ELSEVIER

Cancer Letters 119 (1997) 93–97

CANCER LETTERS

A step toward computer-assisted mammography using evolutionary programming and neural networks

David B. Fogel[a,*], Eugene C. Wasson[b], Edward M. Boughton[c], Vincent W. Porto[a]

[a]Natural Selection, Inc., 3333 N. Torrey Pines Ct., Suite 200, La Jolla, CA 92037, USA
[b]Maui Memorial Hospital, 221 Mahalani, Wailuku, HI 96793, USA
[c]Hawaii Industrial Laboratory, Inc., P.O. Box 1275, Wailuku, HI 96793, USA

Received 24 March 1997; received in revised form 24 April 1997; accepted 24 April 1997

Abstract

Artificial intelligence techniques can be used to provide a second opinion in medical settings. This may improve the sensitivity and specificity of diagnoses, as well as the cost effectiveness of the physician's effort. In the current study, evolutionary programming is used to train artificial neural networks to detect breast cancer using radiographic features and patient age. Results from 112 suspicious breast masses (63 malignant, 49 benign, biopsy proven) indicate that a significant probability of detecting malignancies can be achieved using simple neural architectures at the risk of a small percentage of false positives. © 1997 Elsevier Science Ireland Ltd.

Keywords: Breast cancer; Computer-assisted diagnosis; Artificial neural networks; Evolutionary computation; Evolutionary programming

1. Introduction

Carcinoma of the breast is second only to lung cancer as a tumor-related cause of death in women. There are now more than 180 000 new cases and 45 000 deaths annually in the United States alone [1]. 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 25% if all women in the appropriate age groups were regularly screened [2].

Intra- and inter-observer disagreement and inconsistencies in mammographic interpretation [3,4] have led to an interest in using computerized pattern recognition algorithms, such as artificial neural networks (ANNs) [5], to assist the radiologist in the assessment of mammograms. The second opinion offered by a reliable automated system may be useful in reducing false negative diagnoses [6,7], and other oversights that may result from poor mammographic image quality, physician fatigue, or alternative sources. ANNs hold promise for improving the accuracy of determining those patients where further assessment and possible biopsy is indicated. Furthermore, a reliable automated screening system could provide immediate

* Corresponding author. Tel.: +1 619 4556449; fax: +1 619 4551560; e-mail: dfogel@natural-selection.com

0304-3835/97/$17.00
PII S0304-3835(97)00259-0

results and lower the logistical costs associated with handling mammograms. The eventual cost savings could be passed along to the patient, while simultaneously making the radiologist's time more valuable and improving the quality of patient care.

ANNs are computational models based on the neuronal structure of natural organisms. They are stimulus-response transfer functions that map an input space to a specified output space. They are typically used to generalize such an input–output mapping over a set of specific examples. For example, as will be described here, the input can be radiographic features from mammograms, with the output being an indication of the likelihood of a malignancy.

Given a network architecture (i.e. type of network, the number of nodes in each layer, the weighted 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 potentially multimodal response surface over a multidimensional hyperspace (the dimension is 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 are often unsuitable for practical applications [8]. 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 [9], a population of candidate solutions is maintained, and random variation and selection are imposed on the population to efficiently guide it to appropriate regions of the hyperspace. The use of random mutation avoids entrapment in local optima, and there are several mathematical proofs that variations of these procedures provide asymptotic global convergence, rather than merely local convergence [9,10]. Moreover, there is empirical evidence that the methods are robust to many difficulties in possible response surfaces, including multiple minima or maxima, constraints, disjoint feasible regions and random perturbations [11].

2. Methods

For the current investigation, data were collected by assessing film screen mammograms in light of a set of 12 radiographic features as determined by the domain expert (Wasson) (Table 1). The features selected paralleled those offered in [12], with some modifications to increase the orthogonality of the features, as well as the inclusion of patient age. These features were assessed in 112 cases of suspicious breast mass, all of which were subsequently examined by open surgical biopsy with the associated pathology indicating whether or not a malignant condition had been found. In all, 63 cases were associated with a biopsy-proven malignancy, while 49 cases were indicated to be negative by biopsy.

These data were processed using a simple feedforward ANN restricted to two hidden sigmoid nodes (following the maxim of parsimony, this being the simplest architecture that can take advantage of the non-linear properties of the nodes), with a single linear output node, resulting in 33 adjustable weights. 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 was held out for testing and then

Table 1

The features and rating system used for assessing mammograms in the current study; assessment was made by the domain expert (Wasson)

Mass size	Either zero or in mm
Mass margin	Each subparameter rated as none (0), low (1), medium (2), or high (3) (a) Well circumscribed (b) Microlobulated (c) Obscured (d) Indistinct (e) Speculated
Architectural distortion	None or distortion
Calcification number	None (0), <5 (1), 5–10 (2), or >10 (3)
Calcification morphology	None (0), not suspicious (1), moderately suspicious (2), or highly suspicious (3)
Calcification density	None (0), dense (1), mixed (2), faint (3)
Calcification distribution	None (0), scattered (1), intermediate (2), clustered (3)
Asymmetric density	Either zero or in mm

replaced in a series of 112 separate training procedures, a population of 250 networks of the chosen architecture were initialized 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 incorporated an associated self-adaptive mutational vector used to determine the random variation imposed during the generation of offspring networks (described below). Each of these self-adaptive parameters was initialized to a value of 0.01. Each weight set was evaluated based on how well the ANN classified the 111 available training patterns, where a diagnosis of malignancy was assigned a target value of 1.0 and a benign condition was assigned a target value 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 over the 111 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:

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

where $\tau = 1/\sqrt{(2n)}$, $\tau' = 1/\sqrt{(2\sqrt{n})}$, $N(0,1)$ is a standard normal random variable sampled once for all 33 parameters of the vector σ and $N_i(0,1)$ is a standard normal random variable sampled anew for each parameter. The settings for τ and τ' have been demonstrated to be fairly robust [9]. These updated self-adaptive parameters were then used to generate new weight values for the offspring according to the rule:

$$x'_i = x_i + \sigma'_i C \qquad (2)$$

where C is a standard Cauchy random variable (determined as the ratio of two independent standard Gaussian random variables). The Cauchy mutation allows for a significant chance of generating saltations but still provides a reasonable probability that offspring networks will reside in proximity to their parents. 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. A pairwise tournament was conducted where each candidate weight set was compared against a random sample from the population. The sample size was chosen to be 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 affords a probabilistic selection, not unlike that achieved in annealing methods [13], allowing for the possibility of climbing up and out of hills and valleys on the error response surface.

This process was iterated for 200 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 111 patterns had been classified. Note that each final classification was made using a network that was not trained on the pattern in question.

Each complete cross validation was repeated 16 times with different randomly selected populations of initial weights to determine the reliability of the overall procedure. A typical rate of optimization in each training run is shown in Fig. 1. The probability of detection, P(D), and false positive, P(FP), vary with the discrimination threshold applied to the output of the networks. As the threshold value is lowered, the network can correctly identify a greater number of cancers, but this comes at the expense of a higher false positive rate. Conversely, the false positive rate can be lowered by raising the threshold value, but this in turn decreases the sensitivity of the procedure.

3. Results and discussion

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. A typical ROC curve for the 16 trials is offered in Fig. 2. The area under the curve, typically denoted A_Z, provides a useful measure for assessing the performance of the system. The mean area $\bar{A}_Z$ (determined using polynomial splines) was 0.8982 with a standard error of $s_{\bar{A}=1x(7)_Z} = 0.0098$. The best network achieved $A_Z = 0.9345$.

The average performance of the evolved ANNs in terms of A_Z is comparable to that of [14,15], which also used mammographic features interpreted by a radiologist. The ANN in [15], which used 18 input features (both radiographic and clinical) and possessed 10 hidden nodes, yielded a specificity of 0.62 at a sensitivity of 0.95. By comparison, radiologists attained only a 0.3 specificity on the same data. The evolved networks in the current study yielded a mean specificity of 0.6187 ± 0.0285 at 0.95 sensitivity. Although this result is almost identical to the performance offered in [15], the evolved networks are more parsimonious models (about an order of magnitude fewer degrees of freedom), and may therefore offer greater generalizability while requiring less computational effort.

One criticism of the use of ANNs in medical diagnoses is that they are black box methods, and in general are not explainable [16]. The success of small ANNs in diagnosing breast cancer, as observed here, offers the promise that suitable explanations for the network's behavior can be induced, perhaps leading to a greater acceptance by physicians and ultimately a useful tool.

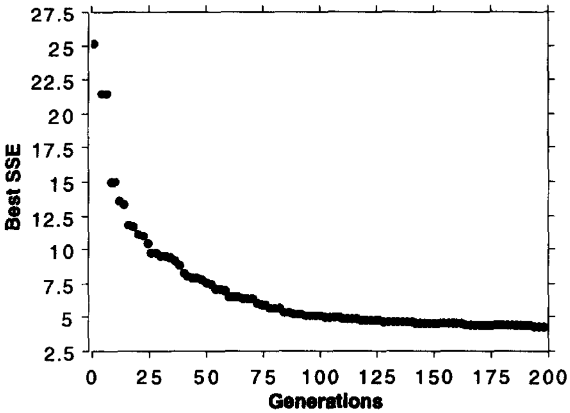

Fig. 1. Typical optimization performance using simulated evolution to train the ANN. The graph depicts the sum of squared error (SSE) of the best network in the population as a function of the number of generations. Training was performed over 111 patterns, with one pattern held out for testing in cross validation. The sufficiency of the number of generations is indicated as the learning curve approaches an asymptote.

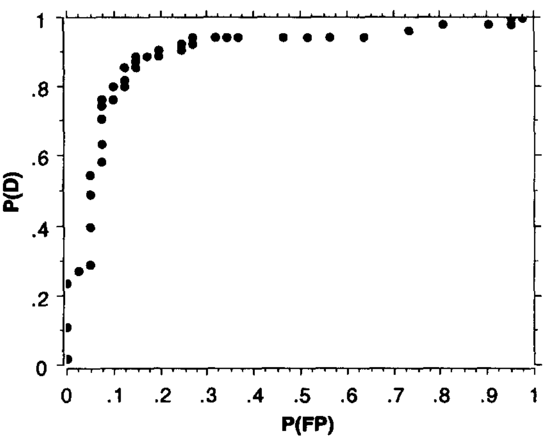

Fig. 2. A typical ROC curve (raw data) generated in one complete cross validation where each of 112 patterns was classified in turn, based on training over the remaining 111 patterns. Each point represents the probability of detection, P(D), and probability of false positive, P(FP), that is attained as the threshold for classifying a result as malignant is increased systematically over [0,1].

It would appear that training by simulated evolution or other stochastic methods is the key to developing these parsimonious networks. Under the more common gradient-based training method of error back propagation, the search for appropriate ANN weight sets can stagnate at local optima. These can be overcome by adding additional nodes and weights, but the resulting networks are no longer as parsimonious as may be possible. Evolutionary algorithms offer the potential for overcoming multiple optima on the error response surface, as well as simultaneously adjusting the ANN topology. Further, the evolutionary training method can be used regardless of the payoffs for correct and incorrect classifications, which may be important in trading off the costs of sensitivity and specificity.

Acknowledgements

The authors thank Peter Angeline, Wirt Atmar, Sal Cianci, Gary Fogel and Larry Fogel for their comments on the paper. This effort is sponsored by the Department of the Army, Grant No. DAMD17-96-1-6312. The content of the information presented does not necessarily reflect the position or the policy of the government, and no official endorsement should be inferred.

References

[1] C.C. Boring, T.S. Squires, T. Tong, Cancer statistics, CA: Cancer J. Clin. 43 (1993) 7–26.

[2] P. Strax, Make Sure that You do not have Breast Cancer. St. Martin's, New York, 1989.

[3] J.G. Elmore, C.K. Wells, C.H. Lee, D.H. Howard, A.R. Feinstein, Variability in radiologists' interpretations of mammograms, N. Engl. J. Med. 331 (1994) 1493–1499.

[4] G. Ciccone, P. Vineis, A. Frigerio, N. Segnan, Inter-observer and intra-observer variability of mammogram interpretation: a field study, Eur. J. Cancer 28A (1992) 1054–1058.

[5] S. Haykin, Neural Networks. Macmillan, New York, 1994.

[6] M. Giger, H. MacMahon, Image processing and computer-aided diagnosis, Imaging Inform. Management: Comput. Systems Changing Health Care Environment 34 (1996) 565–596.

[7] B. Sahiner, H.-P. Chan, N. Petrick, D. Wei, M.A. Helvie, D.D. Adler, M.M. Goodsitt, Classification of mass and normal breast tissue: a convolution neural network classifier with spatial domain and texture images, IEEE Trans. Med. Imaging 15 (1996) 598–610.

[8] V.W. Porto, D.B. Fogel, L.J. Fogel, Alternative neural network training methods, IEEE Expert 10 (1995) 16–22.

[9] D.B. Fogel, Evolutionary Computation. IEEE Press, New York, 1995.

[10] T. Bäck, Evolutionary Algorithms in Theory and Practice, Oxford University Press, New York, 1996.

[11] T. Bäck, D.B. Fogel, Z. Michalewicz (Eds.), Handbook of Evolutionary Computation. Oxford University Press, New York.

[12] C.E. Floyd, J.Y. Lo, A.J. Yun, D.C. Sullivan, P.J. Kornguth, Prediction of breast cancer malignancy using an artificial neural network, Cancer 74 (1994) 2944–2998.

[13] S. Kirkpatrick, C.D. Gelatt, M.P. Vecchi, Optimization by simulated annealing, Science 220 (1983) 671–680.

[14] Y. Wu, M.L. Giger, K. Doi, C.J. Vyborny, R.A. Schmidt, C.E. Metz, Application of neural networks in mammography: applications in decision making in the diagnosis of breast cancer, Radiology 187 (1993) 81–87.

[15] J.A. Baker, P.J. Kornguth, J.Y. Lo, M.E. Williford, C.E. Floyd, Breast cancer: prediction with artificial neural networks based on BI-RADS standardized lexicon, Radiology 196 (1995) 817–822.

[16] C.E. Kahn, Decision aids in radiology, Imaging Inform. Management: Comput. Systems for a Changing Health Care Environment 34 (1996) 607–628.