Neural Network Design for the Recurrence Prediction of Post-Operative Non-Metastatic Kidney Cancer Patients

Baran TANDER[1], Atilla ÖZMEN[2] and Ender ÖZDEN[3]

[1]Kadir Has Vocational School, Kadir Has University, Silivri-İstanbul, Turkey
tander@khas.edu.tr
[2]School of Engineering and Natural Sciences, Kadir Has University, Fatih-İstanbul, Turkey
aozmen@khas.edu.tr
[3]School of Medicine, Ondokuz Mayıs University, Atakum-Samsun, Turkey
eozden@omu.edu.tr

Abstract

In this paper, various post-operative recurrence estimation models called nomograms for the kidney cancer patients without any metastates are introduced and novel systems based on a Multilayer Perceptron Neural Network are designed to simplify and integrate the mentioned techniques which is believed to ease the physician' s post-operative follow up procedures. The parameters effecting the recurrence are the TNM stage, tumor size and nuclear (Fuhrman) grade, the existance of necrosis and vascular invasion. Independent systems for two of the individual prediction methods, as well as a system that combines these are designed and performance analyses are carried out to verify the reliability.

1. Introduction

Re-occurence of cancer which is called recurrence is possible, even a couple of years are passed after the treatment and must be taken into account both by patients and specialists. The probability of recurrence is predicted with special charts, namely nomograms, widely used in prostate and kidney malignities [1,2].

In literature, there are various preoperative [3] and post-operative [2,4,5] nomograms to estimate the tumor recurrence rates which is broadly related with the life expectancy for the non-metastatic kidney cancer patients. All of these are designed for intervals of 1-, 2- 5-, 10- and 12-years after the treatment.

In this study, an alternative method to nomograms based on a Multilayer Perceptron (MLP) Neural Network is designed, to predict the probability of freedom of recurrence for kidney cancers among 5-years after the nephrectomy operations, since a period of 5-years is crucial in most of the cancer follow ups.

The paper is organized as follows: Fist of all, the parameters of kidney tumors that will be employed as the inputs of MLP are introduced. Secondly, two of the popular nomograms and their neural network equivalents for the recurrence estimation are presented. Afterwards, a novel system combining these MLPs is designed. Simulations are carried out for various data to test the performance of the proposed structures.

2. Kidney Tumors

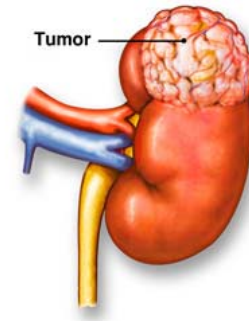Figure 1 shows a tumor in a kidney [6] with the parameters listed below:

Fig. 1. A kidney tumor.

Tumor Size is the diameter of the tumor at its largest dimension which is one of the most important parameters, since it is included as an input at all of the pre- or post-operative nomograms. It is also related with the pathological staging.

Symptoms can be interpreted as a measure of the aggressiveness of the tumor, which have significant effects on prognosis. Tumors without any signs or exhibiting local symptoms are far better than the systematic ones, which make the physicians think about metastases.

Pathological Stage (TNM Stage) is indeed the classification of the tumor size and location. Today, the commonly used staging system is ordered as follows:

T indicates the size and location of the main (primary) tumor and substaged below:

T1: Tumor size < 7cm and limited to the kidney,
T1a: Tumor size <= 4cm,
T1b: 4cm < Tumor size <= 7cm,
T2a: 7cm<Tumor size<=10cm,
T2b: Tumor size>10cm all limited to the kidney.
T3a/b/c: Tumor extends into major veins or perinephric tissues not beyond perirenal (Gerota's) fascia.
T4: Tumor invades beyond Gerota's fascia (Including contiguous extension into the ipsilateral adrenal gland).

N describes the extent of spread to nearby (regional) lymph nodes. Lymph nodes are small bean-sized collections of immune system cells to which cancers often spread first.

M indicates whether the cancer has spread (metastasized) to other parts of the body. All of these stages are shown in figure 2 [7].
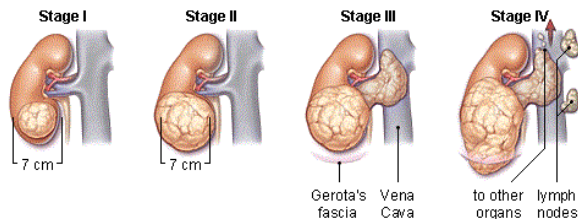

Fig. 2. Pathologic staging of kidney tumors.

Since the nomograms are object to non-metastatic cases only T stage is included, therefore the substagings of N and M are not explained in this paper.

Nuclear Grade (Fuhrman Grade): is a measure of the number and size of the cancerous cells on the tissue, graded between 1 - 4. Grade 1 has the best outcome, in other words, the greater the grade the worse the prognosis.

Histology: is the type of the tumor cell. Nearly 80% of all malignant kidney tumors is the Clear Cell Carsinoma, which has a worse prognosis when compared with the other types, Chromofobe and Papillary carcinoma.

Necrosis/Vascular Invasion: Necrosis is the presence of dead tissues on the tumor on the other hand vascular invasion is the involvement of blood vessels by the tumor cells. Both of their existences have a negative effect on the prognosis.

3. Nomograms

Nomograms are prediction tables that the urologists widely use to predict the freedom of recurrence for kidney and prostate cancers. They are developed by the Cox regression method [8]. Although being popular and pretty reliable, they have the following drawbacks:

1. They are not user friendly,
2. There are many individual nomograms in literature employing different parameters as the inputs, and different periods of freedom of recurrence; therefore a generalized system that will unify these models will be very useful.

In our study, two of the popular nomograms namely, Kattan's and Sorbellini's are considered; MLP Neural Network structures are trained and tested for both, as well as a single MLP is designed to integrate these individual systems.

3.1. Kattan's Nomogram

Kattan introduced the first nomogram for post-operative recurrence-free survival of the kidney cancer patients. The inputs are symptoms, histology, tumor size and TNM stage. A point is assigned for each of these parameters where the physician calculates the total sum to estimate the 60 months of recurrence-free survival as seen in figure 3.
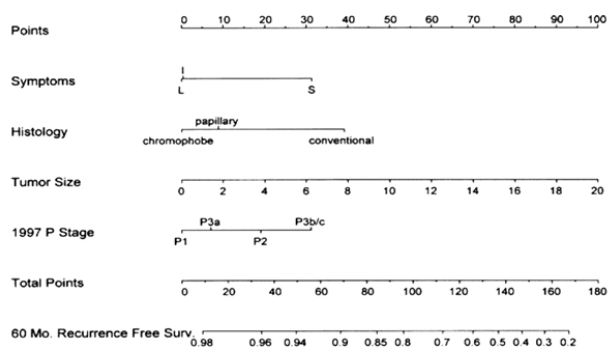

Fig. 3. Kattan's nomogram.

3.2. Sorbellini's Nomogram

Another nomogram presented by Sorbellini is given in figure 4, only for the clear cell carcinoma patients, where the input parameters are different from the Kattan's specifically, Fuhrman grade and existence of necrosis and vascular invasion.
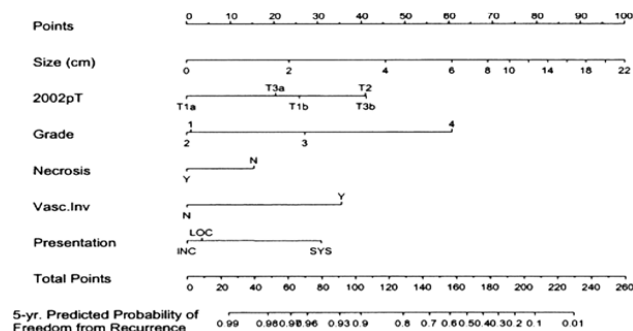

Fig. 4. Sorbellini's nomogram.

There are many other nomograms designed throughout the years such as Karakiewicz' s [5] moreover, some other prediction tools like Leibovich tables [9] however, our goal here is to design the MLP equivalents for the above and to design an integrated system that combines both.

4. MLPs

Three MLPs trained by the Back Propogation Algorithm are designed to succeed these nomograms; two for each individual model and one for the combination of these. Outputs are derived by the random inputs utilizing the nomograms therefore many input/output pairs are obtained which are essential for the training process.

4.1. MLP for the Kattan's Nomogram

In this MLP, as shown in fig. 5, the inputs are the tumor size, T stage, histology and symptoms where the output is the 5-year recurrence-free survival probability. The proposed system has one input, one output and one hidden layer including 4, 30 and 1 neuron, respectively. Tangent sigmoid is employed as the activation function at the hidden layer neurons. To form the input/output pairs for the training set, random inputs are

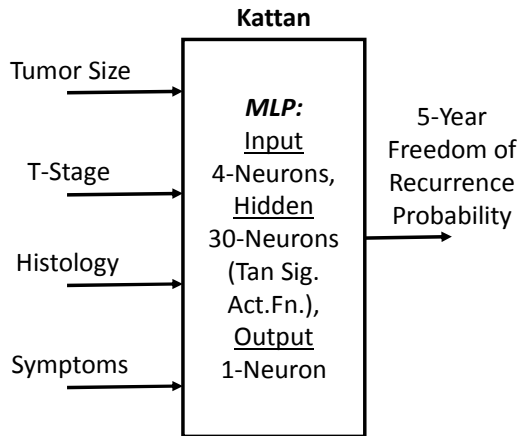introduced to the Kattan's nomogram to generate the corresponding outputs.

Kattan

Fig. 5. MLP for the Kattan's nomogram.

4.2. MLP for the Sorbellini's Nomogram

The input parameters are slightly different from the first one namely, tumor size, T substage, Fuhrman grade and existences of necrosis and vascular invasion. Therefore, 6 input neurons are needed for this MLP. The number of the hidden layer neurons is 45 with tangent sigmoid activation functions, and the output is again generated by a single neuron to give the 5-year recurrence-free survival rate.

4.3. A Generalized MLP

The mentioned nomograms generate close but different outputs therefore a generalized system based on a novel MLP that combines both will be very useful. However, some problems occur in the integration process, since the inputs of Kattan's and Sorbellini's nomograms are not identical. Specifically, Kattan employs a single T stage on the other hand Sorbellini uses T1a and T1b substages. Furthermore, Fuhrman grade, necrosis and vascular invasion parameters don't exist in the Kattan's, similarly histology parameter is excluded in the Sorbellini's. These drawbacks are eliminated by the following procedure:

In the training process by using the nomograms, the input/output pairs are generated as follows: The outputs are generated in three different classes either as the Kattan's, Sorbellini's or their mean. If the T stage input is introduced as T1 or T3c, which only exist in the Kattan's, this means that the output must be Kattan's and inversely, Sorbellini's output is taken into account if it is applied as T1a or T1b. Otherwise, since all other T stages are common for both cases, the output will be their mean. For the rest of the inputs, the mean is computed if any of the parameters is involved in both, or else only for the Kattan's and Sorbellini's outputs are considered.

Once the training is completed, the integrated system shown in fig. 6 can be used having 7 input, 75 hidden and 1 output layer neurons. Again, tangent sigmoid activation function is employed in the hidden layer neurons. The resulting MLP is given in figure 6.
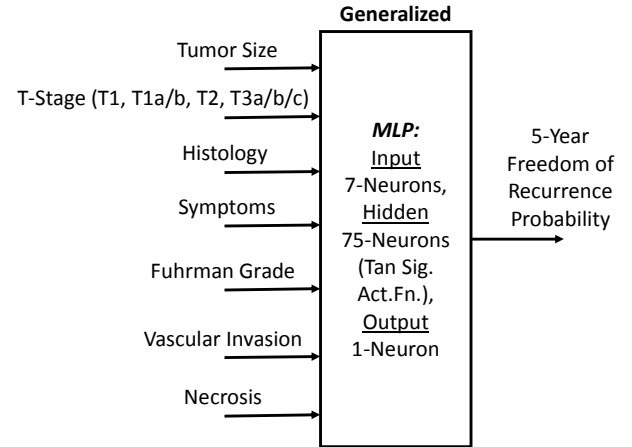
Generalized

Fig. 6. Generalized MLP combining Kattan's and Sorbellini's.

5. Performance Analysis

In the learning process of MLPs, 50000 random samples are run for 500 back propogation algorithm iterations. Once the training is done, another set of 50000 random test inputs is applied to the designed systems as well as to the conventional nomograms to compare the freedom of recurrence probabilities.

Root Mean Square Deviations (RMSD) [10] shown in (1), are employed to measure the performance by computing the total difference between the each MLP and nomogram outputs for the given amount of samples.

$$RMSD = \sqrt{\frac{\sum_{i=1}^{n} \left[NG(i) - MLP(i) \right]^2}{n}} \qquad (1)$$

Here,
n, which is 50000 is the number of samples,
$NG(i)$, is the nomogram output for the ith sample,
$MLP(i)$, is the MLP output for the ith sample.

One can see that, the smaller the RMSD the closer the outputs of each technique in other words, the better the performance. The RMSDs between MLP and nomogram outputs for Kattan and Sorbellini are presented at table 1.

Table 1. Performance analysis of Kattan's and Sorbellini's nomograms and their MLPs for 50000 samples.

Comparison	RMSD for training data	RMSD for test data
Kattan's nomogram & MLP for Kattan	4.09e-4	0.0014
Sorbellini's nomogram & MLP for Sorbellini	0.0063	0.0065

Performance analysis of the generalized MLP is carried out by computing the RMSDs for samples containing only Kattan's or Sorbellini's input parameters which are nearly 15000 for each, thus enabling individual comparisons of the proposed structure with Kattan's and Sorbellini's nomograms are carried out as shown at table 2.

Table 2. Performance analysis of generalized MLP with Kattan's and Sorbellini's nomograms individually for nearly 15000 samples for each.

Comparison	RMSD for training data	RMSD for test data
Kattan's nomogram & Generalized MLP with data for Kattan	0.0047	0.0092
Sorbellini's nomogram & Generalized MLP with data for Sorbellini	0.0057	0.0108

6. Conclusions

In this paper, a novel approach to nomograms, which are widely used by the physicians to predict the freedom of recurrence in 5-years for the post-operative non-metastatic kidney cancer patients, is proposed. The system depends on a very popular neural network: MLP.

Two MLPs are trained and tested with the Kattan's and Sorbellini's nomogram data, furthermore a unified system which combines both is presented. Satisfactory results are found in the simulations. The system is user friendly and is believed to be employed by the physicians for the follow up processes.

As a future work, other neural network architectures can be utilized to improve the predictions and software for mobile applications can be developed.

7. References

[1] Chun F. K. H. et.al., "A Critical Appraisal of Logistic Regression-Based Nomograms, Artificial Neural Networks, Classification and Regression-Tree Models, Look-Up Tables and Risk-Group Stratification Models for Prostate Cancer", *BJU International*, vol. 99, no. 4, pp. 794-800, May, 2007,

[2] Kattan M. et.al., "A Postoperative Prognostic Nomogram for Renal Cell Carsinoma", *The Journal of Urology*, vol. 166, no. 1, pp. 63-67, July, 2001,

[3] Raj G. V. et.al., "Preoperative Nomogram Predicting 12-Year Probability of Metastatic Renal Cancer", *The Journal of Urology*, vol. 179, no. 6, pp. 2146-2151, June, 2008,

[4] Sorbellini M. et.al., "A Postoperative Prognostic Nomogram Predicting Recurrence for Patients with Conventional Clear Cell Renal Cell Carsinoma", *The Journal of Urology*, vol. 173, no. 1, pp. 48-51, Jan., 2008,

[5] Karakiewicz P. I. et.al., "Multi-Institutional Validation of a New Renal Cancer-Specific Survival Nomogram", *Journal of Clinical Oncology*, vol. 25, no. 11, pp. 1316-1322, Apr., 2007,

[6] http://www.laparoboticsurgery.com/resource-pages-for-links/kidney-cancer-staging-and-grading/

[7] http://www.urologist.com.sg/urology-problem/kidney-cancer.html

[8] Cox D. R., "Regression Models and Life-Tables", *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 34, no. 2, pp. 187-220, 1972,

[9] Leibovich B. C., "Prediction of Progression after Radical Nephrectomy for Patients with Clear Cell Renal Cell Carsinoma", *Cancer*, vol. 97, no. 7, pp. 1663-1671, Apr., 2003,

[10] Spiegel M. R., Stephens L. J., "Schaums Outline of Statistics Fourth Edition", McGraw-Hill, USA, 2011.