Time-to-Recur Measurements in Breast Cancer Microscopic Disease Instances

Ioannis Anagnostopoulos[1], Ilias Maglogiannis[1], Christos Anagnostopoulos[2], Konstantinos Makris[3], Eleftherios Kayafas[3] and Vassili Loumos[3]

[1]*Department of Information and Communication Systems Engineering, University of the Aegean, Karlovassi 83200, Samos - Greece, janag@aegean.gr*
[2]*Department of Cultural Technology and Communication, University of the Aegean, Mytilene 81100, Lesvos – Greece, canag@ct.aegean.gr*
[3]*School of Electrical and Computer Engineering, National Technical University of Athens, Heroon Polytechneiou 9, 15773, Athens, Greece, phone/Fax +302107722538, kayafas@cs.ntua.gr*

***Abstract*-**In this paper we present an overview of survival curves measurements using the Kaplan-Meier approximation method. The survival curves were constructed upon the decision of an artificial neural network, which estimates the time-to-recur (TTR) in patients who presented malignant breast tumours with no evidence of distant metastases at the time of diagnosis. For our work we used the Wisconsin Prognostic Breast Cancer (WPBC) data set. The measurements taken over actual data evaluated that the accuracy of our proposed decision system is in high levels and can be used for the breast cancer prognosis problem.

I. Introduction

Breast cancer is the type of cancer with highest incidence rates in women. It is also the most common cause of cancer death in women in many countries, only exceeded by lung cancer in Asian countries and recently in the United States [1].
The economic and social values of breast cancer diagnosis and prognosis are very high. As a result, the problem has attracted many researchers in the area of computational intelligence recently. Due to the importance of achieving highly accurate predictions, Artificial Neural Networks (ANNs) are among the most common methods solving the afore-mentioned two problems, which require complicated nonlinear interactions between the input data and the information to be predicted [2], [3] and [4].
Prediction tasks are among the most interesting activities in which to implement intelligent system. Specifically, prediction is an attempt to accurately forecast the evolution or outcome of a specific situation, using as input information obtained from a concrete set of variables that potentially describe that situation. This paper analyzes the prognosis process existing when patients with primary breast cancer should receive a certain therapy to remove the primary tumour. Once a patient is diagnosed with breast cancer, the malignant lump must be excised. During this procedure, or during a different post-operative procedure, doctors must determine the prognosis of the disease. This is simply a prediction of the expected course of the disease. Prognosis is important because the type and intensity of the medications are based on it.
On the other hand, the prognosis problem is almost identical with a particular class of problems called "analysis of survival or lifetime data" [5], [6]. It is considered as a more difficult problem than that of diagnosis since the data are right-censored in terms of the observation time. That is, there are fewer cases where a recurrence is observed and these patient are classified as recur knowing in parallel the time to recur (TTR). However, in most cases recurrence is not observed and thus, there is no real time point at which we can consider the patient a non recurrent case. For such patients the only time known is the time of their last check-up or the time at which they decided to change doctor/hospital or the time where they passed away from cancer unrelated causes. This time is called disease-free survival time (DFS). In the problem addressed the end of the DFS can be considered as the beginning of the TTR with high probability value.

II. Prognostic dataset

In this work the publicly available by anonymous ftp Wisconsin Prognostic Breast Cancer (WPBC) data set, was used [7]. The specific dataset along with the Wisconsin Diagnostic Breast Cancer (WDBC) data set, are the results of the efforts made at the University of Wisconsin Hospital for the prognosis of breast tumours solely based on the fine needle aspirate (FNA) test. This test involves fluid extraction from a breast lump using a small-gauge needle and then visual inspection of the fluid under a microscope. Figure 1 depicts two images, which were taken from fine needle biopsies of breast tumours [8].

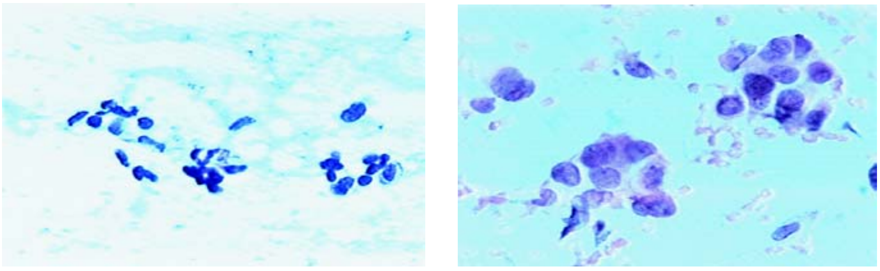

Figure 1. Images taken using the FNA test: (a) Benign, (b) Malignant

The proposed system is a generalised regression neural network, which estimate the recurrence time (TTR, time-to-recur) as well as a period where the patient exceeds her disease-free survival (DFS) time. The prognosis of the specific time interval is considered a difficult problem since the training data are right censored [9], [10].

The Wisconsin Prognostic Breast Cancer (WPBC) dataset consists of 198 instances (151 non-recur - 47 recur), where each one represents follow-up data for one breast cancer case. These were consecutive in-patients at the University of Wisconsin Hospital, the period from 1984 to 1995 and include only those cases exhibiting invasive breast cancer and no evidence of distant metastases at the time of diagnosis. Each case (instance) has 35 attributes, where the first three attributes correspond to a unique identification number and to the prognosis status (recur / non-recur) following by the recurrence time or the DFS time respectively. Then they follow 30 features, which are computations for ten real-valued features along with they mean, standard error and the mean of the three largest values for each cell nucleus respectively. These ten real values (attributes 4-13 at Table 1), are computed from a digitized image of a fine needle aspirate (FNA) of breast tumour, describing characteristics of the cell nuclei present in the image. Finally, the last two attributes are the diameter of the excised tumour (tumour size in centimetres) and the number of positive axillary lymph nodes observed at time of surgery (Lymph node status). All featured values are recorded with four significant digits.

Table 1 depicts the 15 WPBC attributes for the 194 instances, which were used for the purposes of this work. Four instances were not included in the training/testing set since the Lymph node values were missing.

WPBC attributes used
1. ID Number
2. Prognosis (R - recur / N - non-recur)
3. Time (in months)
4. radius [mean of distances from centre to points on the perimeter],
5. texture [standard deviation of grey-scale values],
6. perimeter,
7. area,
8. smoothness [local variation in radius lengths],
9. compactness [((perimeter)^2 / area) - 1],
10. concavity [severity of concave portions of the contour],
11. concave points [number of concave portions of the contour],
12. symmetry,
13. fractal dimension ["coastline approximation" – 1]
14. Tumour size (in centimetres)
15. Lymph node status

Table 1. WPBC data set attributes used

III. System's architecture

For the prognosis problem the neural network belongs to the generalised regression type (Generalised Regression Neural Network - GRNN). These neural networks have the special ability to deal with sparse and non-stationary data where non-linear relationships exist among inputs and outputs. The role of the neural network is to calculate a time interval that corresponds to a possible right end-point of the patient's DFS time.

The topology of the neural network was 14-193-z-z. The input layer consists of 14 nodes, where each node correspond to the prognosis status, the TTR or the DFS time, the ten cell nuclei characteristics attributes of Table 1, the diameter of the excised tumour and the number of positive axillary lymph nodes observed at time of surgery. Due to the small amount of WPBC instances and in order to avoid the "curse of dimensionality" problem, the standard error and the "worst" values from the ten real-valued features were removed during the training phase. Four instances were not included in the training/testing set since their lymph node values were missing. Thus, the second layer consists of 193 nodes, which correspond to the total amount of the patterns for the training epoch, according to the leave-one out method. Finally, the summation/division layer consists of *z* nodes that feed a same amount of processing elements in the output layer, representing the classified time intervals that correspond to a possible right-end of the DFS time.

The training set consists of the WPBC instances, divided in four classes namely C_1, C_2, C_3 and C_4 and the topology of the neural network was 14-193-4-4. This categorisation was made according to the value of the third attribute, which indicates the TTR or DFS time. Thus, C_1 corresponds to the instances where DFS/TTR≤1 year while C_2, C_3 and C_4 correspond to time intervals were 1<DFS/TTR≤3 years, 3<DFS/TTR≤6 years and DFS/TTR>6 years respectively. Table 2 depicts the amount of the WPBC dataset instances in respect to the above-mentioned categorisation. The first column indicates the time interval class, while the second and the third columns present the amount of instances, when the tumour recurred (N_R) and the amount of instances when the tumour did not recur (N_N).

Amount of WPBC instances			
Class	**Interval time**	**N_R**	**N_N**
C_1	Less than 1 year	20	23
C_2	1 year – 3 years	14	34
C_3	3 years – 6 years	7	48
C_4	More than 6 years	5	43

Table 2. WPBC instances and prognosis status

The total WPBC instances were presented to the network in a round-robin manner leave-one out method, while the training ended before the average testing error on the left-out cases began to increase. The total prediction accuracy (TP) is given according to $N^{-1}\sum_{k=1}^{4} p_k$ and the prediction accuracy *(P)* according to p_k / n_k for each location calculated for assessment of the prediction system respectively, where N is the total number of sequences (N_R+N_N), k is the respective class, n_k is the number of instances in class k and p_k is the number of correctly predicted instances in class k.

The accuracy of prediction by leave-one out tests for the WPBC instances and the categorized time intervals are depicted in Table 3. The diagonal cells correspond to the correctly classified WPBC instances for each class respectively, while the other cells present the misclassified instances. Every row expresses the system's ability in terms of correct classification over a tested time interval. The precision rates for the four time intervals were 90.69%, 91.67%, 92.72% and 93.75%, while the respective recall rates were 92.86%, 91.67%, 94.44% and 90%. The overall performance measured at *92.3%*.

confusion matrix		**Predicted**				
		C_1	**C_2**	**C_3**	**C_4**	**Precision (%)**
Actual	**C_1**	39	2	1	1	90.69
	C_2	1	44	1	2	91.67
	C_3	1	1	51	2	92.72
	C_4	1	1	1	45	93.75
Recall (%)		92.86	91.67	94.44	90.00	
Overall Performance = 92.27 %						

Table 3. Testing confusion matrix

IV. Results, measurements and discussion

The recurrence predictions made by the neural network were further examined using survival analysis. The cases were divided in to four aforementioned time intervals according to the predicted DFS time. The actual recurrence probabilities of these four time intervals were then assembled, using the Kaplan-Meier approximation.

Figure 2a presents the survival analysis based on the predicted DFS produced by the proposed neural network, while Figure 2b depicts the survival analysis made from the actual WPBC dataset instances. The y-axis corresponds to the probability of DFS time and the x-axis corresponds to the time in months.

As shown in these figures, the calculated survival probabilities based on the decision of the neural network compose a prediction dataset, which verifies its accuracy in terms of the precision measurements. Particularly, in Figure 2a the DFS probability values for the 24th, 48th, 72nd and 96th month (points p_1, p_2, p_3, and p_4 respectively) were 0.7, 0.5, 0.25 and 0.13. The actual DFS probability values for the same period in months (points a_1, a_2, a_3, and a_4) were 0.77, 0.59, 0.33 and 0.16. However, it was observed that the predicted DFS probabilities are almost identical with the actual lower recurrence probability values as these were measured using the linear Greenwood confidence limits method, with cumulative survival confidence level equal to 95%. The upper and lower recurrence probabilities (predicted and actual) are also depicted in Figures 2a and 2b.

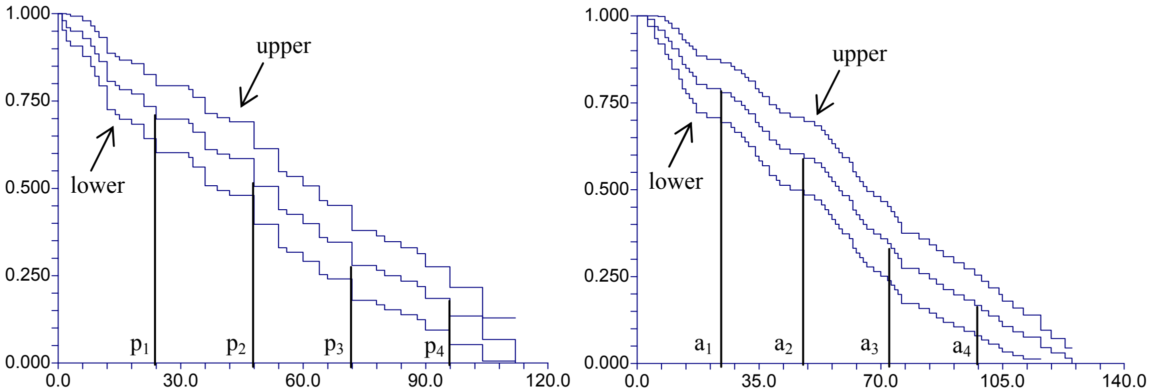

Figure 2. Survival analysis according to: a) the predicted DFS, (b) the actual WPBC dataset instances [y-axis: DFS probability, x-axis time in months]

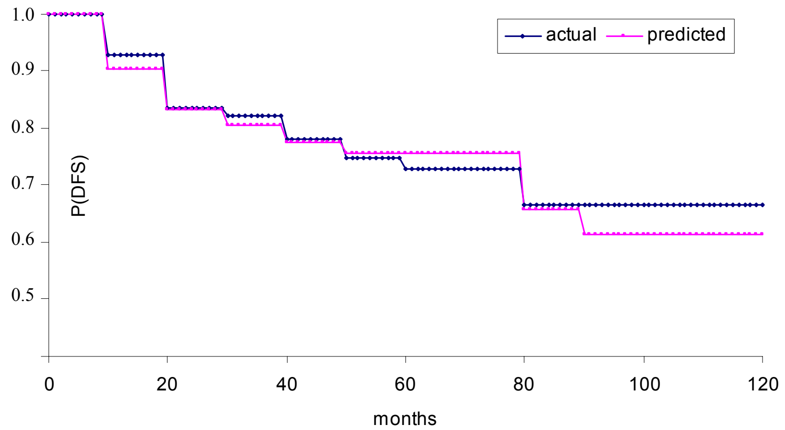

Figure 3. Kaplan-Meier disease-free survival time probabilities based on the predicted TTR

On the other hand, if the predicted right-end points of the Prognosis NN are considered as possible TTR points then the Kaplan-Meier disease-free survival time probabilities based on the predicted TTR can be measured. Thus, Figure 3 presents the survival analysis of the predicted TTR curve produced by the derived results compared to the TTR curve of the actual WPBC dataset instances. The y-axis corresponds to the probability of DFS time and the x-axis corresponds to the time in months.

The two curves are almost identical for time intervals between 0-11 months, 20-28 months, 40-58 months and 80-88 months. The rest predictions are very similar except from the predictions made for the period of more than 90 months. Thus, despite the fact that the neural network had better performance for the classes that correspond to larger time intervals, the survival analysis of the predicted results presented no significant statistical differences especially for the lower time intervals. This is caused by the “unfair” division of the learning set over the categorized intervals and due to the fact that the range of the predicted interval elongates for larger DFS times (classes C_1, C_2, C_3 and C_4 correspond to 12, 24, 36 and nearly 50 months respectively).

V. Conclusions

In this paper the authors describe a biomedical decision support system, which tackles with the breast cancer prognosis problem. The data set used was the Wisconsin Prognostic Breast Cancer that contains processed data obtained from a specific breast cancer diagnosis test. The prognosis is based over an estimation of the right end-point of the patient’s disease free time, after which the tumour may recur. The measurements of the predicted survival probability values over the actual probability values proved the accuracy of the proposed system.

References

[1] http://seer.cancer.gov/cgi-bin/csr/1975_2001/search.pl#results, *Estimated New Cancer Cases and Deaths for 2004.*
[2] Choong P.L, deSilva C.J.S et al., “Entropy maximization networks, An application to breast cancer prognosis”, *IEEE Transactions on Neural Networks, 1996, 7(3):568-577.*
[3] Wang, T.C., Karayiannis N.B., “Detection of microcalcifications in digital mammograms using wavelets”, *IEEE Transactions on Medical Imaging, Vol. 17, Issue. 4, Aug. 1998, pp. 498 – 509.*
[4] Pendharkar P.C., Rodger J.A., Yaverbaum G.J., Herman N. and Benner M., “Association, statistical, mathematical and neural approaches for mining breast cancer patterns”, *Expert Systems with Applications, 17:223–232, 1999.*
[5] Street W. N., “A neural network model for prognostic prediction”, *Proceedings of the 15th International Conference on Machine Learning, Madison, Wisconsin, Morgan Kaufmann, 1998.*
[6] Mangasarian et al, “Breast cancer diagnosis and prognosis via linear programming”, *Operations Research, 43(4), pp. 570-577, July-August 1995.*
[7] http://ftp.ics.uci.edu/pub/machine-learning-databases/breast-cancer-wisconsin/, Wisconsin Breast Cancer (WPBC) Dataset.
[8] Wolberg W.H., Street W.N., Heisey D.M., and Mangasarian O.L., Computer-derived nuclear features distinguish malignant from benign breast cytology, Human Pathology, 26:792--796, 1995.
[9] Mangasarian et al, “Breast cancer diagnosis and prognosis via linear programming”, Operations Research, 43(4), pp. 570-577, July-August 1995.
[10] Street W. N., “A neural network model for prognostic prediction”, Proceedings of the Fifteenth International Conference on Machine Learning, Madison, Wisconsin, Morgan Kaufmann, 1998.