JOURNAL OF ENGINEERING annals OF FACULTY OF ENGINEERING HUNEDOARA

CANCER OUTCOME PREDICTION BY A RADIAL BASIS FUNCTION NEURAL NETWORK USING IMMUNITY BASED APPROACH

Diana TSANKOVA, Vania RANGELOVA

Technical University Sofia - Branch Plovdiv, BULGARIA

ABSTRACT:
Diffuse Large B-cell Lymphoma (DLBCL) is curable in about 40% of patients. The treatment (chemotherapy or stem cell support) of DLBCL depends on the distinction of the significant subtypes of this kind of cancer and influences on the cured/fatal outcome of the disease. Molecular analyses of clinical heterogeneity in DLBCL focus on individual genes, some of which are correlated with DLBCL treatment outcome. The paper proposes a methodology of using a radial basis function (RBF) neural network for classification of cured/fatal lymphomas on the basis of gene expression data. An immunological approach uses the training data in order to initialize the radial basis functions – their number and the positions of centers. Two variants of a cancer outcome predictor (a conventional RBF network and the proposed here immunity based one) are tested in MATLAB environment on 58 data samples (32 cured and 26 fatal) available in the literature. The high prognostic accuracy of the predictors is confirmed by leave-one-out cross-validation test.
KEYWORDS:
gene expression profiling, diffuse large B-cell lymphoma (DLBCL), discrete artificial immune network, radial basis function neural network.

1. INTRODUCTION

Most outcome prediction models of Diffuse Large B-cell Lymphoma (DLBCL) do not identify the molecular basis of clinical heterogeneity of this kind of cancer and could not determine the proper therapeutic regimens (doses of conventional chemotherapeutic agents and stem-cell support). There exist correlations between some of genes and DLBCL treatment outcome. Different machine learning methods have been used for achieving a high classification and prognostic accuracy: hierarchical clustering with unsupervised learning [1]; 'weighted voting' based classification with supervised learning [2]; support vector machines [2]; clustering methods (k-means and c-means), neural networks, and fuzzy-neural hybrid methods [3]; evolving connectionist systems (ECOS) [4], etc. Recently artificial immune networks have been used to predict cured/fatal DLBCL outcome [5, 6, 7, 13]. For solving such significant problem it is helpful to apply a variety of approaches, because each of them has own advantages and drawbacks and is more or less proper to be used in a particular situation.

Radial Basis Function (RBF) neural networks are powerful approximators of multivariable nonlinear continuous functions. Their simple two-layer feed-forward architecture and learning algorithm based on the solution of a linear regression problem lead to a fast training process. The performance of RBF networks depends on the number and positions of the radial basis functions in the hidden layer. De Castro and Von Zuben [8] have proposed an artificial immune network (aiNet) developed to perform automatic data compression, clustering and classification problems. Then they have refined this algorithm and directly applied it to the problem of defining RBF network centers [9].

The purpose of the paper is to propose a methodology for a Lymphoma cancer outcome prediction using a RBF neural network whose hidden layer neurons are tuned by immunity

based algorithm [9]. The prediction model is based on gene expression data. Two variants of a cancer outcome predictor (a conventional RBF network and the proposed here immunity based one) are tested in MATLAB environment on 58 data samples (32 cured and 26 fatal) available in [2]. The high prognostic accuracy of the predictors is confirmed by leave-one-out cross-validation test.

2. GENE EXPRESSION PROFILING OF DLBCL

In this section some gene expression profiles available in [2] are represented in short in order to make easier the describing and understanding the proposed an immunity based RBF network for DLBCL outcome prediction. Twelve-level grey scale modification [6] of the 13 genes, determined in [2] as important and related to the disease outcome, are shown in Figure 1. The genes expressed at higher levels in cured DLBCL are shown on top while the genes that are more highly expressed in fatal disease are shown on the bottom. Black indicates high expression, white - low expression. Each column is a sample (a patient) and each row is a gene. The first rows of the cured and fatal sections show *an idealized expression profile*. Expression profiles of the 32 cured DLBCLs are arranged on the left, while profiles of the 26 fatal tumors are on the right.

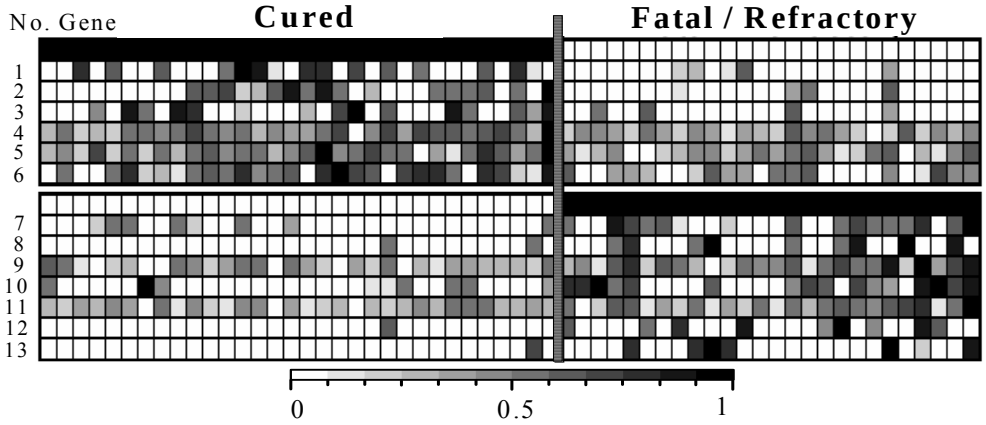

Figure 1. Thirteen genes included in the DLBCL outcome model [2], shown in 12-level grey scale modification [6]. (The description and identifiers of the genes are given in [2])

3. RBF NEURAL NETWORK

A radial basis network is a feed-forward network with two layers: a hidden layer of radial basis neurons and an output layer of linear neurons (Figure 2). The hidden neurons apply a nonlinear transformation from the input space into the hidden space. They are locally tuned and their responses are outputs of radial basis functions. The linear output neurons implement a weighted sum of the hidden neuron responses. The RBF network's excellent approximation capabilities have been studied in [10]. Let the input vector be $\mathbf{x} = (x_1, x_2, ..., x_p)$, and the RBF network outputs are computed as follows:

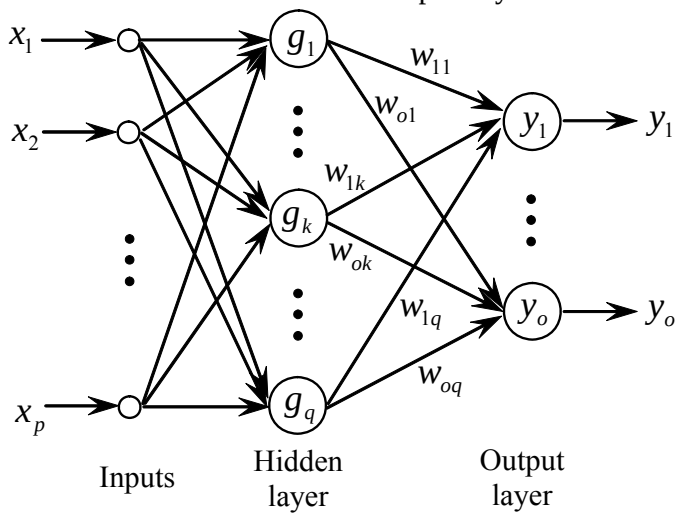

Figure 2. RBF neural network [9]

$$y_i = \mathbf{w}_i^{\mathrm{T}} \mathbf{g} = \sum_{j=1}^{q} w_{ij} g_j, \quad i = 1,...,o, \qquad (1)$$

where $\mathbf{w}_i = (w_{i1}, w_{i2}, ..., w_{iq})^{\mathrm{T}}$ is the

weight vector for the output neuron i, o is the number of network output neurons, and $\mathbf{g} = (g_1, g_2, ..., g_q)^T$ is the vector of basis functions. The output of each radial basis function is

$$g_j = h(\| \mathbf{x} - \mathbf{c}_j \|), \sigma_j), \quad j = 1, ..., q, \tag{2}$$

where $h_j(\cdot)$ is the *basis function*, $\| \cdot \|$ is a norm (the Euclidean norm) defined on the input space, $\mathbf{c}_j \in \Re^p$ - a center (a prototype vector) j, and σ_j - the standard deviation (dispersion).

Let the problem to solve determines an RBF network with p inputs and one output ($o = 1$). Consider a set of R input data points ($\mathbf{x}_i$) and the desired output values are available for all these R data points (d_i). Each basis function can be centered on one of these data points, i.e. the number of centers is equal to the number of data points $q = R$ [9]. The introduction of the full training set into the Equation (1) gives rise to the following matrix notation which shows the linearity of the weights determination problem:

$$\mathbf{H}\,\mathbf{w} = \mathbf{d}, \tag{3}$$

where $\mathbf{d}$ is the desired response vector, $\mathbf{w}$ - the output weight vector, and $\mathbf{H}$ is an $R \times R$ interpolation matrix. The solution of Equation (3) is

$$\mathbf{w} = \mathbf{H}^{-1}\mathbf{d}. \tag{4}$$

The prerequisite for the existence of $\mathbf{H}^{-1}$ is that the R data points are *distinct* [9]. But the setting-up $q = R$ is not a god strategy for the training of RBF networks because of: (1) poor generalization to previously unseen data, (2) high likelihood of obtaining an ill-conditioned matrix $\mathbf{H}$ due to a great amount of redundant data (linearly dependent vectors) [9]. To overcome these drawbacks, a new set of $q_1 < R$ basis functions (assumed to be linearly independent) has to be defined. In this case the solution $\mathbf{w}^*$ is given by [9]

$$\mathbf{w}^* = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{d}. \tag{5}$$

In this paper a fixed radial basis function is chosen to be an activation function of the hidden neurons. The locations of the centers are chosen here from the training data set using an immune-inspired approach [9]. A Gaussian function with standard deviation fixed according to the spread of the centers is used for the RBFs:

$$h(\| \mathbf{x} - \mathbf{c}_j \|), \sigma_j) = \exp\left(\frac{\| \mathbf{x} - \mathbf{c}_j \|^2}{\sigma_j^2} \right), \quad j = 1, ..., q_1, \tag{6}$$

where q_1 is the number of centers (hidden neurons), $\sigma_j = d_{\max} / \sqrt{2q_1}$ - the standard deviation (the same for all basis functions) [9], $d_{\max}$ - the maximum distance between the chosen centers, $\mathbf{x}$ is the input vector, and $\mathbf{c}_j$ is the j-th center location. This choice for σ_j guarantees that the individual radial basis functions are not too peak or too flat [9].

The proper functioning of an RBF neural network depends mainly on the choice of the number and positions of its radial basis function centers. In the paper this choice is made using a preliminary compression and clustering of the input data by an artificial immune network – aiNet, which will be described in the next sections.

4. IMMUNE SYSTEM: AN OVERVIEW

The body maintains a large number of immune cells - *lymphocytes* (mainly *T-cells* and *B-cells*). When an *antigen* (foreign body) invades the body, only a few of these immune cells can recognize the invader. Self/nonself discrimination is the process of distinguishing between our own cells and invaders: the immune system is said to be tolerant to those cells recognized as self, while the others are eliminated. According to Jerne's *idiotypic network hypothesis* [11], lymphocytes communicate with each other through interaction among antibodies. *B-lymphocytes* produce 'Y' shaped *antibodies* that recognize an *antigen* like a *key and lock relationship*. The structure of the antigen and the antibody is shown in Figure 3,

where the part of the antibody that recognizes (matches) the corresponding antigen determinant is known as V-region. It is a variable (V) region, which can alter its shape to achieve a better match (complementarity) with a given antigen, and the affinity of their match measures the strength and specificity of the 'antigen-antibody' interaction. Antibodies also have antigenic characteristic and a key and lock relationship also exists between different species of antibodies. Thus the stimulation and suppression chains among antibodies form a large-scaled network.

The immune system learns to increase the size of population and the affinity of those lymphocytes that have recognized any antigen. Secretion of antibodies requires that B cells become activated, undergo proliferation (cloning) and then differentiate into plasma and memory cells. A *clone* is a cell (or a set of cells), which is the progeny of the same cell. A *memory cell* is the cell with high affinity with the antigen that will be saved for a faster and stronger response to a previously seen (or near) antigen. The initial exposure of immune system to an antigen leads to production of small clones of B cells, each producing antibody with different affinity. The immune response to secondary encounters is enhanced by storing some high affinity antibody producing cells from the first infection (memory cells). They form a large initial high affinity clone for subsequent encounters. Affinity *maturation* is the process of growth in concentration and affinity of those cells that recognize antigens. During maturation random changes (mutations) are introduced into the variable region (V) and an increase in the affinity of the antibody can be occasionally obtained. These high-affinity cells are kept in the pool of memory cells, and those ones possessing receptors with low antigenic affinity, or the self-reactive cells, must be eliminated or may undergo receptor editing. *Hypermutation* is a rapid accumulation of mutations, which lead to a fast maturation of the immune response. The regulation of hypermutation process depends on receptor affinity: cells with low affinity receptors may be mutated, while in cells with high-affinity receptors, mutation may be gradually inactivated.

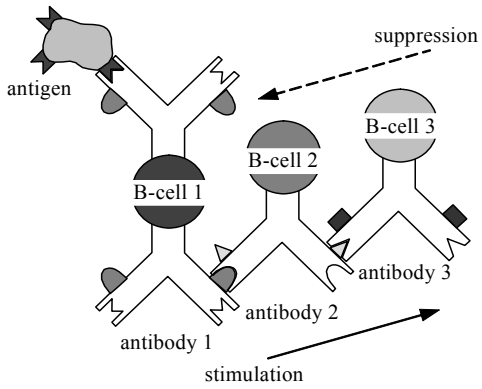

Figure 3. Structure of immune network

The ability of the immune system to adapt its B-cells to new types of antigen is powered by a process known as *clonal selection*, i.e., this is the basic process of pattern recognition and selection [12]. Biological clonal selection occurs to the degree that a B-cell matches an antigen. A strong match causes a B-cell to be cloned many times, and a weak match – few times. These 'clones' are mutated from the original B-cell at a rate inversely proportional to the match strength.

The dynamics of immune network represents the variation in time of the concentrations and affinities of network cells. The *metadynamics* expresses the continuous production of novel antibodies and death of non-stimulated or self-reactive cells. The main purpose of the immune network theory is to determine internal images, corresponding to the individual's molecular structure, which emerge from a network organization after learning the molecular structure of the environment.

5. DE CASTRO & VON ZUBEN'S AINET LEARNING ALGORITHM

In this section, the aiNet learning algorithm is presented in such a way as it has been proposed by De Castro and Von Zuben [8]. The aiNet learning algorithm builds a memory set that recognizes and represents the data structural organization in compressed form. After learning, the aiNet contains clusters which serve as *internal images* (*mirrors*) responsible for mapping real clusters present in the data set into network clusters. The aiNet architecture is represented by final number and spatial distribution of clusters. The interactions of an antibody with an antigen and other antibodies are presented by measures of the similarity degree (affinity). The 'antigen-antibody' ('antibody-antibody') affinity is inversely proportional to the distance between them. Intra-clonal self-recognizing antibodies are

eliminated by a *clonal suppression*, while network *suppression* searches for similarities between different sets of clones. A suppression threshold controls the specificity level of the antibodies and the clustering accuracy: the more specific the antibodies, the less compression rate, and vice versa.

The notation adopted for the description of De Castro & Von Zuben's aiNet learning algorithm, as well as the aiNet algorithm are shown in Table 1 and Table 2, respectively. Each presented antigenic pattern elicits a clonal immune response, according to the aiNet learning algorithm. Matrix $\mathbf{Ab}_{\{m\}}$ represents the network internal images of the antigens presented to the aiNet, and matrix $\mathbf{S}$ determines the connections among network antibodies. These two matrixes can be considered as network outputs. Figure 4 presents hypothetical elements of $\mathbf{Ag}$ (stars) and the corresponding elements of $\mathbf{Ab}_{\{m\}}$ (circles) generated by the aiNet algorithm.

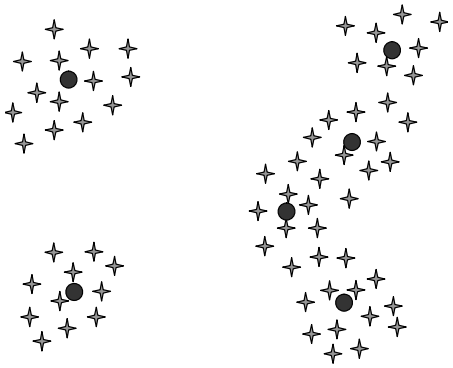

Figure 4. aiNet illustration: learning data (stars) and resulting network antibodies (circles)

The clusters of antibodies $\mathbf{Ab}_{\{m\}}$, obtained after the training process, map those of the original data set $\mathbf{Ag}$ in such a way that the number of antibodies in the network is much smaller (compressed) than the number of data samples ($m < M$ (Table 2)). These antibodies can be used for centers of the radial basis functions of an RBF neural network.

Table 1. Notation adopted for the description of the aiNet learning algorithm [9]

$\mathbf{Ag}$ - population of antigens ($\mathbf{Ag} \in S^{M\times L}$, S is a shape-space [9])
$\mathbf{Ab}$ - available antibody repertoire ($\mathbf{Ab} \in S^{N\times L}$, $\mathbf{Ab} = \mathbf{Ab}_{\{d\}} \cup \mathbf{Ab}_{\{m\}}$)
$\mathbf{Ab}_{\{m\}}$ - total memory antibody repertoire ($\mathbf{Ab}_{\{m\}} \in S^{m\times L}$, $m \le N$)
$\mathbf{Ab}_{\{d\}}$ - d new antibodies to be inserted in $\mathbf{Ab}$ ($\mathbf{Ab}_{\{d\}} \in S^{d\times L}$)
$\mathbf{Ab}_{\{n\}}$ - subset composed of the n highest affinity antibodies is selected ($\mathbf{Ab}_{\{n\}} \in S^{n\times L}$)
$\boldsymbol{f}_j$ - vector containing the affinity of all the antibodies $\boldsymbol{Ab}_i$ $(i = 1,...,N)$ with relation to antigen $\boldsymbol{Ag}_j$. The affinity is inversely proportional to the 'antigen-antibody' distance
$\mathbf{S}$ - similarity matrix between each pair $\boldsymbol{Ab}_i - \boldsymbol{Ab}_j$, with elements $s_{i,j}$ $(i, j = 1,...,N)$
$\mathbf{C}$ - population of N_C clones generated from $\mathbf{Ab}$ ($\mathbf{C} \in S^{N_C\times L}$)
$\mathbf{C}^*$ - vector containing the affinity between every element from the set $\mathbf{C}^*$ with $\boldsymbol{Ag}_j$
$\boldsymbol{d}_j$ - vector containing the affinity between every element from the set $\mathbf{C}^*$ with $\boldsymbol{Ag}_j$
ζ - percentage of the mature antibodies to be selected
$\mathbf{M}_j$ - memory clone for antigen $\boldsymbol{Ag}_j$ (remaining from the process of clonal suppression)
$\mathbf{M}_j^*$ - resultant clonal memory for antigen $\boldsymbol{Ag}_j$
σ_d - natural death threshold
σ_s - suppression threshold

Table 2. De Castro & Von Zuben's aiNet learning algorithm [9]

I. At each iteration, do:
1. For each antigenic pattern $\boldsymbol{Ag}_j$, $j = 1,...,M$, ($\boldsymbol{Ag}_j \in \mathbf{Ag}$), do:
(a) Determine its affinity $f_{i,j}$, $i = 1,...,N$, to all $\boldsymbol{Ab}_i$:
$f_{i,j} = 1/D_{i,j}$, where $D_{i,j} = \| \boldsymbol{Ab}_i - \boldsymbol{Ag}_j \|$, $i = 1,...,N$ (7)
(b) A subset $\mathbf{Ab}_{\{n\}}$ composed of the n highest affinity antibodies is selected;
(c) The n selected antibodies are going to proliferate (clone) proportionally to their antigenic affinity $f_{i,j}$, generating a set $\mathbf{C}$ of clones: the higher the affinity, the larger the clone size for each of the n selected antibodies. The total clone size N_C generated for each of the M antigens is:
$$N_C = \sum_{i=1}^{n} round(N - D_{i,j} N), \quad (8)$$
where N is the total amount of antibodies in $\mathbf{Ab}$, $round(\cdot)$ is the operator that rounds towards the closest integer, and $D_{i,j}$ - the distance between the selected antibody i the antigen $\boldsymbol{Ag}_j$, given by (7).
(d) The set $\mathbf{C}$ is submitted to a directed affinity maturation process (guided mutation) generating a mutated set $\mathbf{C}^*$, where each antibody k from $\mathbf{C}^*$ will suffer a mutation with a rate α_k inversely proportional to the antigenic affinity $f_{i,j}$ of its parent antibody: the higher the affinity, the smaller the mutation rate:

$$\boldsymbol{C}_k^* = \boldsymbol{C}_k + \alpha_k(\boldsymbol{Ag}_j - \boldsymbol{C}_k);\ \alpha_k = 1/f_{i,j};\ k = 1,...,N_C;\ i = 1,...,N \quad (9)$$

(e) Determine the affinity $d_{k,j} = 1/D_{k,j}$ among $\boldsymbol{Ag}_j$ and all the elements of $\mathbf{C}^*$:

$$D_{k,j} = \| \boldsymbol{C}_k^* - \boldsymbol{Ag}_j \|,\ k = 1,...,N_C. \quad (10)$$

(f) From $\mathbf{C}^*$, re-select ζ % of the antibodies with highest $d_{k,j}$ and put them into a matrix $\mathbf{M}_j$ of clonal memory;

(g) *Apoptosis*: eliminate all the memory clones from $\mathbf{M}_j$ whose affinity $D_{k,j} > \sigma_d$:

(h) Determine the affinity $s_{i,k}$ among the memory clones: $s_{i,k} = \| \mathbf{M}_{j,i} - \mathbf{M}_{j,k} \|,\ \forall i,k$ (11)

(i) *Clonal suppression*: eliminate those memory clones whose $s_{i,k} < \sigma_s$.

(j) Concatenate the total antibody memory matrix with the resultant clonal memory $\mathbf{M}_j^*$ for $\boldsymbol{Ag}_j$:

$$\mathbf{Ab}_{\{m\}} \leftarrow [\mathbf{Ab}_{\{m\}}; \mathbf{M}_j^*];$$

2. Determine the affinity among memory antibodies from $\mathbf{Ab}_{\{m\}}$

$$s_{i,k} = \| \boldsymbol{Ab}_{\{m\}}^i - \boldsymbol{Ab}_{\{m\}}^k \|,\ \forall i,k. \quad (12)$$

3. *Network suppression*: eliminate all the antibodies such that $s_{i,k} < \sigma_s$.

4. Build the total antibody matrix $\mathbf{Ab} \leftarrow [\mathbf{Ab}_{\{m\}}; \mathbf{Ab}_{\{d\}}]$

II. Test the stopping criterion.

6. CANCER OUTCOME PREDICTOR USING AN AINET-BASED RBF NETWORK

Consider a cancer outcome prediction task. An aiNet-based RBF network is proposed to predict the cured/fatal DLBCL outcome (and therefore the proper treatment) on the basis of gene expression of 13 genes. The RBF network classifies patients into two categories: category *cured* - people who will survive under conventional chemotherapy treatment, and category *fatal* – people who will die despite of the treatment. An aiNet is used for a preliminary input data processing aiming to determine the number and positions of the radial basis functions of the neural network. Each gene expression sample of DLBCL (a column in Figure 1) works as an *antigen*. The antigen is identified by L-position string ($L = 13$), containing 12-level grey scale colours in each position. The grey colours are scaled between white and black, and they are replaced in calculations with corresponding values, linear-proportional in the interval $[0;1]$. Antibodies are identified by the same type of characteristics (L-position strings). After training the aiNet discovers clusters in the cured and fatal classes DLBCLs. These clusters are compressed images of gene expression data, imprinted on aiNet as memory antibody sets. These memory antibodies can be used to determine the number and positions of the radial basis functions' centers.

As it is proposed in [13] the Euclidean distance measure is used for determining the distance between the antigen and any antibody (and between each two antibodies):

$$D_{i,j} = \| \boldsymbol{Ab}_i - \boldsymbol{Ag}_j \| = \sqrt{\sum_{k=1}^{L} \left(\boldsymbol{Ab}_{i,k} - \boldsymbol{Ag}_{j,k} \right)^2},\ i = 1,...,N. \quad (13)$$

The total clone size N_C for each of the M antigens is calculated as in [13]:

$$N_C = \sum_{i=1}^{n} round\ (N - (D_{i,j} / \sqrt{L})\, N). \quad (14)$$

The mutation rate α_k in (9) is modified increasing search space by introducing a random signal as it is follows: $\alpha_k = D_{i,j} / \sqrt{L} + 0.5(rand - 0.5)$, (15)

where $rand \in [0;1]$ are uniformly distributed random numbers. After mutation, if the digital value corresponding to any position in the strings of $\mathbf{C}^*$ is less than 0 or great than 1, it adopts the boundary value 0 or 1, respectively.

To design the memory cells (internal images), the aiNet learning algorithm (with modifications made above) has to be applied to the gene expression data of cured and fatal patients. The number of distinct clusters of antibodies depends on the choice of the network parameters. The more generalist the antibodies, the more parsimonious the network with relation to the number of antibodies. After training, the memory cells of the aiNet represent the internal images (in compressed form) of the antigens presented as training samples. These memory cells can determine the number and positions of the radial basis functions' centers, i.e., $q_1 = m$ and $\mathbf{c}_j \leftarrow \mathbf{Ab}_{\{m\}j}$, $j = 1,...,q_1$. Then the RBF network can be trained to determine the weights $\mathbf{w}^*$ by (5). The input training (and testing) data consists of the 13-

gene expression samples of the 58 DLBCLs. The corresponding output data is available: the belonging to the two classes of DLBCLs – the 26 cured and 32 fatal patients. But instead of this data, the information containing in the idealized gene expression profile (Figure 1) is used as follows. The idealized profile considers the first six genes as expressed in survival patients and no expressed in fatal patients, and the other seven are determined contrary (Figure 1). This hypothesis gives reason the output training data to be determined as follows:

$$d_i = k_{\text{cured}} \sum_{k=1}^{6} \mathbf{x}_i(k) - k_{\text{fatal}} \sum_{k=7}^{13} \mathbf{x}_i(k), \quad i = 1,2,...,R, \tag{16}$$

where the positive weight coefficients $k_{\text{cured}} = 1$ and $k_{\text{fatal}} = 6/7$ make the influence of the first six genes (high-expressed in cured DLBCLs and low-expressed in fatal ones) commeasure with that of the rest seven (which are expressed contrary). Having in mind the idealized gene expression profiles, it is expected that d_i (16) takes positive values in the cured DLBCLs, and negative – in the fatal ones.

7. RESULTS AND DISCUSSIONS

The proposed method of using an aiNet-based RBF neural network for DLBCL outcome prediction was applied to the 58 samples, shown in Figure 1. Leave-one-out cross-validation built a predictor by removing one sample and then using the rest as an input set to be compressed and clustered by aiNet. The removed sample was used for testing and evaluating the performance of the predictor. Additionally, the results were compared with those obtained by a predictor using a conventional RBF neural network. The two predictors were investigated in MATLAB environment.

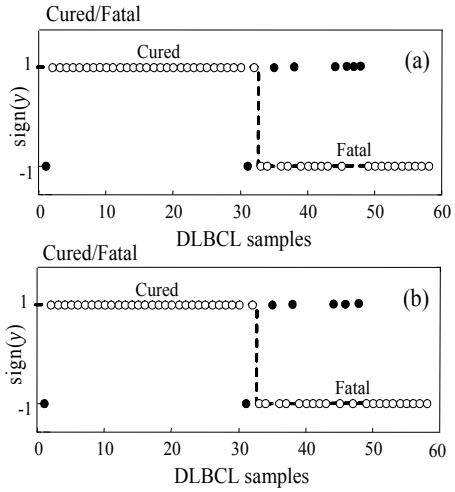

Figure 5. Cured/fatal prediction using: (a) conventional RBF network, (b) aiNet-based RBF network

The conventional RBF network, used here, creates neurons one at a time [14]. At each iteration the input sample which will result in lowering the network error the most, is used to create an RBF neuron. The error of the new network is checked, and if it is low enough the design is finished, otherwise a new neuron is added. This procedure is repeated until the sum-squared error falls below predetermined value ($goal$), or the maximum number of neurons is reached. For simulation of the predictor using a conventional RBF network the following parameters were used: $goal = 0.02$ and $spread = 1$. The $spread$ constant determines the wide of the RBF (the width of an area in the input space to which each neuron responds). The most of the aiNet training parameters were chosen heuristically: $N = 50$, $n = 4$, $d = 5$, $m = 12$, $\zeta = 0.2$, $\sigma_d = 0.5\sqrt{L/2}$, $\sigma_s = 0.25\sqrt{L/2}$, $Generations = 5$. The average compression rate of the resulting aiNet was 46.7%.

Table 3. Predicted class vs. observed one (conventional RBF network)

		Predicted Class		
		Cured	Fatal	
Observed Class	Cured	30	2	32
	Fatal	6	20	26
		36	22	58

Table 4. Predicted class vs. observed one (aiNet-based RBF network)

		Predicted Class		
		Cured	Fatal	
Observed Class	Cured	30	2	32
	Fatal	5	21	26
		35	23	58

The performance of the two DLBCL outcome predictors is shown in Figure 5, Table 3 and Table 4. In Figure 5 the cured cases of DLBCL are represented with values '1', while the fatal cases – with '-1' (The result is obtained as a sign function over the output y of the RBF network, i.e. $\text{sign}(y)$). For both predictors two patients (samples No.1 and 31) from observed class "cured" were predicted wrong as "fatal", while five patients (samples No.35, 38, 44, 46 and 48) from class 'fatal' were predicted as 'cured'. Additionally, the conventional RBF

network predicts wrong one more patient (the sample No.47) from observed class 'fatal' as 'cured'. The proposed aiNet-based RBF network predicts 51 out of 58 samples correctly. 87.9% prognostic accuracy (93.8% class cured and 80.8% class fatal) is achieved. The predictor using a conventional RBF network performs a little worse (86.2%: 93.8% class cured and 76.9% class fatal) than the proposed one. The performance of the aiNet-based RBF network prognostic model is better than the model reported by M. Shipp et al. [2] - 87.9% versus 75%. A little better prognostic results obtained by immune predictors have been reported in the literature, but the predictor either has simultaneously used all the 58 samples for the design and testing (not leave-one-out cross-validation test) [6] or the solution has been very sensitive to some heuristically determined parameters of aiNet (a cutting threshold, which is not used here) [13].

8. CONCLUSIONS

The paper presents a methodology of using an aiNet-based RBF neural network for solving a classification problem oriented to Lymphoma cancer prognosis. A conventional RBF network classifier is additionally simulated for comparison with the proposed one. They both show a high prognostic accuracy, 87.9% for the proposed and 86.2% for the conventional, determined by the 58 leave-one-out cross-validation tests. The preliminary compression and clustering the gene expression data by aiNet determine automatically the appropriate centers of the RBFs and thus improve the performance of the RBF network. The future work will include investigations on the automatic defining not only the number and positions of RBF centers, but also the dispersions σ_j, different for each of the RBF neurons, with a purpose to improve the accuracy of the cancer outcome predictor.

ACKNOWLEDGEMENTS

The paper presents research and development, supported by Ministry of Education and Science in Bulgaria (National Scientific Fund) under the Research Project BY-TH-108/2005.

REFERENCES

[1] Alizadeh, A., et al., Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*, vol.4051, pp.503-511, 2000.

[2] Shipp, M.A., et al., Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning, *Nature Medicine*, vol.8, no.1, pp.68-74, 2002.

[3] Granzow, M., D. Berrar, W. Dubitzky, A. Schuster, F.J. Azuaje, and R. Eils, Tumor classification by gene expression profiling: comparison and validation of five clustering methods, *SIGBIO Newsletter Special Interest Group on Biomedical Computing of the ACM*, ACM Press, vol.21, no.1, pp.16-22, 2001.

[4] Kasabov, N., Bioinformatics: A knowledge engineering approach, *Proc. of the 2nd IEEE Int. Conf. on Intelligent Systems*, Bulgaria, pp.19-24, 2004.

[5] Tsankova, D., V. Georgieva, Cancer Outcome Predicting Model Based on Artificial Immune Network, *Journal of the Technical University of Gabrovo*, vol.31, pp.108-111, 2005.

[6] Tsankova, D., V. Georgieva, and N. Kasabov, Artificial immune networks as a paradigm for classification and profiling of gene expression data, *Journal of Computational and Theoretical Nanoscience*, American Scientific Publishers, vol.2, no.4, pp.543-550, 2005.

[7] Tsankova, D., and V. Rangelova, Cancer outcome prediction by cluster-based artificial immune networks, *Proc. of the 5th IASTED Conference BioMed'07*, Innsbruck, Austria, pp.26-32, 2007.

[8] De Castro, L.N., and F.J. Von Zuben, aiNet: an artificial immune network for data analysis, in *Data Mining: A Heuristic Approach*, Chapter XII, H.A. Abbass, R.A. Sarker, and C.S.Newton, Eds. USA: Idea Group Publishing, pp.231-259, 2001.

[9] De Castro, L.N., and F.J. Von Zuben, An Immunological Approach to Initialize Centers of Radial Basis Function Neural Networks, *Proc. of the 5th Brazilian Conference on Neural Networks*, PUC, Rio de Janeiro - RJ – Brazil, April 2-5, pp. 79-84, 2001.

[10] Park, J., Sandberg, J.W., Universal approximation using radial basis functions network, *Neural Computation*, vol.3, pp.246-257, 1991.

[11] Jerne, N.K., Towards a network theory of the immune system. *Annals of Immunology*, *125C*, 373–389, 1974.

[12] Burnet, F.M., Clonal selection and after, *Theoretical Immunology*, G. I. Bell, A. S. Perelson, & G. H. Pimbley Jr., Eds. Marcel Dekker Inc, pp. 63-85, 1978.

[13] Tsankova, D., V. Rangelova, Modeling Cancer Outcome Prediction by aiNet: Discrete Artificial Immune Network, *Proc. of the 15th Mediterranean Conf. on Control and Automation – MED'07*, Athens, Greece, 2007.

[14] Demuth, H., and M. Beale, *Neural Network Toolbox User's Guide*, Version 3.0, The MathWorks, Inc., 1997.