Neural networks approach to early breast cancer detection

Drasko Furundzic a,b,*., Miodrag Djordjevic c, Ana Jovicevic Bekic c

a Automatic Control Laboratory, Mihailo Pupin Institute, Volgina 15 St., 11000 Belgrade, Yugoslavia
b Experimental Physiology & Pathophysiology Department, Medical School University of Belgrade, Dr. Sahatciu 6 St., 11000 Belgrade, Yugoslavia
c Institute for Oncology and Radiology of Serbia, Pasteurova 14 St., 11000, Belgrade, Yugoslavia

Abstract

Efficient prevention is strongly correlated with an early detection of breast cancer. The common effort of many researchers in this field resulted in the selection of the most prominent risk factors related to breast cancer. In this paper we present a neural network based model for the efficient automated identification of women at high risk of developing breast cancer from a wide, healthy population on the basis of data, referring to a properly selected set of risk factors and symptoms. Using this model we achieved a highly accurate classification and also the initial set of features reduction. © 1998 Elsevier Science B.V. All rights reserved.

Keywords: Neural networks; Breast cancer; Risk factor

1. Introduction

Early detection and diagnosis of breast cancer are high priorities for cancer researchers. Detection is a process which precedes diagnosis. The only valid diagnosis is based on biopsy and histopathological findings, so any other method which precedes biopsy, including mammography, belongs to detection methods. The main goal of breast cancer detection methods is the best possible selection of patients at risk, in other words, the selection of the smallest group with the highest risk of developing breast cancer. In spite of the large number of researchers in the field the breast cancer etiology, it is still insufficiently explored to enable efficient prevention and hence decreased morbidity and mortality rates [32-34]. Efficient prevention assumes both basic and epidemiological researches with a common intention to identify risk factors of great importance and the way to avoid or modify them when it is possible. There

* Corresponding author. E-mail: efurundzic@ubbg.etf.bg.ac.yu
are many research areas that promise to identify the most prominent risk factors and also understand the complex etiology of breast cancer, and develop efficient prevention: hormones, genetics, diet and nutrition, physical activity, environmental risk factors, tumor markers, metastasis and mammography.

The study presented in this paper is our attempt to contribute to reducing the cancer mortality rate. We used an artificial neural network to detect a minimal group of patients with the highest risk of developing, or already suffering from, breast cancer, from a wide, healthy population. The main categories of risk factors of breast cancer used in this study are outlined below. The other classes of features (symptoms and graded risk factors) are described in the following sections.

Hormones: Increased risk of breast cancer may be caused by mother’s hormone levels during pregnancy, menopausal hormonal therapy, increased number of menstrual cycles caused by the onset of menstruation at an early age, late menopause, and a few full-term pregnancies [35-37].

Genetics: Recent studies [40-42] have shown that genetic changes are involved in the etiology of breast cancer. Several genetic processes may be involved in breast cancer, including mutations in normal genes and the loss of tumor suppressor genes, which normally prevent cancer development. Better understanding of genetic processes may be used in early detection and diagnosis of breast cancer, and identification of women at high risk of developing breast cancer. Mark et al. [38] reported important findings about two genes believed to be involved in the development of hereditary breast cancer. First, the investigators pinpointed the precise location of the gene known as BRCA1. BRCA1 and BRCA2 are connected to only 5-10% of all breast cancers, so test results are not beneficial enough for more than 90% of the population.

Diet and nutrition: Scientists are researching the ways in which different foods might protect women from breast cancer. Research is also continuing on obesity. Being seriously overweight may increase a woman’s risk of developing breast cancer [37].

Markers: Both biochemical markers and genetic markers that could be important in diagnosis and treatment are subjects of investigation. These include the her2 neu oncogene, and the p53 tumor suppressor and other breast cancer related oncogenes and surface markers [37].

Physical activity as protector: A few studies have suggested that breast cancer risk is lower among women who have been engaged in physical activity [42].

Environment exposure: There have been an insufficient number of studies of environmental or occupational influence on breast cancer occurrence. Most of them were based on small samples. Alcohol consumption, smoking habit, oral contraceptives, estrogen replacement therapy, and pesticide exposure are some of the risk factors being studied for correlation with breast cancer [35]. Exposure to high doses of ionizing radiation is known to increase the risk of female breast cancer. When scientists understand more about the role of such factors, they may be able to advise as to what environmental, lifestyle, or dietary factors are to be avoided or modified in order to lessen the risk of breast cancer.

Mammography: This is the subject connected to many recent results [19-23]. Many investigators believe that automation of mammogram screening analysis increases the rate of early detection [23]. The current procedure assumes that the image is recorded on an X-ray film and that it is interpreted by a human expert. Automatic digital mammography uses specially designed programs, based on the soft computing, technique, for image interpretation, with increased accuracy. Unfortunately, screening rates for breast cancers remain at levels less than desirable, even in the developed countries.
2. Neural networks and classification

Neurocomputing is one of several new disciplines of information technology collectively known as soft computing, concerned with parallel, distributed, and adaptive information processing structures, developing intelligent behavior in an information environment. The structures known as artificial neural networks are inspired by morphological and functional organization of neural brain structures – biological neural nets – and are an attempt to emulate its high level of organization, based on current understanding of the nervous system [7,8,16]. Emulating some of the human brain characteristics underlying intelligent reasoning, which are “conditio sine qua non” of existence in changeable environmental conditions [6–9]. Neural networks show greater performances, in comparison to conventional approaches based on sequential computing. The characteristics of neural nets can approximately be defined as:

- high degree of fault tolerance and robustness for imprecision and uncertainty in unconstrained information environment
- distributed processing and representation of information
- generalization ability
- massive parallelism
- learning based adaptation.

Neural network development strongly depends on the technology that provides computers with the high performances necessary for processing a large amount of information in a reasonable time. Many researches in different areas of science and technology use neural networks to solve problems in control [10,11], function approximation [14], medical diagnosis and prognosis [19–26], and pattern classification [1–5]. Pattern recognition and classification is a frequent problem in engineering and medicine, where flexible neural net based classifiers provide reduced error rates in comparison to conventional Bayesian classification approaches [3–5]. Neural net based classifiers show fine discrimination ability due to formation of hyperplane decision boundaries in the pattern space. Application of any classifier including the neural net classifier requires appropriate feature selection for specific classes of patterns to be classified. These facts motivated us to use a multilayer neural network with tangent hyperbolic activation function as a confirmed and effective pattern classification tool for early breast cancer detection. The most popular among the multilayer feedforward networks is multilayer perceptron with one or more layers of neurons between input and output nodes. Each unit in hidden layers employs a nonlinear transfer (activation) function. Kolmogorov [14] and Funahashi [15] proved that a three-layered neural net can serve as any continuous function approximator by using a sufficient number of neurons. It is necessary to emphasize that some events appear during the tasks performed using neural nets, which decrease the possibility for practical realization of the mentioned theorem, and demand additional actions to be overcome (local minimum problem).

In recent times neural nets have become very popular in medicine, for clinical diagnosis based on clinical and experimental data, as well as expert (physicians) knowledge which is basically available in a qualitative (linguistic) form. Neural nets are very convenient for various tasks: for diagnosis, healing, and prognosis in hypertension [24], EEG, EMG, ECG analysis and bone fracture healing assessment [30], renal allograft rejection prediction [25] and blood cell classification [31]. Automation of digitized mammogram analysis has been the interest of many researches [19,22,23]. The goal is early detection and classification of tumor cells in a mammogram. Kocur et al. [20], use neural network to choose the best wavelet coefficients for features used in automated breast cancer diagnosis [22]. Recently many papers refer to application of neural networks to cancer
diagnosis and prediction [26,27]. Choong et al. [29] used neural network aiming to derive maximum entropy distributions based on a small data set, for the purpose of making various types of inferences. They tested the efficacy of the proposed model in breast cancer prognosis on the basis of a small group of relevant risk factors, and obtained good results in comparison with probabilistic neural network and multilayer perceptron.

Astion and Wilding [28] investigated the ability of neural networks to differentiate benign from malignant breast conditions on the basis of the set of variables obtained in the laboratory, including the patient’s age. That study has some important limitations in the small number of patients and choice of predictor variables, but the results obtained with their neural network were still good enough. They mentioned the importance of missing data, referring to some risk factors of known diagnostic significance (age at menarche and age at menopause). Those mentioned risk factors are included in the set of factors we used in the present study. This paper presents an attempt to apply neural network for the problem of breast cancer detection on the basis of risk factors and symptoms, both given mostly in a binary form. The problem of feature extraction (input vector reduction) performed by the neural network has also been considered.

3. Selection of features for breast cancer detection

Since 1980, breast cancer has been a leading neoplasm among females in Yugoslavia [43]. For this reason, the project for early breast cancer detection has been initiated, with the aim of reduction of breast cancer mortality. In the first phase of the project, mortality and morbidity data were studied in order to establish the epidemiological situation of breast cancer in Yugoslavia. In the second phase, the significance of the numerous potential risk factors for breast cancer has been investigated by a case-controlled study. On the basis of the data obtained and the results of other studies [34], the project of early detection of breast cancer has been classified as a survey-based selective screening. A special questionnaire form M-17 (Table 1), and a table of risk factors graded by age (Table 2) have been created. The questionnaire consists of four parts: general data, breast cancer risk factors, breast cancer symptoms and results of previous examinations (Table 1). Women patients filled-in the questionnaire with the assistance of the trained nurses in ontological centers and other health institutions. Our previous screening procedure for breast cancer prevention had defined priorities for clinical breast examination as the presence of several risk factors: family history of breast cancer, menarche before the age of 12, menopause after 50, first delivery after 30, and a history of benign tumors or long-lasting fibrocystic dysplasia. But the symptoms are not a reliable indicator of the disease, since there are a large number of asymptomatic patients and vice versa. An important point in breast cancer detection is the avoidance of false negative results, which are frequent in mammogram interpretation (10–30%), that has been proved by retrospective analyses of mammograms [43]. In order to improve the existing screening procedure at the Institute for Oncology and Radiology of Serbia, we incorporated a neural network in it (Fig. 1).

We used data referring to age, risk factors, graded risk factors, and present symptoms as initial features of selected sets of patterns for the neural net classifier. The proper selection of both patterns (patients) and features (risk factors and symptoms) is of crucial importance for valid detection. It is important to select representative examples of patterns to be classified, taking care to include patterns with a small distance in n-dimensional space of features belonging to different classes.
Questionnaire form for breast cancer detection

<table>
<thead>
<tr>
<th>(A) Date</th>
</tr>
</thead>
</table>

| (B) ANSWER THE FOLLOWING QUESTIONS: |
|------------------|------------------|------------------|------------------|------------------|------------------|
| Menarche before the age of 12 | No. of children | First delivery after 30 yrs of age | Breast feeding (3 months at least) | No. of abortions | Use of oral contracept. | Hormonal disorders | Menopause after 50 years of age |
| a) Yes | b) No | a) Yes | b) No | a) Yes | b) No | a) Yes | b) No |

| (C) LATELY, DO YOU HAVE ANY OF THE FOLLOWING SYMPTOMS: |
|------------------|------------------|------------------|------------------|------------------|
| Pain in breasts | Lump in the breast | Nipple changes | Breast swelling | Recent changes in breast size or shape |
| a) Yes | b) No | a) Yes | b) No | a) Yes | b) No | a) Yes | b) No |

If there is any pain, is it:
1. permanent
2. premenstrual
3. occasional

If yes, is it painful?
1. Yes
2. No

If yes, which color?
1. Yes
2. No

| (D) PREVIOUS DIAGNOSTIC METHODS: |
|------------------|------------------|------------------|------------------|
| MAMMOGRAPHY | EXAMINATION OF NIPPLE SECRETION | BIOPSY | EXCISION OF AXILLAR GLANDS |
| No | Yes | No | Yes | No | Yes | No | Yes |

M-17 FORM FOR BREAST CANCER DETECTION
Selection of the main features from available ones provides sufficient information and detects redundant features, and that results in a decrease of computational demands and training sets of patterns.

In classical pattern recognition theory [4,5], a pattern is simply defined as an \( n \)-dimensional feature vector \( \mathbf{X} \), given in the following form:

\[ \mathbf{X} = (X_1, X_2, X_3, \ldots, X_n) \]

where \( \mathbf{X} \in \mathbb{R}^n \) (\( \mathbb{R}^n \) is an \( n \)-dimensional Euclidean space). Formal interpretation of the initial selected features as neural network inputs that we used in the study is presented below: \( F_{ri} \in \{20, 21, \ldots, 70\} \), the age of patients in years, given in area (A) in the M-17 form (Table 1), a risk factor of particular importance which strongly influences other risk factors.

\( F_{ri} \in \{0, 1\} \) (i = 2, 3, \ldots, 19), risk factors given in area (B) in the M-17 form, where (1) and (0) assign the presence or absence of the \( i \)-th risk factor, respectively.

No risk. **Low risk. ** Medium high risk. *** High risk.

<table>
<thead>
<tr>
<th>S No.</th>
<th>Risk Factor</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Age</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Menarche before the age of 12</td>
<td>**</td>
</tr>
<tr>
<td>3</td>
<td>First delivery after the age of 30</td>
<td>**</td>
</tr>
<tr>
<td>4</td>
<td>Having children</td>
<td>**</td>
</tr>
<tr>
<td>5</td>
<td>Breast feeding</td>
<td>**</td>
</tr>
<tr>
<td>6</td>
<td>Abortions</td>
<td>**</td>
</tr>
<tr>
<td>7</td>
<td>Contraception</td>
<td>*</td>
</tr>
<tr>
<td>8</td>
<td>Hormonal disorders</td>
<td>**</td>
</tr>
<tr>
<td>9</td>
<td>Menopause after the age of 50</td>
<td>**</td>
</tr>
<tr>
<td>10</td>
<td>Breast injuries</td>
<td>*</td>
</tr>
<tr>
<td>11</td>
<td>Mastitis</td>
<td>***</td>
</tr>
<tr>
<td>12</td>
<td>Benign breast tumor</td>
<td>***</td>
</tr>
<tr>
<td>13</td>
<td>Cancer in family</td>
<td>**</td>
</tr>
<tr>
<td>14</td>
<td>Obesity over 10%</td>
<td>*</td>
</tr>
<tr>
<td>15</td>
<td>Psycho-stress</td>
<td>***</td>
</tr>
<tr>
<td>16</td>
<td>Migration rural urban</td>
<td>*</td>
</tr>
<tr>
<td>17</td>
<td>Born and living in the city</td>
<td>*</td>
</tr>
<tr>
<td>18</td>
<td>Thyroid disorders</td>
<td>**</td>
</tr>
<tr>
<td>19</td>
<td>Allergy</td>
<td>*</td>
</tr>
</tbody>
</table>

Fig. 1. Neural net in breast cancer diagnosis.
Gr_i \in \{0.1/3, 2/3, 1\}, (j = 1.2, \ldots , 19), the risk factors graded by age, given in Table 2, where (0) represents zero level of risk ("-"), (1/3) represents low level of risk ("("), (2/3) represents medium level ("="), and (1) a high level of risk ("*").

S_i \in \{0, 1\}, (k = 1.2, \ldots , 9), symptoms given in area (C) of the M-17 form, where (1) and (0) assign the presence or absence of the kth symptom.

In the present study the initial input vector X is hence defined as

\[ X = (F_{r1}, F_{r2}, \ldots , F_{r19}, G_{r1}, G_{r2}, \ldots , G_{r19}, S_1, S_2, \ldots , S_9)^T, \]

with dimension of \( n = 47 \) expressing the number of all initially selected features \( (19 + 19 + 9) \) (Table 3).

4. A neural network architecture and learning algorithm for pattern classification

4.1. A neural network architecture for pattern classification

A multilayer perceptron architecture (Fig. 2) is defined by input, output and at least one hidden layer of processing neurons, where the neurons of two consecutive layers are completely connected by synapses formally represented by the weight matrix between mentioned layers. Bias members, represented by neurons having no inputs and a fixed output equal to a value of 1, are joined to the input and hidden layers, consequently influencing their performance.

The number of neurons in each layer without the bias neuron is \( N(i) \), where \( i = 1, \ldots , L \) and \( L \) represents the number of layers. The neurons of the input layer are linear, and they distribute input signals from the information environment to the neurons of the first hidden layer. The neurons belonging to the hidden layers and the output layer transform incoming signals by a nonlinear differentiable sigmoid logistic function or tangent hyperbolic function.

The output vector \( Y^{i+1}(t) \) of the \((i+1)\)-th layer at the time step \( t \) is expressed as follows:

\[ Y^{i+1}(t) = F(Z^{i+1}(t), A^{i+1}(t-1)) \]

where

\[ Z^{i+1}(t) = W^i(t)X^i(t) \quad \text{for} \quad i = 1, \ldots , L - 1, \]

\[ X^{i+1}(t) = Y^i(t) \quad \text{for} \quad i = 1, \ldots , L - 1, \]

where \( X^i(t) = [x_1^i(t), \ldots , x_{N(i)}^i(t), 1]^T \), is the \( i \)-th layer input vector. \( Z^{i+1}(t) = [z_1^{i+1}(t), \ldots , z_{N(i+1)}^{i+1}(t)]^T \), is the net input of the \((i+1)\)-th layer. \( Y^i(t) = [y_1^i(t), \ldots , y_{N(i)}^i(t)]^T \), is the output of the \((i+1)\)-th layer. \( A^{i+1}(t) = Z^{i+1}(t) + A^{i+1}(t-1) \) = \([a_1^{i+1}(t), \ldots , a_{N(i+1)}^{i+1}(t)]^T \), is the state of activation of the \((i+1)\)-th layer, \( F(A^{i+1}(t)) = [F(a_1^{i+1}(t)), \ldots , F(a_{N(i+1)}^{i+1}(t))]^T \), is the activation function of the \((i+1)\)-th layer, and

\[ W^i = \begin{bmatrix}
  w_{11} & \cdots & w_{1N(i+1)} \\
  \vdots & \ddots & \vdots \\
  w_{N(i)+11} & \cdots & w_{N(i)+1,N(i+1)}
\end{bmatrix}, \]

where \( W^i \) represents the weight matrix between two consecutive layers, \( i \) and \((i-1)\), \( w_{ij} \) is the weight between the \( j \)-th neuron of the \((i+1)\)-th layer and \( i \)-th neuron of the \( i \)-th layer. Initial weights \( w_{ij} \) assigned randomly between \(-1\) and \( 1 \) are later determined by the back propagation learning procedure [7].

Neural network architectures used for pattern classification are generally but not always designed to have the same number of output neurons as the number of classes where the \( p \)-th neuron corresponds to the \( p \)-th class. The target vectors, \( Y^p_o = [y^p_1, \ldots , y^p_{N(i)}]^T \), are defined so that for classes \( p \) and \( k \) \((k \neq p)\), the next expressions are valid: \( y^p_k = 1 \) and \( y^p_k = 0 \). The input vectors (patterns to be classified) \( X^q = [x_1^q, \ldots , x_{N(i)}^q]^T \), presented to the instructed neural classifier produce the output vector, \( Y^q = [y_1^q, \ldots , y_{N(i)}^q]^T \), which takes real bounded values as components. The input vector
which produces the highest value at the $p$th component of the output vector ($p$th output neuron) of the neural network is classified as class $p$.

The three-layer neural network used in this paper (Fig. 2) is designed to have only one output neuron which is determined by the actual tangent hyperbolic activation function to produce the output $Y^{q} \in [-1,1]$, never reaching boundary values ($-1$ and $+1$). The output vector and the target vector in this case are reduced to scalar values, $Y^{q} = y^{q}$, where $y^{q} \in [-1,1]$, and $Y^{q} - \gamma_{y}^{q}$, where $\gamma_{y}^{q} = \{-1,1\}$. The values $-1$ and $1$ in the target
vector formula correspond to class I (healthy patients) and class II (ill and high risk patients), respectively. We define decision rule for classification of the \( q \)th input vector Eq. (4) with the main intention of avoiding false negative results, i.e., high risk patients are not to be classified as healthy group:

\[
\text{If } Y^q > 0.90, \text{ then } X^q \in \text{class I.}
\]

\[
\text{otherwise } X^q \in \text{class II.}
\]

4.2. Applied learning algorithm

Back propagation is an iterative learning procedure that adjusts weights through a gradient descent with respect to cost function \( E \) which should be minimized to the assigned positive value \( E_j \). The introduced cost function:

\[
E = \frac{1}{2} \sum_{q} \| Y^q_j - Y^q \|^2.
\]

is the mean square error between the actual output and the desired output. This algorithm requires a continuous differentiable nonlinear activation function, for example, a tangent hyperbolic,

\[
y = f(x) = \frac{1 - e^{-x + \theta}}{1 + e^{-x + \theta}}
\]

(which has been used in designed neural network classifier), where \( x \) is an activation of the considered neuron, and \( \theta \) is a threshold.

The learning algorithm then assumes several steps:

Step 1. Set the initial weights and thresholds to small random values between -1 and 1.

Step 2. Present randomly chosen training data pair from the set of \( Q \) pairs of \( N(1) \)-dimensional input vectors and associated \( N(L) \)-dimensional desired output vectors,

\[
\{ (x', y'_q) \} \text{ for } i = 1, \ldots, N(1),
\]

\[
j = 1, \ldots, N(L), \quad q = 1, \ldots, Q.
\]

to a selected neural network.

Step 3. Use the specified function (5) and formulas as in Eq. (1) to calculate the actual output of the network \( Y^q \) and cost function \( E \). If \( E < E_j \), then stop the learning procedure; otherwise go to Step 4.

Step 4. Adapt weights using recursive algorithm starting at the output layer and propagating back through the hidden layers, adjusting the weights in the following way:

\[
w^r_s(i + 1) = w^r_s(i) + \eta \delta^r_{i+1} Y^q_s
\]

for \( i = 1, \ldots, L - 1 \).

where \( w^r_s \) is the weight from the \( s \)th neuron of the \( i \)th layer to the \( r \)th neuron of the \((i+1)\)th layer in the time step \( t \eta \) (eta) is a learning gain term, and \( \delta^r_{i+1} \) is an error term obtained for neuron \( r \) from the \((i+1)\)th layer. For the neuron \( r \) which belongs to the output layer \((L)\) the error term is defined by the desired and actual outputs:
If neuron \( s \) belongs to the \( i \)th hidden layer, then

\[
\delta_i^s = y_i^s (1 - y_i^s) (y_i^L - y_i^s).
\]  

(7)

where \( k = N(i + 1) \) is the number of neurons in the \((i + 1)\)th layer. To speed up convergence, the next formula has been adopted for updating the weights:

\[
w_{ij}^{(t+1)} = w_{ij}^{(t)} + \eta \delta_i^{L-1} \sum_{j=1}^{k} \delta_j^{(t+1)} w_{ij}^{(t)}.
\]  

(8)

where \( 0 < \eta < 1 \). The other way to speed up convergence is to use the adaptive learning rate parameter

\[
\eta(t+1) = \begin{cases} 
\rho \eta(t) & \text{if } E(t+1) > \xi E(t) \\
\sigma \eta(t) & \text{otherwise}, 
\end{cases}
\]  

where \( \eta \) takes small initial values from 0.01 to 0.001, \( \rho \approx 1.1 \) denotes the multiplier to increase learning rate, \( \sigma \approx 0.7 \) is the multiplier used to decrease the learning rate parameter, and \( \xi \approx 1.04 \) is the error ratio—the maximum ratio of new to old error allowed for acceptance of new weights and biases.

**Step 3.** Go back to Step 2.

### 5. Breast cancer detection

We first selected a set of relevant features (risk factors and symptoms) referring to breast cancer, which, we expected, would help us identify the critical class of patients. Then we collected data referring to the mentioned features, for two classes of patients: a class of healthy patients with a risk, and patients with already established diagnosis of breast cancer. Later on, we used the prepared data as patterns to instruct a neural network to classify them in two corresponding groups. The second task described in this paper, performed by the neural network, was a feature extraction procedure based on the evaluation of impact of every single feature (input) to the output of previously properly instructed neural network.

#### 5.1. A neural network implementation

In the first phase a feedforward neural network with 47 input neurons (Table 3), five neurons in the hidden layer and one output neuron was exposed to the back propagation learning algorithm. PC/AT 486 at 133 MHz was used for all simulations. One characteristic of the back propagation algorithm is long training time. particularly when the neural network contains many neurons in the hidden layers, that usually depend on the complexity of the task to be performed. We began the heuristic approach to the selection of the number of hidden neurons by taking two neurons, and finally accepted five neurons in the hidden layer. To speed up convergence we used adaptive learning rate as specified in Section 4.2 \((\eta = 0.01, \rho = 1.1, \sigma = 0.7 \text{ and } \xi = 1.04)\).

In the first phase, using random pattern selection procedure, we got randomly chosen sets of learning and test patterns. The selected network has been trained with training sets containing the initial set of features (47 inputs). The training sets contained different number of available patterns while the rest of the patterns were used for the test. The goal of the training/testing procedure was to achieve reliable classification of the patients into two exclusive categories: sick and healthy. The results obtained in the first phase with different training/testing sets are shown in Table 4. The network has also been trained with specific training patterns, one at a time, consisting of risk factors and symptoms (see Tables 5 and 6 for the results).

As we can see, the symptoms in this case are not a reliable set of features for proper selection of patients, that is caused by the presence of asymptotic
patients in the selected data. The instructed neural network showed high accuracy in the classification of patients belonging to the asymptomatic group, which is frequently classified as group at low risk, even by the experts.

5.2. Neural network for feature selection

The initial set of features can be redundant or useless for classification. It is beneficial to reject those features that are redundant, achieving decreasing complexity and computational demands. Hence, the objective is to detect the minimum number of features that can ensure accuracy of classification, close to the one computed from the initial set of features. There are many different approaches to perform the feature extraction procedure, for example [19–21], and we propose yet another. In order to detect the main features and decrease the initial number of these for further observations, we used the neural network, instructed to appropriately classify the full set of patterns (200 patients), where patterns contained the initial number of features.

In Fig. 3, the classification results of the instructed neural network obtained with the training set of patterns are presented. Network trained in this way was used to test the influence of every single feature on the decision process (output). We achieved this by presenting the simulation input vectors to the instructed neural network. Aiming to examine relevancy of the $i$th feature for the classification, where $i = 1, \ldots, 47$, let us consider related factors:

$$X_i^y = \left[ x_{i1}^y, \ldots, x_{iR}^y, \ldots, x_{iL}^y \right]^T,$$

$$Y_q^y = \left[ y_q^y \right],$$

$$F_q = \frac{1}{2} \sum_{j=1}^{Q} \| Y_q^j - Y_q^y \|^2 \quad \text{for} \quad q = 1, \ldots, Q.$$
where $X'_i$ is the simulated input vector, and differs from the original only for the $i$th component $x'_i \neq x''_i$. The $i$th component of $X'_i(x'_i)$, takes some of the subsequent simulated values; maximal, minimal or random. In the present study maximal and minimal values were 1 and 0, respectively. $Y'_q$ is the output produced by the simulated $q$th input vector, $E'_q$ is the cost function produced by the set of $Q$ simulation input vectors with $x'_q = 1$ for $q = 1, \ldots, Q$, and $Q$ represents the number of patterns (patients) used for the simulation task.

Exposing the instructed neural net to the simulated input vectors containing the $i$th component $x'_i = 1$, in one case or $x'_i = 0$, for $q = 1, \ldots, Q$, in the other we obtain different values for $E'_q$. Repeating this procedure $k$ times (for $k$ sets of patterns which were randomly generated) we obtain mean value of the relevant cost function:

$$E'_s = \frac{1}{k} \sum_{i=1}^{k} E'_i$$  \hspace{1cm} (12)

It is easy to see that for $i \in \{1, 2, \ldots, 47\}$ there exists $p = i$ such that the next relation is satisfied:

$$\overline{E}_s(p - i) - \overline{E}_s(p) = \min(\overline{E}_s).$$  \hspace{1cm} (13)

The index $p$ denotes the position of the feature in the input vector that influenced output the least, and that should be rejected (Fig. 4). After that a new set of input vectors was created, with dimension decreased for 1. This procedure of feature elimination continued repeatedly until classification accuracy was achieved on the reduced set of features and remained close to the classification accuracy based on the original set of features.

![Fig. 3. Output of the neural net, trained with 200 patterns. All patients classified properly in two exclusive classes. A value of +1 denotes a healthy group.](image)

![Fig. 4. Output of the trained network submitted to a simulated maximal input value instead the feature signal which slightly influences the output.](image)
Fig. 5. Output of the instructed network submitted to a simulated maximal (+1) input value instead of the original feature signal (age) which strongly influences classification.

Fig. 6. Output of the instructed network submitted to a simulated minimal (-1) input value replacing the original feature signal (age) which strongly influences the output.

Fig. 7. Output of the trained network submitted to a randomly selected sample of binary values instead of original signal which strongly influences output.

Fig. 5 represents the output obtained by the set of simulated input vectors, presented to the trained network, which causes the healthy population to become classified closer to the sick one in proportion with the relevancy of the input feature. Replacing the same feature with minimal simulated input (absence of risk) results in the sick population becoming classified closer to the healthy one (Fig. 6). Features that influenced output the least were rejected (Fig. 7), and the network was trained again with this reduced set of features. The criterion for rejection of features, as previously defined, was the minimal influence of a particular feature on the instructed network output. This
procedure of reduction of features was repeated iteratively until we finally reduced the set of features to 29. Results of training/testing procedure performed with the reduced set of features are given in Table 7. The features identified by the presented procedure corresponded to the group of

Table 7
Neural network classification results obtained extracted (29) set of features

<table>
<thead>
<tr>
<th>Exp. no.</th>
<th>Overall success rates (%) for different training/testing number of patterns</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td>80/120</td>
<td></td>
</tr>
<tr>
<td>Malign</td>
<td>97.6</td>
</tr>
<tr>
<td>Benign</td>
<td>91.2</td>
</tr>
</tbody>
</table>

Table 8
Reduced set of features

<table>
<thead>
<tr>
<th>S. No</th>
<th>Feature</th>
<th>Description</th>
<th>Num. values</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Fr1</td>
<td>Age</td>
<td>20, ..., 70</td>
</tr>
<tr>
<td>2.</td>
<td>Fr2</td>
<td>Menarche before the age of 12</td>
<td>0, 1</td>
</tr>
<tr>
<td>3.</td>
<td>Fr3</td>
<td>First delivery after the age of 30</td>
<td>0, 1</td>
</tr>
<tr>
<td>4.</td>
<td>Fr4</td>
<td>Having children</td>
<td>0, 1</td>
</tr>
<tr>
<td>5.</td>
<td>Fr5</td>
<td>Breast feeding</td>
<td>0, 1</td>
</tr>
<tr>
<td>6.</td>
<td>Fr6</td>
<td>Abortions</td>
<td>0, 1</td>
</tr>
<tr>
<td>7.</td>
<td>Fr7</td>
<td>Contraception</td>
<td>0, 1</td>
</tr>
<tr>
<td>8.</td>
<td>Fr8</td>
<td>Hormonal disorders</td>
<td>0, 1</td>
</tr>
<tr>
<td>9.</td>
<td>Fr9</td>
<td>Menopause after the age of 50</td>
<td>0, 0.5, 1</td>
</tr>
<tr>
<td>10.</td>
<td>Fr12</td>
<td>Benign breast tumors</td>
<td>0, 1</td>
</tr>
<tr>
<td>11.</td>
<td>Fr13</td>
<td>Cancer in family</td>
<td>0, 1</td>
</tr>
<tr>
<td>12.</td>
<td>Fr18</td>
<td>Thyroid disorders</td>
<td>0, 1</td>
</tr>
<tr>
<td>13.</td>
<td>Fr19</td>
<td>Allergy</td>
<td>0, 1</td>
</tr>
<tr>
<td>14.</td>
<td>Gr1</td>
<td>Fr1 Graded by age</td>
<td>0, 1/3, 2/3, 1</td>
</tr>
<tr>
<td>23.</td>
<td>Gr10</td>
<td>Fr10 Graded by age</td>
<td>0, 1/3, 2/3, 1</td>
</tr>
<tr>
<td>24.</td>
<td>Gr12</td>
<td>Fr12 Graded by age</td>
<td></td>
</tr>
</tbody>
</table>

Risk Factor

<table>
<thead>
<tr>
<th>S. No</th>
<th>Feature</th>
<th>Description</th>
<th>Num. values</th>
</tr>
</thead>
<tbody>
<tr>
<td>25.</td>
<td>S4</td>
<td>Breast swelling</td>
<td>0, 1</td>
</tr>
<tr>
<td>26.</td>
<td>S6</td>
<td>Enlarged axillary nodes</td>
<td>0, 1</td>
</tr>
<tr>
<td>27.</td>
<td>S7</td>
<td>Nipple secretion</td>
<td>0, 1</td>
</tr>
<tr>
<td>28.</td>
<td>S8</td>
<td>Blood from nipple</td>
<td>0, 1</td>
</tr>
<tr>
<td>29.</td>
<td>S9</td>
<td>Redness of the breast skin</td>
<td>0, 1</td>
</tr>
</tbody>
</table>
risk factors which has been defined as priority for clinical breast examination [36,42]. According to our results the group of main risk factors for the breast cancer, besides age, were pointing to hormones (menarche before the age of 12, menopause after 50, first delivery after 30 and thyroid disorders), history of benign tumors or long-lasting fibrocystic dysplasia and family history of breast cancer. The reduced set of risk factors is given in Table 8.

6. Conclusions

The model proposed in this paper, based on the case-controlled study, gave encouraging results for early breast cancer detection in comparison to existing survey-based selective screening and other methods. What we achieved with this model is considerable reduction of the false negative results which frequently appear in detection through mammogram interpretation. Existence of many approaches to breast cancer detection and prediction increase the chances for an efficient solution to the problem. We suggest the neural network model which will use the data obtained by mammography, unified with data about relevant risk factors, and present the probable manifestations of the disease.

Acknowledgements

The authors are grateful to Professor Vuceta Mandic and Vlada Radivojevic for their helpful advice. We would like to thank the anonymous reviewers for their suggestions and corrections.

References

Drasko Furundzic was born in Rasovo, The Old Rascia Area, Yugoslavia, in 1958. He received his BS degree in mechanical engineering at Belgrade University. He was a graduate student at the Automatic Control Department, High School of Electrical Engineering, and also at the Experimental Physiology & Pathophysiology Department, Medical School of Belgrade University. He is currently researcher at the Milutin Pupin Institute, Automatic Control Department/Intelligent Systems Division, Belgrade. His research interests include neural network theory and application as one of exploration tools for the conscious matter phenomenon.

Ana Jocićević Bekić was born in Belgrade, Yugoslavia, in 1963. She received her BS in medicine (MD) at the Medical School University of Belgrade in 1987. She specialized in epidemiology for her postgraduate studies at the same University, in 1996. She is currently Head of the Department of Epidemiology and Prevention at the Institute of Oncology and Radiology of Serbia, Belgrade. She is also a member of the Serbian Medical Association (SMA) and the Balkan Union of Oncologou (BUON).
Miodrag Dj. Djordjevic was born in Pirot, Yugoslavia, in 1938. He received his BS in medicine (MD) at the Medical School University of Belgrade, in 1963. He specialized in epidemiology for his postgraduate studies, and submitted a thesis titled "Breast Cancer Risk Factors" at the same University, in 1976 and 1986, respectively. He worked at the Institute of Oncology and Radiology of Serbia, Belgrade, for 26 years, and is now at the Clinic Bazariska Kosa, Belgrade. He has published many papers and monographs. He is a member of several international associations for breast cancer epidemiology, and one of the reviewers of the European Journal of Cancer.