A neural network approach to breast cancer diagnosis as a constraint satisfaction problem

Georgia D. Tourassi
Department of Radiology, Duke University Medical Center, Durham, North Carolina 27710

Mia K. Markey
Department of Biomedical Engineering, Duke University, Durham, North Carolina 27710

Joseph Y. Lo and Carey E. Floyd, Jr.
Department of Radiology, Duke University Medical Center, and Department of Biomedical Engineering, Duke University, Durham, North Carolina 27710

(Received 25 September 2000; accepted for publication 12 February 2001)

A constraint satisfaction neural network (CSNN) approach is proposed for breast cancer diagnosis using mammographic and patient history findings. Initially, the diagnostic decision to biopsy was formulated as a constraint satisfaction problem. Then, an associative memory type neural network was applied to solve the problem. The proposed network has a flexible, nonhierarchical architecture that allows it to operate not only as a predictive tool but also as an analysis tool for knowledge discovery of association rules. The CSNN was developed and evaluated using a database of 500 nonpalpable breast lesions with definitive histopathological diagnosis. The CSNN diagnostic performance was evaluated using receiver operating characteristic analysis (ROC). The results of the study showed that the CSNN ROC area index was 0.84±0.02. The CSNN predictive performance is competitive with that achieved by experienced radiologists and backpropagation artificial neural networks (BP-ANNs) presented before. Furthermore, the study illustrates how CSNN can be used as a knowledge discovery tool overcoming some of the well-known limitations of BP-ANNs.

Key words: constraint satisfaction, neural networks, data mining, breast cancer

I. INTRODUCTION

Computer assisted decision making (CAD) is an active field of research with numerous applications in the diagnosis of breast cancer.1–6 The end product is typically aimed to provide physicians with a reliable second opinion during their medical decision process.7,8 CAD tools have explored a wide range of artificial intelligence (AI) techniques as potential decision algorithms. However, the backpropagation artificial neural network (BP-ANN) is, by far, the most popular AI algorithm utilized in breast cancer diagnosis. BP-ANNs have proven to be effective for a variety of prediction problems. However, after many years of successful application, BP-ANNs still carry the stigma of operating as black boxes. Even when they produce reliable predictions, BP-ANNs are unable to explain how they reach a particular decision. Consequently, some physicians are uncomfortable using CAD tools without justifications of their suggestions.

Although CAD focuses primarily on developing strict prediction and classification tools, it is part of a more general field called data mining.9 Data mining aims to gain insight into relationships between elements contained in large databases by following a systematic approach. While CAD typically targets only diagnostic performance, clinical data mining emphasizes knowledge discovery and its translation to clinical practice. BP-ANNs certainly fall short as data mining tools because they do not provide comprehensible models that account for the predictions they make. Some studies have addressed this weakness by attempting to extract decision rules from the weights of a BP-ANN.10–12 However, the extraction processes are tedious and the decision rules are often complex. A promising alternative to BP-ANNs is a case based reasoning (CBR) approach.13 CBR can offer predictions that are easily explained in the form of decision rules (such as if-then statements). Furthermore, CBR predictions can be interpreted as odds ratios, a familiar concept for the practicing clinician. However, CBR is based on case-matching decision rules that are predetermined by the operator.

The purpose of this study is to introduce an innovative ANN that can be used as a data mining tool with clinical databases. It is the constraint satisfaction neural network (CSNN), an associative memory type network suitable for solving optimization problems.14 Previously, constraint satisfaction networks have been primarily explored for medical image segmentation tasks.15–17 In addition, the CSNN was recently introduced for the analysis of drug vulnerability18 and treatment evaluation in drug addicts.19 Our study shows how breast cancer diagnosis can be formulated as a constraint satisfaction problem suitable for the application of a CSNN approach. The study illustrates the flexibility of the network to operate as both a prediction and decision analysis tool overcoming some of the limitations of other AI algorithms.
The activation level of each neuron depends on the influence of the other neurons and the external information available.

II. MATERIALS AND METHODS
A. Definition and dynamics of the CSNN

The CSNN is a Hopfield-type network, shown in Fig. 1. The network consists of neurons arranged in a nonhierarchical structure. The neurons are highly interconnected with symmetrical, bidirectional weights (\( w_{ij} = w_{ji} \)). Similar to the Hopfield network, there are no reflexive weights (\( w_{ii} = 0 \)). The weights describe fuzzy interactions among the neurons.

The CSNN network operates as a nonlinear, dynamic system aimed to achieve global stability by assigning values to its neurons while the weights remain fixed. Given an optimization problem, the CSNN weights can be interpreted as the problem constraints and every network state can be viewed as a possible solution to the problem. A problem is solved when the network achieves a globally stable state without violating the constraints. The individual neuron states describe the solution found. Therefore, the CSNN is suitable for optimization tasks that can be formulated as “constraint satisfaction problems.”

To achieve global stability, the CSNN employs a dynamic and iterative mechanism. The mechanism assumes that the activation level of all neurons can take any value in the range [0, 1]. At each iteration, the activation level \( u_i \) of every CSNN neuron \( i \) is determined by its external influence (\( \text{Ext}_i \)) and by the activation levels of all other neurons. Specifically, the activation level \( u_i \) of each CSNN neuron \( i \) is calculated as follows:

\[
\text{Net}_i = \frac{1}{N} \left( \sum_{j=1}^{N} w_{ij} \cdot u_j + \text{Bias}_i \right) + \frac{1}{\sqrt{N}} \cdot (\text{Ext}_i),
\]

where \( N \) is the total number of CSNN neurons, \( \text{Bias}_i \) is the bias term associated with the \( i \)th neuron, and \( \text{Ext}_i \) is the external influence applied to the \( i \)th neuron.

The updated activation level of the \( i \)th neuron is

\[
u_{i(n+1)} = u_{i(n)} + \Delta_i, \quad \text{where } n \text{ denotes the iteration. (5)}
\]

With this update rule, the network will restrict the activation levels to the [0, 1] range and will evolve so that all neurons achieve their maximum possible activation while still satisfying the constraints imposed by the weights. Similar to a Hopfield network, the measure of global stability is a Lyapunov function \( E \) (referred to as “energy function”) often used to describe the state of nonlinear dynamic systems:

\[
E(n) = -\frac{1}{2} \sum_{i} \sum_{j} \Lambda_{ij} \cdot u_{i(n)} \cdot u_{j(n)} - \sum_{i} \text{Bias}_i \cdot u_{i(n)} + \sum_{i} \text{Ext}_i \cdot u_{i(n)}. \quad \text{(6)}
\]

A dynamic system achieves a stable state when this function is minimized. In the CSNN context, the energy function can be viewed as a measure of constraint satisfaction. The first two terms in Eq. (6) describe the internal dynamics of the network. The last term is the penalty imposed by any external influences. The network achieves stability when it reaches a low energy state. It has been theoretically proven that the iterative approach presented above converges to a low energy state, though not necessarily to the global minimum. From that aspect, CSNN and BP-ANN are both subject to being trapped into local minima.

An important component of developing a CSNN is determining its weight matrix. As explained before, the weight matrix contains the relationships or constraints among all neurons. Most of the heuristic approaches that have been used in the past are based on approaches used with the Hopfield network. For this study we explored an autoassociative backpropagation (auto-BP) scheme, a technique that showed promising results before. The auto-BP network is a simple perceptron without hidden layers. The input and output layers have an equal number of nodes (\( N \)). During the training phase, the auto-BP learns to map any given pattern to itself using the backpropagation technique for gradient descent with the sigmoid activation function. However,
during training the reflexive weights are forced to be 0. Thus, if \( w_{ij} \) denotes the weight connection between input neuron \( i \) and output neuron \( j \), then the weights \( w_{ij} = 0 \) for \( i = j \). Furthermore, to force the bidirectional weights to be equal (\( w_{ij} = w_{ji} \)), the previous interactive values of these terms are averaged at the end of each training iteration:

\[
    w_{ij}(n) = w_{ij}(n-1) + w_{ji}(n-1)
    \]

\[= \frac{w_{ij}(n-1) + w_{ji}(n-1)}{2}, \tag{7}
\]

where \( n \) denotes the training iteration. When the training phase is complete, the autoassociative BP weights act as the CSNN constraints.

**B. CSNN as a data mining tool**

The CSNN weight matrix can be used as an associative memory to solve optimization problems. A clinical decision such as breast cancer risk assessment, mammographic diagnosis, or treatment planning can be approached as an optimization problem. As such, a patient is viewed as a ‘‘puzzle’’ composed of many pieces (e.g., clinical findings, personal and family history, mammographic findings, presence or absence of breast cancer, optimal treatment). All pieces can be coded into variables that are interconnected with constraints to keep the puzzle intact and stable. For a typical clinical decision, there is information about some patient variables (e.g., clinical and mammographic findings) and some questions to be answered (Is there breast cancer? If yes, what is the appropriate treatment?). Answering the above questions is equivalent to completing the ‘‘patient puzzle’’; that is, finding the optimal values of the remaining variables so that all constraints are satisfied. The CSNN is designed to solve this type of optimization problems.

A very attractive aspect of the CSNN network is its flexibility. Given its nonhierarchical architecture, the CSNN has no designated input and output variables. Therefore, the same CSNN weight matrix can be used to reproduce any missing components of a given pattern. The following section illustrates the development and utilization of a CSNN for data mining a breast cancer database.

**C. CSNN application in a breast cancer database**

A CSNN approach was applied to the prediction of the outcome of breast biopsy for patients with mammograms suspicious for breast cancer (BC). The database used in this study has been described in detail in a previous publication.23 Specifically, the database consisted of mammographers’ descriptions of 500 nonpalpable breast lesions from 478 consecutive patients who underwent excisional biopsies resulting in definitive histopathological diagnosis. Of these 500 lesions, 174 (35%) were found to be malignant at biopsy. The dataset contained a variety of lesions: 232 lesions were characterized by mass findings alone, 192 were characterized by calcification findings alone, 29 lesions were characterized by combinations of masses and calcification findings, and 47 lesions represented special cases such as architectural distortion, regions of asymmetric breast density, areas of focal asymmetric density, and areas of asymmetric breast tissue.

The prevalence of breast cancer was approximately the same for cases with masses only (30%), calcifications only (37.5%), and no masses or calcifications (32%). BC prevalence was significantly higher (62%) when both masses and calcifications were present.

For each lesion, expert mammographers retrospectively viewed the patient films and reported the mammographic findings according to the BI-RADS lexicon.24 Patient age and history findings were also collected. In total, 16 mammographic and clinical findings were recorded for each patient. Table I lists the findings included in the study.

The 16 BI-RADS findings were converted into a binary input vector. For each patient, the input vector consisted of 82 exclusively binary nodes 0 or 1 representing if a particular finding was present or not. The input findings were coded so that one neuron was assigned to each possible description for every finding. For example, according to the BI-RADS lexicon, there are four possible types of ‘‘mass shape.’’ Four nodes were assigned to this BI-RADS finding, each one corresponding to a different type of mass shape (i.e., round, oval, lobulated, irregular). In addition, four separate neurons were assigned to correspond to the presence of masses, microcalcifications, special findings, and associated findings. The continuous findings such as patient age and mass size were represented as categorical data. Mass size was coded in seven possible nodes. Each node corresponded in mass size increments of 10 mm. Similarly, patient age was coded in five nodes (<40 yrs, 40–50, 50–60, 60–70, and >70 yrs old). One extra neuron was added to constitute the diagnosis. The diagnosis neuron took the value of 1 if breast cancer was present and the value of 0 if breast cancer was absent. We used only a single diagnosis node so that the CSNN can be used as a predictive rather than a classification tool. In total, 83 CSNN neurons were used to represent the problem.

During the development or ‘‘training’’ phase, the CSNN constraints were determined using the backpropagation autoassociative (auto-BP) network described before. The auto-BP network had an input layer and an output layer of 83 nodes each. Initially, the weights were randomly selected inside the interval \([-1/N, 1/N]\) where \(N = 83\). The biases were set to 0. The auto-BP was then trained according to the backpropagation algorithm. After the auto-BP weights and biases were determined, the weights were converted according to Eq. (7) to serve as the CSNN constraints.

<table>
<thead>
<tr>
<th>Table I. List of mammographic and clinical findings used as CSNN inputs.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mammographic findings</strong></td>
</tr>
<tr>
<td>1. Calcifications distribution</td>
</tr>
<tr>
<td>2. Calcifications number</td>
</tr>
<tr>
<td>3. Calcifications description</td>
</tr>
<tr>
<td>4. Quadrant location of abnormality</td>
</tr>
<tr>
<td>5. Associated findings</td>
</tr>
<tr>
<td>6. Special cases</td>
</tr>
<tr>
<td>7. Mass margin</td>
</tr>
<tr>
<td>8. Mass shape</td>
</tr>
<tr>
<td>9. Mass density</td>
</tr>
<tr>
<td>10. Mass size</td>
</tr>
</tbody>
</table>
Next, the CSNN was applied as a predictive tool. For each test case, the CSNN network was used to predict the diagnosis based on the network’s constraints (the weight matrix determined by auto-BP) and the external inputs (the available medical findings for each case). If a particular finding was present, then the corresponding external influence was active and set equal to 1.0. The testing phase is an iterative process that determines the activation values of all neurons so that the network achieves a lower energy state. Initially, the activation level of all CSNN neurons was set at 0. The activation level achieved by the diagnosis neuron represents the optimal value that makes the “patient puzzle” complete and stable. The activation level achieved by the diagnosis neuron was used as the decision variable for further analysis.

Custom software written in the C programming language was used to implement the Constraint Satisfaction network. The software ran on an Ultra SPARC workstation (Sun Microsystems, Mountain View, CA). Typical CSNN case evaluation required approximately 0.5 s for 200 iterations.

III. RESULTS

A. Determining the CSNN constraints

Results are presented based on the 50%–50% cross-validation sampling technique. Initially, one-half of the cases were randomly selected and used with the auto-associative BP network to determine the CSNN constraints. Subsequently, the CSNN weights remained fixed and the CSNN was tested on the other half of the database. The whole process was then reversed.

The auto-BP networks were trained for 300 iterations. The root mean squared (rms) training error achieved by the networks was approximately 0.12. Determining the training ending point for the auto-BP network was a tricky task. It was found that although the autoassociative perceptron was able to converge and generalize, it resulted in weights that did not perform as well when used with the CSNN. Therefore, the auto-BP training phase was terminated earlier than expected even though both training and testing rms errors were still descending. The training ending point (300 iterations) was determined empirically.

B. Using the CSNN as a predictive tool for breast cancer diagnosis

As mentioned before, the dataset was randomly divided into two subsets (A and B) with 250 patient cases each. Initially, subset A was used to train the auto-BP network and determine the CSNN constraints. Then, the predictive ability of the CSNN was tested on subset B. For each test case, CSNN proceeded iteratively until its energy function was stabilized.

It was found that 200 iterations were sufficient for the task. At the end of the iterative process, the activation level achieved by the designated diagnosis neuron was used as a decision variable for receiver operating characteristics (ROC) analysis. Then, the whole process was reversed so that subset B was used to determine the CSNN constraints and subset A was used to test the CSNN as a predictive tool.

We used the ROCKIT software package developed by Metz et al. (http://xray.bsd.uchicago.edu/krl/toppage11.htm) to fit ROC curves to the activation level achieved by the CSNN diagnosis neuron. Specifically, the ROC area index $A_z$ achieved by the network was:

- Subset A: $A_z = 0.85 \pm 0.03$
- Subset B: $A_z = 0.84 \pm 0.03$
- Overall: $A_z = 0.84 \pm 0.02$

The same experiment was repeated using a different randomized split of the dataset. Minor $A_z$ differences were observed for each subset and the overall ROC performance of the network remained unchanged.

Table II compares the CSNN diagnostic performance to that achieved by experienced mammographers and a previously published BP-ANN. Several indices of diagnostic performance are presented: overall ROC area index, specificity at 95% sensitivity level, and the corresponding PPV at the same operating point. The ROC evaluation of the radiologists’ performance was based on a gestalt, five-point scale, categorical assessment of the likelihood of malignancy (not the BI-RADS categorical assessment). The table shows that the CSNN overall diagnostic performance is competitive to that achieved by the other decision models.

Subsequently, the results were analyzed based on the type of mammographic lesions present for each patient. Table III

<table>
<thead>
<tr>
<th>Type of lesions</th>
<th>No. of cases (no BC+BC)</th>
<th>$A_z$ (CSNN)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Masses only</td>
<td>233 (163+69)</td>
<td>0.93±0.02</td>
</tr>
<tr>
<td>Calcifications only</td>
<td>192 (120+72)</td>
<td>0.65±0.04</td>
</tr>
<tr>
<td>Masses+Calcifications</td>
<td>29 (11+18)</td>
<td>0.83±0.08</td>
</tr>
<tr>
<td>No masses or Calcifications</td>
<td>47 (32+15)</td>
<td>0.70±0.09</td>
</tr>
<tr>
<td>Total</td>
<td>500 (326+174)</td>
<td>0.84±0.02</td>
</tr>
</tbody>
</table>

The previously published performances of experienced mammographers and a BP-ANN are included for comparison purposes. PPV: positive predictive value.
shows that there is great variability in the CSNN predictive ability depending on if there are masses or calcifications present. However, similar discrepancy has been reported with a traditional BP network and clinicians themselves.\textsuperscript{25}

In addition, we analyzed the energy level achieved by the neural network for each test case. Since the CSNN is designed to evolve toward a lower energy state, we tested if the energy level achieved can be used as a quality measure for the CSNN predictions. The CSNN predictions were grouped into two subsets as follows. One subset contained all the test cases for which the network converged to the correct diagnosis within an absolute distance of 0.5 from the target value. The second subset contained all test cases for which the network predicted incorrectly the diagnosis (i.e., the activation level achieved by the diagnosis neuron had an absolute distance larger than 0.5 from the target value). The average energy level achieved by the network when it was correct (386/500 cases) was statistically significantly lower than the average energy level achieved by the network when it was incorrect (114/500 cases). Specifically, the average CSNN energy for each subset was:

\[ E \text{ (CORRECT)} = -7.19 \pm 2.37 \]

\[ E \text{ (INCORRECT)} = -6.55 \pm 2.42. \]

The difference is statistically significant at the 95\% confidence level with a confidence interval (0.13, 1.14).

C. Using CSNN for data mining hidden associations

An exciting aspect of the constraint satisfaction network is its ability to be used not only as a prediction tool but also as a data mining tool to discover trends and associations among clinical findings and diagnosis. By selecting the neurons that accept external information, the network can follow the iterative approach presented before to find the activation values of the remaining neurons so that the energy function is stabilized while satisfying all constraints. We demonstrate this quality by asking the network three prototypical questions. Given our database (i.e., cases which were sufficiently suspicious of breast cancer to require the mammographers to recommend biopsy), what is the profile of a patient with BC? What is the profile of a patient without BC? What is the profile of a “confusing” patient? To answer these questions, initially the auto-associative perceptron is used to determine the CSNN weights given all 500 cases. The auto-BP network was trained for 300 iterations (as in the cross-validation experiment) until it reached a MSE of approximately 0.10. The auto-BP weights and biases were then used as the CSNN constraints.

To answer the first question (“What is the profile of a patient with BC?”), the activation level of the diagnosis neuron was set to 1.0 and the remaining 82 neurons were left free to evolve until the network reached a stable state. None of the 82 neurons accepted external information. Table IV shows which neurons were activated and reached maximum values indicating strong association with breast cancer. Specifically, the table shows that the mammographic variables strongly correlated with breast cancer are architectural distortion and small, spiculated masses with irregular shape, and high density. Furthermore, history findings that are risk factors for BC are older age and family history of BC. Similarly, the profile of a patient without breast cancer, is a female much younger (40–50 yrs old), menopausal, without any family or personal history of BC, and without any lesions. It is apparent that these patient profiles are prototypical and as such they provide a more qualitative form of information. Although it is not clinically surprising that the patient without BC has no masses or calcifications present, it is unexpected given that the majority of the negative patients in the database had some type of lesion present. This is an indication that the CSNN gives responses that are not necessarily statistical in nature. The same quality can be further demonstrated with the profile of a “confusing” patient. To acquire this profile, the activation level of the diagnosis neuron was set at 0.5. After 1000 iterations, the neurons that were activated were the same ones as in the BC prototype with the exception of masses. Therefore, older age and family history of BC alone increase the risk of breast cancer as it is clinically known.

The ability to use the network from “bottom-up” is very attractive compared to the backpropagation network. Hidden associations can be discovered and used as decision rules. By controlling several neurons at a time, the CAD tool operator can interrogate the constraint satisfaction network and get more detailed explanations of its decision reasoning. Furthermore, by observing the activation order of the various neurons, we can evaluate the relative importance of the various variables during the decision process. For example, Fig. 2 shows the evolution of representative neurons when the network was asked to create the profile of a patient with breast cancer. The graph shows that the first neurons that fired up were: family history of BC, older age (>70 yrs old), and spiculated mass, immediately followed by architectural distortion. They all reached their maximum value at 200 iterations. Although it activates early, the variable mass shape evolves at a slower rate. It is interesting that the neurons describing the density and size of the spiculated mass fire up in a delayed fashion (at about 300 iterations). However,
when these neurons finally become fully active, the architectural distortion neuron starts decaying. This result indicates how the CSNN decides the relative significance of the various findings associated with breast cancer. The CSNN considers the presence of architectural distortion as a red flag for breast cancer. However, the presence of a small, spiculated mass, with irregular shape and high density reduces the diagnostic importance of architectural distortion for the CSNN. In the absence of masses, CSNN considers architectural distortion as diagnostically important regardless of the presence of calcifications, or other special findings (Table V). Each column in Table V shows how the network answers prototypical questions by controlling more than one variable at a time. For example, the second column shows the findings that indicate breast cancer in a patient without masses or calcifications present. To answer this question, three neurons (BC present, mass absent, calcification absent) were externally controlled and the remaining 80 neurons evolved until the network reached a stable state. In the end, the following variables were activated: architectural distortion, focal asymmetric density, older age, and family history of breast cancer.

### IV. DISCUSSION

In this paper, we have proposed a CSNN for the prediction of breast cancer from mammographic and history findings. The diagnostic performance of the network was comparable to that of experienced mammographers and conventional backpropagation networks previously presented for the same problem. However, our study showed that the CSNN structure lends itself to some exciting data mining possibilities not easily available with conventional ANNs or statistical techniques. Some of them are highlighted below.

By formulating a medical decision task as a constraint satisfaction problem, we allow the application of optimization algorithms as predictive tools. The constraint satisfaction network is basically an associative memory. As such, when the CSNN is given partial knowledge of a pattern, it tries to guess the pattern’s missing components while satisfying the constraints that exist among them. The network operates as a dynamic, nonlinear system trying to achieve a state of global stability without violating the constraints. The main hypothesis is that when the missing components are replaced by the proper values the network finally settles into a stable, low energy state. Therefore, conceptually, the CSNN operates differently than traditional statistical techniques or backpropagation networks. It is generally accepted that there is no statistical basis for associative networks.20 Considering the recent trend to boost CAD performance by combining several decision models26 CSNN is a great candidate to join a pool of experts such as BP-ANNs and linear statistical techniques.

Since the constraint satisfaction network tries to converge to a stable state, the final energy state in which the network settles can be exploited in two ways. First, the activation level achieved by the diagnosis neuron is used as the final prediction. Second, the energy level achieved by the network can be used as a measure of how confident the network is about its response. Lower energy indicates a more stable network state, and probably a more accurate response. Our results certainly support the hypothesis. Given the breast cancer database, the network had a tendency to settle into more stable states when its diagnosis prediction was closer to the truth. The potential to assign a confidence measure to a prediction is very attractive when applying a CAD tool.

The nonhierarchical architecture of the CSNN makes it a very flexible CAD tool. As our study illustrated, the CSNN can be applied to reconstruct simultaneously not only the correct diagnosis but also any missing components of a given clinical case. Contrary, the hierarchical architecture of the BP-ANN requires the CAD user to determine a priori which variables \( X = (x_1, x_2,...) \) will comprise the input information and which variables \( Y = (y_1, y_2,...) \) will comprise the output information. Then, the backpropagation network estimates the predictive function \( Y = f(X) \). If the CAD user changes the output variables, then a different BP network needs to be

---

**Table V. Example of data mining showing the variables associated with the presence of breast cancer for various patient subgroups. Gray-shaded cells represent variables that remained inactive. BC: Breast Cancer.**

<table>
<thead>
<tr>
<th>Activated variables</th>
<th>Patients with no masses</th>
<th>Patients with no masses and no focal asymmetric density</th>
</tr>
</thead>
<tbody>
<tr>
<td>Califications</td>
<td></td>
<td></td>
</tr>
<tr>
<td>description</td>
<td></td>
<td></td>
</tr>
<tr>
<td>distribution number</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>number</td>
<td>clustered</td>
<td>clustered</td>
</tr>
<tr>
<td>number</td>
<td>&gt;10</td>
<td>&gt;10</td>
</tr>
<tr>
<td>number</td>
<td>pleomorphic</td>
<td>pleomorphic</td>
</tr>
<tr>
<td>number</td>
<td>architectural</td>
<td>architectural</td>
</tr>
<tr>
<td>number</td>
<td>distortion</td>
<td>distortion</td>
</tr>
<tr>
<td>Associate findings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>architectural</td>
<td></td>
<td></td>
</tr>
<tr>
<td>distortion</td>
<td>focal asymmetric</td>
<td>focal asymmetric</td>
</tr>
<tr>
<td>density</td>
<td>density</td>
<td>density</td>
</tr>
<tr>
<td>Special findings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;70 yrs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family Hx of BC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;70 yrs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;70 yrs</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

**Fig. 2. Dynamics of the variables activated during the first 1000 iterations when the CSNN was asked to determine the prototype of a patient with breast cancer (BC). The variables are listed in the order they activated.**
developed and stored. Since the CSNN does not have designated input and output variables, the same network can perform any prediction task selected given the same variables \((X + Y)\). Consequently, it appears that the CSNN might be able to impute missing data and perform diagnosis simultaneously. This is an exciting possibility for clinical databases with missing data. Imputing missing data is an important issue that tends to compromise the performance of a decision model. In the future, we will explore if the CSNN can deliver this promise.

An added bonus of the nonhierarchical CSNN architecture is its ability to be used as a data mining tool for knowledge discovery. It is similar to trying using a backpropagation network from the bottom up to deplete its decision process. We demonstrated this quality by asking the CSNN a series of questions. The questions were prototypical and the answers gave us some insight on how the network forms predictions. For example, the network appears to consider seriously the age, family history, and the presence of a speculated mass when trying to decide the presence of breast cancer. Furthermore, the presence of architectural distortion or a focal asymmetric density are also red flags for breast cancer although diagnostically not as important in the presence of a speculated mass with irregular shape and high density. The ability to interrogate the network makes it very attractive for a potential CAD user who can formulate many clinically meaningful questions and discover hidden associations. Furthermore, given the profile of the patient without breast cancer, our results indicate that some discovered associations are not obviously statistical in a nature. It remains to be seen if thorough interrogation of the network can lead to clinically applicable decision rules.

Several issues related to the development of the network need thorough investigation. For example, the parameters determining the relative importance of external and internal influences when forming the activation of the CSNN neurons can affect somewhat the overall diagnostic performance of the network. A more critical issue though is how to determine the CSNN weight matrix. The auto-associative perception scheme followed in our study worked well, although other techniques can be used for the task as long as they can produce a symmetric matrix with no reflexive weights. Choosing the end point of the training session for the perceptron was more challenging than originally expected. However, following a trial-and-error approach with all above issues worked well. This heuristic approach should not be considered a weakness since we tend to do the same when selecting the training parameters or number of hidden nodes with feed-forward neural networks. Finally, from our experience, another issue that needs further investigation is how to initialize the activation levels of the CSNN neurons. In this study, all neurons were initialized at 0.0. We experimented with random initialization but the overall diagnostic performance was not affected. However, initialization appears to be more important when the network is used for prototype analysis. As shown in the study, in prototype analysis a very small number of neurons accept external information. Consequently, the CSNN can evolve and settle in states that satisfy sufficiently the overall constraints but are different from each other depending on the network’s starting point.

To summarize, our study showed that the constraint satisfaction neural network is a very promising CAD tool for data mining. Our results indicate that the CSNN is a competitive predictive model in the diagnosis of breast cancer from mammographic and history findings. Furthermore, our study demonstrated how the dynamic and nonhierarchical nature of the CSNN can be exploited to overcome the limitations of other popular decision models. Specifically, the CSNN appears to be able to make predictions of different variables at a time without restructuring and retraining of the network. It can discover hidden associations that can be easily formulated into decision rules. Finally, CSNN not only can provide a prediction but also a confidence measure on it. These qualities make the CSNN an exciting CAD alternative worthy of further investigation.