A Neural Network that Learns to Interpret Myocardial Planar Thallium Scintigrams[*]

Charles Rosenberg, Ph.D.[†]
Department of Computer Science
Hebrew University
Jerusalem, Israel
`crr@cogsci.psych.utah.edu`

Jacob Erel, M.D.
Department of Cardiology
Sapir Medical Center - Meir General Hospital
Kfar Saba, Israel

Henri Atlan, M.D., PhD.
Department of Biophysics
and Nuclear Medicine
Hadassah Medical Center
Jerusalem, Israel

June, 1992

Abstract

The planar thallium-201 (^{201}Tl) myocardial perfusion scintigram is a widely used diagnostic technique for detecting and estimating the risk of coronary artery disease. Interpretation is currently based on visual scoring of myocardial defects combined with image quantitation and is known to have a significant subjective component.

Neural networks learned to interpret thallium scintigrams as determined by both individual and multiple (consensus) expert ratings. Four different types of networks were explored: single-layer, two-layer back-propagation (BP), BP with weight smoothing, and two-layer radial basis function (RBF). The RBF network was found to yield the best performance (94.8% generalization by region) and compares favorably with human experts.

We conclude that this network is a valuable clinical tool that can be used as a reference "diagnostic support system" to help reduce inter- and intra-observer variability. This system is now being further developed to include other variables that are expected to improve the final clinical diagnosis.

1 Introduction

Coronary artery disease (CAD) is one of the leading causes of death in the Western World. The planar thallium-201 is considered to be a reliable diagnostic tool in the detection of CAD. Thallium is a radioactive isotope that distributes in mammalian tissues after intervenous administration and is imaged by a gamma camera. The resulting scintigram is visually interpreted by the physician for the presence or absence of defects — areas with relatively lower perfusion levels. In myocardial applications, thallium is used to measure myocardial ischemia and to differentiate between viable and non-viable (infarcted) heart muscle [16].

Diagnosis of CAD is based on the comparison of two sets of images, one set acquired immediately after a standard effort test (BRUCE protocol), and the second following a delay period of four hours. During this delay, the thallium redistributes in the heart muscle and spontaneously decays. Defects caused by scar tissue are relatively unchanged

---

[*] This paper is to appear in *Neural Computation*.

[†] Supported by grants from the Ministry of Science and Technology, Israel and from the Golda Meir Fund. Current address: GRECC 182, VA Medical Center, 500 Foothill Blvd., Salt Lake City, Utah 84148

over the delay period (fixed defect), while those caused by ischemia are partially or completely filled-in (reversible defect) [2].

Image interpretation is difficult for a number of reasons: the inherent variability in biological systems which makes each case essentially unique, the vast amount of irrelevant and noisy information in an image, and the "context-dependency" of the interpretation on data from many other tests and clinical history. Interpretation can also be significantly affected by attentional shifts, perceptual abilities, and mental state [10, 6].

Neural networks have been applied to several problems in cardiology including the prediction of stenosis [17, 5, 4, 12]. These studies encouraged us to explore the problem of visual interpretation using neural networks which, to our knowledge, has not been previously addressed.

2 Data

Scintigraphic images were acquired for each of three views: anterior (ANT), left lateral oblique (LAO 45), and left lateral (LAT)[1] for each patient case. Each image was first pre-processed[2] and presented to the networks as circumferential profiles [9, 7][3], in which maximum pixel counts within each of 60, 6° contiguous segmental regions are plotted as a function of angle [8] (see Figure 1 A).

Cases were pre-selected based on the following criteria [2]:

- **Insufficient exercise.** Cases in which the heart rate was less than 130 b.p.m. were eliminated, as this level of stress is generally deemed insufficient to accurately distinguish normal from abnormal conditions.
- **Positional abnormalities.** In a few cases, the "region of interest" was not positioned or aligned correctly by the technician.
- **Increased lung uptake.** Typically in cases of multi-vessel disease, a significant proportion of the perfusion occurs in the lungs as well as in the heart, making it more difficult to determine the condition of the heart due to the partially overlapping positions of the heart and lungs.

Twenty percent of the cases were eliminated due to insufficient heart rate, and an additional 5 percent due to either positional abnormalities or increased lung uptake. A set of one hundred usable cases resulted.

Each case was visual scored for each of nine anatomical regions generally accepted as those that best relate to the coronary circulation: *Septal: proximal* and *distal*, *Anterior: proximal* and *distal*, *Apex*, *Inferior: proximal* and *distal*, and *Posterior-Lateral: proximal* and *distal*. Scoring for each region was from *normal* (1) to *severe* (4), indicating the level of the observed perfusion deficit.

Approximately 80% of the cases were visually interpreted by a single expert alone. Ten percent were read by all three experts, and 10% by two experts, with agreement in the scoring reached by consensus.

2.1 Methods

Four different types of networks were explored: a one-layer (sigmoidal) network trained using the delta rule, a two-layer network trained using back-propagation (BP), a two-layer network trained using BP combined with weight smoothing (to be described), and a two-layer network combining one layer of competitive radial-basis function (RBF) units and a second layer of (sigmoidal) units trained using the delta rule.

The general two-layer network architecture is depicted in Figure 1 B. All networks consisted of 180 input units and 9 output units. The input units were divided into three groups. Each group of sixty units encoded the circumferential profile for each view. A relative coding scheme was used in which the sixty ordered segments constituting each view were divided by the segment with the highest scintigraphic count. Each input pattern was then normalized to unit vector

---

[1] Also sometimes referred to as LAO 70 or the "steep LAO" view.

[2] Pre-processing involved positioning of the region of interest (ROI), interpolative background subtraction, smoothing and rotational alignment to the heart's apex [8].

[3] The profiles were generated using the Elscint CTL software package for planar quantitative thallium-201 based on the Cedars-Sinai technique [9, 13, 1].

length. Target numerical values were assigned to the categorical visual scores to make the data suitable for network learning: *normal* = 0.0, *mild defect* = 0.3, *moderate defect* = 0.7, and *severe defect* = 1.0.

2.1.1 Weight smoothing

Weight smoothing “penalized” adjacent fan-in weights to the hidden units in proportion to their difference:

$$C_{ij} = (w_{ij} - w_{i,j-1}) + (w_{ij} - w_{i,j+1}) \tag{1}$$

where $j-1$ and $j+1$ are the left and the right “neighbors” of j. This rule is a special case of one proposed elsewhere [11]. The value of C_{ij} was subtracted from Δw_{ij} on each weight update:

$$w_{ij} = w_{ij} + \Delta w_{ij} - \beta C_{ij} \tag{2}$$

In the present experiments, the smoothing parameter β, was set to 0.01.

2.1.2 Radial Basis Function Networks

RBF networks [14, 15, 3] typically consist of two layers of units: one layer of radial units, typically gaussian, and a second layer of semi-linear units. In our experiments, the activation value of a gaussian unit, O_j, is given by:

$$net_j = \sum_i (w_{ij} - \frac{v_i}{||v||})^2 \tag{3}$$

$$O_j = exp(-\frac{net_j}{\omega}) \tag{4}$$

where j is an index to a gaussian unit, i is an input unit index, and $||v||$ is the length of the current input vector v (for a single view). The width of the gaussian, given by ω, was fixed at 0.25 for all units.

The gaussian units were trained using a competitive learning scheme that moves the center of the unit closest to the current input pattern (O_{max}, i.e. the “winner”) closer to the input pattern:

$$\Delta w_{i,winner} = \eta(\frac{v_i}{||v||} - w_{i,winner}) \tag{5}$$

In addition, the other units were also pulled towards the input pattern, although to a much smaller extent[4]. We used a ratio of 1 in 100.

The second layer was trained using the delta-rule [20]. Input to the second layer consisted of the activations of the entire set of nine gaussian units plus an additional unit per view to encode the scaling factor of the input patterns lost as a result of input vector normalization. This additional unit was clamped to the minimum value of the input pattern for that view[5].

In our experiments, the competitive and the supervised layers were trained sequentially: first competitive then supervised.

2.1.3 Training

Because of the limited number of cases available, we employed the “leave-one-out” or “jack-knife” method of training. We trained 100 separate networks; each network was trained on a subset of 99 of the 100 cases and tested on the remaining one. This procedure enabled us to determine the generalization performance for each case.

Each of the 100 networks was initialized in the normal way; Weights were initialized to small random values between 0.5 and -0.5 and trained with learning rate (η) = 0.05 and momentum (α) = 0.9.

---

[4] Referred to as the *leaky learning model* in [18]

[5] The scaling factor was given as additional input to the output units in the other networks as well to ensure that the that the comparisons among them were fair.

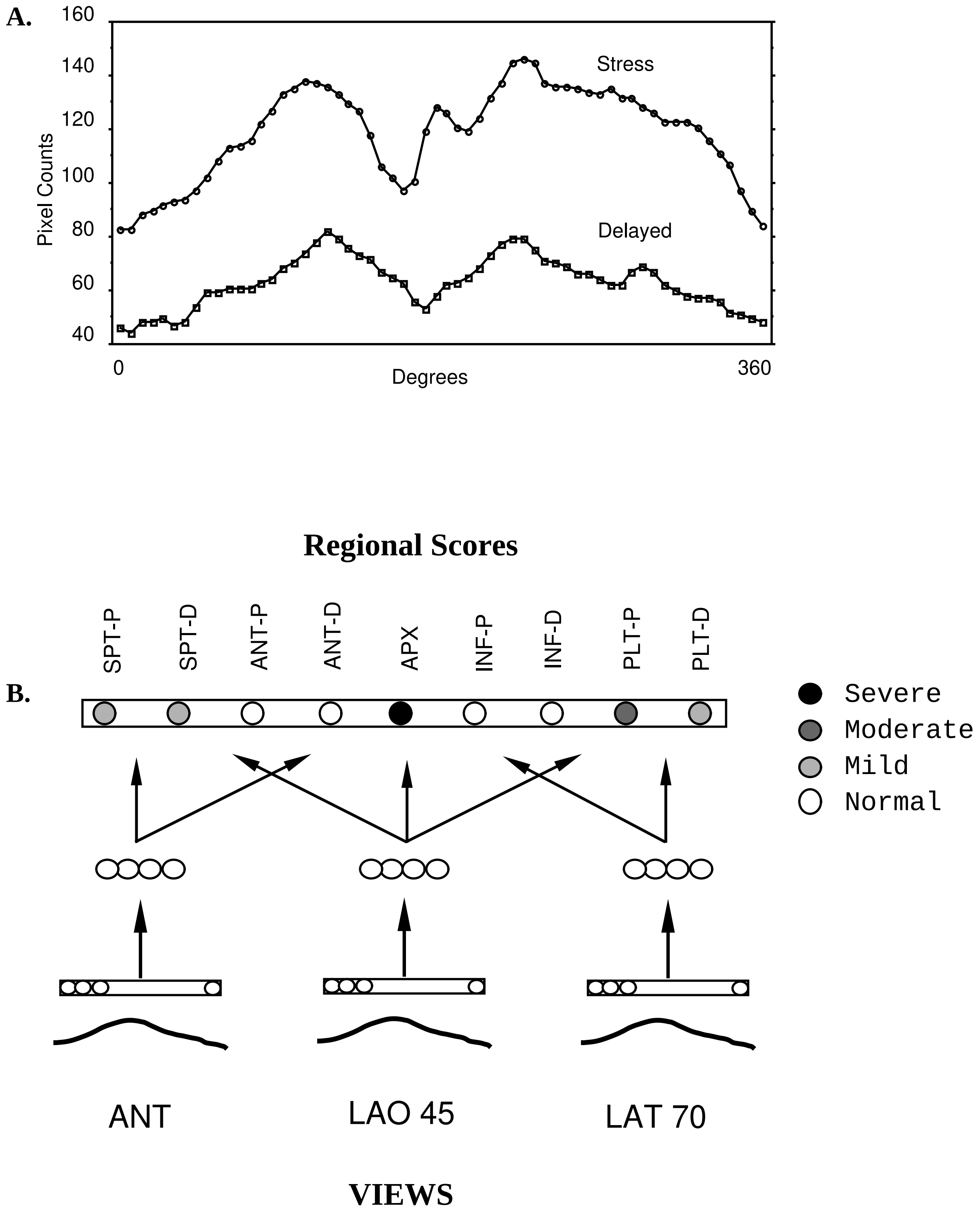

Figure 1: **A)** Circumferential profiles for stress (top) and delayed (bottom) distributions for the left lateral view (LAT). The relatively lower levels of perfusion in the center of both graphs is evidence of a mild-to-moderate defect in the apical region. The defect remains after the delay period (bottom curve), suggesting that the defect is “fixed” and likely due to the presence of scar tissue. **B)** The general two-layer network architecture. See text for explanation.

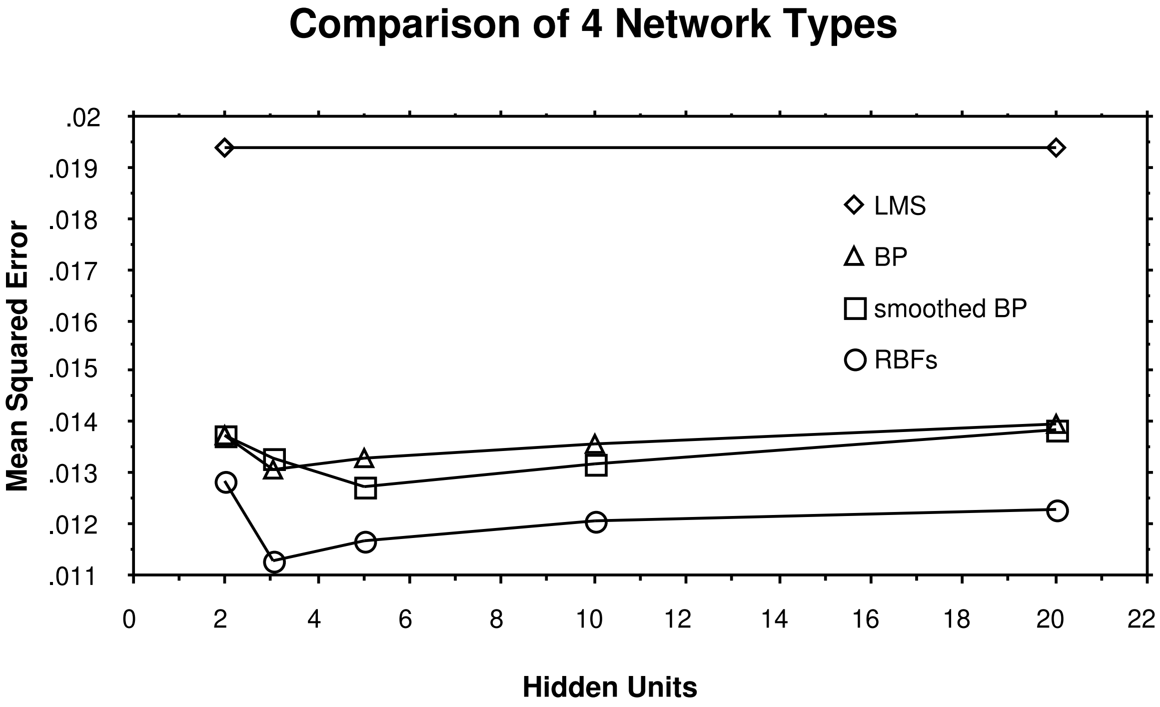

Figure 2: Average mean squared error (MSE) of the four network types as a function of the number of hidden units.

Training was terminated based on cross-validation testing on the remaining case. An optimal level of training was observed for all network types. This was found to be around 40 epochs for the back-propagation and back-propagation with smoothing networks, 100 epochs for the delta rule networks, and 300 epochs for the RBF networks (supervised training portion). Further training always led to over-training and poorer generalization.

The RBF competitive layer was trained for 20 epochs at a learning rate (η) of 0.1, and then for 80 epochs at 0.01. Momentum (α) was not used for competitive learning.

2.2 Results

Performance was examined as a function of the number of hidden (or RBF) units (Figure 2). For all networks, a optimal number of units was found. For the RBF and the BP networks, three hidden units per input group was optimal. For the BP with smoothing networks, 5 was optimal[6].

In order to compare the performance more directly to the target consensus scores, we found decision thresholds which best separated normal and mild defects from moderate and severe, as measured by percent agreement with the consensus scores. In all cases (except for the single-layered network), this decision threshold value turned out to be 0.50 (± 0.01)[7].

Based on these decision thresholds, the performances of the networks in separating normal and mild defects from moderate and severe defects were computed (Figure 4). The best performance was attained by the RBF network. This network agreed with the target consensus scores on 94.8% (853/900) of the regions, sensitivity 57.0% (47/82) and specificity 98.5% (806/818).

---

[6] Although we will not pursue this here, the optimal number of hidden units for the smoothing networks seemed to interact somewhat with smoothing level β. With larger values of β, larger numbers of hidden units were found to be optimal.

[7] The threshold for the single-layered network was found to be 0.78.

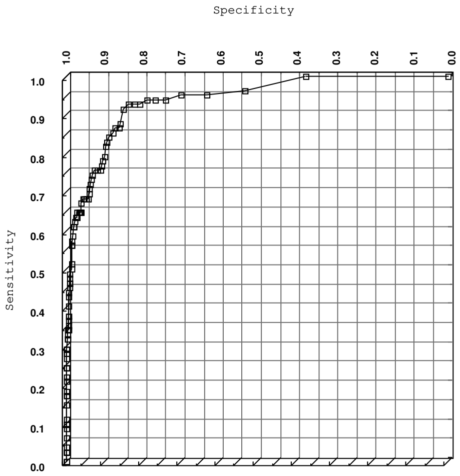

Figure 3: ROC curve (Sensitivity vs. Specificity) for the RBF network with 3 hidden units.

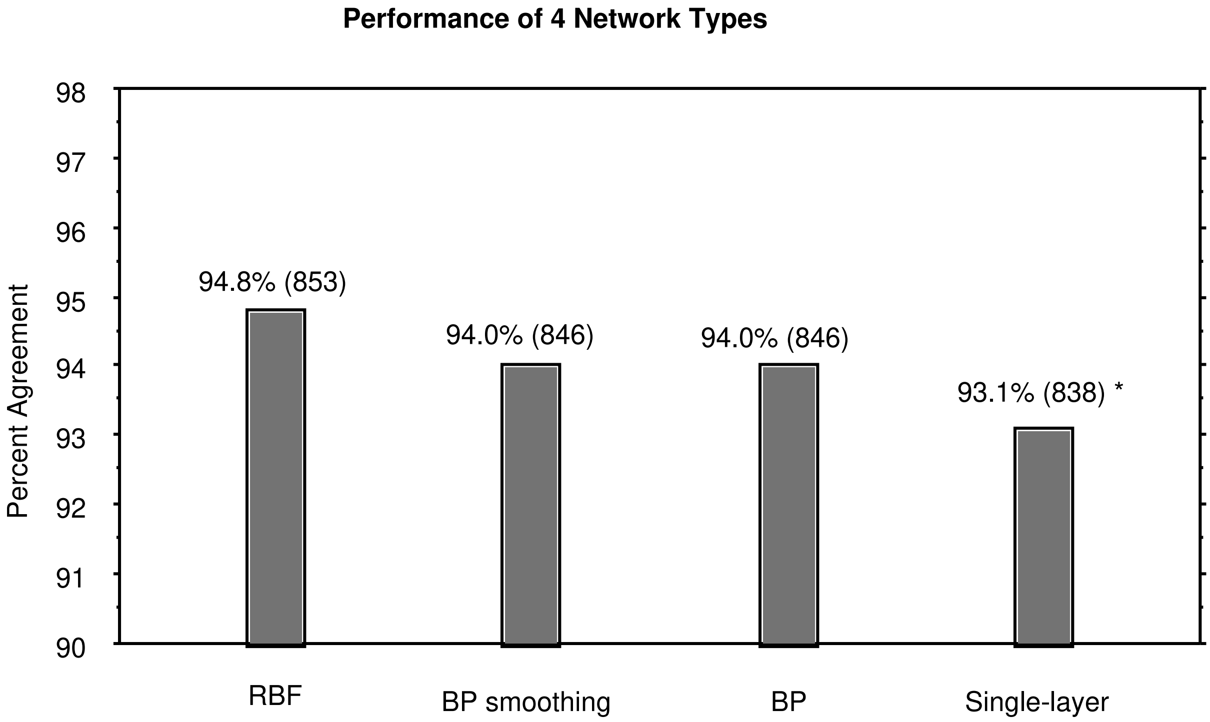

Figure 4: Percent agreement (by region) with the target consensus scores of the best network of each type. There were 9 regions for each of 100 cases or 900 regions in total. The '*' indicates a statistically significant difference, based on MSE, at the $p < 0.05$ level.

Differences among the networks were tested using the non-parametric Friedman test with further evaluation using the two-tailed Wilcoxon signed-rank test for significance ($p < 0.05$). Analysis was based on mean squared differences between outputs and targets (control).

Network type was a significant factor ($\chi_r^2 = 11.244, Df = 3, p < 0.025$). The single-layer network was significantly inferior to all the other networks ($p < 0.001$). However no significant differences were found among the RBF, BP and BP with smoothing networks. The BP with smoothing network was more accurate than the BP network, although this difference was not significant ($p < 0.1$).

2.2.1 Comparison with an Individual Expert

A randomly selected set of fifty-two cases were re-interpreted by one of the participating physicians. Scoring was performed without regard to the original scores. Nine of the cases were previously rated by this physician, either individually or consensus.

In normal or mild versus moderate or severe defect discrimination, this physician's scores agreed with the consensus scores on 97.6% (457/468) of the regions (sens: 89.5% (34/38), spec: 98.4% (423/430)). On the same set of cases, the RBF network's scores agreed with the consensus scores on 95.9% (449/468) of the regions (sens: 63.2% (24/38), spec: 98.8% (425/430)), and agreed with the physician's scores for 94.9% (444/468) of the regions (sens: 56.1% (23/41), spec: 98.6% (421/427)). These differences were not found to be significant in a paired sign test.

3 Conclusions and Future Directions

A network with one hidden layer of RBF units was compared with networks trained with back-propagation, back-propagation with weight smoothing, and single-layer networks trained with the delta-rule. The best performance (94.8% generalization) was attained by an RBF network with 3 hidden units per view (nine in total). This level of performance did not differ significantly from that of an expert. Performance of the single-layer delta rule network was significantly poorer than that of the other networks. The BP with weight smoothing yielded a moderate, though not statistically significant, improvement over standard BP.

Improvements and extensions include:

- **Standardize visual scores.** The data currently is mixture of single expert and consensus scores. Given the likelihood of individual differences in interpretation [10, 6], improved performance should be attained if single expert *or* standardized group data are used.
- **Elicit confidence ratings.** Expert visual interpretations could be augmented by degree of confidence ratings. Highly ambiguous cases could be reduced in importance or eliminated. The ratings could also be used as additional targets for the network[8]: cases indicated by the network with low levels of confidence would require closer inspection by a physician.
- **Provide additional information.** We have not yet incorporated the delayed distribution profiles, clinical history, gender, and examination EKG. The inclusion of these variables should allow the network to approximate more closely a complete diagnosis, and boost the utility of the network in the clinical setting.
- **Add constraints.** Currently we do not utilize the angles that relate the three views. It may be possible to build these angles in as constraints and thereby cut down on the number of free network parameters.
- **Expand application.** Besides planar thallium, our approach may also be applied to non-planar 3-D imaging technologies and other nuclear agents or stress-inducing modalities. Preliminary results are promising in this regard.

---

[8] See [19] for a related idea.

We believe that our study is the first to successfully apply neural networks to the prediction of the human expert's visual interpretation of myocardial scintigraphy. Our network can be used *today* as a standard reference tool to help the physician reduce response bias and other types of scoring variability and may, with the improvements suggested, eventually equal or surpass the interpretive abilities of the physician.

Acknowledgements

The authors wish to thank Prof. Benny Shanon for much kind assistance and support throughout this research project, Mr. Haim Karger for technical assistance, and the Departments of Computer Science and Psychology at the Hebrew University for computational support. We would also like to thank Drs. David Shechter, Moshe Bocher, Roland Chisin and the staff of the Department of Medical Biophysics and Nuclear Medicine for their help, both large and small. This paper has also benefitted from the comments of two anonymous reviewers. Terry Sejnowski suggested our usage of RBF units.

References

[1] J. Areeda, K. Van Train, E. V. Garcia, J. Maddahi, A. Rosanki, A. Waxman, and D. Berman. Improved analysis of segmental thallium-201 myocardial scintigrams: Quantitation of distribution, washout, and redistribution. In P. D. Esser, editor, *Digital Imaging*. Society of Nuclear Medicine, New York, 1982.

[2] G. A. Beller. Myocardial perfusion imaging with thallium-201. In M. L. Marcus, H. R. Schelbert, D. J. Skorton, and G. L. Wolf, editors, *Cardiac Imaging*. W. B. Sanders, 1991.

[3] D.S. Broomhead and D. Lowe. Multivariable functional interpolation and adaptive networks. *Complex Systems*, 2:321–355, 1988.

[4] D. Cianflone, O. Carandente, G. Fragasso, A. Margononato, C. Meloni, E. Rossetti, P. Gerundini, and S. L. Chiechia. A neural network based model of predicting the probability of coronary lesion from myocardial perfusion spect data. In *Proceedings of the 37th Annual Meeting of the Society of Nuclear Medicine*, page 797, May 1990.

[5] K. J. Cios, L. S. Goodenday, M. Merhi, and R. Langenderfer. Neural networks in detection of coronary artery disease. In *Computers in Cardiology Conference*, September 1989.

[6] A. Cuarón, A. Acero, M. Cárdena, D. Huerta, A. Rodríguez, and R. de Garay. Interobserver variability in the interpretation of myocardial images with tc-99m-labeled diphosponate and pyrophosphate. *Journal of Nuclear Medicine*, 21(1):1–9, 1980.

[7] D. A. Francisco, S. M. Collins, and R. T. Go et al. Tomographic thallium-201 myocardial perfusion scintigrams after maximal coronary artery vasodiliation with intravenous dipyridamole: Comparison of qualitative and quantitative approaches. *Circulation*, 66(2), 1982.

[8] E. V. Garcia. Physics and instrumentation of radionuclide imaging. In M. L. Marcus, H. R. Schelbert, D. J. Skorton, and G. L. Wolf, editors, *Cardiac Imaging*. W. B. Sanders, 1991.

[9] E. V. Garcia, J. Maddahi, D. S. Berman, and A. Waxman. Space-time quantitation of thallium-201 myocardial scintigraphy. *Journal of Nuclear Medicine*, 22:309–317, 1981.

[10] E. A. Franken Jr. and K. S. Berbaum. Perceptual aspects of cardiac imaging. In M. L. Marcus, H. R. Schelbert, D. J. Skorton, and G. L. Wolf, editors, *Cardiac Imaging*. W. B. Sanders, 1991.

[11] K. J. Lang and G. E. Hinton. Dimensionality reduction and prior knowledge in e- set recognition. In David S. Touretzky, editor, *Advances in Neural Information Processing Systems 2*, pages 178–185, San Mateo, 1990. Morgan Kaufmann.

[12] S.C. Lee. Using a translation-invariant neural network to diagnose heart arrhythmia. In David S. Touretzky, editor, *Advances in Neural Information Processing Systems 2*, pages 240–247, San Mateo, 1990. Morgan Kaufmann.

[13] J. Maddahi, E. V. Garcia, D. S. Berman, A. Waxman, H. J. C. Swan, and J. Forrester. Improved noninvasive assessment of coronary artery disease by quantitative analysis of regional stress myocardial distribution and washout of thallium-201. *Circulation*, 64:924–935, 1981.

[14] J. Moody and C. Darken. Fast learning in networks of locally-tuned processing units. *Neural Computation*, 1:281–294, 1989.

[15] T. Poggio and F. Girosi. Regularization algorithms for learning that are equivalent to multilayer networks. *Science*, 247:978–982, 1990.

[16] G. M. Pohost and M. J. Henzlova. The value of thallium-201 imaging. *New England Journal of Medicine*, 323(3):190–192, 1990.

[17] G. Porenta, S. Kundrat, G. Dorffner, P. Petta, J. Duit, and H. Socher. Computer based image interpretations of thallium- 201 scintigrams: Assessment of coronary artery disease using the parallel distributed processing approach. In *Proceedings of the 37th Annual Meeting of the Society of Nuclear Medicine*, page 825, May 1990.

[18] D. E. Rumelhart and D. Zipser. Feature discovery by competitive learning. In D. E. Rumelhart and J. L. McClelland, editors, *Parallel Distributed Processing*, volume 1, chapter 5, pages 151–193. MIT Press, Cambridge, Mass., 1986.

[19] G. Tesauro and T. J. Sejnowski. A parallel network that learns to play backgammon. Technical Report CCSR-88-2, University of Illinois at Urbana-Champaign Center for Complex Systems Research, February 1988.

[20] B. Widrow and M.E. Hoff. Adaptive switching circuits. In *1960 IRE WESCON Convention Record*, volume 4, pages 96–104. IRE, New York, 1960.