An improved Artificial Neural Network based model for Prediction of Late Onset Heart Failure

Nader Salari[1, 2], Shamarina Shohaimi[1], Farid Najafi[2], Meenakshii Nallappan[1], Isthrinayagy Karishnarajah[3]

[1]Department of Biology, Faculty of Science, University Putra Malaysia, Serdang, Selangor, Malaysia.
[2]Department of Biostatistics and Epidemiology, School of Public Health, Kermanshah University of Medical Sciences, Kermanshah, Iran.
[3]Department of Mathematics, Faculty of Science, University Putra Malaysia, Serdang, Selangor, Malaysia.
Corresponding Author E-mail: nader.salari.1344@gmail.com

Abstract: ***Background and Objective:*** The present study aims to present an artificial neural network (ANN)-based model for prediction of Late Onset Heart Failure (LOHF) in patients, with no previous Heart Failure (HF) history, who experienced non-fatal, first-ever Acute Myocardial Infarction (AMI) without previous history of heart failure. ***Methods*****:** Two models of multilayer perceptron (MLP) and Radial Basis Function (RBF) neural network approaches based on decision support system were developed. The MLP model was used to optimize the predicting algorithm based on the conjugate gradients descent method. To design the RBF network, K-Means clustering technique was used to select the centers of RBFs, and k-nearest neighbourhood to define the spread and forward selection for determining the optimum number of RBFs. To assess the generalization of the network, K-fold cross-validation test was used. A total of 3,109 medical records containing 19 main clinical parameters were used to train and test the networks. ***Results*****:** The findings indicate a reliable performance of the proposed system. The MLP based model yields a sensitivity, specificity, and an area under the receiver/relative operating characteristic (ROC) curve (AUC) of 87.1%, 90%, and 0.887 ± 0.02, respectively. However, the RBF network shows the above parameters as 84.4%, 94.3%, and 0.905 ± 0.017, respectively. ***Conclusions*****:** The proposed intelligence system achieved a high degree of diagnostic accuracy (92.9% for MLP and 93.7% for RBF) indicating its high efficiency for clinical diagnosis of LOHF.

[Nader Salari, Shamarina Shohaimi, Farid Najafi, Meenakshii Nallappan, Isthrinayagy Karishnarajah. **An improved Artificial Neural Network based model for Prediction of Late Onset Heart Failure.** *Life Sci J* 2012;9(4):3684-3689]. (ISSN: 1097-8135). http://www.lifesciencesite.com. 546

Keywords: Heart failure, late-onset heart failure, acute myocardial infarction, artificial neural network, multilayer perceptron, radial basis function.

1. Introduction

Heart failure (HF) is defined as an inability of the heart to supply sufficient blood flow as well as oxygen to meet the body's requirements. There are several risk factors can cause HF such as ischemic heart disease because of heart valve dysfunction, hypertension, heart attack (myocardial infarction), and cardiomyopathy. Heart attack or acute myocardial infarction (AMI) is simply cellular death of part of heart muscle due to the interruption of blood flow, resulting in shortage of oxygen in myocardium. Without prompt treatment, this condition is potentially fatal (Nadar et al., 2005, Bordier, 2009). It is noteworthy that over one million cases of hospital discharge with HF were registered in the United States in 2005, illustrating a 171% rise compared to 1979 (Cohn et al., 2000). Furthermore, approximately 3.8 million cases of HF were diagnosed in American hospitals in 2004(Gheorghiade and Pang, 2009) .

The high rate of morbidity and mortality following an episode of heart attack is a major concern of physicians and cardiologists necessitating development of an efficient approach to prevent the heart attack or control its consequent complications through analysis of information obtained from patients. This process typically consists of numerous clinical visits, sometimes involving many paraclinical assessments, as well as the physicians own clinical experiences and insight to predict the event.

Nevertheless, there is a growing effort to find more accurate and faster solutions for prediction at the macro level. In addition to physicians' efforts, some prediction models, based on mathematical and statistical methods, have been developed during the recent years. The most commonly used methods include Bayesian network, logistic regression, and neural networks (Barlow et al., 1984, Fabbri et al., 2008, Hsu et al., 2005, Lang et al., 1997, Pang et al., 2007, Smits et al., 2010). Particularly, artificial neural networks (ANNs) have been successfully used for surgical decisions and mortality prediction based on initial clinical data (Li et al., 2000).

Although the characteristics of HF have been studied extensively (Bordier, 2009, Hamner and Ellison, 2005, Najafi et al., 2007, Najafi et al., 2008), there are only few studies conducted on the prediction of HF using ANN in MI patients (Eggers et al., 2007). Furthermore, previous studies showed that an efficient prediction of HF following AMI should be based on the time of its occurrence.] The importance of time-based prediction is originated from the different mechanisms responsible for early-onset HF (EOHF) (HF complicating an index AMI within the first 28 days) and late-onset HF (LOHF) (HF developing beyond 28 days after an index AMI) (Cohn et al., 2000, Stone et al., 1988). The present study was mainly aimed to propose an ANN-based model for prediction of Late Onset Heart Failure (LOHF) in patients, with no previous HF history, who experienced non-fatal, first-ever AMI without previous history of heart failure.

2. Materials and Methods

Study populations and data collection

The Perth MONItoring trends and determinants of CArdiovascular disease (MONICA) Register covered the entire residents of the Perth Statistical Division, in an effective manner the metropolis district of Perth, aged 25-64 years (Tunstall-Pedoe et al., 1994). We have been using the Western Australian Linked Database System (WALDS)(Holman et al., 1999) and MONICA data from Perth, Western Australia. The register consisted of all main coronary events that have occurred, and used the 'cold pursuit' method to clarify non-fatal potential instances of AMI through surveillance of hospital discharge codes(Tunstall-Pedoe et al., 1994). Data were extracted and compiled from medical records for each aspect of hospitals with International Classification of Diseases (9th revision, clinical modification) (ICD-9-CM) code for AMI or somewhat acute coronary heart disease (codes 410 and 411, respectively). The analysis relates to all patients with events who met the following criteria: the patients had no history of previous HF or AMI and no evidence of EOHF; the event fulfilled the MONICA criteria for 'definite AMI' (Stone et al., 1988); and the patients were alive 28 days after the onset of AMI symptoms. Using the Western Australian Linked Database System WALDS (Holman CD et al., 1999), we followed up all patients included in our study sample for a subsequent admission to the hospital with a diagnosis of HF. To capture all cases of HF (even those complicating a recurrent AMI), we defined a patient as having HF when he/she had an electronic record for a new hospital admission including the ICD-9-CM code for HF (428) in either the first or the second diagnostic position. We refer to such cases as late-onset HF, as opposed to EOHF indicating HF complicating the first-ever AMI within the first 28 days.

A total of 19 variables including demographic information, clinical history, symptoms, lab results and physical examinations were collected from each patient's paper record (3109 number of the patients were used (2652 male and 457 female)) and then normalized under the supervision of cardiologists (see Table 1). Patient records often contain missing values. In this study, missing values were replaced with the normal values assuming that if the readings were abnormal they would have been recorded.

Table 1.Characteristics of patients with first-ever non-fatal myocardial infarction, Perth MONICA

Variables	Comment
-Basic information	
Age	Normalize on (0 1)
Sex	Male=0, Female=1
-Medical history	
History of diabetes	Absence = 0, presence = 1
History of hypertension	Absence = 0, presence = 1
History of angina	Absence = 0, presence = 1
Current smoker	Absence = 0, presence = 1
Recurrent MI	Absence = 0, presence = 1
-Presenting characteristics	
Mean of CPK ratio	Normalize on (0 1)
Systolic blood pressure	Normalize on (0 1)
Pulse	Normalize on (0 1)
Shock	Absence = 0, presence = 1
Syncope	Absence = 0, presence = 1
ALVF	Absence = 0, presence = 1
Complication of infarction	Absence = 0, presence = 1
Length of hospital admission	Normalize on (0 1)
-ECG findings	
ST-elevation	Absence = 0, presence = 1
ST- depression	Absence = 0, presence = 1
Q-wave	Absence = 0, presence = 1
Anterior MI	Absence = 0, presence = 1

Multilayer perceptron model

The ANNs are strong tools for prediction, classification, generalization, simulation, etc. in different applications. Furthermore, neural network approach is a way of modelling data, based on computer learning which are basically trained to perform complex functions in various fields of applications including pattern recognition, identification, classification, speech, vision , control systems, and etc.(Gallant and White, 1992) The multi-layer perceptron (MLP) networks are one of the most widely used neural networks consisting of a great deal number of processing elements called neuron. The neurons are connected to each other through a set of weights. These weights are adjusted based on an error-minimization technique called back-propagation rule. A diagram of the used MLP model with one hidden layer is shown in Fig. 1. The specified network consists of three layers named as input layer, hidden layers and output layer. Each layer has its own number of neurons. The input to the node l in the hidden layer is given by

$$\eta\_in_l = \sum_{u=1}^{q}(x_u v_{ul}) + \theta_l \; ; \qquad l = 1,2,\dots,s \qquad (1)$$

where s is the number of neurons in the hidden layer, q is the number of neurons in the input layer, θ_l is the bias term of the l^{th} neuron of hidden layer, and $v_{ul}$is the weighting factor between u^{th} input neuron and the l^{th} hidden one.(Gallant and White, 1992). The output from l^{th} neuron of the hidden layer is given by

$$\eta\_out_l = f_1(\eta\_in_l) \ ; \qquad l = 1,2,\ldots,s \qquad (2)$$

where f_1 is the transfer function of the hidden layer. Some of the commonly used transfer functions are threshold, Gaussian, logarithm-sigmoid and tan-sigmoid functions. Because of the bipolar advantages, MLP neural networks often make use of the tan-sigmoid transfer function in the hidden layer; therefore, the output from l^{th} neuron of the hidden layers is given by

$$o_l = Tansig(\eta\_in_l) = \frac{2}{1+\exp(-2(\eta\_in_l))} - 1 \ ; \qquad (3)$$

$$D_{o_l}: [-\infty\ \infty] \rightarrow R_{o_l}: [-1\ 1];\ activation\ domain\ \approx [-2\ 2]$$

The input of the j^{th} neuron in the output layer is given by

$$y\_in_j = \sum_{u=1}^{s} (o_u w_{uj}) + b_j \ ; \qquad j = 1,2,\ldots,m \qquad (4)$$

where b_j is the bias term of j^{th} hidden neuron, w_{uj} is the weighting factor between u^{th} neuron of hidden layer and the j^{th} neuron of output layer, and m is the number of neurons in the output one (Taylor, 1997).

And then, the output of j^{th} neuron is represented as follows:

$$y\_out_j = f_2\left(y\_in_j\right) \ ; \qquad j = 1,2,\ldots,m \qquad (5)$$

where f_2 is the logarithm-sigmoid transfer function of the output layer.

$$Logsig\left(y\_in_j\right) = \frac{1}{1+\exp\left(-(y\_in_j)\right)} \ ; \qquad (6)$$

$$D_{f_2}: [-\infty\ \infty] \rightarrow R_{f_2}: [0\ 1] \ ;\ activation\ domain \approx [-4\ 4]$$

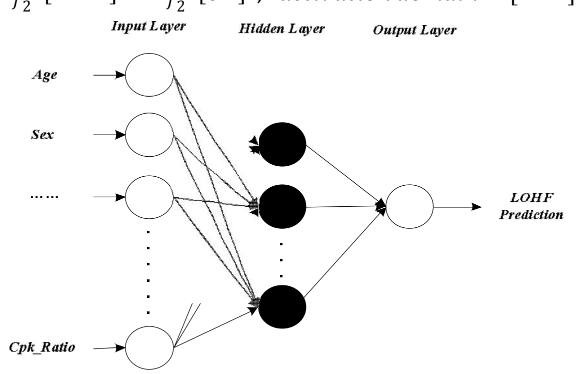

Fig 1. System architecture.

The training of a network is a process where the set of adjusted parameters (weights and biases) is optimized to make the best prediction of the target variable based on background variables. MLP networks are trained with the standard back-propagation algorithm (Gallant and White, 1992). Back-propagation algorithm basically consists of two steps: a forward step where the signal propagates through the computational units until it gets to the output layer and a backward step where all synaptic weights are adjusted accordingly to an error correction rule. In this method, often the adjusted parameters are determined iteratively to achieve a minimum mean square error between the network output and the target values. In our study, a three-layer MLP network, with tangent-sigmoid transfer functions in the hidden layer as well as a logarithm-sigmoid transfer function in the output layer was constructed.

It should be noted that a complicated network with more neurons is capable of solving more sophisticated problems. However, increasing the number of neurons makes the system vulnerable to the noises present in the training data resulting in a condition called over-fitting. Therefore, a network with fewer neurons increases the generalization capability of the system as well as weights' size reduction and more convergence of the network to its desired output. On the other hand, a network with few neurons can't learn as good as it should. Therefore, there is a challenge between the power of generalization and preventing of over-fitting. Consequently, one of the main problems with implementation of an ANN model is selecting the correct number of the neurons. One method to determine the optimum number of the nodes in the hidden layer is implementing different networks with one to thirty nodes, and then the minimum error on test data calculates. Number of neurons with the minimum error on test data was chosen as the optimum network.

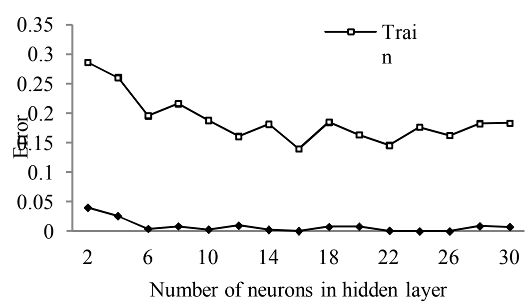

Fig 2 . Train and test errors of MLP vs. number of hidden nodes.

Radial Basis Function Model

This network also consists of input, hidden and output layers. Input to each hidden neuron is the distance between the network inputs and centre of that neuron's transfer function. The network's output is the weighted combination of hidden neurons' output. The network parameters are centre and spread of hidden layer's transfer functions and weights of output layer. There are different forms of transform functions but the most widely used one is the Gaussian function.

Setting the centers randomly to the training inputs is the simplest method of defining the centers but this approach is prone to overfitting. An alternative is to cluster the training patterns into groups according to some similarity measurement and then assigning nodes to each cluster. The typical method to determine such clusters is the k means clustering algorithm. Since it is an unsupervised method and data belonging to two different groups can be clustered as one, application of supervised methods like genetic algorithm (Billings and Zheng, 1995), supervised fuzzy c-means (Pedrycz, 1998) decision tree(Kubat, 1998), have been investigated recently.

Although supervising methods like those mentioned can improve the RBF neural network classification performance, they are slow to train which can be a disadvantage comparing with the fast learning of the MLP achieved through combining an unsupervised with a supervised method is a disadvantage of RBF against the MLP approach. Forward selection was used to determine the number of hidden layer neurons where the number of neurons was changed from two to thirty and considering the error on test data, optimum number of neurons and therefore number of clusters were determined (Fig. 3). Spread or width of Gaussian transfer functions was determined from the k-nearest neighbour heuristic according to the following formula:

$$\beta = \left(\frac{1}{k} \sum_{j=1}^{k} \left\| c - c_j \right\|^2 \right)^{1/2} \quad (7)$$

where cj is the k-nearest neighbour of c. K = 2 was used as suggested previously (Moody and Darken, 1988). Weights in the output layer were found using the pseudo inverse method.

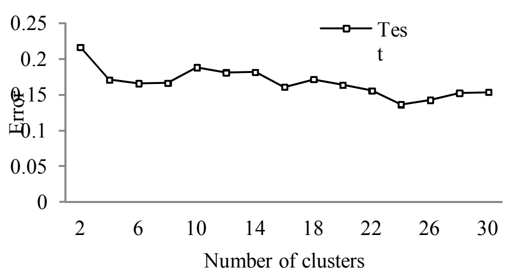

Fig 3. Test errors of RBF vs. number of clusters.

3. Results

Experimental Results and Performance Assessment

To estimate the performance of the model, a 10-fold cross validation method was used. The data set was randomly divided to 10 roughly equal parts (approximately 310 patients with 6218 data points per set). The network was trained on nine parts and tested on the remaining part. This procedure was repeated 10 times so the network error is the average of these 10 errors. To evaluate the performance of network, sensitivity and specificity were calculated. The values of sensitivity and specificity obtained for a test with continuous output depend on the particular cut-off value chosen to distinguish normal and abnormal results. Selecting a threshold for a clinical decision support depends on the disease in question and the purpose of testing, if the disease is serious and lifesaving therapy is available, then falsely diagnosing a patient as healthy should be minimized (increasing sensitivity) and if disease is not serious and the therapy is dangerous, then falsely diagnosing a healthy individual as a patient should be minimized (increasing specificity). So sensitivity and specificity alone do not indicate the performance of an experiment and the chosen value of threshold is also effective. Altering the threshold value does not have any impact on the experiment and only provides a balance between sensitivity and specificity. Therefore, the best way to demonstrate the performance of a test is determining values of sensitivity and specificity for all cut off points obtained by an ROC curve.

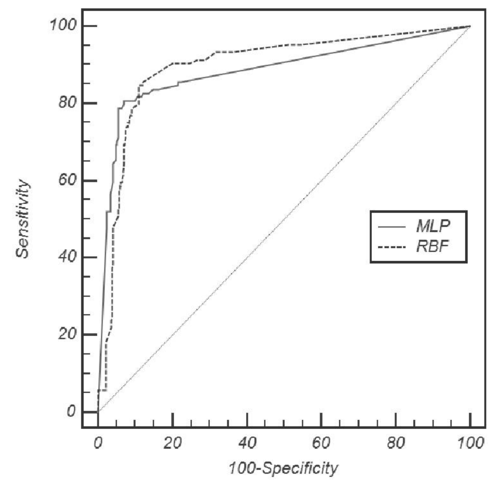

Fig 4. ROC curve comparison of MLP and RBF.

Table 2. Performance evaluation MLP and RBF networks

	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	AUC ± SE
MLP	87.1	90	88.4	86.2	0.887±0.020
RBF	84.4	94.3	93.1	87.7	0.905±0.017

The network output values range 0-1. Zero corresponds to non-LOHF and one to LOHF disease. The desired output was based on the final diagnosis of physicians according to the patients' record. MedCalc software (v11.1.6) was used to find the threshold value which maximizes both sensitivity and specificity (Moody and Darken, 1988). Application of MLP with 16 neurons in hidden layers resulted in sensitivity of 87.1%, specificity of 90% and ROC of 0.887 ± 0.02 and using RBF neural networks with 6 neurons in hidden layer resulted in specificity of 94.3%, sensitivity of 84.4% and ROC of 0.905 ± 0.017 (see Fig. 4-Table 2).

4. Discussion

As mentioned before heart attack is a major concern for physicians and cardiologists that leads to their efforts to prevent the complications through analysis of information obtained from patients to make sound decision and suitable therapy and care. This process typically may consume long time which consists of numerous clinical visits, sometimes involving many Para clinical tests, as well as the physician's own experiences and insight to predict the event. So, there is a growing effort to find more accurate and faster solutions for prediction. In addition to physicians' efforts, some prediction models were created. Mathematical and statistical methods have been used to develop models for prediction. The most commonly used methods include Bayesian network, logistic regression, and ANN models.

Nowadays, ANN model because of its high efficiency and accuracy is the most popular tool for predicting hence this method have been used in very large amount of decision systems. The present study corroborates that ANNs can be trained from clinical data available for the diagnosis of the disease. ANNs have the ability to learn classification or pattern recognition tasks thus AMI patients are classified into two categories, namely LOHF and non-LOHF from complex data sets. Furthermore, we presented a medical decision support system based on the MLP and RBF neural networks architecture to predict the LOHF for patients who have experienced their first-ever, non-fatal AMI but who had never experienced heart failure and EOHF.

Conclusions

In our study, we presented a medical decision support system based on the MLP and RBF neural network architectures for the LOHF prediction. In particular, we identified 19 input variables critical to LOHF prediction and encoded them accordingly. The system is trained through an improved BP algorithm. A database consisting of 3109 cases was used in this study and 10-fold cross validation was applied to assess the generalization of the models resulting in sensitivity of 87.1% and specificity of 90%, whereas for the RBF network, these values were 94.3% and 84.4%, respectively. The Findings of our study show that the proposed systems can achieve very high diagnostic accuracy and comparably small intervals, proving their efficiency as an alternative and adjunctive option in clinical diagnosis decision of LOHF.

Acknowledgements

The authors would like to thank the School of Population Health at the University of Western Australia for providing us with a complete copy of MONICA data linked to Mortality and Hospital Morbidity data.

Corresponding Author:
Nader Salari, Ph.D. Candidate at, University Putra Malaysia;
Department of Biology, Faculty of Science, University Putra Malaysia, Serdang, Selangor, Malaysia.
He is a permanent lecturer at the department of biostatistics and epidemiology, School of Public Health, Kermanshah State University of Sciences, Kermanshah, Iran.
nader.salari.1344@gmail.com

References:

1. Nadar, S., Prasad, N., Taylor, R.S., Lip, G.Y.H. Positive pressure ventilation in the management of acute and chronic cardiac failure: a systematic review and meta-analysis. International journal of cardiology, 2005. 99(2): p. 171-185.
2. Bordier, P. Sleep apnoea in patients with heart failure. Part I: Diagnosis, definitions, prevalence, pathophysiology and haemodynamic consequences. Archives of Cardiovascular Diseases, 2009. 102(8-9): p. 651-661.
3. Cohn, J.N., Ferrari, R., Sharpe, N. Cardiac remodeling--concepts and clinical implications: a consensus paper from an international forum on cardiac remodeling* 1. Journal of the American College of Cardiology, 2000. 35(3): p. 569-582.
4. Gheorghiade, M.,Pang, P.S. Acute Heart Failure Syndromes. Journal of the American College of Cardiology, 2009. 53(7): p. 557-573.
5. Barlow, P., Teasdale, G., Jennett, B., Murray, L., Duff, C., Murray, G. Computer-assisted prediction of outcome of severely head-injured patients. Journal of microcomputer applications, 1984. 7(3-4): p. 271-277.
6. Fabbri, A., Servadei, F., Marchesini, G., Stein, S., Vandelli, A. Early predictors of unfavourable outcome in subjects with moderate head injury in the emergency department. Journal of Neurology, Neurosurgery & Psychiatry, 2008. 79(5): p. 567.
7. Hsu, M.H., Li, Y.C., Chiu, W.T., Yen, J.C. Outcome prediction after moderate and severe head injury using an artificial neural network. Studies in health technology and informatics, 2005. 116: p. 241.
8. Lang, E.W., Pitts, L.H., Damron, S.L., Rutledge, R. Outcome after severe head injury: an analysis of prediction based upon comparison of neural network versus logistic regression analysis. Neurological research, 1997. 19(3): p. 274.
9. Pang, B.C., Kuralmani, V., Joshi, R., Hongli, Y., Lee, K.K., Ang, B.T., Li, J., Leong, T.Y., Ng, I. Hybrid outcome prediction model for severe

traumatic brain injury. Journal of Neurotrauma, 2007. 24(1): p. 136-146.
10. Smits, M., Dippel, D.W.J., Nederkoorn, P.J., Dekker, H.M., Vos, P.E., Kool, D.R., van Rijssel, D.A., Hofman, P.A.M., Twijnstra, A., Tanghe, H.L.J. Minor Head Injury: CT-based Strategies for Management—A Cost-effectiveness Analysis1. Radiology, 2010. 254(2): p. 532.
11. Li, Y.C., Liu, L., Chiu, W.T., Jian, W.S. Neural network modeling for surgical decisions on traumatic brain injury patients. International journal of medical informatics, 2000. 57(1): p. 1-9.
12. Hamner, J.B.,Ellison, K.J. Predictors of hospital readmission after discharge in patients with congestive heart failure. Heart & Lung: The Journal of Acute and Critical Care, 2005. 34(4): p. 231-239.
13. Najafi, F., Dobson, A.J., Hobbs, M., Jamrozik, K. Temporal trends in the frequency and longer-term outcome of heart failure complicating myocardial infarction. European journal of heart failure, 2007. 9(9): p. 879.
14. Najafi, F., Dobson, A.J., Hobbs, M., Jamrozik, K. Late-onset heart failure after myocardial infarction: Trends in incidence and survival. European journal of heart failure, 2008. 10(8): p. 765.
15. Eggers, K.M., Ellenius, J., Dellborg, M., Groth, T., Oldgren, J., Swahn, E., Lindahl, B. Artificial neural network algorithms for early diagnosis of acute myocardial infarction and prediction of infarct size in chest pain patients. International journal of cardiology, 2007. 114(3): p. 366-374.
16. Stone, P.H., Raabe, D.S., Jaffe, A.S., Gustafson, N., Muller, J.E., Turi, Z.G., Rutherford, J.D., Poole, W.K., Passamani, E., Willerson, J.T., Sobel, B.E., Robertson, T., Braunwald, E. Prognostic significance of location and type of myocardial infarction: Independent adverse outcome associated with anterior location. Journal of the American College of Cardiology, 1988. 11(3): p. 453-463.
17. Tunstall-Pedoe, H., Kuulasmaa, K., Amouyel, P., Arveiler, D., Rajakangas, A.M., Pajak, A. 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, 1994. 90(1): p. 583-612.
18. Holman, C.D.A.J., Bass, A.J., Rouse, I.L., Hobbs, M.S.T. Population based linkage of health records in Western Australia: development of a health services research linked database. Australian and New Zealand Journal of Public Health, 1999. 23(5): p. 453-459.
19. Holman CD, Bass AJ, Rouse IL, Hobbs MS Population-based linkage of health records in Western Australia: development of a health services research linked database. 1999. 23: p. 453–9.
20. Gallant, A.R.,White, H. On learning the derivatives of an unknown mapping with multilayer feedforward networks. Neural Networks, 1992. 5(1): p. 129-138.
21. Taylor, J. Neural networks and their applications. Computers & Mathematics with Applications, 1997. 33(5): p. 131.
22. Billings, S.A.,Zheng, G.L. Radial basis function network configuration using genetic algorithms. Neural Networks, 1995. 8(6): p. 877-890.
23. Pedrycz, W. Conditional fuzzy clustering in the design of radial basis function neural networks. Neural Networks, IEEE Transactions on, 1998. 9(4): p. 601-612.
24. Kubat, M. Decision trees can initialize radial-basis function networks. Neural Networks, IEEE Transactions on, 1998. 9(5): p. 813-821.
25. Moody, J.,Darken, C. *Learning with localized receptive fields*. 1988. Morgan Kaufmann.

11/20/2012