Individual and Collective Prognostic Prediction

W. Nick Street
Departments of Computer Sciences and Surgery
University of Wisconsin
1210 West Dayton Street, Madison, WI 53706
street@cs.wisc.edu

O. L. Mangasarian
Department of Computer Sciences
University of Wisconsin
1210 West Dayton Street, Madison, WI 53706
olvi@cs.wisc.edu

W. H. Wolberg
Department of Surgery
University of Wisconsin
600 Highland Avenue, Madison, WI 53792
wolberg@eagle.surgery.wisc.edu

January 4, 1996

Abstract

The prediction of survival time or recurrence time is an important learning problem in medical domains. The Recurrence Surface Approximation (RSA) method is a natural, effective method for predicting recurrence times using censored input data. This paper introduces the Survival Curve RSA (SC-RSA), an extension to the RSA approach which produces accurate predicted rates of recurrence, while maintaining accuracy on individual predicted recurrence times. The method is applied to the problem of breast cancer recurrence using two different datasets.

1 Introduction

A common prediction problem in many different fields is the analysis of *survival* or *lifetime data* [13, 15], in which the objective can be broadly defined as predicting the time of a particular event. Specifically, in medicine, we are concerned with prognosis, that is, predicting the course of a disease based on known patient characteristics. The "event time" to be estimated could be either the death of the patient or the recurrence time of the disease. Accurate prognosis allows the physician and the patient to make more informed treatment decisions, especially in cases involving difficult quality-of-life versus probability of survival decisions. This paper describes a prognosis prediction method based on nonlinear programming, and applies the method to breast cancer prognosis, to predict how long after surgery one can expect the disease to recur.

In applying inductive machine learning to prognosis, we first note that this problem does not fit into any of the classic learning paradigms of classification, function approximation, or time series prediction. While a patient can be classified "recur" if the disease is observed, there is no real cutoff point at which the patient can be considered a non-recurrent case. The data are therefore *censored* in that we know a time to recur (TTR) for only a subset of patients; for the others, we know only the time of their last check-up, or disease-free survival time (DFS). In particular, recurrence or survival data is *right censored*, i.e., the right endpoint (recurrence time) is sometimes unknown, since some patients will inevitably move away, change doctors, or die of unrelated causes. Further, we are not considering here a time series, as the prediction is made at a particular point in time (e.g., diagnosis or surgery) based on characteristics known about the patient at that time.

A significant body of work concerning survival analysis exists in the statistics literature; see, for instance, the summary review by Henderson [8] and references therein. However, the application of machine learning methods to problems involving censored data has been rare. Schenone *et al.* [21] used a self-organizing neural network to find classes of cases with similar expected recurrence times. However, they did not directly address the problem of using censored data, that is, cases which have not been followed to recurrence / death. Burke [4] used artificial neural networks (ANNs) to approach prognosis as a separation problem, as was done in previous work at Wisconsin [26, 28]. This is done by choosing one or more endpoints and separating sets such as "patients who recurred in less than two years." The work of Ravdin and colleagues [7, 19, 20] used ANNs to generate survival curves, which plot the probability of survival or disease-free survival against time. This work uses the the trained network's output as an approximation of recurrence probability. While their cumulative results closely fit the population recurrence characteristics of the test cases, we believe that better individual predictions may be obtained by directly predicting survival time, while still maintaining accurate group recurrence characteristics.

2 The Recurrence Surface Approximation (RSA)

We approach the prediction of time to recur (TTR) as a function estimation problem, a mapping of an n-dimensional input of cytological and other features to a one-dimensional time output. Our original solution to this estimation problem is termed the recurrence surface approximation (RSA) technique [23, 14, 27]. RSA uses linear programming to determine a linear combination of the input features that accurately predicts TTR. Intuitively, we wish to fit the observed recurrences as closely as possible, and use the disease-free survival time (DFS) of the censored cases as a lower bound on the recurrence time of that patient. These assumptions can be formulated into the following linear program for a given training set:

$$\begin{array}{lrcl} \underset{w,\gamma,y,z}{\text{minimize}} & e^T y + e^T z & & \\ & -y \leq & Mw + \gamma e - t & \leq y \\ & & -Nw - \gamma e + r & \leq z \\ \text{subject to} & & Mw + \gamma e & \geq 0 \\ & & Nw + \gamma e & \geq 0 \\ & & y, z & \geq 0 \end{array} \tag{1}$$

The purpose of this linear program is to learn the weight vector w and the constant term γ. These parameters determine a recurrence surface $s = xw + \gamma$, where x is the n-dimensional vector of measured features and s is the surface (in this case, a plane defined on the feature space) that predicts recurrence times. Here M is an $m \times n$ matrix of the m recurrent points, with times to recur (TTRs) given by the m-dimensional vector t. Similarly, the k non-recurrent points are collected in the $k \times n$ matrix N, and their last known disease-free survival (DFS) times are in the k-dimensional vector r. The vectors y and z represent the errors for recurrent and non-recurrent points, respectively. Any difference between observed TTR and predicted time to recur $s = xw + \gamma$ is an error, while predicting a recurrence time smaller than an observed DFS is also an error. The objective merely minimizes the sum of these errors, using e, a vector of 1's of appropriate dimension. We also require that all time-of-recurrence predictions are non-negative. In the terminology of [23] this formulation was referred to as the "pooled error" RSA variation.

This prediction method has been applied, with considerable success, to breast cancer data collected at the University of Wisconsin Hospital. The current version of this dataset (the Wisconsin Prognostic Breast Cancer (WPBC) data [25], also available at the UCI Machine Learning Repository [16]) consists of 198 cases, each containing 32 input features measuring cytological and other prognostic factors [24, 28]. The RSA technique was shown to outperform prediction techniques which rely solely on recurrent cases, and to reliably differentiate cases with a good prognosis from those with an intermediate or poor prognosis.

However, while RSA successfully "rank-ordered" the cases based on expected outcome, the predicted recurrence times did not closely follow the observed outcomes *as a group*. Figure 1 shows the disease-free survival curve obtained using the RSA recurrence estimates (using leave-one-out testing [12]), along with the curve plotting actual outcomes (as estimated using the Kaplan-Meier method [10]). Error bars representing 95% confidence intervals around the Kaplan-Meier curve are shown at each 12-month interval. The RSA method tends to cluster the predicted recurrences close together, and shows an overall bias toward pessimistic (that is, early) predictions. In fact, RSA predicted all of the recurrence times to be less than 20 years; in reality, the expected recurrence rate at 20 years could be expected to be around 65% to 70% [9]. This follows the bias toward early recurrences which necessarily exists in all such datasets. Not all cases can be followed to an observed recurrence; patients move, change physicians, or die of unrelated causes. Therefore, the observed recurrences are unavoidably biased toward those which occur early in the study. Ideally, a predictive method would account for this bias and follow the group recurrence characteristics, while still maintaining accuracy on individual cases.

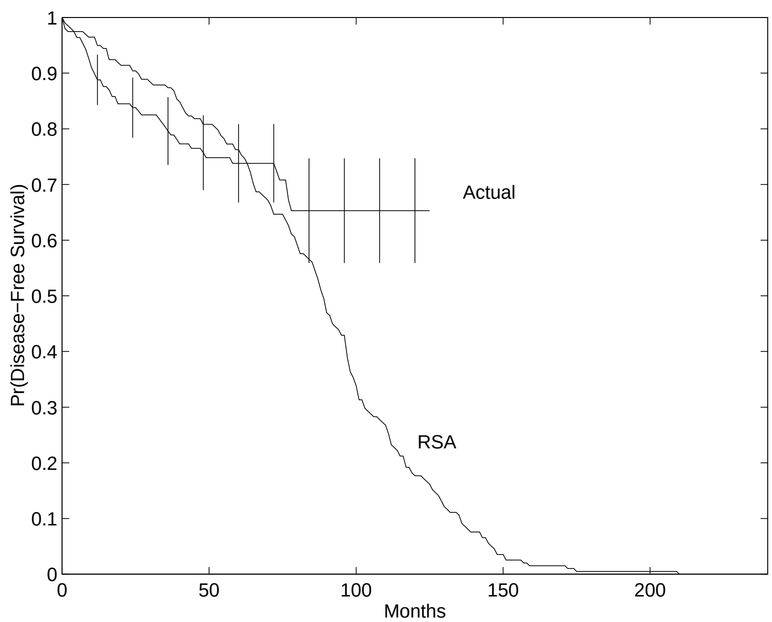

Figure 1: Disease-free survival curve predicted by RSA method on WPBC data compared to curve using actual outcomes. Error bars represent 95% confidence intervals.

3 Predicting Collective Recurrence Rates

In order to get better agreement between the actual and RSA-predicted survival curves, we modify our RSA surface as follows. We sample the actual survival curve at discrete intervals, construct a term which approximates the value of the RSA-predicted curve, and modify our RSA by minimizing the difference of the two curves. First, the time of the study is divided into equal time intervals and the values stored as the vector s; specifically, $s = [12, 24, 36, \ldots, 240]^T$, where each element represents months. At each time step s_i, we wish to know the percentage of cases which should be expected to recur at some time $\tau > s_i$. This is done by computing the corresponding Kaplan-Meier estimate s_i of the actual disease-free survival curve that our modified RSA will try to fit. These values are stored in the vector p, e.g., $p = [0.93, 0.86, 0.80, \ldots, 0.30]^T$. Thus, for a given training set, we expect the probability of disease-free survival at time s_i to be p_i.

Now we want our modified RSA to predict a survival curve in close agreement with the true one. At each time step s_i, we will count the number of cases with RSA-predicted recurrence time $\tau > s_i$, and compare that to the expected number. For a particular training point, consider the value

$$(x^T w + \gamma - s_i)_* \tag{2}$$

where $(\zeta)_*$ is the step function, defined as one if $\zeta > 0$ and zero otherwise. The value (2) is one if the predicted time of recurrence for the case with feature vector x is greater than a particular time step s_i. Now let $X := \begin{bmatrix} M \\ N \end{bmatrix}$, all of the training points collected into one matrix. The total number of cases with predicted TTR greater than s_i is

$$e^T(Xw + \gamma e - s_i e)_*, \tag{3}$$

where again e represents a vector of ones of appropriate dimension. When divided by the total number of points, $m + k$, this gives the desired height of the predicted survival curve at time s_i.

Enforcing closeness over the k different time steps results in the following nonlinear optimization problem that generates a modified RSA, which we term the Survival Curve RSA (SC-RSA):

$$\begin{array}{ll} \underset{w,\gamma,y,z}{\text{minimize}} & (1-\lambda)(e^T y + e^T z) + \dfrac{\lambda}{2}\displaystyle\sum_{i=1}^{k}(e^T(X^T w + \gamma e - s_i e)_* - (m+k)p_i)^2 \\ & \begin{array}{rcccl} -y & \leq & Mw + \gamma e - t & \leq & y \\ & & -Nw - \gamma e + r & \leq & z \\ \text{subject to} & & Mw + \gamma e & \geq & 0 \\ & & Nw + \gamma e & \geq & 0 \\ & & y, z & \geq & 0 \end{array} \end{array} \tag{4}$$

The parameter $\lambda \in [0, 1]$ controls the relative weight given to accurate individual predictions versus accurate recurrence percentages, allowing us to trade off the two objectives of the resulting model. In order to solve this nonlinear program with a discontinuous objective by using standard techniques of smooth optimization, the step function can be replaced by its smooth approximation, a sigmoid function. We used the error function, that is, the integral of the standard normal curve. [1]

[1] It may be necessary to multiply the error function by some positive parameter δ to control the steepness of the resulting sigmoid. However, in practice, our results were obtained with $\delta = 1$.

Considering the two terms of the objective function in (4) separately is closely related to the field of multi-objective mathematical programming [22]. This area of optimization considers problems of the form

$$\underset{x\in\mathcal{S}}{\text{vector min}} \quad \begin{bmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_k(x) \end{bmatrix}. \tag{5}$$

Here, multiple, possibly competing objectives are to be optimized on some feasible set $\mathcal{S}$. The possible solutions to such a problem are known as *efficient* or *Pareto optimal* [18] points. At a Pareto optimal point, none of the $f_i(x)$ can be further lowered without increasing the value of some other $f_j(x)$. Pareto optimal points can be obtained by solving

$$\begin{array}{ll} \underset{x\in\mathcal{S}}{\text{minimize}} & \sum_{i=1}^{k} \lambda_i f_i(x) \\ \text{subject to} & \lambda_i \geq 0, \sum_{i=1}^{k} \lambda_i = 1 \end{array} \tag{6}$$

The SC-RSA mathematical program can be considered as having define two competing objective functions, with the first one based on predicting individual prognosis while the second predicting collective recurrence rate. Solutions to (4) can therefore be viewed as Pareto optimal points, in this case, optimal for a given value of λ. Choosing a particular solution from the various Pareto optimal points (that is, choosing λ) assigns relative weights to the two objectives. This choice of λ must be done on a domain-dependent basis, as the relative weight given to the two parts of the objective could vary significantly depending on the application. In our case of medical prognosis, we are willing to sacrifice only a small amount of accuracy on the individual predictions in order to obtain satisfactory predicted recurrence rates.

4 Computational Results

This section describes computational experiments that generate an SC-RSA for each of two breast cancer prognosis databases. Computations were carried out using the GAMS modelling language [3] and the MINOS nonlinear program solver [17].

4.1 Wisconsin Prognostic Breast Cancer (WPBC) Data

The first set of experiments was performed on the Wisconsin Prognostic Breast Cancer (WPBC) data cited above. As depicted in Figure 1, the follow-up times for these patients are available for ten years only. In order to compensate for this relatively short time interval and produce more realistic recurrence characteristics, we extend the actual disease-free survival curve by assuming a 7% per year recurrence rate between ten and twenty years, based on medical literature [9].

Figure 2 shows the true Kaplan-Meier estimate of these cases along with the 7% per year recurrence extension. This estimate of the true recurrence rate is compared to the predicted recurrence rate of SC-RSA using three different values of the parameter λ: 0.0 (corresponding to the RSA plot from Figure 1), 0.1 and 0.9. The 0.9 case gives primary importance to the nonlinear curve-fitting term of the objective function, resulting in a very good collective recurrence rate. Assigning the small weight, 0.1, to the nonlinear term still results in a reasonable fit to the desired survival curve, but emphasizes the correctness of individual prognostic prediction.

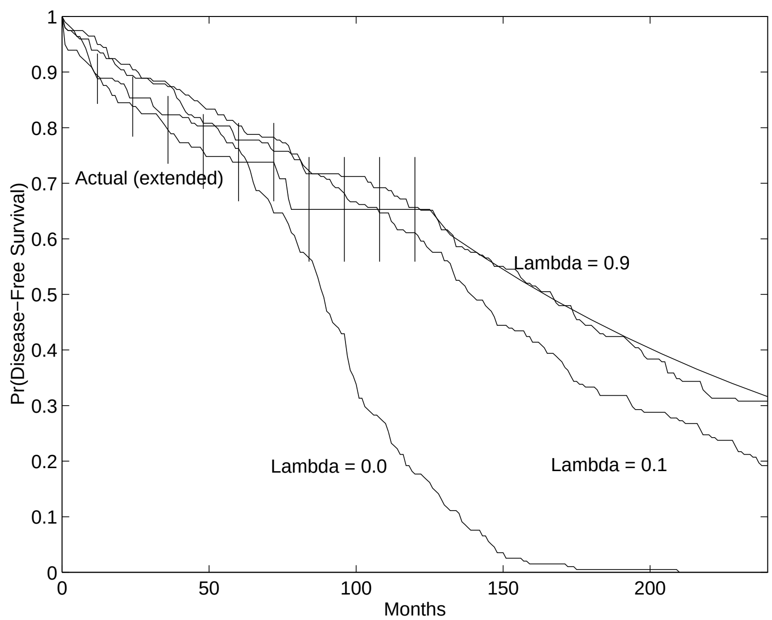

Figure 2: **Actual disease-free survival curve compared to those predicted by SC-RSA using various lambda values.**

Table 1 shows three different metrics comparing how well the three different prediction methods – original RSA ($\lambda = 0$) and SC-RSA with $\lambda = 0.1$ and $\lambda = 0.9$ – fit the true disease-free survival curve. For each of these metrics, the curves are sampled at one-year intervals up to ten years. The measurements are: cumulative difference of the curves at these ten points, worst percentage difference between the curves, and number of the ten time intervals at which the predicted curve lies within the 95% confidence interval of the true curve. Table 1 shows that no matter what error metric is used, a small value of the parameter λ decreases the error markedly over the $\lambda = 0$ case, while larger values of λ have only marginal effects.

The above results indicate that SC-RSA improves the original RSA model by generating recurrence predictions at approximately the appropriate rate. Table 2 shows that the predictive accuracy on individual cases has not been sacrificed in order to achieve this gain. This

λ	Error in Survival Curve		
	Cumulative Difference	Maximum % Difference	Number of Points in CI
0.0	1.53	72.9	3
0.1	0.53	10.8	7
0.9	0.32	9.8	10

Table 1: **Different error measurements (cumulative absolute error, maximum percentage error, and number of points in confidence interval) comparing the true Kaplan-Meier recurrence rate to the recurrence rate predicted by SC-RSA, using various settings of the parameter λ.**

table shows the generalization estimates of the three leave-one-out tests plotted in Figure 2, considering only the estimated prediction error, that is, the original objective function of Equation (1). Surprisingly, for small values of λ, the SC-RSA method appears to predict recurrences slightly *better* than the original RSA, although this difference is not statistically significant. For larger values of λ, the prediction accuracy deteriorates significantly, reflecting the diminished importance placed on individual prognoses. We note that significantly better generalization results can be obtained on this data using simple feature selection techniques such as that proposed in [23] and [2].

λ	Average Error (months)		
	All Points	Non-recur	Recur
0.0	17.4	8.0	47.6
0.1	16.4	3.4	57.9
0.9	34.9	7.1	121.3

Table 2: **Mean error (objective of Equation (1)) on Wisconsin prognostic data using SC-RSA (Equation (4)). Results obtained using leave-one-out testing.**

In addition to maintaining the quality of the individual predictions, SC-RSA is also replicates the success of RSA in rank-ordering the cases. Figure 3 shows the Kaplan-Meier curves of four different subsets, using the true outcomes: cases predicted by the SC-RSA ($\lambda = 0.1$) to recur before five years (39 cases), five to ten years (38), ten to fifteen years (55), and later than fifteen years (66) following surgery. Here we see that, not only are the cases ordered correctly by the SC-RSA prediction, but also that in the two categories for which actual follow-up data are available, many of the observed recurrences take place around the time they are predicted by SC-RSA.

4.2 SEER Data

We repeated the above tests on a subset of the breast cancer outcomes data from the Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute [5], which contains follow-up data for over 24,000 breast cancer patients. The SEER program

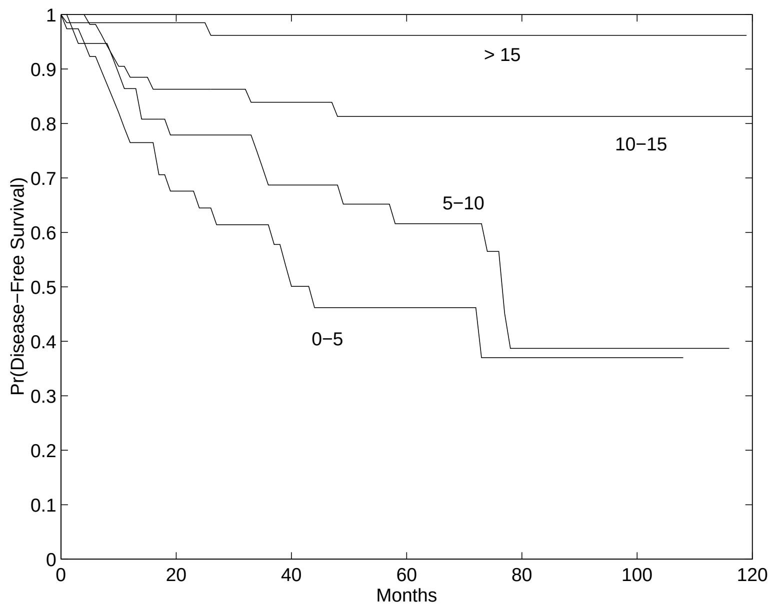

Figure 3: **True disease-free survival curves (using Kaplan-Meier estimates) for those cases from the WPBC dataset with various predicted recurrence times.**

uses survival time as the endpoint, rather than recurrence time. These cases contain the following five integer-valued input features: Histological Grade, Tumor Size, Tumor Extension, Number of Axillary Lymph Nodes Positive and Number of Axillary Nodes Examined. There is a high incidence of missing feature values in this data; for instance, histological grade is recorded for only about 17% of the cases. For the purposes of this experiment, we randomly selected a set of 1,195 cases that had no missing feature values.

Figure 4 shows the actual survival curve of this subset of the SEER dataset, again with 95% confidence intervals and again extended with a 7% per year death rate (rather than recurrence rate) between ten and twenty years. The predicted survival curves are shown using the same three values of λ: 0.0 (original RSA), 0.1 and 0.9. Although the SC-RSA method does not bring the curves into such close agreement on this data, Figure 4 and the error metrics shown in Table 3 indicate that the survival curves resulting from the SC-RSA predictions are markedly better than that of the original formulation. Further, Table 4 shows that the mean prediction error using SC-RSA generally degrades gradually as λ increases.

Finally, we stratified the SEER predictions based on the predicted time of death. The results in Figure 5 show the true Kaplan-Meier survival estimates for four groups: those with a predicted survival (using SC-RSA with $\lambda = 0.1$) between zero and five years (64 cases), 5-10 years (171), 10-15 years (396), and greater than 15 years (562). Again, the SC-RSA procedure successfully rank-ordered the cases.

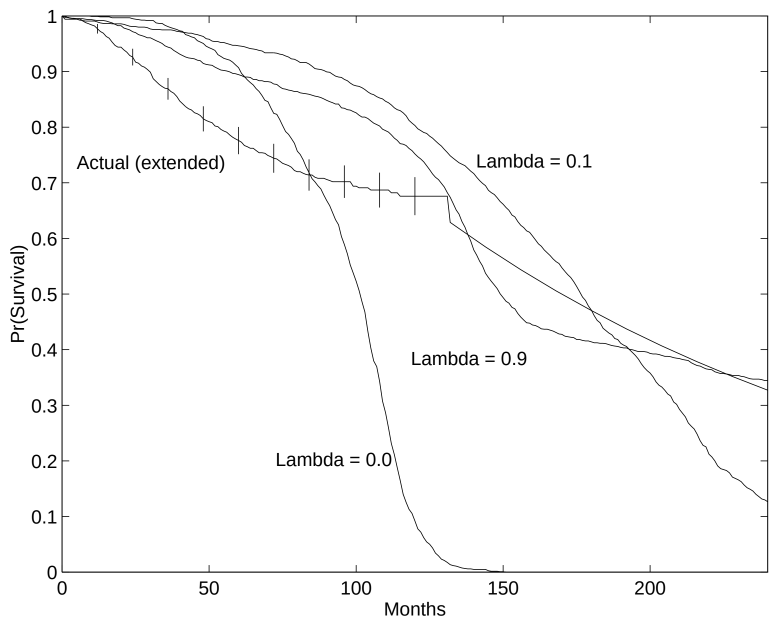

Figure 4: **Actual survival curve of selected SEER cases, augmented with 7% per year death rate between ten and twenty years, compared to the curves predicted by SC-RSA using various lambda values. Error bars represent 95% confidence intervals around the survival probability.**

λ	Error in Survival Curve		
	Cumulative Difference	Maximum % Difference	Number of Points in CI
0.0	1.60	86.4	1
0.1	1.36	28.0	0
0.9	0.95	20.3	0

Table 3: **Different error measurements comparing the true Kaplan-Meier recurrence rate to the recurrence rate predicted by SC-RSA, using various settings of the parameter λ, on the SEER dataset.**

λ	Average Error (months)		
	All Points	Non-recur	Recur
0.0	15.8	6.1	43.0
0.1	29.2	1.9	105.7
0.9	31.8	3.4	111.4

Table 4: **Mean error (objective of Equation (1)) on SEER data using SC-RSA (Equation (4)). Results obtained using leave-one-out testing.**

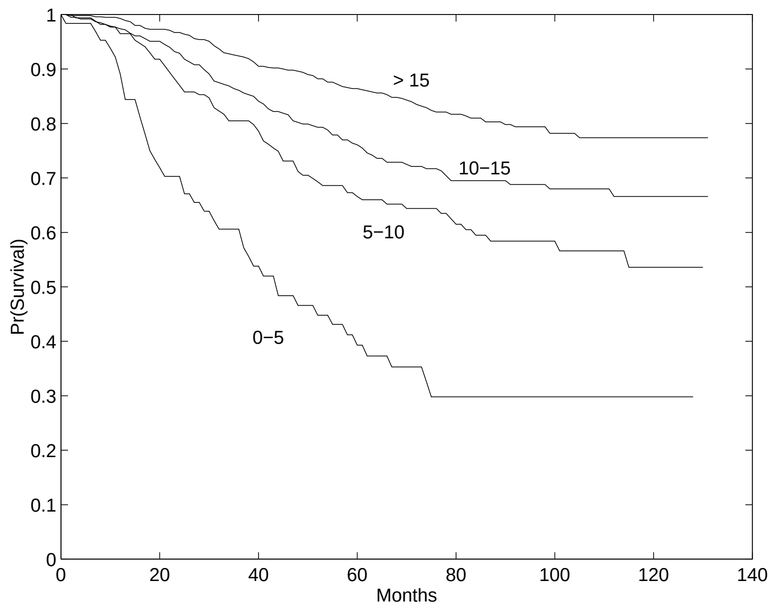

Figure 5: **True disease-free survival curves (using Kaplan-Meier estimates) for those cases from the SEER dataset with various predicted recurrence times.**

5 Clinical Application

Both medical and personal decisions hinge on the projected future course of the breast cancer. Decisions whether or not chemotherapy is needed and the intensity of such therapy are based on the anticipated course of the cancer. For example, patients with favorable outlooks may forego chemotherapy entirely, whereas those with less favorable outlooks will undergo varying intensity of chemotherapy even to bone marrow transplantation for those with the least favorable outlook. This chemotherapy is given immediately after surgery even in the absence of demonstrable tumor spread and is based solely on the expected future course of the cancer. However, there are three serious considerations: 1) chemotherapy's toxicity, 2) the morbidity associated with removing axillary (armpit) lymph nodes, and 3) the inaccuracy of prognostic estimations. The mental state of the patient, as well as personal and career plans are greatly affected by the anticipated course of the disease. Hence, improved prognostic prediction is an important goal for cancer treatment. Our current approach has the potential of eliminating the need for routine axillary lymph node removal and for increasing prognostic accuracy.

In current traditional medical practice, the strongest available prognostic features are tumor size and the extent to which cancer is present in the lymph nodes. The presence of cancer in the lymph nodes is determined by microscopic examination of lymph nodes after they have been surgically removed from the patient's armpit. However, prognostic determinations based on tumor size and lymph node involvement are inaccurate. For, 10%

of patients in the most favorable category will die of breast cancer and 40% of those in the most unfavorable category will survive.

The surgical removal of lymph nodes leaves the patient more susceptible to infection and the arm frequently develops lymphedema, a potentially severe swelling of the arm [1, 11]. Hence, one of the goals of our work is to generate accurate predictive models that are obtained by using only nuclear features, and are not improved by including lymph node status. Previous results with the RSA formulation indicate that precise nuclear features indeed do provide the most useful prognostic information [23]. This result is corroborated to a certain degree in this paper, as SC-RSA was able to match the true survival curve much better using the WPBC dataset with its nuclear features versus the SEER data which relies much more on lymph node information. If further studies confirm these findings, the routine and potentially hazardous and debilitating removal of lymph nodes from the armpit of breast cancer patients for prognostic purposes can be avoided.

The predictive model with the best estimated accuracy is now in clinical use at the University of Wisconsin Hospital. Using this model, we predict a time of recurrence for patients who have been diagnosed with a malignant tumor. In addition, the patient's probability of disease-free survival is estimated in the form of a Kaplan-Meier survival curve. The disease-free survival curve for the individual patient is based on those training cases that had a similar predicted time of recurrence.

6 Conclusions and Future Work

The Survival Curve RSA method presented here uses a mathematical programming approach to function approximation to produce a predictive model for recurrence data. The two parts of the objective function trade off apparent accuracy of individual predictions versus the overall rate of recurrence, in order to adjust for unavoidable biases in real datasets. Our computational results indicate that assigning a small weight to the survival-curve term of this dual objective results in considerably more accurate overall recurrence rates while maintaining an acceptable loss in estimated predictive accuracy.

The next issue we intend to address is that of estimating the predictive error of the SC-RSA surface. The RSA objective function makes no assumptions about the recurrence times of censored cases beyond the obvious one, that is, that the observed disease-free time is a lower bound on possible recurrence. However, it is reasonable to assume, as we have done in this work, that recurrences will occur at a rate consistent with the training data. This assumption can be incorporated into the original RSA objective function (the first term of the SC-RSA objective) by revising the error term associated with censored cases. This will result in a more accurate predictive surface as well as giving more reliable estimates of the expected error of the system in practice. Further, this will allow for easier comparisons of our work with that in the statistical literature, most of which is based on the Cox proportional hazards model [6]. Computational comparisons on both clinical and simulated datasets will be performed. We will also explore variations involving nonlinear predictive surfaces, which may have a more appropriate inductive bias for some learning situations; for example, some prognostic factors may exhibit a nonlinear relationship to expected recurrence, which is lost in the current model. Further, we will continue to pursue the goal of accurate breast cancer prognostic prediction without using lymph node status.

A potentially important research direction is the use of the RSA method to evaluate the relative effectiveness of different treatment strategies. By building separate prognostic prediction surfaces using patients who received different treatments, we can determine which regions of the feature space correspond to an improved prognosis for each particular treatment. The portion of the feature space where one treatment is clearly favored would represent a patient profile, a set of characteristics which would indicate use of that treatment. As with all of our medical applications, the goal is to provide doctors and patients with predictive models that increase the accuracy with which clinicians can plot the future course of diseases, resulting in treatment decisions in which both doctor and patient have the most reliable information.

Acknowledgements

The authors would like to thank Don Henson for use of the SEER database, and Dennis Heisey for his suggestions regarding methods for comparing survival curves. This research was supported by Air Force Office of Scientific Research Grant F–49620–94–1–0036 and National Science Foundation Grant CCR–9322479.

References

[1] R. J. Aitken, M. N. Gaze, A. Rodger, U. Chetty, and A. P. M. Forrest. Arm morbidity within a trial of mastectomy and either nodal sample with selective radiotherapy or axillary clearance. *British Journal of Surgery*, 76:568–571, 1989.

[2] P. S. Bradley, O. L. Mangasarian, and W. N. Street. Feature selection via mathematical programming. Technical report, Computer Sciences Department, University of Wisconsin, Madison, Wisconsin, 1995. To appear.

[3] A. Brooke, D. Kendrick, and A. Meeraus. *GAMS—A User's Guide*. The Scientific Press, 1988.

[4] H. B. Burke. Artificial neural networks for cancer research: Outcome prediction. *Seminars in Surgical Oncology*, 10:73–79, 1994.

[5] C. L. Carter, C. Allen, and D. E. Henson. Relation of tumor size, lymph node status, and survival in 24,740 breast cancer cases. *Cancer*, 63:181–187, 1989.

[6] D. R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society*, B 34:187–202, 1972.

[7] M. De Laurentiis and P. M. Ravdin. A technique for using neural network analysis to perform survival analysis of censored data. *Cancer Letters*, 77:127–138, 1994.

[8] R. Henderson. Problems and prediction in survival-data analysis. *Statistics in Medicine*, 14:161–184, 1995.

[9] H. Joensuu and S. Toikkonen. Cured of breast cancer? *Journal of Clinical Oncology*, 13(1):62–69, January 1995.

[10] E. L. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53:457–481, 1958.

[11] M. W. Kissin, G. Querci della Rovere, D. Easton, and G. Westbury. Risk of lymphoedema following the treatment of breast cancer. *British Journal of Surgery*, 73:580–584, 1986.

[12] P. Lachenbruch and P. Mickey. Estimation of error rates in discriminant analysis. *Technometrics*, 10:1–11, 1968.

[13] E. T. Lee. *Statistical Methods for Survival Data Analysis*. John Wiley and Sons, New York, 1992.

[14] O. L. Mangasarian, W. N. Street, and W. H. Wolberg. Breast cancer diagnosis and prognosis via linear programming. *Operations Research*, 43(4):570–577, July-August 1995. Available from ftp://ftp.cs.wisc.edu/math-prog/tech-reports/.

[15] R. G. Miller Jr. *Survival Analysis*. John Wiley and Sons, New York, 1981.

[16] P. M. Murphy and D. W. Aha. UCI repository of machine learning databases [http://www.ics.uci.edu/~mlearn/MLRepository.html], 1994. Irvine, CA: University of California, Department of Information and Computer Science.

[17] B.A. Murtagh and M.A. Saunders. MINOS 5.0 user's guide. Technical Report SOL 83.20, Stanford University, December 1983. MINOS 5.4 Release Notes, December 1992.

[18] V. Pareto. *Manuel d'Economie Politique*. Giard, Paris, 1909.

[19] P. M. Ravdin and G. M. Clark. A practical application of neural network analysis for predicting outcome of individual breast cancer patients. *Breast Cancer Research and Treatment*, 22:285–293, 1992.

[20] P. M. Ravdin, A. K. Tandon, D. C. Allred, G. M. Clark, S. A. W. Fuqua, S. H. Hilsenbeck, G. C. Chamness, and C. K. Osborne. Cathepsin D by western blotting and immunohistochemistry: Failure to confirm correlations with prognosis in node-negative breast cancer. *Journal of Clinical Oncology*, 12:467–474, 1994.

[21] A. Schenone, L. Andreucci, V. Sanguinetti, and P. Morasso. Neural networks for prognosis in breast cancer. *Physica Medica*, IX(Supplement 1):175–178, June 1993.

[22] R. E. Steuer. *Multiple Criteria Optimization: Theory, Computation, and Application*. John Wiley and Sons, 1986.

[23] W. N. Street, O. L. Mangasarian, and W. H. Wolberg. An inductive learning approach to prognostic prediction. In A. Prieditis and S. Russell, editors, *Proceedings of the Twelfth International Conference on Machine Learning*, pages 522–530, San Francisco, 1995. Morgan Kaufmann.

[24] W. N. Street, W. H. Wolberg, and O. L. Mangasarian. Nuclear feature extraction for breast tumor diagnosis. In *IS&T/SPIE 1993 International Symposium on Electronic Imaging: Science and Technology*, volume 1905, pages 861–870, San Jose, California, 1993.

[25] W. N. Street W. H. Wolberg and O. L. Mangasarian. WPBC: wisconsin prognosis breast cancer database. Computer Sciences Department, University of Wisconsin, Madison, ftp://ftp.cs.wisc.edu/math-prog/cpo-dataset/machine-learn/WPBC/., 1995.

[26] W. H. Wolberg, K. P. Bennett, and O. L. Mangasarian. Breast cancer diagnosis and prognostic determination from cell analysis. Manuscript, Departments of Surgery and Human Oncology and Computer Sciences, University of Wisconsin, 1992.

[27] W. H. Wolberg, W. N. Street, D. H. Heisey, and O. L. Mangasarian. Computerized breast cancer diagnosis and prognosis from fine needle aspirates. *Archives of Surgery*, 130:511–516, 1995.

[28] W. H. Wolberg, W. N. Street, and O. L. Mangasarian. Machine learning techniques to diagnose breast cancer from image-processed nuclear features of fine needle aspirates. *Cancer Letters*, 77:163–171, 1994.