Detection of acute coronary syndromes in chest pain patients using neural network ensembles

Green, Michael; Björk, Jonas; Lundager Hansen, Jakob; Ekelund, Ulf; Edenbrandt, Lars; Ohlsson, Mattias

Published in:
Second International Conference on Computational Intelligence in Medicine and Healthcare

2005

DETECTION OF ACUTE CORONARY SYNDROMES IN CHEST PAIN PATIENTS USING NEURAL NETWORK ENSEMBLES

M. Green*, J. Björk†, J. Hansen†, U. Ekelund†, L. Edenbrandt‡, M. Ohlsson*
*Lund University, Sweden
†Lund University Hospital, Sweden
‡Malmö University Hospital, Sweden

Abstract— **Patients with suspicion of acute coronary syndrome (ACS) are difficult to diagnose and they belong to a very heterogenous group of patients. Some require immediate treatment while others, with only minor disorders, may be sent home. Detecting ACS patients using a machine learning approach would be advantageous in many situations.**

This study is based on patients with chest pain attending the emergency department of Lund University Hospital. A total of 915 cases were incorporated of which 190 were diagnosed as ACS and 725 as non ACS. We have developed classifiers using neural network ensembles that can provide a prediction of ACS for patients with chest pain at an emergency department. We compared two different ensemble strategies, Bagging and K-fold cross splitting. The obtained results were also compared with the results of a standard multiple logistic regression model.

Our results show that it is possible to construct a machine learning tool that can predict the presence of ACS among patients with chest pain at a ROC area of 77.8%, corresponding to a level of 40% specificity and 95% sensitivity.

Keywords: acute coronary syndrome, neural networks, logistic regression, ensemble methods.

INTRODUCTION

Patients who come to emergency departments with suspicion of infarction or unstable angina pectoris (acute coronary syndrome (ACS)) are common. This group of patients represent a very heterogeneous group, some have a severe ischemic heart disease that, without treatment, may lead to serious complications. Others have only minor disorders and may, without any risk, be sent home. It is important to be able to identify patients at the emergency department that have a need for immediate treatment.

A number of methods have been developed to support the physicians in their decision making regarding patients presenting the emergency department with chest pain. Goldman et al. [1] developed a model to estimate the relative risk of major cardiac events within 72 hours after arrival at the emergency department. The independent variables used included age, gender and electrocardiographic (ECG) findings, all available at presentation. This model have the potential to be used at emergency departments to decide a suitable level of patient treatment, ultimately being able to reduce the number of intensive care admissions.

There are also a number of approaches that have been developed to predict the presence of acute myocardial infarction based on a full range of clinical data Baxt [2], Baxt et al. [3], Hollander et al. [4] and data limited to the 12-lead ECG only Hedén et al. [5], Ohlsson et al. [6]. Many of these methods used artificial neural networks as the classification tool. The performance is usually good compared to interpretation made by experienced physicians. An example of ACS prediction can be found in Xuea et al. [7] where a hybrid machine learning approach was used, combining neural networks and decision trees.

Artificial neural networks (ANN) represents one machine learning tool that have turned out to be useful for complex pattern recognition problems. ANN is also widely used for medical applications (see e.g. Lisboa et al. [8]). Ensemble learning for ANN is standard procedure to increase the generalization performance by combining several individual networks trained on the same task. The ensemble approach has been justified both theoretically Hansen and Salamon [9], Krogh and Vedelsby [10] and empirically Opitz and Maclin [11]. Combining the outputs is clearly only relevant when they disagree on some or several of the inputs. This insight was formalized by Krogh et al. [10], who showed that the squared error of the ensemble when predicting a single target is equal to the average squared error of the individual networks, minus the diversity defined as the variance of the individual network outputs. Thus, to reduce the ensemble error, one tries to increase the diversity (called “ambiguity” in [10]) without increasing the individual network errors too much. The simplest method for creating diverse ensemble members is to train each network using randomly initialized weights. A more elaborate approach is to train the different networks on different subsets of the training set. An example is Bagging Breiman [12] where each training set is created by resampling (with replacement) the original one, with uniform probability.

In this study we will look into the idea of using **K-fold cross splitting** when training the ANN ensemble.

This approach is similar to the K-fold cross validation method that can be used to estimate the generalization performance. We will compare this approach with ANN Bagging and standard multiple logistic regression. Our main focus is to develop a classification method that can predict ACS as good as possible on our study population.

MATERIALS AND METHODS

Study Population

This study is based on patients with chest pain attending the emergency department of Lund University Hospital, Sweden, from July 1 to November 20 1997. One thousand consecutive visits were recorded where 813 patients had only one visit and 68 patients had two or more visits Hansen et al. [13]. For patients with two consecutive visits within 20 days one of these visits were removed (randomly), in order to have as independent data as possible. This reduced the dataset to 942 visits. Furthermore, 27 visits where removed because of missing values. The final dataset consisted of 915 cases, where 190 were diagnosed with ACS and 725 cases with no ACS. Table I shows the independent variables used in this study. The diagnosis of ACS is defined as one of the following discharge diagnoses for the patient: acute myocardial infarction and angina pectoris. The discharge diagnoses were made by the attending senior ward physicians and also reviewed by an experienced research nurse and classified according to the following criteria: acute myocardial infarction was defined by the WHO criteria Tunstall-Pedoe et al. [14] where the biochemical criterion was at least one measurement of CK-MB$>$10 μg/l or Troponin T$>$0.1 μg/l. The criteria for unstable angina were (i) together with (ii) or (iii):

(i) Ischemic symptoms: chest pain $>$15 min., syncope, acute heart failure or pulmonary oedema
(ii) Electrocardiogram (ECG) changes: transient or persisting ST segment depression ($\geq$1 mm) and/or T-wave inversion ($\geq$1 mm) without developing Q waves or loss of R wave height.
(iii) Biochemical markers: CK-MB 5-10 μg/l or Troponin T$>$0.05 μg/l if CK-MB$<$10 μg/l.

The non ACS cases consists of patients with the diagnosis of stable and suspected angina pectoris, together with the category "other diagnosis". Out of the 725 non ACS cases, 380 correspond to discharge diagnoses other than stable or suspected angina pectoris.

TABLE I
CHARACTERISTICS OF THE INDEPENDENT VARIABLES USED TO TRAIN THE CLASSIFICATION METHODS. THERE ARE 190 CASES OF ACS AND 725 CASES WITHOUT ACS.

Input variable	ACS n (%)	No ACS n (%)
Age	69.8*(12.9)†	60.9*(17.9)†
Gender		
Male	120 (63)	392 (54)
Female	70 (37)	333 (46)
Symptom duration		
0-6 hours	145 (76.3)	381 (52.6)
7-12 hours	21 (11.1)	94 (13.0)
13-24 hours	10 (5.3)	60 (8.3)
$>$ 24 hours	14 (7.4)	190 (26.2)
Chest discomfort at presentation		
Yes	127 (66.8)	340 (46.9)
No	63 (33.2)	385 (53.1)
Diabetes		
Yes	33 (17.4)	69 (9.5)
No	157 (82.6)	656 (90.5)
Diastolic blood pressure	84.5*(15.1)†	83.4*(12.3)†
Congestive heart failure		
Yes	29 (15.3)	101 (13.9)
No	161 (84.7)	624 (86.1)
Angina pectoris		
Yes, $\leq$ 1 month	7 (3.8)	4 (0.6)
Yes, $>$ 1 month	82 (43.2)	236 (32.6)
No	101 (53.2)	485 (66.9)
Previous myocardial infarction		
Yes, $\leq$ 6 months	18 (9.5)	27 (3.7)
Yes, $>$ 6 months	54 (28.4)	148 (20.4)
No	118 (62.1)	550 (75.9)
Previous PTCA		
Yes	6 (3.2)	38 (5.2)
No	184 (84.7)	687 (94.8)
Previous CABG		
Yes	14 (7.4)	70 (9.7)
No	176 (92.6)	655 (90.3

* Mean.
† Standard deviation.

Neural Network Ensembles

We considered neural networks in the form of feed-forward multilayer perceptrons (MLP) with one hidden layer and no direct input-output connections. The hidden unit activation function was the hyperbolic tangents and the output activation function was the standard logistic function. We used the standard cross-entropy error function for two classes. In addition we introduced a weight elimination term Hanson and Pratt [15], controlled by a tunable parameter λ, to possibly regularize the network.

$$E_{\text{reg}} = \lambda \sum_{c} \frac{\omega_i^2}{1 + \omega_i^2} \tag{1}$$

The sum for network runs over the weights of the connections i between the layers, but not over the thresholds. The total error is just the sum of the cross-entropy part and E_{reg} for the case when using regularized MLPs. The minimization of the error function was accomplished using the quasi-Newton variable metric

method (see e.g. Bishop [16]).

Two methods were used to construct an ensemble of MLPs, Bagging and K-fold cross splitting. In Bagging we start with a given training set and then create new training sets by resampling, with replacement, the original one. Thus, the Bagging ensemble contains MLPs trained on *bootstrap* samples of the original training set. The ensemble output y^{ens} is simply taken as the mean of the individual ensemble members, i.e.

$$y^{\text{ens}} = \frac{1}{C}\sum_{i=1}^{C} y_i \ , \tag{2}$$

where y_i is the output of the i:th MLP in the ensemble and C is the Bagging ensemble size.

Another approach to construct different, but yet similar, training sets for the ensemble members is to split the original training set into K disjoint sets of approximately equal size. From these K parts we construct K MLPs by removing part k, ($k = 1, .., K$), when training the k:th MLP. This results in K MLPs trained on different (although similar) parts of the full training set. Repeating this procedure N times results in an ensemble of NxK MLPs. As for Bagging the ensemble output for the NxK-fold cross splitting ensemble is computed as the mean of the NxK MLP outputs (see Eq. (2)). Clearly, for small values of K the training sets differs more compared to large values of K where only a small part ($1/K$) of the original training set is removed. For the efficiency of the ensemble we therefore only consider small values of K (e.g. $2 \leq K \leq 10$). This approach to ensemble creation can be found in the work of Krogh et al. [10], but used in a different content.

The ensemble size, C for Bagging and KxN for K-fold cross splitting, influences the performance of the ensemble method compared to single MLP classifiers. In this study we will use ensemble sizes up to 50, which is found to be enough in numerical studies (see e.g. [11],West et al. [17]).

Ensemble Model Selection

It is important that the ensemble members disagree in their predictions, if anything is to be gained by combining them. This fact becomes apparent when decomposing the mean square ensemble error into the average squared error of the individual networks minus the average diversity, defined as variance of the individual network outputs (see e.g. [10]). This suggests that over-fitting of the individual ensemble members might be advantageous since this will most likely increase the diversity. Still, we must be able to do model selection for our ensemble classifiers to determine an optimal architecture and possibly a regularization parameter for the weight elimination error term, Eq. (1). For the Bagging ensemble one approach would be to use the 0.632 bootstrap estimator Efron [18]. Let θ denote our measure of performance. The 0.632 bootstrap estimator $\hat{\theta}_B$ is then defined as,

$$\hat{\theta}_B = \frac{1}{C}\sum_{i=1}^{C} 0.632\hat{\theta}_i + 0.368\hat{\theta}_i^a \ , \tag{3}$$

where $\hat{\theta}_i$ is computed from the remaining instances of the original data set not present in bootstrap training set i and $\hat{\theta}_i^a$ is the performance computed on the i:th training set (often called the apparent performance). We will use the $\hat{\theta}_B$ estimator in order to do model selection for the Bagging ensemble.

For the K-fold cross splitting ensemble the first approach would be to compute an estimate of θ using the parts of the K-fold split that was left out during training, exactly the way one would do in K-fold cross validation. However, this is not optimal from the ensemble point of view, since this estimate of the generalization performance is not based on any ensemble average. Thus to validate the K-fold cross splitting ensemble we need to compute an estimate of θ based on a validation ensemble. We can do this in a convenient way using the fact that we have a disjoint split of the training set into K parts. Training an ensemble for a given K, assuming $N = 1$, gives rise to K different validation sets and the corresponding sets of MLP validation outputs. These K parts can be concatenated to a full validation set, identical in size to the original training set. Repeating this procedure N times results in N such full validation sets. It is now straight forward to average these lists to produce the final *validation ensemble* from which we can compute the estimate $\hat{\theta}_{CS}$, the K-fold splitting ensemble estimate of the generalization performance.

The model selection performed in this study, both the Bagging and the K-fold cross splitting ensemble, is based on $\hat{\theta}_B$ and $\hat{\theta}_{CS}$, respectively.

Multiple logistic regression

To compare against a standard statistical classification method the probability of ACS was also predicted using multiple logistic regression Hosmer and Lemeshow [19]. In this analysis all independent variables except age were entered in the regression models as categorical variables as shown in table I. The variable diastolic blood pressure were categorized into two categories, either < 70 mmHg or ≥ 70 mmHg, as opposed to the ANN model where it was used as a numerical value. In this model we also allowed for synergetic effects with age, implemented as cross-product terms, for the following variables: gender, chest discomfort at presentation, symptom duration and previous myocardial infarction.

TABLE II

THE TEST PERFORMANCE OBTAINED FROM THE DIFFERENT METHODS. THE NUMBERS ARE MEDIANS (2.5, 97.5 PERCENTILES) OVER THE 100 TEST SETS DEFINED BY THE CROSS TESTING PROCEDURE.

Method	Test Roc Area (%)
Single ANN	75.0 (62.0, 87.5)
ANN Ensemble (CS $N = 25, K = 2$)	77.3 (64.9, 86.7)
ANN Ensemble (Bagging, $C = 50$)	78.2 (65.8, 88.5)
Multiple Logistic Regression	77.6 (65.4, 88.0)

Performance estimation

All classification models in this study provides an estimate of the posterior probability for ACS given a data record. It is therefore straight forward to construct receiver operating characteristics (ROC) curves for all methods. As a measure of performance we will use the area under the ROC curve. This area has the usual interpretation that a randomly chosen patient with ACS has a larger outcome probability than a randomly chosen patient without ACS (see e.g. Hanley and McNeil [20]).

To estimate the generalization performance of the classification methods we used a 10-fold cross testing procedure, repeated 10 times, resulting in 100 test sets on which the area under the ROC curve was calculated. The procedure is similar to the cross validation method used for model selection and is accomplished by dividing the data set into 10 parts of (approximately) equal size. A classification model is constructed on all parts except one, which is used as the independent test set. The average of the 100 ROC areas is used as the test performance for a given classification method. It is important to stress that a given test set was never part of the model selection or model construction procedure. The same $10x10$-fold split was used for all methods tested, which allows for a statistical comparison of the performances using e.g. a paired t-test. Strictly, it is only within each 10-fold split we, by definition, have independent data. The p-value calculations are therefore applied for each of the 10 splittings. The median p-value is used throughout this paper.

RESULTS AND DISCUSSION

The best results obtained for the different classification methods are summarized in Table II. Here we also included results from a classifier based on a single MLP. The model selection carried out for this single network was based on 10-fold cross validation, repeated 5 times. As expected there is an advantage of using an ensemble of networks compared to the single MLP classifier, where p-values for the differences are p=0.03 and p=0.1 for Bagging and K-fold cross splitting, respectively.

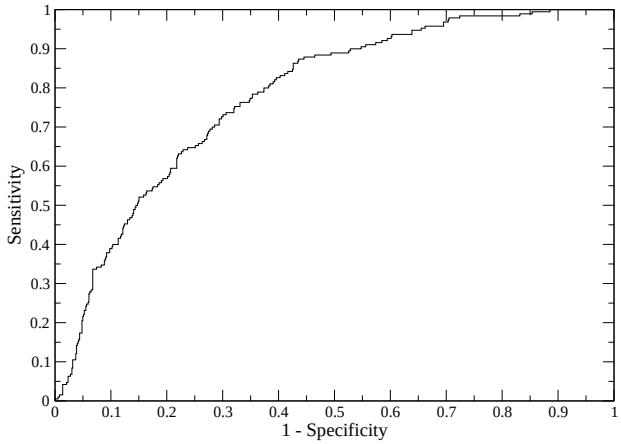

Fig. 1. The ROC curve for the Bagging ensemble. This curve is based on the complete data set obtained from one of the 10 cross testing splits. The area under this curve is close to the median of 78.2%.

Comparing with the multiple logistic regression we can conclude that this method works very well. There is only a small increase in performance in favor of the Bagging ensemble (p=0.7). It should be noted that the logistic regression method is not linear since it contains cross terms as described earlier. Figure 1 shows the ROC curve, for the Bagging ensemble, of the full test set, corresponding to one of the test splits that obtained a ROC area closest to the median (78.2%).

Although the K-fold cross splitting ensemble did not obtain the best test result it is interesting to compare the results for different values of K and N, both for unregularized and regularized MLPs. For the K-fold cross splitting ensemble we defined an estimate of the generalization performance based on a validation ensemble. Is is interesting to see how well this estimate correlates to the test performance given by the $10x10$-fold cross testing procedure. Table III shows the results obtained for different combinations of K and N. For the unregularized ensemble the number of hidden nodes in the MLPs varied from 3-5, based on the validation ensemble results. To allow for a comparison between the validation result and the test result, the size of the validation ensemble was increased as to match the size of the ensemble used for the testing (KxN). As expected, the results from combinations where the KxN is small did not perform as well as for larger values of KxN. The $2x1$ ensemble resulted in a validation (test) ROC area of 69.5% (73.1%), compared to the best result of 75.7% (76.8%) obtained by the $2x25$ ensemble. The test difference is statistically significant (p=0.02). Comparing the three different settings all having a total ensemble size of 50 ($2x25, 5x10, 10x5$) we find a small increase in performance for smaller values of K.

Turning to the regularized K-fold cross testing ensemble

TABLE III

THE VALIDATION AND TEST PERFORMANCE (ROC AREA) FOR BOTH REGULARIZED AND UNREGULARIZED K-FOLD CROSS SPLITTING ENSEMBLES. THE NUMBERS ARE MEDIANS (2.5, 97.5 PERCENTILES) OVER THE 100 TEST SETS DEFINED BY THE CROSS TESTING PROCEDURE.

		Unregularized Ensemble	
K	N	Validation	Test
2	1	71.5 (68.3, 74.6)	73.1 (55.5, 83.2)
2	5	74.6 (72.2, 76.6)	74.8 (64.1, 84.4)
2	10	76.0 (74.2, 77.7)	76.3 (63.3, 86.3)
2	25	76.5 (74.2, 78.0)	76.8 (61.0, 88.1)
5	1	73.8 (70.9, 76.0)	74.6 (60.1, 85.1)
5	5	75.1 (73.2, 76.7)	76.0 (63.4, 86.3)
5	10	76.1 (74.3, 77.5)	76.6 (61.4, 89.3)
10	1	74.0 (71.5, 76.7)	74.5 (62.9, 85.9)
10	5	75.5 (73.1, 77.4)	76.1 (64.3, 85.4)
		Regularized Ensemble	
K	N	Validation	Test
2	1	74.1 (71.0, 76.8)	75.9 (59.3, 86.7)
2	5	76.1 (74.3, 77.8)	76.4 (64.8, 87.8)
2	10	76.2 (74.8, 77.8)	77.2 (64.0, 88.1)
2	25	76.7 (75.2, 77.9)	77.3 (64.9, 86.7)
5	1	75.2 (71.1, 77.1)	76.1 (64.4, 85.4)
5	5	76.2 (74.7, 78.0)	77.2 (64.6, 87.4)
5	10	76.3 (74.8, 77.8)	77.1 (64.8, 86.8)
10	1	75.9 (74.0, 77.6)	76.6 (64.4, 86.5)
10	5	76.2 (74.7, 77.8)	76.3 (63.3, 86.5)

(bottom part of Table III) we observe a consistent increase of the performance when comparing the median ROC areas for the same (K, N). None of these differences are however statistically significant. The best result was obtained for the $(2x25)$ regularized ensemble with median test ROC area of 77.3%. Again we notice an increase using large enembles with small K. Regardless of whether we use regularization or not, the estimate of the generalization performance as given by the validation ensemble is only slightly pessimistic compared to the "true" generalization performance.

Table IV shows the result for the Bagging ensemble for different values of the ensemble size C, both for unregularized and regularized MLPs. Here we also observe a difference between regularizing or not. For the C=50 ensemble we have an increase of the test performance from 76.6% to 78.2% (p=0.3). Our approach to validate the Bagging ensemble using the 0.632 bootstrap rule (Eq. 3) results in too optimistic estimates of the generalization performance. However, for our medical prediction task, the Bagging ensemble receives the best test set result compared to both the K-fold splitting ensemble and logistic regression, although not significant.

In conclusion we find that it is possible to construct a machine learning tool that can predict the presence

TABLE IV

THE VALIDATION AND TEST PERFORMANCE (ROC AREA) FOR BOTH REGULARIZED AND UNREGULARIZED BAGGING ENSEMBLES. THE NUMBERS ARE MEDIANS (2.5, 97.5 PERCENTILES) OVER THE 100 TEST SETS DEFINED BY THE CROSS TESTING PROCEDURE.

	Unregularized Ensemble	
C	Validation	Test
2	78.3 (74.1, 81.7)	71.8 (56.9, 81.3)
5	78.1 (76.1, 80.6)	73.9 (62.2, 86.0)
10	78.2 (76.4, 79.9)	76.0 (61.9, 85.7)
25	78.1 (76.9, 79.5)	75.6 (65.2, 87.3)
50	78.1 (76.8, 79.4)	76.6 (65.1, 87.0)
	Regularized Ensemble	
C	Validation	Test
2	78.9 (75.7, 81.6)	71.4 (62.0, 82.1)
5	79.2 (76.7, 81.0)	75.3 (63.1, 86.0)
10	79.0 (77.6, 80.5)	76.9 (64.0, 86.6)
25	78.9 (77.7, 80.2)	77.6 (64.8, 87.3)
50	79.0 (77.9, 80.5)	78.2 (65.8, 88.5)

of ACS among patients with chest pain at a ROC area level of about 78%. This result is limited to our study population and to our choice of input variables. We find the K-fold splitting ensemble technique interesting since it provides an estimate of the generalization performance that is in good agreement of the test performance.

From a medical point of view the evaluation of patients with suspected ACS in the emergency department is very important. Since the ACS diagnosis is a difficult one, there is a certain level of overadmission of patients. In a study of a Swedish emergency department it was found that almost 3/4 of the patients with suspected ACS were incorrectly hospitalized Ekelund et al. [21]. On the other hand, patients are also incorrectly sent home from the emergency department, where a 5% level has been reported Lee et al. [22], Pope et al. [23]. It is therefore important to develop medical decision support for the prediction of ACS in the emergency department. The machine learning approach used in this study allows for a sensitivity and a specificity of about 95% and 40% respectively. There still is a lot to be accomplished before one can apply our method in a clinical setting, but the level of accuracy is promising.

ACKNOWLEDGMENTS

This study is in part supported by the Swedish Knowledge Foundation (2208/04-504).

REFERENCES

[1] L. Goldman, E. F. Cook, P. A. Johnson, D. A. Brand, G. W. Rouan, and T. H. Lee. Prediction of the need for intensive care in patients who come to emergency departments with acute chest pain.

The New England Journal of Medicine, 334(23): 1498–1504, 1996.

[2] W.G. Baxt. Use of an artificial neural network for the diagnosis of myocardial infarction. Annals of Emergency Medicine, 115:843–848, 1991.

[3] W.G. Baxt, F.S. Shofer, F.D. Sites, and J.E. Hollander. A neural computational aid to the diagnosis of acute myocardial infarction. Annals of Emergency Medicine, 34:366–373, 2002.

[4] J. E. Hollander, K. L. Sease, D. M. Sparano, F.D. Sites, F. S. Shofer, and W.G. Baxt. Effects of neural network feedback to physicians on admit/discharge decision for emergency department patients with chest pain. Annals of Emergency Medicine, 44(3): 199–205, 2004.

[5] B. Hedén, H. Öhlin, R. Rittner, and L. Edenbrandt. Acute myocardial infarction detected in the 12-lead ECG by artificial neural networks. Circulation, 96 (6):1798–1802, 1997.

[6] M. Ohlsson, H. Öhlin, S.M. Wallerstedt, and L. Edenbrandt. Usefulness of serial electrocardiograms for diagnosis of acute myocardial infarction. American Journal of Cardiology, 88:478–481, 2001.

[7] J. Xuea, T. Aufderheide, R.S. Wright, J. Klein, R. Farrell, I. Rowlandson, and B. Young. Added value of new acute coronary syndrome computer algorithm for interpretation of prehospital electrocardiograms. Journal of Electrocardiology, 37: 233–239, 2004.

[8] P.J.G. Lisboa, E.C. Ifeachor, and P.S. Szczepaniak, editors. Artificial neural networks in biomedicine. Springer-Verlag, London, 2000.

[9] L. K. Hansen and P. Salamon. Neural network ensembles. IEEE Transactions on Pattern Analysis and Machine Intelligence, 12:993–1001, 1990.

[10] A. Krogh and J. Vedelsby. Neural network ensembles, cross validation, and active learning. In G. Tesauro, D. Touretzky, and T. Leen, editors, Advances in Neural Information Processing Systems, volume 2, pages 650–659, San Mateo, CA, 1995. Morgan Kaufman.

[11] D. Opitz and R. Maclin. Popular ensemble methods: An empirical study. Journal of Artificial Intelligence Research, 11:169–198, 1999.

[12] L. Breiman. Bagging predictors. Machine Learning, 24:123–140, 1996.

[13] J. L. Hansen, J. Björk, M. Ohlsson, L. Edenbrandt, H. Öhlin, and U. Ekelund. Triage decisions, outcome and important prediction variables in unselected emergency department chest pain. Manuscript in preparation, 2005.

[14] H. Tunstall-Pedoe, K. Kuulasmaa, P. Amouyel, D. Arveiler, A.M. Rajakangas, and A. Pajak. Myocardial infarction and coronary deaths in the world health organization monica project. registration procedures, event rates, and case-fatality rates in 38 populations from 21 countries in four continents. Circulation, 90:583–612, 1994.

[15] S. J. Hanson and L. Y. Pratt. Comparing biases for minimal network construction with back–propagation. In D. S. Touretzky, editor, Advances in Neural Information Processing Systems 1, pages 177–185. Morgan Kaufmann, 1989.

[16] C. M. Bishop. Neural Networks for Pattern Recognition. Oxford University Press, 1995.

[17] D. West, P. Mangiameli, R. Rampal, and V. West. Ensemble strategies for a medical diagnostic decision support system: A breast cancer diagnosis application. European Journal of Operational Research, 162(2):532–551, 2005.

[18] B. Efron. Estimating the error rate of a prediction rule: some improvements on cross-validation. Journal of the American Statistical Association, 78: 316–331, 1983.

[19] D.W. Hosmer and S. Lemeshow. Applied logistic regression. Wiley, New York, 1989.

[20] J. A. Hanley and B. J. McNeil. The meaning and use of the area under the receiver operating characteristic (roc) curve. Radiology, 143:29–36, 1982.

[21] U. Ekelund, H-J. Nilsson, A. Frigyesi, and O. Torffvit. Patients with suspected acute coronary syndrome in a university hospital emergency department: an observational study. BMC Emergency Medicine, 2:1–7, 2002.

[22] T.H. Lee, G.W. Rouan, M.C. Weisberg, D.A. Brand, D. Acampora, C. Stasiulewicz, J. Walshon, G. Terranova, L. Gottlieb, B. Goldstein-Wayne, and et al. Clinical characteristics and natural history of patients with acute myocardial infarction sent home from the emergency room. American Journal of Cardiology, 60(4):219–224, 1987.

[23] J.H. Pope, T.P. Aufderheide, R. Ruthazer, and et al. Missed diagnoses of acute cardiac ischemia in the emergency department. The New England Journal of Medicine, 342(16):1163–1170, 2000.