 12-lead ECG. Furthermore, to explain the reasoning behind the ANN output, a method was developed that aims at showing regions of the ECG important for this particular case. The key ingredients of our approach are: a representation of the ECG using Hermite functions, and a combinatorial optimization problem formulation for finding important ANN inputs.

Each lead of the ECG is expressed as a series of Hermite basis functions [10,11], and the coefficients in this series are used as inputs for the ANN. After the training, a small number of the coefficients for a given ECG are perturbed within a limited interval; these perturbations are selected so as to maximize the corresponding change in diagnosis (ANN output). A perturbed ECG is reconstructed from the perturbed coefficients, and by comparing this perturbed ECG to the original one some insight can be gained about the reasoning behind this particular ANN output. The possibility of constructing perturbed ECGs from perturbed ANN inputs is one of the virtues of Hermite decomposition; no such reconstruction is possible in the traditional approach where the feature extraction consists of measuring amplitudes and slopes.

The purposes of this study were to:

- Train ANNs to detect AMI using ECGs represented by Hermite expansion coefficients.
- Use a case-based sensitivity analysis to indicate what regions of an ECG that are most important for the ANN in calculating its output.

2. Materials and methods

2.1. Study population

This study was based on 12-lead ECGs recorded at the emergency department of the University Hospital in Lund, Sweden, from July 1990 to June 1997. A total of 1119 suitable ECGs were found to originate from patients that were admitted to the coronary care unit and discharged with the diagnosis AMI. This group of ECGs was defined as the *AMI* group. A random selection of 1119 ECGs among the remaining ones was defined as the *control* group. This group contained ECGs, recorded at the emergency department, from patients with no diagnosis of ''acute myocardial infarction''. Thus, the control group contains patients admitted to the coronary care unit, but discharged without the AMI diagnosis, and patients that were not admitted to the coronary care unit. Patients with pacemaker and ECGs with severe technical deficiencies were excluded from the study.

This study population was also studied in Hedén et al. [6] where a detailed definition of the criteria for diagnosing acute myocardial infarction can be found.

The acute infarction group consisted of 699 ECGs recorded on men and 420 recorded on women. The mean age for this group was 70.8 ± 12.4. The

corresponding numbers for the control group was 578 men and 541 women with a mean age of 63.8 ± 19.0.

The total population of 2238 ECGs was further randomly divided into one training set and one test set of 1499 and 739 ECGs, respectively. The training set was used for all algorithm designs and for training the ANN classifiers. No tuning of parameters was based on the test set, whose only purpose was to assess the performance of the methods.

2.2. Electrocardiography

The 12-lead ECGs were recorded by the use of computerized electrocardiographs (Siemens-Elema AB, Solna, Sweden). The following six standard measurements, taken from each of the 12 leads, were selected for further analysis using neural networks: 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 ST-J point and the end of the T wave into eight 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. These measurements were obtained from the computerized ECG recorders using their measurements program.

To facilitate for later modeling of the ECGs using Hermite functions the signals were re-sampled from the original 500 Hz sampling rate to 1 kHz. Part of the analysis includes using software for QRS detection [12].

2.3. Hermite decomposition of ECGs

In this section we describe the decomposition of the ECGs using Hermite functions. This method was first used by Sörnmo et al. [10] for evaluation of QRS shape features. It has also been used as a method for ECG data compression [11] and as a preprocessing tool for clustering ECG complexes with self-organizing maps [13]. Following these ideas we define two windows in each beat, the QRS complex and the ST segment together with the T wave (ST-T complex). For the current application of detecting AMI, the P wave is not important and was consequently not part of the Hermite modeling. To prepare each window for the Hermite expansion a linear trend subtraction was used such that the start and end of each complex was zero. The subtracted amplitude at the start of the ST-T complex was part of the parameter set used to characterize the beats. Furthermore, each window was centered, in time, around the largest amplitude.

Hermite basis functions (Fig. 1) have the property that an arbitrary signal which is bounded in time can be approximated by a unique sum of such functions. The error in this approximation can be made arbitrarily small by increasing the number of basis functions used in the expansion.

The Hermite basis functions $\phi_n(t,\sigma)$ are given by the following expression:

$$\phi_n(t,\sigma) = \frac{1}{\sqrt{\sigma 2^n n! \sqrt{\pi}}} \mathrm{e}^{-t^2/2\sigma^2} H_n\left(\frac{t}{\sigma}\right), \tag{1}$$

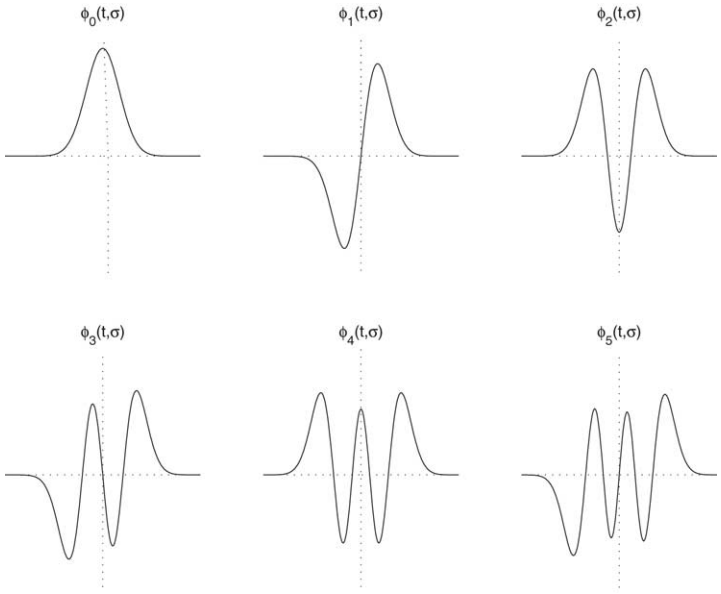

Figure 1 The first six Hermite basis functions (Eqs. (1) and (2)) plotted as a functions of t, for a fixed value of the width σ. The same scale is used in all figures.

where the width σ approximates the half-power duration. $H_n(t/\sigma)$ are the Hermite polynomials, given recursively by $H_0(x) = 1, H_1(x) = 2x$, and

$$H_n(x) = 2xH_{n-1}(x) - 2(n-1)H_{n-2}(x). \quad (2)$$

The QRS and ST-T complexes, generically denoted as $\xi(t)$, can be expressed as a sum of these functions:

$$\xi(t) = \sum_{n=0}^{N-1} c_n(\sigma)\phi_n(t,\sigma) + e(t,\sigma), \quad (3)$$

where the error $e(t,\sigma) \to 0$ as $N \to \infty$. The uniqueness of this expansion is guaranteed by the orthonormality:

$$\sum_{t=-\infty}^{\infty} \phi_m(t,\sigma)\phi_n(t,\sigma) = \delta_{mn}. \quad (4)$$

This allows for an easy evaluation of the coefficients in Eq. (3); multiplying by $\phi_m(t,\sigma)$ and summing over time, one obtains immediately

$$c_n(\sigma) = \sum_{t=-\infty}^{\infty} \phi_n(t,\sigma)\xi(t). \quad (5)$$

The discrepancy between the $\xi(t)$ and its expansion (Eq. (3)) is measured by the summed square error

$$\sum_t |e(t,\sigma)|^2 = \sum_t \left| \xi(t) - \sum_{n=0}^{N-1} c_n(\sigma)\phi_n(t,\sigma) \right|^2. \quad (6)$$

This error depends on the choice of σ. The optimal σ was found by stepwise incrementation of σ in order to find the expansion that resulted in the lowest possible error.

There is still a choice of how many Hermite functions to use in the expansion (Eq. (3)). Adding more functions will decrease the error between the original and the reconstructed ECG, at the expense of having more expansion coefficients. Fig. 2 shows the original V2 lead of an AMI ECG together with the reconstructed one, using 3, 5, 11 and 20 Hermite functions for each complex. The normalized root mean squared (RMS) deviation between the reconstructed and the original ECG leads were 0.19, 0.11, 0.054 and 0.040 for the expansions using 3, 5, 11, 20 basis functions, respectively. The RMS values were averaged over all leads and all ECGs in the full data set.

We decided to use 11 Hermite functions for both the QRS and the ST-T complex. This generally allowed for reconstructed leads that visibly were very similar to the original ones. However, not all of the Hermite components were actually used as feature variables in the classification. The reason for using more Hermite functions than actually needed for the classification of the ECG is that ECGs will be reconstructed from the Hermite coefficients c_n in the case-based sensitivity analysis (cf. Section 2.5). The selection of components used

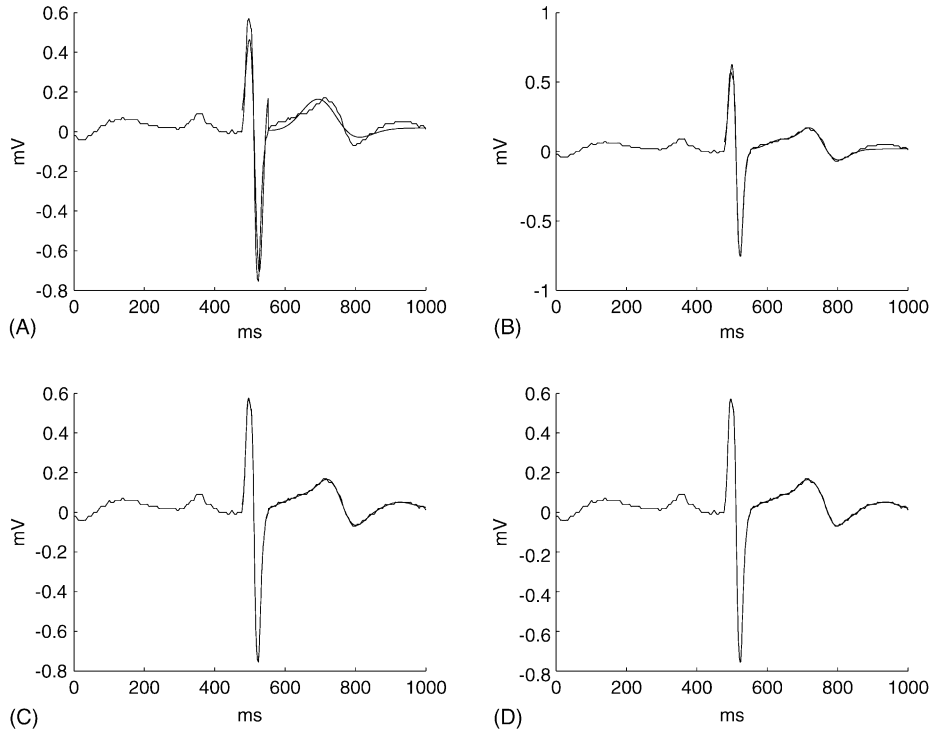

Figure 2 Reconstruction of lead V2 (thin line) for an AMI ECG. Graphs A—D shows the reconstruction (thick line) using 3, 5, 11 and 20 Hermite functions, respectively.

in the classification of the ECGs were determined using cross-validation on the training set.

2.4. Artificial neural networks

In this study we have used Bayesian inference techniques for ANN learning. Bayesian training of ANNs consists of picking a number of independent ANNs, each defined by its set of weights $\mathbf{w}$, from a probability distribution. This posterior distribution $p(\mathbf{w}|D)$, which is a function of the weights $\mathbf{w}$, represents the degree of belief for the different ANNs using the observed patient data D and is constructed using Bayes' theorem. To make a prediction for a test case one has to, theoretically, *integrate* over a multi-dimensional weight space. Practically, one approximates the integration with a finite sum, i.e. the average output $\langle y(\mathbf{x}_{\text{test}})\rangle$ for a test case $\mathbf{x}_{\text{test}}$, is computed as

$$\langle y(\mathbf{x}_{\text{test}})\rangle = \frac{1}{L}\sum_{i=1}^{L} y(\mathbf{w}_i, \mathbf{x}_{\text{test}}), \tag{7}$$

where $y(\mathbf{w}_i, \mathbf{x}_{\text{test}})$ is the output of the ANN defined by the set of weights $\mathbf{w}_i$ that is drawn from the probability distribution $p(\mathbf{w}|D)$ using Markov Chain Monte Carlo methods [14]. The training of the ANNs was accomplished using the software package of Neal [15,16].

The ANNs consisted of one input layer, one hidden layer and one output layer. The number of input nodes ranged from 48 to 108 depending on the set of input variables used. The hidden layer contained eight nodes and the output layer consisted of one node that encoded whether the ECG originated from a patient that suffered from AMI or not. The number of hidden nodes was selected using three-fold cross-validation on the training set.

When training Bayesian ANNs some parameters have to be set, including the number of networks L to average over and hyper-parameters to use in the posterior distribution. This model selection was also accomplished using three-fold cross-validation on the training set. The result of the Bayesian training procedure is L ANNs, whose outputs are averaged according to Eq. (7) to obtain the final output. The averaged output of these L networks is, throughout the paper, referred to as an ANN classifier.

2.4.1. Measuring the performance

The performance of the ANN classifiers was assessed using the test set comprising 739 randomly selected patients from the total population of 2238 cases. This test set was not part in any algorithm design or model selection. For each test case in the test set the ANN classifier presents an output value between 0 and 1. A threshold in this interval was used above which all values were regarded as consistent with AMI. By varying this threshold, a receiver operating characteristic (ROC) curve was obtained. The performance of each ANN classifier was measured as the area under the ROC curve. The 95% confidence limits of the area were estimated by a bootstrap technique [17].

The difference in performance between two ANN classifiers was measured as the difference in area under the ROC curves. The statistical significance of such an observed area difference was assessed by means of a permutation test [17] as follows:

A new classification list was created by randomly selecting for each of the 739 test cases either the classification made with the first ANN or the classification made with the other ANN. A second list was created from the classification not included in the first one. The two lists were used to construct two ROC curves, and the areas under the curves were calculated, as was the area difference (test statistic). The procedure was repeated 50,000 times. The relative frequency of area differences that had an absolute value greater than the actual difference was taken as the probability of obtaining at least the actual area difference if no true difference existed.

2.5. Case-based sensitivity analysis

In a case-based analysis we are interested in a single ECG and the diagnosis it received using a trained ANN classifier. By case-based *sensitivity* analysis for ECGs we aim at finding regions of a 12-lead ECG that are critical to the output of the trained ANN classifier. There are situations where such a case-based analysis of ECGs may be of interest, namely:

- Explaining the functioning of the ANN classifier by showing regions of the ECG critical to the ANN output.
- Understanding misclassified ECGs by ''reverse engineering''.
- Feature selection.

The approach taken in this paper for the sensitivity analysis is based on the notion of *causal importance*, since the method will monitor the ANN output response when manipulating the ANN inputs. This is different from the notion of *predictive importance* which aims at monitoring the generalization performance when deleting ANN inputs, where the latter often requires retraining of the ANN classifier.

The proposed method will perturb a small number, N_{mod}, of ANN inputs. The size of each

perturbation, measured in units of standard deviation over the ANN training set, is constrained not to be larger than some constant. The aim is to select the perturbation, following these constraints, that maximizes the change in ANN output.

To see why this is useful, consider the case when the ANN classifier has diagnosed a certain ECG as showing signs of myocardial infarction. By producing a perturbed ECG which the ANN classifier classifies as healthy, and comparing this perturbed ECG to the original one, one may reveal information about what parts of the ECG that the ANN classifier found critical when computing the output. Notice that this reconstruction of perturbed ECGs from perturbed ANN inputs is made possible by the fact that the Hermite decomposition is invertible; no such reconstruction is possible in the traditional approach where the feature extraction consists of measuring amplitudes and slopes.

To proceed in finding the optimal perturbation, we discretize the perturbation into M steps. M should be large enough that the output varies smoothly as the inputs move from one perturbation to the next. For the current application, $M = 10$ was used.

Let ϵ_{il} denote the set of possible perturbations of input i:

$$\epsilon_{il} = \left(-\frac{M}{2} + l\right)\frac{\sigma_i}{M} \quad (l = 0, \ldots, M), \tag{8}$$

where σ_i is the standard deviation of input i. Note that the extremal perturbations are $\epsilon_{i0} = -\sigma_i/2$ and $\epsilon_{iM} = \sigma_i/2$, and that $\epsilon_{i(M/2)} = 0$ which means that $l = M/2$ corresponds to no perturbation.

The actual perturbation dx_i of input i is expressed as

$$dx_i = \sum_{l=0}^{M} s_{il}\epsilon_{il}, \tag{9}$$

where s_{il} are binary decision variables defined such that only the one corresponding to the chosen perturbation is turned on:

$$s_{il} = \begin{cases} 1 & \text{if } x_i \to x_i + \epsilon_{il}, \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

Obviously, the variables must fulfill the *Potts* normalization condition:

$$\sum_{l=0}^{M} s_{il} = 1 \quad (i = 1, \ldots, N). \tag{11}$$

By this construction, the input vector $\mathbf{x}$ can be changed into one out of possible $(M+1)^N$ new input vectors.

This is a NP-hard combinatorial optimization problem, which we address by heuristic methods. An energy (or cost) function is defined as

$$E(\{s_{il}\}) = \pm y(\mathbf{x} + d\mathbf{x}(\{s_{il}\})) + \alpha\left(\sum_{i=1}^{N} s_{i(M/2)} - (N - N_{\text{mod}})\right)^2. \tag{12}$$

The first term serves to maximize the decrease of the output $y(\mathbf{x})$ (or increase, in the case of the minus sign). The second term favors solutions where N_{mod} inputs are perturbed; α is a tunable parameter.

To find a good minimum in this energy landscape, i.e. to find a good solution, we change the binary decision variables s_{il} into fuzzy, real-valued decision variables, and perform the optimization by means of mean field annealing [18]. The result of the minimization of Eq. (12) is a set of binary variables that defines the optimal perturbation, i.e. the perturbation of the original input that will have the largest effect on the output of the ANN classifier. This modified input also corresponds to a new ECG by means of the Hermite expansion.

Apart from the requirement that exactly N_{mod} inputs be perturbed, we imposed the additional constraint that the perturbed inputs must all come from two leads. This constraint is optional; gathering all the perturbations into two leads has the advantage of making these perturbations easier to see, and of increasing the detail in the perturbation of the beat. To impose this constraint, the energy function (Eq. (12)) is modified by adding the term

$$\beta\left(\Theta - \sum_{e=1}^{12}\theta(n_e)\right)^2, \tag{13}$$

where β is another tunable parameter and n_e the number of modified inputs for lead e:

$$n_e = \frac{N}{12} - \sum_{i \in \text{lead } e} s_{i(M/2)}. \tag{14}$$

Recall that $N/12$ is the number of inputs per lead. $\theta(n_e)$ is a *soft* measure of whether lead e is perturbed or not, defined as

$$\theta(n_e) = \frac{1}{1 + \exp(-2(n_e - 0.5))}. \tag{15}$$

The reason for using this soft sigmoid rather than a step function with threshold 0.5, comes from the mean field annealing procedure that uses fuzzy decision variables. Θ is the value of $\sum_{e=1}^{12}\theta(n_e)$ when the desired solution is reached, $\Theta = 2\theta(N_{\text{mod}}/2) + 10\theta(0)$, for the 2-lead constraint.

Table 1 Number of input variables used, for each lead, when training different ANN classifiers

Input variables	QRS complex	ST-T complex
	Hermite expansion ANNs	
Hermite components	2(c_0, c_1)	5 (c_0, c_1, c_2, c_3, c_4)
Amplitude shift	–	1
QRS duration	1	–
	ST-T amplitude ANNs	
Amplitudes	–	5 (ST-J amp, ST-2/8 amp, ST-3/8 amp, T_+ amp and T_- amp)
Slope	–	1 (ST slope)

3. Results

3.1. Selecting input variables

The QRS and ST-T complexes were decomposed using 11 Hermite components each. Of these, only a few were used when training the ANN classifiers. The selection of inputs to the ANN was performed using three-fold cross-validation on the training set which resulted in the following set: $\{c_0, c_1\}$ for the QRS beat and $\{c_0, c_1, c_2, c_3, c_4\}$ for the ST-T beat (see Table 1). The amplitude shift defined in the preprocessing step prior to the Hermite expansion and the QRS duration were also part of the input variable set. The variables not used in the training were, however, stored in order to be used when reconstructing the modified ECGs.

To compare with the results found by Hedén et al. [6] we have also used amplitude and slope measurements from the ST-T complex (see Table 1).

3.2. Comparing ROC area performance

Table 2 shows the performance of the ANN classifiers, in terms of the area under the ROC curve. We present results using all 12 leads and results using a subset of eight leads only. This reduction is supported by the fact that the extremity leads III, aVF, aVL and aVR are constructed using leads I and II, e.g. lead III = II − I.

Table 2 The results of the ANN classifiers using different sets of inputs

Input data set	ROC area (%)
Hermite expansion (12 leads)	83.4 (80.6, 86.3)
ST-T amplitudes (12 leads)	84.3 (81.6, 87.1)
Hermite expansion (8 leads)	82.7 (79.8, 85.6)
ST-T amplitudes (8 leads)	82.2 (79.4, 85.3)

The 95% confidence intervals, given in parentheses, are estimated using a bootstrap technique. For each input set 10 independent ANN classifiers were trained. The median ROC area and its corresponding confidence interval are presented in this table.

When using all 12 leads, a marginally larger ROC area (84.3%) is obtained for the representation using ST-T amplitudes compared to the Hermite expansion representation (83.4%). This difference is not statistically significant ($P = 0.4$). Hedén et al. [6] reported an ROC area of 86% using the ST-T amplitudes as inputs. The data set used in their study contained a larger group of normals (10,452 ECGs), which may explain their somewhat better performance.

Using only eight leads (V1-V6, I and II) as inputs to the ANN classifiers, the performance is almost the same for the Hermite and the ST-T representation, 82.7 and 82.2% respectively ($P = 0.8$). For ANNs trained with ST-T amplitudes, the ROC area is significantly larger when using 12 leads than when using eight leads ($P = 0.04$). Looking at a particular point of the ROC curve we obtain sensitivity values of 56.3 and 63.3%, at a specificity level of 85.0%, for ANN classifiers using the ST-T representation with 8 and 12 leads, respectively. The corresponding values for the Hermite classifiers were 59.3% for 8 leads and 61.5% for 12 leads.

One can conclude that representing the ECGs in terms of Hermite expansion coefficients together with preprocessing parameters (QRS duration and the amplitude shift) is almost as good the ST-T amplitudes, when detecting AMI using ANN classifiers. The former representation, however, has the advantage of being invertible, i.e. one can recover the complete ECG signal from it. This is further pursued in the case-based sensitivity analysis below.

3.3. Sensitivity analysis

Here we present some results using the case-based sensitivity method described in Section 2. Our optimization procedure perturbs a small number of the

Hermite components fed to the ANN, such that the change of the ANN response is as large as possible. These modified Hermite coefficients, along with the unperturbed components, are then used to reconstruct a modified ECG. This modified ECG is displayed along with the original ECG reconstructed from the unperturbed Hermite coefficients. The modifications are clearly visible, and indicate the ECG regions critical for ANN performance.

The ANNs trained with 12 leads were used throughout; 6 inputs from 2 leads were allowed to be modified by up to half a standard deviation each (cf. Section 2.5). The sign in Eq. (12) was selected so as to *change* the diagnosis; ECGs that gave an ANN output larger than 0.5 were modified so as to decrease the output, and ECGs that gave an output smaller than 0.5 were modified so as to increase the output. Only the five Hermite coefficients corresponding to the ST-T complex, and the amplitude shift, were allowed to be modified. The QRS durations were kept constant as changing them would compress or expand the entire beat in time in an curious manner. The two Hermite coefficients corresponding to the QRS complex were also kept fixed as they are less important for the diagnosis of AMI; we chose to focus on the ST-T complex in this analysis.

Fig. 3 shows an AMI ECG from the test set that was correctly classified by the ANN. The original ANN output was 0.94; the ANN output when fed with the optimally perturbed coefficients was 0.76.

The interpretation is as follows: reducing the ST-depression of the ST complex in the V2 and V3 leads produces an ECG that is, according to the ANN, more ''healthy'' i.e. showing less signs of AMI. These are thus the regions that were, according to our automatic sensitivity analysis, most important for the ANN in making its decision. This corresponds very well to what an experienced ECG reader would select as being the important part of the ECG for a diagnosis of AMI.

Another example can be found in Fig. 4 which shows another AMI ECG correctly classified by the

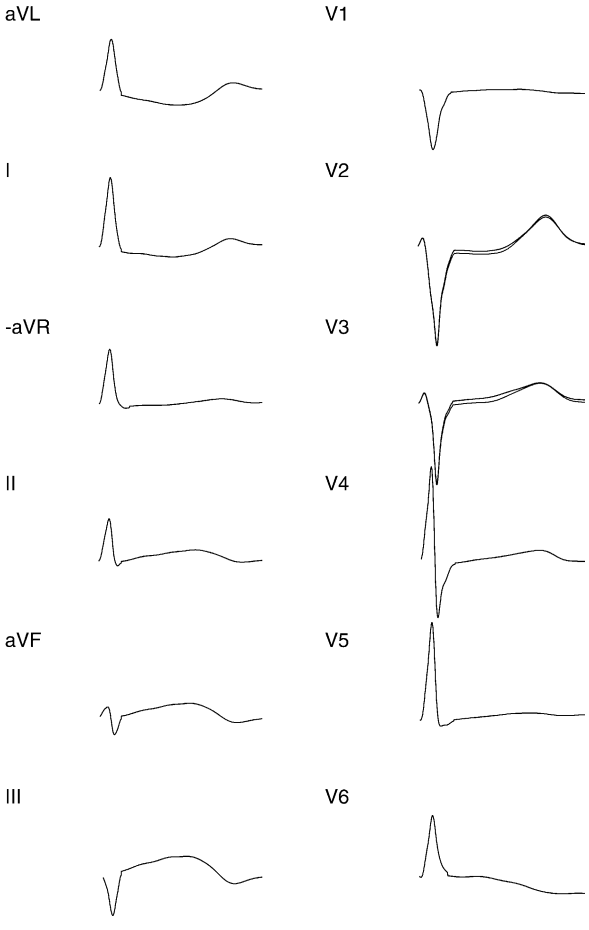

Figure 3 An ECG from the AMI test set. The thick line corresponds the modified ECG and the thin line is the ECG seen by the ANN classifier. A difference can be seen in the ST-T complex of leads V2 and V3, which are the regions critical for the ANN classification.

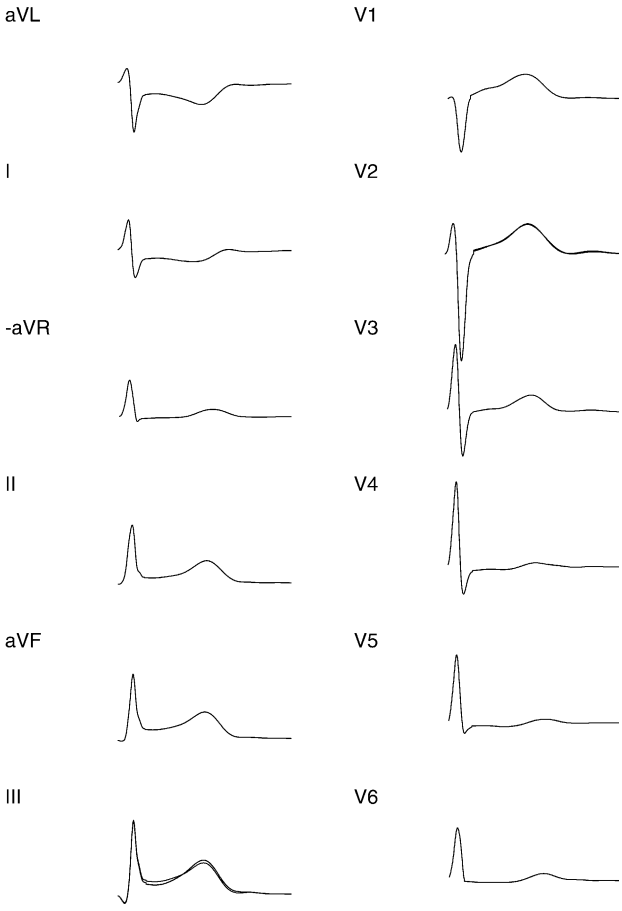

Figure 4 An ECG from the AMI test set. The thick line corresponds the modified ECG and the thin line is the ECG seen by the ANN classifier. A noticeable difference can be seen in lead III.

ANN. For this ECG the case-based sensitivity method resulted in a modified ECG with a noticeable difference in lead III. (There is also a small, hardly visible, change in lead V2.) The difference corresponds to a lowering of the ST elevation in lead III, thus making the signs of AMI less pronounced. (The ANN output was lowered from 0.93 to 0.84.) This is again in accordance with what a human expert would say.

To what extent do the regions found by the case-based sensitivity method correspond to regions that an experienced ECG reader would classify as important? To answer that question an experienced ECG reader examined all AMI cases from the test set. The ECG reader was instructed to select the two leads where the most important signs of AMI was found. The ECG reader managed this task for 146 of the 371 AMI cases in the test set; the others were dismissed as inconclusive.

These selected leads were then compared to the two leads that the case-based sensitivity method found for the same set of test ECGs. In 64 of these 146 cases (46%) at least one of the leads agreed. The experienced ECG reader most often selected leads V2, V3, III and aVF, whereas our method focused on all pre-cordial leads together with lead III.

4. Discussion

4.1. Main findings

The results show that the 12-lead ECG can be represented using Hermite polynomials when diagnosing AMI using ANN classifiers. The performance was slightly better using traditional amplitude (and slope) measurements from the ST-T complex (84.3% compared to 83.4%, $P = 0.4$). The performance using Hermite polynomials was also in accordance with previous findings [6].

An important aspect of ANN learning is to be able to explain the way ANNs work. In this paper we have developed an approach that implicitly tries to explain the functioning of the ANN by finding small perturbations of the inputs that correspond to large changes in ANN output. This together with a Hermite representation of the leads results in a modified ECG that can be visualized. Differences between the original and the modified ECG represent critical regions for the ANN classifier.

The comparison between the experienced ECG reader and the case-based sensitivity method showed only a limited agreement (46% where at least one of the two leads agreed). This may seem like a large discrepancy, but bear in mind that the leads are correlated, the AMI information is not resident in only one or two of the directly measured leads, and there is additional redundancy in the ECG as four of the leads are constructed as linear combinations of leads I and II. Therefore, one can never hope to obtain an exact or near-exact match.

The leads focused on by the human reader depends to some extent on his or her habits; the physician is used to focus on leads V2 and V3, but in some cases there may be more information in the other pre-cordial leads, picked up by the ANN but not easily discernible for the human eye. A more systematic investigation with at least two experienced ECG readers is needed in order find intra-observer and inter-observer variabilities among them. As for the ANNs, there is a semi-degeneracy in the optimal perturbations: there is one perturbation found to give the largest change in ANN output, but as the network uses information from all 12 leads there are often other perturbations involving other leads that also give a fairly large change in output. What's more, there is nothing magic about our selection of the number of ANN inputs to vary or the maximum perturbation of each input, and changing these parameters sometimes results in different leads being selected. In addition, the minimization of the cost function (Eq. (12)) is a difficult problem, and one may not always find a good solution; further developments with respect to the construction and minimization of this cost function may therefore be needed.

Further refinements of the method are needed, but we believe that what we have already looks promising and much more attractive than the ''black box'' operation often characteristic of ANNs and other data mining tools.

4.2. Clinical implications

Previous studies (e.g. [6,19,20]) indicate that ANN can be used to improve automated ECG interpretation for AMI. Even experienced cardiologists can benefit from using ANNs as decision support when diagnosing AMI [6]. This improvement could lead to more accurate early diagnosis of AMI for patients attending the emergency department. The lack of an explanation for each diagnosis made by the ANN classifier has often been regarded as a drawback. The proposed case-based sensitivity method together with the Hermite representation of the 12-lead ECG is an approach to eliminate that drawback. This method could indicate ECG leads, and part of these leads, important for a particular diagnosis. For the treating physician such an explanation could improve his/her confidence in the decision support system. Also, and equally important, it could help the physician to reject a diagnosis that is incorrect.

In an emergency department a decision support system for the interpretation of ECGs has the advantage of being fully automated and the proposed method for explanation support is not a limitation in that respect. A disadvantage of this type of decision support lies in the fact that the ECG is only used in a limited part of the diagnostic decision, regarding AMI. In a recent retrospective study, based on a study population of 2204 patients, by Baxt et al. [5] an ANN system was constructed for the diagnosis of AMI. Their system used 40 variables from patient histories, physical examination, ECG recordings and chemical cardiac markers as inputs to the ANN. The results found in this study were excellent, despite the fact that the patient material contained missing data. The accuracy on an independent test set was 94.0% for the AMI cases and 93.3% for patients without AMI. This system is, however, not fully automated and requires pre-classifications made by physicians for some of the input variables.

4.3. Conclusions

We have presented a method for automated detection of AMI patients using the 12-lead ECG. Each lead was decomposed using Hermite basis functions and the resulting coefficients were used as inputs to a Bayesian ANN classifier that were trained to detect AMI. The performance was compared to that of using traditional ST-T amplitudes and slopes on a study population of 2238 ECGs. No significant difference was found in favor for the amplitude/slope measurements. The area under the ROC curve was 83.4% for the Hermite representation using 12 leads.

Furthermore, we have also presented a novel method for case-based sensitivity analysis that was used to show regions of the ECG critical for the ANN response. This method was based on: (i) a cost function used to find local ECG perturbations leading to the largest possible change in ANN output and (ii) a minimization scheme for this cost function using mean field annealing.

With proper further developments, we believe the proposed method has potential as a decision support system that can provide a good suggestion for diagnosis, as well as providing the physician with insight into the reason underlying the advice.

Acknowledgements

We thank Carsten Peterson for fruitful discussions. This work was in part supported by the Swedish Natural Science Research Council and the Swedish Foundation for Strategic Research.

References

[1] Lisboa P, Ifeachor E, Szczepaniak P, editors. Artificial neural networks in biomedicine. London: Springer-Verlag, 2000.
[2] Baxt W. Use of an artificial neural network for the diagnosis of myocardial infarction. Ann Int Med 1991;115:843—8.
[3] Harrison R, Marshall S, Kennedy R. The early diagnosis of heart attacks: a neurocomputational approach. In: Proceedings of the International Joint Conference on Neural Networks, Seattle, WA, 1991. p. 1—5.
[4] Kennedy R, Harrison R, Burtonc A, Fraser H, Hamer W, MacArthur D, et al. An artificial neural network system for diagnosis of acute myocardial infarction (AMI) in the accident and emergency department: evaluation and comparison with serum myoglobin measurements. Comput Meth Programs Biomed 1997;52:93—103.
[5] Baxt W, Shofer F, Sites F, Hollander J. A neural computational aid to the diagnosis of acute myocardial infarction. Ann Emerg Med 2002;34:366—73.
[6] Hedén B, Öhlin H, Rittner R, Edenbrandt L. Acute myocardial infarction detected in the 12-Lead ECG by artificial neural networks. Circulation 1997;96(6):1798—802.
[7] Ohlsson M, Öhlin H, Wallerstedt S, Edenbrandt L. Usefulness of serial electrocardiograms for diagnosis of acute myocardial infarction. Am J Cardiol 2001;88:478—81.
[8] Baxt WG, White H. Bootstrapping confidence intervals for clinical input variable effects in a network trained to identify the presence of acute myocardial infarction. Neural Comput 1995;4:772—80.
[9] Caruana R, Case-based explanation for artificial neural nets. In: Malmgren H, Borga M, Niklasson L, editors. Artificial neural networks in medicine and biology. London: Springer-Verlag, 2000. p. 303—8.
[10] Sörnmo L, Börjesson P, Nygårds M, Pahlm O. A method for evaluation of QRS shape features using a mathematical model for the ECG. IEEE Trans Biomed Eng 1981;28:713—7.
[11] Jane R, Olmos S, Laguna P, Caminal P. Adaptive Hermite models for ECG data compression: performance and evaluation with automatic wave detection. In: Proceedings of the Computers in Cardiology. IEEE Computer Society, 1993. p. 389—92.
[12] Nygårds M, Sörnmo L. Delineation of the QRS complex using the envelope of the ECG. Med Biol Eng Comput 1983;21:538—47.
[13] Lagerholm M, Peterson C, Braccini G, Edenbrandt L, Sörnmo L. Clustering ECG complexes using Hermite functions and self-organizing maps. IEEE Trans Biomed Eng 2000;47:838—48.
[14] Neal RM, Bayesian training of backpropagation networks by the hybrid Monte Carlo method, Technical Report CRG-TR-92-1. Department of Computer Science, University of Toronto, Canada, 1992.
[15] Neal RM. Software for flexible Bayesian modeling, 1999. http://www.cs.toronto.edu/~radford/fbm.1999-03-13.doc/index.html.
[16] Neal RM. Bayesian learning for neural networks. New York: Springer-Verlag, 1996.
[17] Wehrens R, Putter H, Buydens L. The bootstrap: A tutorial. Chemometr Intell Lab Syst 2000;54:35—52.
[18] Peterson C, Söderberg B. A new method for mapping optimization problems onto neural networks. Int J Neural Syst 1989;1:3—22.
[19] Bortolan G, Willems JL. Diagnostic ECG classification based on neural networks. J Electrocardiol 1993;26(Suppl 1):75—9.
[20] Olsson S, Ohlsson M, Öhlin H, Edenbrandt L. Neural networks—A diagnostic tool in acute myocardial infarction with concomitant left bundle branch block. Clin Physiol Funct Imaging 2002;22:295—9.