Control and Cybernetics

vol. **26** (1997) No. 3

Initial results of training neural networks to detect breast cancer using evolutionary programming

by

David B. Fogel*, Eugene C. Wasson**, Edward M. Boughton***,
Vincent W. Porto*, Jamen W. Shively*

* Natural Selection, Inc., 3333 N. Torrey Pines Ct., Suite 200,
La Jolla, CA 92037
E-mail: {dfogel,bporto,jshively}@natural-selection.corn

** Maui Memorial Hospital, 221 Mahalani,
Wailuku, HI 96793
E-mail: wasson@maui.net

*** Hawaii Industrial Laboratory, Inc., P.O. Box 1275,
Wailuku, HI 96793
E-mail: boughton@maui.com

Abstract: Artificial neural networks arc applied to the problem of detecting breast cancer from radiographic features and patient age. Evolutionary programming is used to train neural networks based on sigmoid or Gaussian kernel functions. Preliminary results on 96 biopsy-proven cases (62 malignant, 34 benign) indicate that a reasonable probability of detecting malignancies can be achieved using simple neural architectures. The features appear to be more amenable to discrimination by partitioning functions than to clustering functions, although final analysis remains for larger sample sizes.

Keywords: breast cancer, neural networks, evolutionary programming

1. Introduction

Carcinoma of the breast is second only to lung cancer as a tumor-related cause of death in women. There arc now more than 180,000 new cases and 45,000 deaths annually in the United States alone. It begins as a focal curable disease, but it is usually not identifiable by palpation at this stage, and mammography remains the mainstay in effective screening. It has been estimated that the mortality from breast carcinoma could be decreased by as much as one-third if all women in the appropriate age groups were regularly screened.

Computer technology offers many potential benefits to the radiologist, including computer-aided diagnosis. There is currently considerable intra- and inter-observer disagreement or inconsistencies in mammographic interpretation. This has led to an interest in the possibility of utilizing computerized pattern recognition algorithms, such as artificial neural networks (ANNs), to assist in the decision-making required in the assessment of mammograms. ANNs have been demonstrated to be useful in many engineering pattern recognition applications and these techniques hold promise for improving the accuracy of determining those patients where further assessment and possible biopsy is indicated. Furthermore, there should also be an eventual cost saving when a reliable automated screening system can be developed. The successful development of a neural network that is capable of reliably assessing the potential for the existence of breast carcinoma based on radiographic features of mammograms would make the radiologist both more efficient and more effective.

ANNs are models based on the neuronal structure of natural organisms (Haykin, 1994). They arc stimulus-response transfer functions that accept some input and yield some output. They are typically used to learn an input-output mapping over a set of examples. For example, as will be described here, the input can be radiographic features from mammograms, with the output being a decision regarding the likelihood of a malignancy. Hornik et al. (1989) and Poggio and Girosi (1990) have proved that neural networks with sigmoid or Gaussian basis functions in a single hidden layer can in principle generate any measurable mapping, indicating the versatility of these functions.

Given a network architecture (i.e., type of network, the number of nodes in each layer, the connections between the nodes, and so forth), and a training set of input patterns, the collection of variable weights determines the output of the network to each presented pattern. The error between the actual output of the network and the desired target output defines a response surface over a hyperspace having a dimension equal to the number of weights. A commonly employed method for finding weight sets in such applications is error *back propagation*, which is essentially a gradient method. As such, it is subject to entrapment in locally optimal solutions, and the resulting weight sets arc often unsuitable for practical applications. Numerical optimization techniques that do not suffer from such entrapment can be used to advantage in these cases.

Evolutionary algorithms offer one such technique. In these stochastic optimization methods, a population of candidate solutions is maintained, and random variation (m1ltation and/or recombination) and selection are imposed on the population to guide it to appropriate regions of the hyperspace. The use of random variation to bias the search avoids entrapment in local optima, and there arc several mathematical proofs that variations of these procedures provide asymptotic global convergence, rather than merely local convergence (Fogel, 1994; Rudolph, 1994; Back, 1996). Moreover, there is empirical evidence that the methods arc robust to many pathologies in possible response surfaces, including multiple minima or maxima, constraints, disjoint feasible regions, and random perturbations (Schwefel, 1995; Fogel, 1995; Michalewicz, 1996; and others).

There have been many efforts to train neural networks using evolutionary algorithms (Fogel et al., 1990; Angeline et al., 1994; McDonnell and Waagen, 1994; Yao and Liu, 1996; and many others). This paper describes the results of preliminary efforts to use evolutionary programming to train simple ANNs to respond to a set of radiographic features from film screen mammograms, along with the patient's reported age, to make a determination regarding the presence or absence of a malignant condition. It begins with a brief review of selected efforts to use neural networks in breast cancer detection, before describing the current methods and results.

2. Background

Neural networks have been recelving recent attention in medical diagnostics (Brotherton and Simpson, 1995; Rizki et al., 1995; and others). With regard to detecting breast cancer, efforts have been directed at classifying histologic data f 'om cells removed by fine needle aspiration (Wolberg et al., 1994, 1995) and radiographic features from film screen mammography (Kocur et al., 1995; and others). Three of these efforts are reviewed here.

The investigation of Wu et al. (1993) used 43 preselected features related to density, microcalcification, parenchymal distortion, skin thickening, correlation with clinical findings, and so forth. Data was taken from 133 textbook cases in Tabar and Dean (1985). For each mammogram, each of the selected features was rated by an experienced mammographer on a scale of 0-10, and this served as the vector input to a multilayer perceptron neural network (i.e., feedforward and fully connected). The network possessed 10 hidden units and a single output unit which was trained to yield a value of 0.0 for a benign case and 1.0 for a malignancy. Training was accomplished using back propagation. The results of this preliminary study and other described experiments indicated the suitability of this approach. By pruning the feature set to a more reasonable, smaller collection, the neural network was able to statistically outperform an attending radiologist and residents in assessing patterns of mammographic image features that arc associated with benign and malignant lesions. There was no statistically significant difference between the performance of the network and the experienced mammographer used to rate each of the image features.

Floyd et al. (1994) used back propagation on multilayer perceptrons to predict breast cancer from mammographic findings from patients who were scheduled for biopsy. They used only eight input parameters (mass size, mass margin, asymmetric density, architectural distortion, calcification number, calcification morphology, calcification density, and calcification distribution) and each of these was paramcterized less subjectively than in Wu et al. (1993). There were

260 cases used for training and testing. Data was not separated into complementary training and testing sets; all of the exemplars were processed using a jackknife statistical procedure. After significant training, the results indicated that if a threshold value of 0.1 were used (output on a scale from [-1, 1]), 38 out of 168 benign cases and all 92 malignancies would be identified. The authors compared this performance to that of radiologists and suggested that these results were statistically significantly better than radiologists at a P < 0.08 level. Although their results do appear fairly impressive with regard to detecting maligrnmcy, the number of false alarms is somewhat high (23%), and the statistical validity of the hypothesis test carried out can be questioned because the threshold of 0.1 was chosen after the authors reviewed the data and statistics were compiled on those same data. Thus, the data did not reflect a random sample, but rather a biased sample. New data would have to be tested at the threshold value of 0.1 to ensure a sound statistical procedure.

Wilding et al. (1994) used back propagation on multilayer perceptrons to assess both breast and ovarian cancer. Their procedure was similar to Wu et al. (1993) and Floyd et al. (1994), except that their input parameters consisted mainly of objective blood specimens and analyses (maximum of 10 total input parameters) from 104 patients. Unfortunately, Wilding et al. (1994) reported that the neural network was able to "provide little improvement on the sensitivity of testing comparing to the use of [the tumor marker] CA 15-3 only. Furthermore, it would appear that none of the networks appear to identify any worthwhile parameters or operating conditions with clinical utility".

Wu et al. (1993), Floyd et al. (1994), and Wilding et al. (1994) each used back propagation to determine the weights of their neural networks. But networks trained by gradient methods may require many more hidden nodes to train to a tolerable level of error than are actually required because the method may converge at suboptimal weight sets. Adding more weights (i.e., degrees of freedom) can help overcome local optima and offer the possibility for suitable training, but overparameterized networks may not generalize well on new data. These concerns were specifically discussed in Floyd et al. (1994) and Wilding et al. (1994). The most effective methods employed in these investigations for limiting the number of nodes and network parameters were based on sensitivity analysis and ad hoe pruning. Sensitivity analysis is problematic on nonlinear transfer functions (such as neural networks) and ad hoe pruning can be largely unproductive. Despite directly mentioning concerns about overfitting their data, Floyd et al. (1994) found that their best performance occurred when using 177 weights (16 hidden nodes), but they used only 260 samples. Wilding et al. (1994) used networks with as few as 38 weights and as many as 132 weights, and despite having 104 samples were still unable to generate satisfactory performance. Even if the blood statistics that were being used were not particularly relevant to the classification task at hand, the failure to find suitable networks with more parameters than data indicates the limitations of the training method and suggests alternative methods for optimizing classification networks, such as evolutionary computation.

3. Method

Data for the current effort were collected by assessing film screen mammograms in light of a set of radiographic features as determined by the domain expert (Wasson). The features selected paralleled those of Floyd et al. (1994) with some important modifications. Under the system of Floyd et al. (1994), certain features were described as lying on a continuum when it appeared more useful to rate these features independently. For example, Floyd et al. (1994) rated mass margin with six categories: (1) no mass (value 0.0), (2) well circumscribed (value 0.2), (3) microlobulated (value 0.4), (4) obscured (value 0.6), (5) indistinct (value 0.8), and (6) and spiculated (1.0). In contrast, the current parameterization rated the five categories of masses (see (2)-(6) in Table 1) in four levels [0, 1, 2, 3] as none, low, medium, and high. The complete set of radiographic features used appears in Table 1. In addition, the age of the patient was considered leading to a total of 13 input features. These features were assessed in 96 cases all of which subsequently had open surgical biopsy of the area of concern, with the associated pathology indicating whether or not a malignant condition had been found. In all, 62 cases were associated with a biopsy-proven malignancy, while 34 cases were indicated to be negative by biopsy (although the possibility remains that such an indication may be in error).

These data were processed using two forms of neural networks: (1) multilayer perceptrons and (2) receptive fields. Each network architecture was restricted to two hidden nodes, with a linear output node, resulting in 31 adjustable weights (see Figure 1). The perceptron network used a sigmoid filter on each hidden node of *J(/3)* = (1+exp(-fJ))-^{1}, where (3 is the sum of a bias term and the clot product of the input feature vector and the associated weight vector. The receptive field network used a Gaussian filter on each hidden node *off*(/3) = (27r)-$^{0.5}$ exp(-(3^{2}). Evolutionary programming was used to train the networks in a leave-one-out cross validation procedure.

Specifically, for each complete cross validation where each sample pattern in turn was held out for testing then replaced in a series of 96 separate training procedures, a population of 250 networks of the chosen architecture were selected at random by sampling weight values from a uniform random variable distributed over [-0.5, 0.5]. Each weight set (i.e. candidate solution) also had an associated self-adaptive mutational vector used to determine the random variation imposed during the generation of offspring networks (described below). Each of the self-adaptive parameters was initialized to a value of 0.01. Each weight set was evaluated based on how well the network classified the 95 remaining available training patterns (with one "left out" for testing), where a malignant condition was assigned a target value of 1.0 and a benign condition was assigned a target of 0.0. The performance of each network was determined as the sum of the squared error between the output and the target value taken

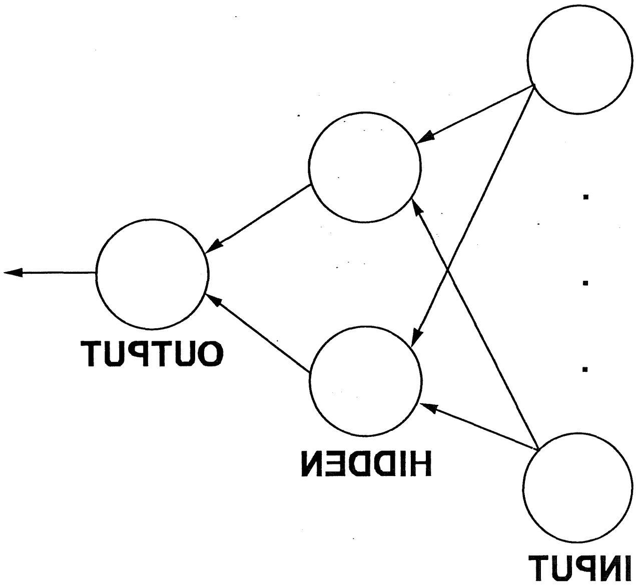


Figure 1. The design used for processing data, both for the multilayer perceptron and receptive field neural architectures. Input data are weighted in connections to the two hidden nodes. Each hidden node passes the sum of a bias term (not shown) and the dot product of the weights and inputs through a nonlinear filter. The filter is f(β) = 1/(1 + e-fl) for the multilayer perceptron, and f(β)= (27r)$^{-0.5}$e-fl^{2} for the receptive field. The output node is a linear filter which performs the sum of a bias term with the dot product of filtered hidden nodes and their associated weights. There are 31 weights given 13 inputs.

1. Mass size: either zero or in mm.
2. Mass margin: (each subparameter rated as none (0), low (1), medium (2), or high (3))
   (a) Well circumscribed
   (b) Microlobulated
   (c) 0 bscured
   (d) Indistinct
   (e) Spiculated
3. Architectural distortion: none or distortion
4. Calcification number: none (0), < 5 (1), 5 - 10 (2), or > 10 (3).
5. Calcification morphology: none (0), not suspicious (1), moderately suspicious (2), or highly suspicious (3)
6. Calcification density: none (0), dense (1), mixed (2), faint (3)
7. Calcification distribution: none (0), scattered (1), intermediate (2), clustered (3)
8. Asymmetric density: either zero or in mm.

Table 1. The features and rating system used for assessing mammograms in the current study. Assessment was made by the domain expert (Wasson).

over the 95 available patterns.

After evaluating all existing (parent) networks, the 250 weight sets were used to generate 250 offspring weight sets (one offspring per parent). This was accomplished in a two-step procedure. For each parent, the self-adaptive parameters were updated as:

$$\sigma_i' = \sigma_i \exp\left(\tau N(0,1) + \tau' N_i(0,1)\right) \qquad (1)$$

where τ = $\frac{-b:-}{Vh}$, τ' = $\frac{-b,}{y2'1,i}$ N(0, 1) is a standard normal random variable sampled once for all n = 31 parameters of the vector Q, and Ni(0, 1) is a standard normal random variable sampled anew for each parameter. The settings for T and τ' have been recommended as robust in Back and Schwefel (1993). These updated self-adaptive parameters were then used to generate new weight values for the offspring according to the rule:

$$x = xi + CJC \qquad (2)$$

where C is a standard Cauchy random variable

$$f(y) = \frac{1}{7r(1 +y2)'} \; -oo < Y < 00 \qquad (3)$$

(determined as the ratio of two independent standard Gaussian random variables). Traditional methods in evolutionary programming and evolution strategies have relied on Gaussian mutation, however, recent research in Yao and Liu

(1996), Saravanan and Fogel (1997), and others, have suggested a possible benefit of using Cauchy variation because it has a greater probability to generate longer jumps than the Gaussian. This offers a greater chance of escaping local optima on a error surface at the expense of poorer fine tuning. Initial observations with both Gaussian and Cauchy mutations on the existing data appeared to favor the Cauchy distribution, however a more careful analysis remains for future study. All of the offspring weight sets were evaluated in the same manner as their parents,

Selection was applied to eliminate half of the total parent and offspring weight sets based on their observed error performance. Following typical methods in evolutionary programming (Fogel 1995), a pairwise tournament was conducted where each candidate weight set was compared against a random sample from the population. The sample size was chosen as 10 (a greater sample size indicates more stringent selection pressure). For each of the 10 comparisons, if the weight set had an associated classification error score that was lower than the randomly sampled opponent it received a "win". After all weight sets had participated in this tournament, those that received the greatest number of wins were retained as parents of the next generation.

This process was iterated for 100 generations, whereupon the best available network as measured by the training performance was used to classify the held out input feature vector. The result of this classification was recorded (i.e., the output value of the network and the associated target value) and the process was restarted by replacing the held out vector and removing the next vector in succession until all 96 patterns had been classified.

Each complete series of cross validation was repeated 10 times for both the multilayer perceptron and receptive field networks.

4. Results

A typical rate of optimization in each training run is shown in Figure 2. The overall training error often fell as a nearly linear function of the number of generations without saturation. This suggests that further training time might be warranted.

The probability of detection, *P(D),* and false alarm, *P(FA),* vary as a function of the discrimination threshold applied to the output of the networks. As the threshold value is lower, the network can correctly identify a greater number of cancers, but this comes at the expense of a higher false alarm rate. Conversely, the false alarm rate can be lowered by raising the threshold value, but this in turn decreases the sensitivity of the procedure.

The effectiveness of the classification procedures can be assessed using receiver operating characteristic (ROC) analysis, where the probability of detecting a malignancy is traded off as a function of the likelihood of a false positive result. Typical ROC curves for the multilayer perceptron and receptive field

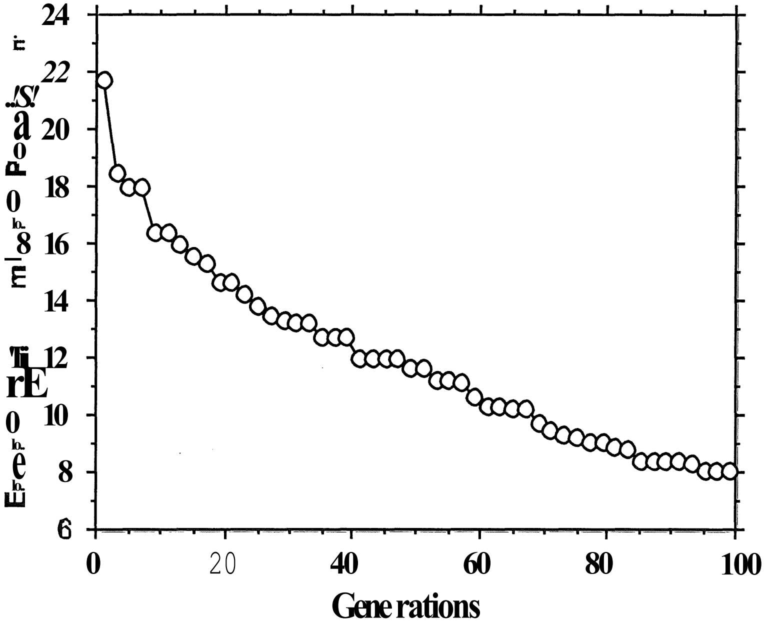


Figure 2. A typical rate of optimization in an evolutionary training of the neural networks on 95 patterns (one held out) over 100 generations. The error of the best member (weight set) in the population is seen to decrease nearly linearly as function of the number of generations. With further training, the observed best error would eventually saturate at an asymptote. The error is taken as the sum of the squared difference between the target value and the realized output from the network generated as a result of each input pattern.

networks are shown in Figure 3. The area under the curve provides a useful measure for comparison.

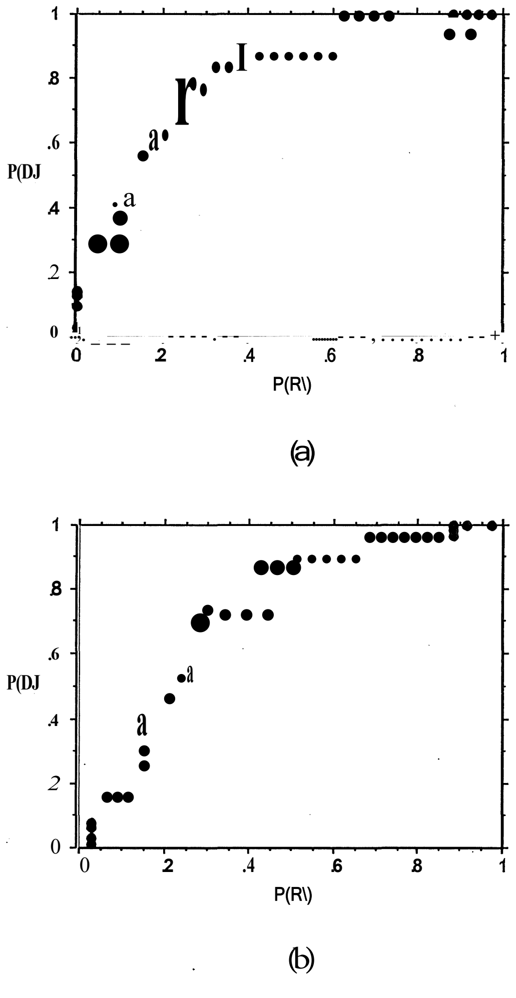
To compare the effectiveness of the multilayer perceptron architecture with the receptive field, the area under the ROC curve for each of the 10 trials of cross validation with each method was estimated. This was accomplished by performing a polynomial regression of at least third order to the available samples of fraction of false alarms versus fraction of detections in each ROC curve. Regression models were determined by choosing the lowest order polynomial that provided (1) an R^2 value of at least 0.99, and (2) was non-decreasing over false alarm rates from zero to one (see Figure 4). The models were integrated over the range [0, 1] to compute the desired areas. The mean area under the ROC curve (and standard deviation) for the perceptron and receptive field networks, respectively, were 0.787611 ± 0.022346 and 0.739060 ± 0.035574. Under a two-sample meant-test (which assumes populations of normally distributed values), these data indicate statistically significant evidence in favor of the perceptron (sigmoid) networks ($P < 0.01$).

5. Conclusions

Under the assumptions of normally distributed integrals of the ROC curves, the data suggest that partitioning functions (as offered by sigmoid filters) may be more useful than clustering functions (as offered by Gaussian filter:s) for classifying the radiographic features and patient age as being indicative of a breast malignancy. The longer-term relevance of this result is yet unclear due at least to (1) the relatively small sample size, and (2) the constraint that all patterns were derived from mammograms that presented sufficient radiographic findings to suggest biopsy. Current efforts are directed to obtaining a larger sample.

Comparisons of the overall performance offered here with the results offered in Floyd et al. (1994) must be made with caution. The composition of the 260 samples in Floyd et al. (1994) was 64.6% benign cases, with only 35.4% malignancies. In contrast, the current data set offered almost the obverse conditions. Further, the demographics between the studies were different. The data in Floyd et al. (1994) were derived from examinations at Duke University Medical Center, whereas the data for the current study were collected from radiology centers on the island of Maui, which can be expected to provide a more diverse racial mix (24% part-Hawaiian, 22% Caucasian, 17% Japanese, and so forth). This greater diversity might be expected to pose a more significant challenge for a classification algorithm.

In separate analysis, evolutionary programming was used to train a two-hidden node perceptron over the entire 96 available patterns. Using an output threshold of 0.5 (i.e., greater than 0.5 indicates a diagnosis of a malignancy), it was possible, after more than 2000 generations, to find a weight vector that misclassified only 3 of the 96 patterns (i.e., it was in error on two malignancies



(a)

(b)

Figure 3. Typical ROC curves for the (a) perceptron and (b) receptive field neural networks. As the probability of a false alarm (i.e., indication of malignancy when none is present) increases, so does the probability of detection.

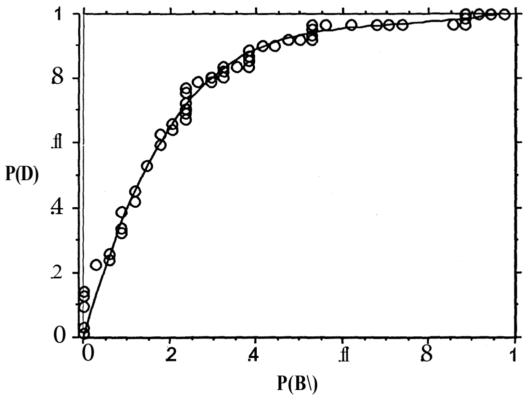


Figure 4. An example of using polynomial regression to estimate the ROC curve based on the observed pairs of probabilities for false alarm and detection. The equation shown is $P(D) = 4.742 \text{ x } P(FA) - 8.897 \text{ x } P(FA)^2 + 7.452 \text{x} P(FA)^3 - 2.295 \text{ X } P(FA)^4$. The goodness-of-fit is $R^2 = 0.999$. Note that the regression equation is constrained to pass through the origin. Regressions were conducted for each of the 10 complete cross validation studies with both the perceptron and receptive field networks in order to determine the area under the approximate ROC curve.

and one benign case). Yet when this same architecture was used in the cross validation trials, this degree of overall performance was not attained. This suggests that (1) further training in the cross validation trials may be useful, and/or (2) the current neural architecture overfits the available data, in which case future analysis on a larger collection of samples should yield a closer correspondence between the error rates when training on all available data and when training/testing in cross validation. Other possibilities for improving the discrimination performance of the evolved networks include imposing small amounts of random noise on the input patterns to increase the possible generalizability (i.e., essentially creating a larger sample size) and reducing the degrees of freedom of the neural networks by limiting the input parameters to a subset ofthe current assortment.

Acknowledgments

This effort is sponsored by the Department of the Army, Grant No. DAMD17-96-1-6312. The content of the information presented docs not necessarily reflect the position or the policy of the government, and no official endorsement should be inferred. Thanks are owed to M.A. Williams for facilitating the use of the SP-2 at the Maui High Performance Computing Center. Research sponsored in part by the Phillips Laboratory, Air Force Materiel Command, USAF, under cooperative agreement number F29601-93-2-0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of Phillips Laboratory or the U.S. Government.

References

ANGELINE, P.J., SAUNDERS, G.M. and POLLACK, J.B. (1994) An evolutionary algorithm that constructs recurrent neural networks. *IEEE Trans. Neural Networks,* 5, 1, 54-65.

BACK, T. (1996) *Evolutionary Algorithms in Theory and Practice,* Oxford, NY.

BACK, T. and ScHWEFEL, H.-P. (1993) An overview of evolutionary algorithms in parameter optimization. *Evolutionary Computation,* 1, 1, 1-24.

BROTHERTON, T.W. and SIMPSON, P.K. (1993) Dynamic feature set training of neural nets for classification. *Evolutionary Programming IV: Proceedings of the Fourth Annual Conference on Evolutionary Programming,* McDonnell, J.R., Reynolds, R.G. and Fogel, D.B., eds., MIT Press, Cambridge, MA, 83-94.

FOGEL, D.B. (1994) Asymptotic convergence properties of-genetic algorithms and evolutionary programming: Analysis and experiments. *Cybernetics and Systems,* 25, 3, 389-407.

FOGEL, D.B.(1995) *Evolutionary Computation: Toward a New Philosophy of Machine Intelligence.* IEEE Press, NY.

FOGEL, D.B., FOGEL, L.J. and PORTO, V.W. (1990) Evolving neural networks. *Biological Cybernetics,* 63, 6, 487-493.

FLOYD, C.E., Lo, J.Y., YUN, A.J., SULLIVAN, D.C. and KORNGUTH, P.J. (1994) Prediction of breast cancer malignancy using an artificial neural network. *Cancer,* 74, 2944-2998.

HAYKIN, S. (1994) *Neural Networks: A Comprehensive Foundation.* MacMillan, NY.

HORNIK, K., STINCHCOMBE, M. and WHITE, H. (1989) Multilayer feedforward networks are universal approximators. *Neural Networks,* 2, 359-366.

KOCUR, C.M., ROGERS, S.K., BAUER, K.W. and STEPPE, J.M. (1995) Neural network feature selection for breast cancer diagnosis. *Applications and Science of Artificial Neural Networks,* Rogers, S.K. and Ruck, D.W., eds., Proc. SPIE 2492, 905-918.

McDONNELL, J.R. and WAAGEN, D. (1994) Evolving recurrent perceptrons for time-series modeling. *IEEE Trans. Neural Networks,* 5, 1, 24-38.

MICHALEWICZ, Z. (1996) *Genetic Algorithms + Data Structures = Evolution Programs.* 3rd ed., Springer, Berlin.

POGGIO, T. and GIROSI, F. (1990) Networks for approximation and learning. *Proc. of the IEEE,* 78, 9, 1481-1497.

RIZKI, M.M., TAMBURINO, L.A. and ZMUDA, M.A. (1995) Evolution of Morphological Recognition Systems. *Evolutionary Programming IV: Proceedings of the Fourth Annual Conference on Evolutionary Programming,* McDonnell, J.R., Reynolds, R.G. and Fogel, D.B., eds., MIT Press, Cambridge, MA, 95-106.

RUDOLPH, G. (1994) Convergence analysis of canonical genetic algorithms. *IEEE Trans. Neural Networks,* 5, 1, 96-101.

SARAVANAN, N. and FOGEL, D.B. (1997) Multi-operator evolutionary programming. *Evolv.tionary Programming VI: Proceedings of the Sixth Annv.al Conference on Evolutionary Pro_qramming,* Angeline, P.J., Eberhart, R.C., Reynolds, R.G. and McDonnell, J.R., eds., Springer, Berlin, in press.

SCHWEFEL, H.-P. (1995) *Evolv,tion and Optimum Seeking,* John Wiley, NY.

TABAR, L. and DEAN, P.B. (1985) *Teaching Atlas of Mammography,* 2nd ed., Thieme-Stratton, NY.

WILDING, P., MORGAN, M.A., GRYGOTIS, A.E., SHOFFNER, M.A. and ROSATO, E.F. (1994) Application of backpropagation neural networks to diagnosis of breast and ovarian cancer. *Cancer Letters,* 77, 145-153.

WOLBERG, W.H., STREET, W.N., HEISEY, D.M. and MANGASARIAN, O.L. (1995) Computerized breast cancer diagnosis and prognosis from fine-needle aspirates. *Arch. Surg.,* 130, 511-516.

WOLBERG, W.H., STREET, W.N. and MANGASARIAN, O.L. (1994) Machine learning techniques to diagnose breast cancer from image-processed nuclear features of fine needle aspirates. *Cancer Letters,* 77, 163-171.

Wu, Y.Z., GIGER, M.L., DOI, K., VYBORNY, C.J., SCHMIDT, R.A. and METZ, C.E. (1993) Artificial neural networks in mammography: Application to decision making in the diagnosis of breast cancer. *Radiology,* 187, 1, 81-87.

YAO, X. and LIU, Y. (1996) Evolving artificial neural networks through evolutionary programming. *Evolutionary Programming V: Proceedings of the Fifth Annual Conference on Evolntionary Programming,* Fogel, L.J., Angeline, P.J. and Back, T., eds., MIT Press, Cambridge, MA, 257-266.