Prostate Cancer Prognosis Evaluation Assisted by Neural Networks

CORINA BOTOCA[1], RAZVAN BARDAN[2], MIRCEA BOTOCA[2] and FLORIN ALEXA[1]
[1]Communications Department
"Politehnica" University Timisoara
Bv.V.Pârvan, no.2
[2] Department of Urology
"Victor Babes" University of Medicine and Pharmacy, Timisoara
ROMANIA
corina.botoca@etc.upt.ro http://www.etc.upt.ro

Abstract: - Neural networks (NN) are new promising tools that can assist the clinicians in the diagnosis process and in therapy decision making, because they can deal with a great number of parameters, learning from examples and assessing any nonlinear relationships between inputs and outputs. In this paper, the problem of prostate cancer evolution prediction is approached using NN. The original database contained 650 records of patients, which underwent radical prostatectomy for prostate cancer. The NN variables were the parameters with the highest prognostic value selected and pre-processed from the original database. Different NN architectures and NN parameters have been tested in order to obtain the best complexity/accuracy ratio. The input data were structured, according to the latest statistical and representation concepts used in the current medical practice, aiming to improve the global performance. Different experiments were done using the rough database and the structured database. The NN performances were compared with the most widely used prediction statistical method, the logistic regresion. All NN models performed better than the logistic regression. The best obtained global prediction of correct classification 96.94% is better than the results of similar experiments available in literature. The NN prediction performance might be improved, because, in our opinion, its limits are given by the relatively small number of cases and the methods of collecting data.

Key Words: - neural networks, prostate cancer, prediction, capsule penetration

1 Introduction

Neural networks are nonlinear dynamical systems consisting of a large number of relative simple processing elements, named neurons that operate simultaneously, in parallel. The neurons interact through excitatory and inhibitory connections, which have associated weights, in a similar manner to the biological neurons. Learning is performed by weights changes conform to a learning rule. N.N. can learn from examples and deal with a great number of parameters, assessing non-linear relationships between any of the input to any of the desired output variable, with faster and better results, if compared with traditional statistical methods. NN are recommended especially in cases were a conventional process is not suitable, can't be easy defined or cannot fully capture the data complexity and where stochastic behavior is important.

The advantage of NN over conventional techniques lies in their ability to solve problems that can not be mathematically modeled or the solution is too complex to be found.

In medicine, especially in oncology, prediction is very important. For instance, treatment decision making is based on the ability to predict therapy outcome. New tools that could improve prediction are always welcome. These may assist the clinicians in the diagnosis process and in therapy decisions. Although the Tumor-Node-Metastasis (TNM) classification system is widely used, describing the anatomical level of lesion extent, it has some limitations, because it does not include the new tumor markers, or other pathological elements, which are necessary for a specific diagnosis, leading to the most appropriate therapy. Using new tools that can deal with a large amount of data and find relationships among them opens new perspectives. NN offer a viable solution, prediction being a straight forward application, on the basis of examples defined as pairs of input model to output model.

Recent developments, as is the improved computational power and the advances in neural networks software encouraged us to investigate new potential applications, as the finding of correlations between different predicting factors, the prognosis prediction for different groups of patients, or as the risk group stratification.

Various fields of oncology have been targeted by researchers in the domain of NN in order to assist cancer diagnosis and prognosis, as for example: gastric cancer [1] breast cancer [2], [3], lung cancer [4] and prostate cancer [5].

In urology the largest number of experiments using the neural networks was performed for prostate cancer, a domain of great clinical interest, with a large number of validated diagnosis and prognostic parameters, but still hampered by uncertainties [6]. The necessity of early diagnosis, correct staging and of choosing the best therapy, eliminating the potential risk factors, could all benefit from neural networks. An explored application is the staging process, using the results of transrectal prostate biopsy, along other non-invasive parameters, in order to avoid radical therapy indications for advanced forms of cancer, which would not have significant benefits for the patients [7],[8].

Possible metastases prognosis was approached by Tewari et al with a NN [9] on a group of 1200 patients with prostate cancer, using: age, race, tumor size, Gleason score, digital rectal examination and the number of positive biopsy cores as input parameters. The results were promising at that time: the neural network had a specificity of over 72% and sensitivity higher than 81%.

In 2001, Han et al used the clinical stage, Gleason score, preoperative Prostate Specific Antigen (PSA) level and age as input parameters for a multilayer perceptron (MLP), trying to predict lymph node involvement. At a specificity of 90%, the neural network detected 34% of patients with localized prostate cancer and 59% of patients with lymph node involvement [10]. These predictions were much better than those obtained by using the Partin nomograms of that time [11].

In a paper published in 2001 by Borque et al., the performance of several types of neural networks with logistic regression in the prediction of organ confinement after radical prostatectomy was compared [12]. The selected input parameters were: patients' age, clinical stage, biopsy Gleason score and preoperative PSA. The results, obtained with the artificial neural networks simulation software available at that time, were not superior to logistic regression.

In 2003, Zlotta et al. used the data from the European Prostate Cancer Detection Database to train a MLP, trying to predict the pathological stage for 200 patients with total PSA values lower than 10 ng/mL, before radical prostatectomy [13]. The NN input variables were: age, total serum PSA values, free/total PSA ratio, PSA velocity, PSA density, transition zone PSA, digital rectal examination and total Gleason score in trans-rectal prostate biopsy. The accuracy of the obtained prediction was of 92.7% in localized prostate cancer cases and of 84.2% in locally advanced prostate cancer cases. These results were much better than those obtained using the logistic regression.

The prediction accuracy of the LR-based nomogram versus NN was studied considering age, digital rectal examination, PSA, percent-free PSA and prostate volume of 3980 patients, who underwent multicore systematic prostate biopsy [14]. At that time, the accuracy of the nomogram was 71%, versus 67% for the NN.

In paper [15] the probability of prostate cancer was studied based on initial biopsy results. The input variables were age, PSA, digital rectal examination (DRE) and prostate volume, available in 843 cases. The conclusion was that the predictive capacity of the NN was significantly improved in this case, if compared with the previous models, considering only PSA or prostate volume. The NN performance was similar to that of logistic regression.

The above papers and others more recent [16], [17], [18], [19] show continuous interest for the use of NN in everyday urological practice.

2 Problem Formulation

The single most important factor in the prognosis of prostate cancer is capsule penetration [6]. The performances obtained in uro-oncology by neural networks, comparably with those of traditional statistical methods encouraged us to study the problem of prediction of prostate capsule penetration using NN.

2.1 Database

With kind permission, we had access to the database records of the Department of Urology of the St. Radboud University, Nijmegen, the Netherlands. The original database contained 650 records of patients which underwent radical prostatectomy for prostate cancer between 1 January 1992 and 31 December 2005. The recorded parameters were:

- Date of birth;
- Age at diagnosis (years);
- Height (cm);
- Weight (kg);
- Date of initial diagnosis of prostate cancer (confirmed by pathological exam);
- Total Gleason Score in initial specimen (from transrectal prostate biopsy or transurethral resection of the prostate - TURP). The pathological grading is very important, helping to evaluate the aggressiveness of prostate cancer and especially to

predict the future evolution of the patients. The Gleason score is a pathology grading system which has five levels of tumor aggressiveness (grade 1 is the best structured, thus the least aggressive, while grade 5 is the most anaplasic, thus the most aggressive). In our database the total Gleason score, which is the sum of the two most prominent patterns in the pathology specimen was used, with values between 2 and 10.

- Primary, secondary, tertiary and quaternary Gleason paterns;
- The value of preoperative total serum PSA. Prostate specific antigen (PSA) is the most widely used marker for prostate cancer, with an acceptable specificity and sensitivity in the detection of the malignant disease. Values ranging from 0 to 4 ng/ml are generally considered as low risk for prostate cancer, values between 4 and 10 ng/ml raise the suspicion of malignancy, while values over 10 ng/ml are mostly associated with a positive cancer diagnosis.
- The TNM staging is the method used to standardize the prostate cancer diagnosis. It includes three parameters: the tumor stage T has values from T_0 (no tumor) to T_4 (tumor extending into the adjacent organs); lymph node invasion is assessed by the N parameter (N_0 – negative lymph nodes, N_1 – positive lymph nodes); distant metastasis is described by the M parameter (M_0 – no metastasis, M_1 – metastasis present). The T stages present in our database were: T_{2a} , T_{2b} , $T_{2c,}$ T_{3a} and T_{3b}. Taking into account the fact that this parameter was the symbol type in the original database, we have converted it into five distinct values, from 1 to 5.
- Prostate volume, as measured by trans-rectal ultrasound;
- Date of radical prostatectomy (the surgical intervention that removes the prostate, with the intention to cure the prostate cancer, with no need for additional, adjuvant therapy);
- Capsule penetration, which is a sign of poor prognosis: if capsule penetration is found on the prostatectomy pathology specimen, then the intervention generally loses its radical intention, triggering the need for further forms of therapy);
- Seminal vesicles invasion, which means that the tumor is not any longer confined inside the prostate;
- PSA recurrence (quantified by the rise of the postoperative total seric PSA values over 0.1 ng/ml – normally the PSA value is at one month after the surgery below 0.1 ng/ml). The PSA rise, defining the PSA recurrence, is the single most significant prediction factor of postoperative recurrence of prostate cancer. When we have analyzed the ratio between preoperative and postoperative PSA values and we have compared them with the presence of capsule penetration, we have found that from the 203 patients with capsule penetration 148 (72.9%) had also PSA recurrence, while from the 337 patients without pathological capsule penetration, only 4 (1.19%) had subsequently PSA recurrence. Our findings thus confirmed that the value of the PSA recurrence is highly correlated with capsule penetration, making it a very good prognosis factor.
- Time interval from prostatectomy to PSA recurrence (in months).

As it is known, the success of a NN application significantly depends on the database gathering [20]. In this designing phase expertise in the field of the problem to be solved, respectively in urology, was absolutely necessary and invaluable. We selected the parameters with the highest prognostic value for prostate cancer diagnosis, excluded the cases with incomplete information and obtained the a reliable database of 548 cases, for our retrospective study.

The parameters finally selected were:

- Preoperative stage (according to the 2002 TNM classification): data from patients in stages T_{2a}, $T_{2b,}$ T_{2c}, T_{3a}, T_{3b} were included (T_4 cases were excluded, because if the cancer is too advanced, radical prostatectomy alone has not curative intent);
- Age (42 – 73 years);
- Preoperative total PSA value (range: 0-84 ng/mL);
- Total Gleason Score in initial specimen (primary + secondary, range: 3-9);
- Capsule penetration (No / Yes).

2.2 Neural networks models

During the years various types of NN architectures have been developed, which, from the point of view of information flow, can be classified in feed forward NN and recurrent NN [20], [21].

2.2.1 Multilayer perceptron

Multilayer perceptron (MLP) is the most usual type of feed forward NN, which propagates the information from the input towards the output layer. MLP has an input layer of neurons, a number of hidden layers and an output layer. It is a universal approximator. This means that MLP can map any nonlinear input–output function to an arbitrary degree of accuracy, provided that a sufficient number of hidden neurons are used. The architecture of a MLP with a single hidden layer is represented in Fig.1.

Typically, all the neurons from a layer have the same continuous activation function. The continuous activation functions allow the use of gradient based methods to train the NN parameters. Usually a sigmoid function is used [20], for example given by the folowing relation:

$$o_{pj} = \frac{1}{1 + e^{-\beta(\sum_i w_{ji} x_{pi}(t) + \theta_j)}} \qquad (1)$$

where

- x_{pi} is the input to the i neuron for the p input model;
- o_{pj} is the output of the neuron j for the p input model;
- θ_j is the bias of the j neuron;
- w_{ji} is the weight of the interconnection between the i input and the j neuron;
- β is a proportionality parameter, chosen in the (0,1) interval;

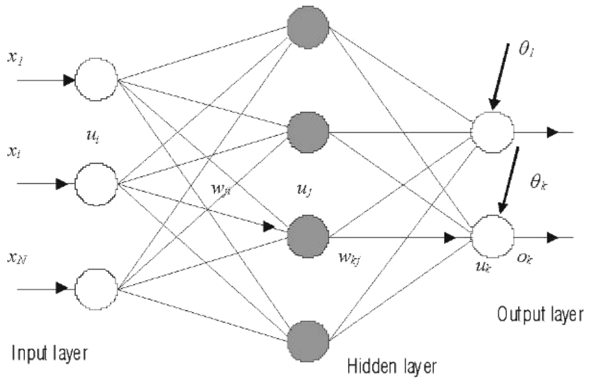


Fig.1 The architecture of a multilayer neural network

To train the NN parameters, respectively the weights and biases, usually the back-propagation error algorithm (BKP), a supervised training algorithm is used. It requires pairs of input model-output desired model ($\boldsymbol{x}_p$, $\boldsymbol{d}_p$) and it has two stages. During the first stage the information propagates through layers, from input towards output. In the second stage, the errors between the desired output and the current output propagate from output towards input, determining changes in the NN parameters. The BKP algorithm minimizes the overall mean square error [20].

There are different stages in developing a MLP application: the training stage, the testing stage and the practical operation.

Gathering the data base and splitting it into the training and testing sets have a major influence on the success of the NN application [20]. During training each pair of input model–desired model (x_p,d_p) are repeatedly applied to the NN, in order to minimize the error. This conducts to a relatively long training time, which is one of the drawbacks of the BKP algorithm. Another BKP deficiency is that it can get stuck into a local minimum. There are several methods to avoid this, but usually the cross-validation is used [20]. During the testing stage, new input models are applied to the NN and its performance is verified.

2.2.2 Radial basis function neural networks

Recently, the radial basis function (RBF) neural network received considerable attention, since it can avoid the above mentioned drawbacks of the MLP network. The RBF network is able to approximate, as the MLP does, any arbitrary nonlinear function in the complex multi-dimensional space with a reduced computing complexity comparative with other NN [21]. In Fig.2 the RBF structure, which contains an input layer, one single hidden layer and an output layer, is represented.

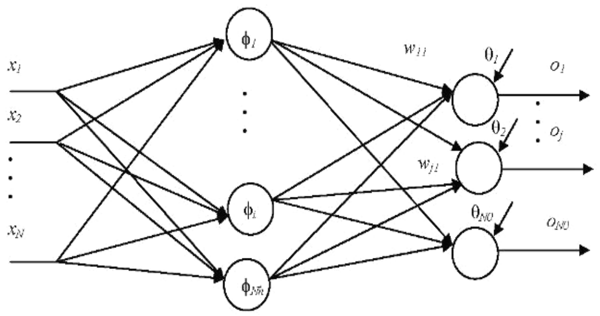


Fig. 2 The structure of a radial basis function neural network

The RBF NN output is determined with the relation:

$$o_{pj} = \sum_{i=1}^{N_h} w_{ji} \phi_i \left(\left\| \boldsymbol{x}_p - \boldsymbol{c}_i \right\| \right) + \theta_j \qquad (2)$$

where:

- o_j is the neuron j output;
- w_{ji} is the weight interconnection from the j neuron to the i neuron;
- θ_j is the bias of j neuron;
- $\phi_i(\bullet)$ is the radial function of the i hidden neuron;
- x_p is the p input model;

- $\boldsymbol{c}_i$ is the centre vector associated to the i neuron;
- $\| \|$ is a metric distance, between its arguments;
- N_h is the number of the hidden neurons;

The RBF NN parameters are the centers vectors $\{\boldsymbol{c}_i\}$ and the weights vectors $\{\boldsymbol{w}_j\}$. Each of the neurons of the hidden layer calculates a metric distance between the input model x and its centre vector $\boldsymbol{c}_i$. This distance can be of different types, but usually the Euclidian distance is used [21].

The closer the input x_p is to the vector centre c_i, the smaller will be the distance. In case that x is identical with c_i the Euclidian distance is zero. The result is passed through a real, nonlinear, continuous activation function, symmetrical to the centre $\phi_i(\bullet, \rho_i)$, $\phi : R^+ \to R$, named radial function. This function also gives the name to the NN.

Several radial functions can be used, for example: the Gauss function, the multiquadratique function, the inverse multiquadratique, the Cauchy function [21]. Usually the Gauss function is used:

$$\phi(x) = e^{\frac{-x^2}{\rho^2}} \tag{3}$$

where ρ is a parameter named radius, width or spread. The radius may be chosen proportionally with the input dispersion.

For the inputs close to the centre, the radial basis function output is greater, approaching to 1 as the Euclidian distance is decreasing to 0 and it is converging to zero as the distance between the input and the centre is increasing. Thus the RBF –NN is capable of a local modeling of the inputs.

Theoretical and practical studies prove that the type of radial function does not essentially influence the performances of RBF-NN.

Training a RBF NN means to determine the number of basis functions (hidden neurons), the centers vectors, the radius of the radial functions and the output interconnections weights. For some algorithms these steps are separately treated, while in others the parameters are found simultaneously. An unsupervised learning algorithm is usually used to find the number and positions of the hidden neurons and a supervised learning algorithm is used to determine the weights of the output layer [21].

The design and training of RBF-NN essentially depend on the centers, thus a lot of studies deal with this problem. The following strategies imposed into practice: randomly choosing a subset of fixed centers from the database [20], the standard competitive algorithm (the k mean) [22], the competitive algorithm with conscience [23], the rival penalization competitive algorithm [24], the dynamic rival penalization algorithm [25], a subset selection from the database using the orthogonal least squares method [26] and the supervised centers selection [20].

The basic approaches to determine the output layer weights can be classified into off line and on line methods. Off line methods include the usual least squares method, while the others include the least mean square algorithm (LMS) or recursive least square (RLS). The simplest way is to use the LMS algorithm [21], which updates the weight interconnections with the relation:

$$\Delta_p \mathbf{w}_{ji} = \eta e_{pj} \phi_i \tag{4}$$

where: e_{pj} is the difference between the desired output d_{pj} and the current NN output o_{pj} for the p input model:

$$e_{pj} = d_{pj} - o_{pj} \tag{5}$$

This algorithm minimizes the mean square error:

$$MSE = \frac{1}{P} \sum_{p=1}^{P} e_p^2 \tag{6}$$

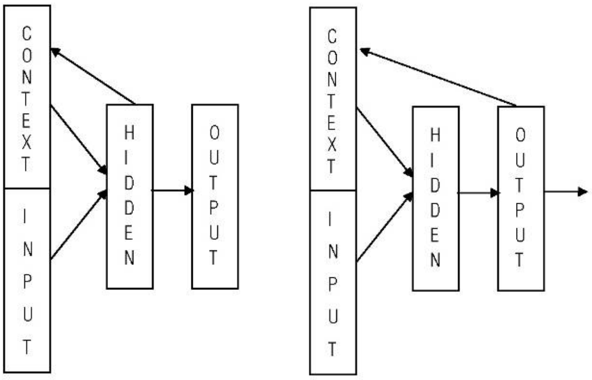
where P is the number of data samples .

2.2.3 Recurrent neural network

Recurrent NN (RNN) is the most general case of NN which stores a record of the prior inputs and factor them with the current data to produce the output. In a RNN, information about past inputs is fed back into and mixed with the inputs, through recurrent or feedback connections, for hidden or output neurons. In this way, the neural network contains a memory of the past inputs via the activations (see Fig.3).

Two types of RNN are known, partial and fully RNN.

Two major architectures for limited recurrent networks are widely used, the Elman structure and the Jordan structure. Elman [27] suggested allowing feedback from the hidden neurons to a set of additional inputs called context neurons. The output neurons function may be linear or nonlinear. Earlier, Jordan [28] described a network with feedback from the output back to a set of context neurons. This form of recurrence is a compromise between the simplicity of a feed-forward network and the complexity of a fully recurrent neural



network because it still allows the popular back propagation training algorithm to be used.

Fig.3 Partial recurrent neural networks

Fully recurrent NN have each neuron connected to all the others. The output of a neuron depends not only on the outputs of the other neurons, but also on theirs previous outputs. The RNN neurons have nonlinear activation functions and a complex dynamic behaviour, so these networks are especially recommended for real time applications. RNN with the same structure have different dynamic evolution if they are trained with a different algorithm. Thus a RNN is completely defined by both architecture and training algorithm. Some of the training algorithms are back propagation through time with its variants and real time recurrent learning [29]. These algorithms minimize the gradient of the global error. The numerical complexity of calculus is N^4, where N is the number of the RNN neurons. The reduced dimension of the RNN if compared with the MLP and RBF-NN, in conditions of the same performance, is the major argument of using RNN in real time applications.

2.3 Performance parameters

As performance parameters the mean square error (MSE), the positive predictive value (PPV), the negative predictive value (NPV) and global percentage of correct classification (GPCC) were considered relevant.

The mean square error was defined in the previous section, by relation (6).

The positive predictive value represents the proportion of patients actually having the feature of interest among those classified as positive.

The negative predictive value means the proportion of patients not having the feature of interest among those classified as negative. Ideally one would like to have high values for both predictive values.

The global percentage of correct classification was calculated as the proportion of correctly classified patients (regardless of the positive or negative value), from the entire testing dataset.

3. Problem Solution

As NN application development environment we used Neurosolutions 5.0 for Excel. We have pre-processed data from the original SPSS format, a statistical analysis software program and exported them into the Microsoft Excel format, compatible with the Neurosolutions software. The cohort contained 365 cases with no prostate capsule penetration and 183 cases diagnosed after radical prostatectomy with positive capsule penetration.

The input data were randomized and pre-processed as follows:

- Age, total Gleason Score and preoperative PSA were considered in absolute value;
- Preoperative TNM stages were assigned with the following values: $T_{2a} = 1$; $T_{2b} = 2$; $T_{2c}= 3$; $T_{3a} = 4$; $T_{3b} = 5$;
- To the predicted output, the capsule penetration, values of 0 and 1 were assigned, for the cases without prostate capsule penetration, respectively for the cases with penetration.

The partition of the database into the training, validation and testing group has a major influence on the NN overall performance [20]. We used different partitions of our database according to some empirical common sense, in order to improve the prediction accuracy.

Several models of prostate penetration prediction have been developed, using various NN architectures: MLP, radial basis function NN and recurrent NN.

In order to obtain the best complexity/accuracy ratio, we have used various structures (different number of neurons on the hidden layers), different activation functions and different training algorithms.

3.1 Example 1

We have used the following data sets:

- 300 cases for training;
- 70 cases for cross-validation;
- 178 cases for testing.

The MLP NN with the best performance had a structure of 4 neurons on the first hidden layer, 8 neurons on the second one and 1 neuron as output. A non-linear activation function, the hyperbolic tangent, was used to update all the neurons and the back-propagation algorithm with momentum to train the NN parameters. The selected momentum was 0.7.

Another NN with a good performance was a RBF NN combined with a MLP. The RBF–MLP NN had a structure of 4 neurons on the first hidden layer, 8 neurons on the second one (the MLP layer) and 1 neuron as output. The Gaussian activation function and Euclidean distance were used for the RBF neurons and the hyperbolic tangent for the other ones. The competitive algorithm with conscience was used to obtain the RBF centers and the back-propagation algorithm with momentum was used to determine the MLP parameters. The chosen momentum was 0.7.

The RNN was of Jordan-Elman type and it had a structure of 4 neurons on the first hidden layer, 8 neurons on the second one and 1 output neuron.

The performance parameters, respectively the mean squared error (MSE), positive predictive value (PPV) and negative predictive value (NPV) of these networks are shown in Table 1. The MSE, the PPV and NPV in our study were better than the results of similar experiments available in literature [9], [10], [12].

Table 1 NN performance parameters

Type of NN/ Performance parameter	MSE	PPV[%]	NPV[%]
MLP (4-8-1)	0.05970759	90.74074074	97.58064516
RBF-MLP (4-8-1)	0.073881769	90.74074074	96.77419355
RNN (4-8-1)	0.042889437	92.59259259	97.58064516

Table 2 represents the NN performance in terms of the global percentage of correct prediction.

For comparison, the statistical logistic regression (LR) applied to the same database, in order to predict the prostate capsule penetration, has given a percentage of correctly classified cases of 94.89%.

It can be observed that all NN prediction models performed better than the logistic regression.

In the best case it can be noticed an improvement of 1.18 %. This can be interpreted as a promising result.

Table 2 Global performance

Type of NN	Global percentage of correct prediction
MLP (4-8-1)	95.51%
RBF-MLP (4-8-1)	94.94%
RNN (4-8-1)	96.07%
LR	94.89%

3.2. Example 2

The following partition of the database was used:

- 300 cases for training;
- 150 cases for cross-validation;
- 98 cases for testing.

The MLP NN with the best performance had a structure of 4 neurons on the first hidden layer, 8 neurons on the second one and 1 neuron as output. A non-linear activation function, the hyperbolic tangent, was used to update all the neurons and the back-propagation algorithm with momentum to train the NN parameters. The chosen momentum was 0.7 as well.

Another NN with a good performance was a RBF NN combined with a MLP. The RBF–MLP NN had a structure of 4 neurons on the first hidden layer, 8 neurons on the second one (the MLP layer) and 1 neuron as output. The Gaussian activation function and Euclidean distance were used for the RBF neurons and the hyperbolic tangent for the other ones. The competitive algorithm with conscience was used to obtain the RBF centers. The back-propagation algorithm, with a chosen momentum of 0.7, was used to determine the MLP parameters.

Various RNN with different memory constants have been tested: partially RNN and fully RNN. The best one had a structure of 4 neurons on the first hidden layer, 8 neurons on the second one and

1 output neuron. The back-propagation through time algorithm with the chosen momentum of 0.7 was used to train the RNN parameters. The memory constant was 0.8.

Table 3 represents the NN performance in terms of positive predictive value, negative predictive value and global percentage of correct prediction. The PPV, NPV and GPCC in our study were better than the results of similar experiments available in literature [9], [10], [12].

Table 3. NN performance parameters

Type of NN/ Performance parameter	PPV[%]	NPV[%]	GPCC[%]
MLP (4-8-1)	96.88	95.45	95.92
RBF-MLP (4-8-1)	96.88	93.94	94.9
Partial RNN (4-8-1)	96.88	95.45	95.92
Fully RNN (4-8-1)	100	95.45	96.94
LR	-	-	94.89%

It can be observed that all NN prediction models performed better than the statistical method. In the best case it can be noticed an improvement of 2.05% of the global percentage of correct classification. This can also be interpreted as very encouraging.

3.3. Example 3

In our final experiments, we have compared the performance of a MLP NN, using the original database and new dataset, resulting from a stratification of the parameters which will be presented below.

In both cases, we have used a MLP NN with the following structure: only one hidden layer, with 8 neurons, hyperbolic tangent activation function and the back-propagation algorithm, with a momentum of 0.8. The training set included 350 patients, the cross-validation had 50 patients, while the testing set was composed of 148 patients.

For the stratification of the input parameters we have used the selection criteria recently published by Makarov et al [30], criteria which have been used for the development of the last edition of the Partin nomograms.

The first stratified parameter was the PSA value, which was divided into the following categories, with the assigned values in brackets: 0 – 2.5 ng/ml (1), 2.6 – 4 ng/ml (2), 4.1 – 6 ng/ml (3), 6.1 – 8 ng/ml (4), 8.1 – 10 ng/ml (5), > 10 ng/ml (6).

The second parameter, the T stage, was stratified as follows: T2a (1), T2b + T2c (2), T3a + T3b (3).

Finally, the total Gleason score was stratified in the following categories: Gleason score 5 and 6 (1), Gleason score 7 from 3+4 (2), Gleason score 7 from 4+3 (3), Gleason score ≥ 8 (4).

We have obtained the following results: for the both databases mean square error was relatively high (5.72% – rough database vs. 5.45% – stratified database), but the predictive values were very good, as we can observe from Table 4 overall prediction accuracy was the same, 96.59%, with small differences between PPV and NPV.

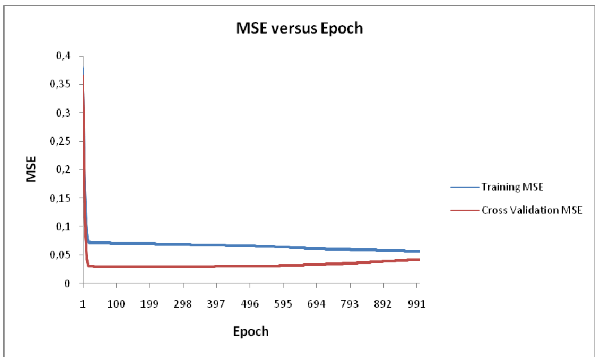


Fig. 4. The evolution of MSE for the rough database

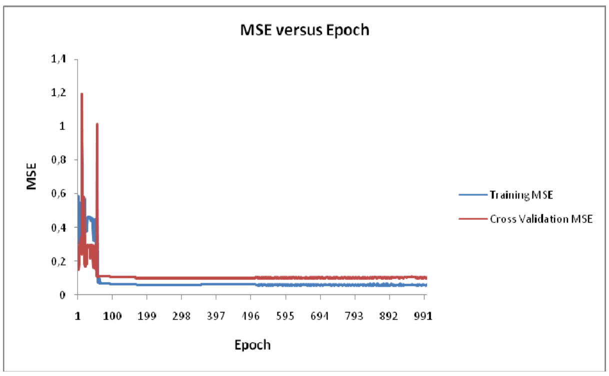


Fig. 5 The evolution of MSE for the stratified database

Table 4. NN performance parameters

Database/ Performance parameter	PPV[%]	NPV[%]	GPCC[%]
Rough	96.55	96.63	96.59
Stratified	94.83	97.75	96.59

These results proof once again that NN can manage well sets of data distributed in a very large range of values.

4. Conclusions

Neural networks (NN) are promising tools that can assist the clinicians in the diagnosis process and in therapy decision making, because they can deal with a great number of parameters, learning from examples and assessing any nonlinear relationships between inputs and outputs.

The NN performance in prostate cancer prediction was better than one obtained with the statistical logistic regression. The best gain of improvement by using NN is 2.05%, quite a significant difference. In clinical terms this is beneficial, avoiding over-treatment in a number of cases with prostate capsule penetration, when radical prostatectomy is no longer an option. This also has a beneficial psychological impact on the patient, by avoiding unnecessary surgery.

Using a stratified database, according to the latest statistical and representation concepts used in the current medical practice, did not improve the global percentage of correct classification.

The performance limits of the neural network prediction, in our opinion, are given by the rather reduced dimension of the database and by its retrospective modality of collecting.

Taking into account a larger number of parameters, as for instance the third and fourth grade Gleason patterns might improve the overall NN prediction.

References:

[1] Afshar S., Abdolrahmani F., Tanha F V., Seaf M.Z., Taheri K., Quick and reliable diagnosis of stomach cancer by artificial neural network, *Proceedings of the 10th WSEAS International Conference on Mathematics and Computers in Biology and Chemistry (MCBC'09),* Prague, Czech Republic, 2009, pp.30-35.

[2] Sudhir D. S. , Ashok A. G. , Amol P. Pande, Neural network aided breast cancer detection and diagnosis, *Proceedings of the 7th WSEAS International Conference on Neural Networks* , Cavtat, Croatia, 2006, pp.158-163.

[3] Gulzar A. Khuwaja Computer-Aided Breast Cancer Detection with Neural Networks using Digital Mammography, *WSEAS Transactions On Signal Processing*, 2006, Vol.2, No.5, pp.811-818.

[4] Kim Le, Automated Detection of Early Lung Cancer and Tuberculosis Based on X-Ray Image Analysis, *Proceedings of the 6th WSEAS International Conference on Signal, Speech and Image Processing*, Lisbon, Portugal, 2006, September 22-24, pp.110-115.

[5] Botoca C, Bardan R, Botoca M., Alexa F, Prediction of Prostate Capsule Penetration Using Neural Networks, *Recent Advances in Computational Intelligence, Man-machine Systems and Cybernetics Proceedings of the 8th WSEAS International Conference on Computational Intelligence, Man-machine Systems and Cybernetics*, Puerto De La Cruz, Canary Islands, Spain, 2009, pp.108-111.

[6] Thompson I, Thrasher JB, Aus G, Burnett AL, Canby-Hagino ED, Cookson MS, D'Amico AV, Dmochowski RR, Eton DT, Forman JD, Goldenberg SL, Hernandez J, Higano CS, Kraus SR, Moul JW, Tangen CM., AUA Prostate Cancer Clinical Guideline Update Panel. Guidelines for the management of clinically localized prostate cancer, *Journal of Urology* , Vol 177, No 6, 2007, pp.2106-2131.

[7] Snow X PB, Smith DS, Catalona WJ, Artificial neural networks in the diagnosis and prognosis of prostate cancer: a pilot study, *Journal of Urology,* Vol.152, 1994, pp.1923-1926.

[8] Anagnostou A, Remzi M, Lykourinas M, et al Artificial neural networks for decision-making in urologic oncology, *European Urologyl,* Vol.43, No.6, 2003, pp.596-603.

[9] Tewari A, Narayan P., Novel staging tool for localized prostate cancer: a pilot study using genetic adaptive neural networks*, Journal of Urology*, Vol 60, No.2, 1998, pp.430-436

[10] Han M, Snow PB, Brandt JM, Partin A.W., Evaluation of artificial neural networks for the prediction of pathologic stage in prostate carcinoma, *Cancer*, Vol. 91, No.8, 2001, pp.1661-1666.

[11] Partin AW, Kattan MW, Subong EN, et al Combination of prostate-specific antigen,

clinical stage, and Gleason score to predict pathological stage of localized prostate cancer: a multi-institutional update, *JAMA* , Vol.277, No.18, 1997, pp.1445-1451.

[12] Borque A, Sanz G, Allepuz C et al., The use of neural networks and logistic regression analysis for predicting pathological stage in men undergoing radical prostatectomy: a population based study, *Journal of Urology*, Vol.166, 2001, pp.1672-1678.

[13] Zlotta AR, Remzi M, Snow PB, Schulman C., Marberger M., Djavan B, An artificial neural network for prostate cancer staging when serum prostate specific antigen is 10 ng/ml or less, *Journal of Urology,* Vol.169, No.5, 2003, pp.1724-1728.

[14] Chun F., Graefen M., Briganti A., Gallina A., Hopp J., Kattan M., Huland H., Karakiewicz P., Initial Biopsy Outcome Prediction—Head-to-Head Comparison of a Logistic Regression-Based Nomogram versus Artificial Neural Network *European Urology*, Vol. 51, No 5, pp.1236-1243.

[15] Rodríguez Alonso A, Pértega Díaz S, González Blanco A, Pita Fer̕nandez S, Suárez Pascual G, Cuerpo Pérez MA The utility of artificial neural networks in the prediction of prostate cancer on transrectal biopsy, *Actas Urol Esp.*, Vol.30, No.1, 2006, pp.18-24.

[16] Stephan C, Meyer HA, Kwiatkowski M, Recker F, Cammann H, Loening SA, Jung K, Lein M, A (-5, -7) proPSA based artificial neural network to detect prostate cancer, *European Urology*, Vol.50, 2006, pp.1014–1020.

[17] Stephan C, Xu C, Finne P, Cammann H, Meyer HA, Lein M, Jung K, Stenman UH, Comparison of two different artificial neural networks for prostate biopsy indication in two different patient populations, *Urology,* 2007, Vol.70, No.3, pp.596-601.

[18] Stephan C., Cammann H., Deger S , Schrader M , Meyer H A., Miller K., Lein M , Jung K, Benign Prostatic Hyperplasia-associated Free Prostate-specific Antigen Improves Detection of Prostate Cancer in an Artificial Neural Network*, Urology*, Vol.74, No.4, 2009, pp.873-877.

[19] Stephan C., Cammann H., Bender M. , Miller k. , Miller K., Lein M , Jung K, Meyer H A., Internal validation of an artificial neural network for prostate biopsy outcome, *International Journal of Urology,* Vol. 17, No. 1, 2010, pp.62-68.

[20] Haykin S, *Neural Networks*. Mcmillan Publishing Co.,Englewood Cliffs, 1994.

[21] Hu Y.H., Hwang Jenq-Neng, *Handbook of Neural Networks Signal Processing*, New York, CRC Press, 2002.

[22] Hecht-Nielsen, *Neurocomputing* Addison-Wesley Publishing Company, 1990

[23] Ahalt S.C, Krishnamurty A.K., Chen P.,D. Melton E., Competitive algorithms for vector quantization, *Neural Networks*, Vol.3, No.3, 1990, pp. 277-290.

[24] Xu L., Krzyzak A., Oja A.E., Rival penalized competitive learning for clustering analysis. RBF net and curve detection, *IEEE Trans. on Neural Networks*, Vol.4, No.4, 1993, pp.636-648,

[25] Botoca C., Budura G., Complex data clustering using a new competitive learning algorithm, *Proceedings of International Conference on Applied Electronics,* Pilsen, Czech Republic, *2006,* pp.23-26

[26] Chen S.. Cowan C.F., Grant P.M., Orthogonal least squares learning algorithm for radial basis function networks", IEEE. Transactions on Neural Networks, Vol. 2, No.2, 1991, pp.302-309.

[27] Elman J.L., Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning*, Vol.7, No.2-3, 1991, pp.195 – 225.

[28] Jordan, M. Attractor dynamics and parallelism in a connectionist sequential machine. *Proc. Eighth Annual Conf. Cognitive Sci. Soc.*, New York, 1986, pp. 531–546.

[29] Haykin, S., *Neural Networks - A Comprehensive Foundation,* Prentice-Hall, Upper Saddle River, NJ, 1998.

[30] Makarov DV, Trock BJ, et al., Updated nomogram to predict pathologic stage of prostate cancer given prostate-specific antigen level, clinical stage, and biopsy Gleason score (Partin tables) based on cases from 2000 to 2005. *Urology*, Vol. 69, No.6, 2007, pp.1095-1101.