CANCER PROGNOSTIC EVALUATION
VIA SUPPORT VECTOR MACHINES

Domenico Conforti [1], Domenico Costanzo [2], Rosita Guido [1]

[1] D.E.I.S., Università della Calabria, Ponte Pietro Bucci 41C, 87036 Rende (CS), Italy , mimmo.conforti@unical.it,
http://www.deis.unical.it/deis1.0/portale/home/conforti/
[2] Dipartimento di Medicina Sperimentale e Clinica, Università degli Studi "Magna Graecia", Via Tommaso
Campanella 115, 88100 Catanzaro, Italy, domcostanzo@unicz.it

Abstract: *In this paper we considered a very challenging medical decision making problem: the short-term prognosis evaluation of breast cancer patients. In particular, the oncologist has to predict the more likely outcome of the disease in terms of survival or recurrence after a given follow-up period: "good" prognosis if the patient is still alive and has not recurrence after the follow-up period, "poor" prognosis if the patient has recurrence or dies within the follow-up period. This prediction can be realized on the basis of the execution of specific clinical tests and patient examinations. The relevant medical decision making problem has been formulated as a supervised binary classification problem. By the framework of generalized Support Vector Machine models, we tested and validate the behavior of four kernel based classifiers: Linear, Polynomial, Gaussian and Laplacian. The overall results demonstrate the effectiveness and robustness of the proposed approaches for solving the relevant medical decision making problem .*

Keywords: *Medical Decision Making, Breast Cancer Prognosis, Support Vector Machine, Kernel Functions*

1. INTRODUCTION

The aim of this paper is to analyze and validate kernel based Support Vector Machine learning methods as basic tools for developing reliable medical decision support systems.

Typically, the core function of any medical decision making process is represented by the correct and accurate identification of the clinical conditions of the patient, in order to effectively make the recognition of the relevant disease, the prediction of the evolution and outcome of the disease, the planning of the optimal therapy.

Under this respect, medical decision making can be suitable modeled within the framework of pattern classification problem, since the core process is to predict a class membership. To discriminate patterns is a fundamental and well known problem, which can be solved by knowledge discovery methodologies, on the basis of which it is possible to learn general knowledge from some specific cases (i.e. learning from experience). In fact, by solving the classification problem, we reach the decision of assigning an object to a suitable class, based on a set of features describing the same object and related to the characteristics of the class. If we model the object as a vector of variables, each representing a specific feature, then the classification problem becomes discriminating the belonging of a new vector of variables on the basis of the "training data", that is the set of historically observed instances of the target vector of variables, which can be considered, to some extent, the relevant knowledge base. Obviously, since our aim is to develop effective decision support systems suitable for the medical domain, we strictly require that classifiers be able to correctly recognize a new and unseen case; this is called "generalization" property. Hence, we seek classifiers with very good generalization property.

Kernel base learning methods represent new, efficient and effective approaches for solving classification problems. Their are based on a suitable transformation of the training data, embedding the original data into a transformed space, in which the classification can be performed by a simple linear separation of the classes. This transformation can be carried out by the so-called kernel function, and, in order to get good performance, in terms of generalization property, it is crucial to select the more appropriate kernel function. Moreover, the kernel function can be used in combination of the technique of Support Vector Machine, a recently established inference engine derived from some fundamental results of the statistical learning theory.

We apply this new methodology to a challenging medical domain: the prognostic evaluation of breast cancers. More specifically, the physician has to predict the possible outcome of the disease in terms of survival or recurrence after a given follow-up period: "good" prognosis if the patient is still alive and has not recurrence after the follow-up period, "poor" prognosis if the patient has recurrence or dies within the follow-up period. This prediction can be realized on the basis of the execution of specific clinical tests and patient examinations.

Within the wide literature in this field, strongly related medical prognosis applications has been considered in [1] and [2].

The main goal of [1] was to identify breast cancer patients for whom adjuvant chemotherapy could prolong survival time. This has been achieved by considering a suitable combination of Support Vector Machine classification with Gaussian kernel and clustering techniques, on the basis of 5 cytological features and only one prognostic factor (tumor size).

In [2], artificial neural networks were developed with the aim to produce quantitative estimates for risk of relapse of breast cancer over 5 years. To this end, the standard TNM (Tumor–Node–Metastasis) staging system was considered, plus some markers of the biological aggressiveness of breast cancer. In particular, the number of axillary metastasis was evaluated.

Our approach here is quite different and original, both in terms of medical decision making problem formulation and in terms of definition and use of the relevant learning algorithms. In fact, we will face here two specific and new problems:

1. the definition and formulation of the relevant medical decision making problem as a binary classification problem, by further developing the relevant knowledge base. Under this respect, our original contribution stays on the particular selection of the clinical features. In fact, we considered a total of 16 features, as a suitable combination of well established prognostic factors and cyto-morphological features from biopsy samples. It is worth while to remark that we did not evaluate the possible axillary metastasis; this aspect could give a strong impact on the practical clinical value of our approach. We will give motivations about that in the sequel.

2. the developing and implementation of the Support Vector Machine learning algorithms, by embedding and evaluating four kernel functions, which could be considered quite suitable for the specific medical domain.

In this way, we were able to achieve, as best testing correctness, the figure of **82.5%**, which can be considered a good estimate of the "generalization" property of the proposed classifiers and a quite interesting performance in terms of accuracy and reliability of the prognosis evaluation.

The paper is organized as follows. In section 2 we will illustrate the relevant medical decision making problem, by defining the basic issues of the breast cancer prognosis. In section 3, we will introduce some fundamental results of the learning methodologies that we will develop and use for our medical problem. Section 4 is devoted to amply describe the experimental testing and validation of the proposed classifiers for effectively solving the prognosis problem. Some concluding remarks complete the paper.

2. PROGNOSIS OF BREAST CANCERS

The prognosis of breast cancer usually refers to the likely outcome of breast cancer. In general, the prognosis of breast cancer may include the duration of disease, chances of complications of breast cancer, probable outcomes, prospects for recovery, recovery period for breast cancer, survival rates, death rates, and other outcome possibilities in the overall prognosis of breast cancer.

The prognosis evaluation of a breast cancer patient is typically performed by measuring and interpreting the prognostic factors, that is those variables which influence the outcome of the patients once they have developed breast cancer. Interest in prognostic factors has been stimulated by the success of the systematic adjuvant therapy for early-stage breast cancers. Patients destined for recurrence can be selected for systematic adjuvant therapy, while patients who will not have a recurrence can be spared the morbidity of a treatment that offers no benefit. Prognostic factors may be clinical (e.g. sex and family history), histologic (obtained from the pathology report and strongly related to the histo-pathologic structure of the tumor and its anatomic staging), or specifically associated to the patient (host prognostic factors: age, menopausal status, previous cancer history, etc.).

In particular, the *lymph node status* is one of the most important prognostic factor for breast cancer: from prognostic point of view, cancer spread to lymph node is worse. In fact, all the analysis regularly indicates that the presence or absence of metastasis of axillary lymph nodes is the single most influential predictor of post treatment recurrence and death. Under this respect, some attempts have been made to identify prognostic factors which would predict axillary node involvement. In fact, the examination of the lymph node status has several

disadvantages: the information is available only postoperatively, the axillary dissection (surgery to remove all the lymph nodes under the arm to see if they contain cancer spread) is very invasive for the patient and causes arm morbidity.

Hence, starting from this well posed motivation, we consider here the problem of making a reliable short-term prognosis prediction without the histo-pathological examination of axillary limph nodes. This could result in a prognosis evaluation procedure which is characterized by a low invasiveness for the patient and, typically, low costs for the health care organization.

To this end, the study reported in this paper is based on the clinical data collected from *84* patients, with confirmed diagnosis of breast cancer and at least three years of follow-up period, during which they were differently treated by some adjuvant therapy. *42* of the patients had **"Good"** prognosis, in the sense that they were still alive and had not recurrence after the follow-up period, whereas the remaining *42* patients had **"Poor"** prognosis, since they died or had recurrence within three years of diagnosis.

Each patients has been described by *16* features, clinical variables measured from the patient and related to the prognostic evaluation process. In particular, we remark that all these features are extracted by histopathological analysis of biopsy samples of the tumor. More specifically, we selected the following *4* tumor prognostic factors:

1. **Tumor size**: anatomic staging classification of the tumor from *1* (best prognosis) to *4* (worst prognosis), on the basis of the dimension of the tumor localized and confined to breast tissue.
2. **Histologic/Nuclear grade**: degree of cell differentiation, in terms of morphology (in particular degree of tubule formation, number of mitoses, nuclear pleomorphism), measured from *1* (high differentiation) to *3* (low differentiation). Well differentiated cells is better than poorly differentiation.
3. **Estrogen receptor status**: presence of estrogen hormone receptor on the tumor cells. Positive status (present, measured by *1*) is better than negative status (not present, measured by *-1*).
4. **Progesterone receptor status**: presence of progesterone hormone receptor on the tumor cells. Positive status (present, measured by *1*) is better than negative status (not present, measured by *-1*).

Moreover, the remaining *12* features were selected by considering morphological characteristics of cell nuclei from biopsy tissue and surgical remove specimens. This features are specifically related to the shape, size and level of chromatine of the nuclei, and automatically measured by effective image processing tools [3]. In particular, we considered 7 features related to the shape, 2 features related to the size and 3 features related to the level of chromatine of the cell nuclei.

All the clinical data have been collected and created at the "Cosenza General Hospital" at Cosenza, Italy, within the clinical activity of the Oncology unit of the same hospital.

For more details on the prognosis of breast cancers, the interest reader is referred to [4].

3. SUPPORT VECTOR MACHINE CLASSIFIERS

In this section, we sketch the basic methodology that has been implemented for solving our prognosis evaluation problem.

Let us consider *m* individuals (patients), each of one with a diagnosis of breast cancer, and assume that for *p* of them we exactly know that they show good prognosis ("Good" class), whereas the remaining *q* individuals show poor prognosis ("Poor" class). Moreover, each individual is characterized by *n* features strictly related to the prognosis evaluation process of breast cancers. Hence, the prognosis procedure can be stated as the problem of classifying the *m* individuals in the *n*-dimensional real space (compactly represented by the *m*n* matrix *A*), by detecting the membership of each individual x^i, *i=1,...,n*, in the class "Good", labeled by $y_i = 1$, or in the class "Poor", labeled by $y_i = -1$.

Following [5] and [6], this problem can be formulated as the following convex quadratic programming problem (linear Support Vector Machine)

$$\min_{\omega,\gamma,\varepsilon} \left(Ce^T\varepsilon + \frac{1}{2}\omega^T\omega \right) \qquad (1)$$

$$s.t.$$

$$D(A\omega - e\gamma) + \varepsilon \geq e$$

$$\varepsilon \geq 0$$

where *C* is a strictly positive parameter, *D* is a *m*n* diagonal matrix of labels *1* and *-1* on the diagonal, *A* is the *m*n* data set matrix and *e* is the *m*-dimensional vector of ones. The optimal solution of the above optimization problem provides the *n*-dimensional vector *ω* and the scalar *γ*, which represent, respectively, the normal vector and the position relative to the origin of the bounding hyperplanes

$$\omega^T x = \gamma + 1 \qquad (2)$$

$$\omega^T x = \gamma - 1 \qquad (3)$$

If the *m*-dimensional vector of the slack variables ε is zero at the optimum, then the "Good" and "Poor" classes will be strictly linearly separated by the bounding hyperplanes. In this case the separating surface is the hyperplane

$$\omega^T x \;=\; \gamma \qquad (4)$$

midway between the bounding hyperplanes.

If the two classes are linearly inseparable then the optimization problem determines the separating hyperplane which minimizes the number of misclassified individuals. To tune the weight of the classification error in the objective function, we can vary the value of *C*. Moreover, the quadratic term in the objective function maximizes the margin, i.e. the distance between the bounding hyperplanes. The maximization of the margin is strictly related to the enhancement of the generalization property of the classifier [7].

However, in the most of the real situations, the classes are linearly inseparable. In this case, a nonlinear separating surface could be much more effective in terms of classification performance. To this end, the Generalized Support Vector Machine model (nonlinear Support Vector Machine) [8]

$$\min_{u,\gamma,\varepsilon} \; \left(Ce^T \varepsilon + \frac{1}{2} u^T u \right) \qquad (5)$$

s.t.

$$D(K(A,A^T)Du - e\gamma) + \varepsilon \geq e$$
$$\varepsilon \geq 0,$$

generates the nonlinear surface

$$K(x^T,A^T)Du \;=\; \gamma, \qquad (6)$$

by using the *m*m* Kernel matrix $K(A,A^T)$.

The structure of the kernel matrix is determined by the relevant kernel function, which, in general, defines a similarity measure of the form

$$(x,x') \;\rightarrow\; K(x,x'), \qquad (7)$$

where

$$K : X \times X \;\rightarrow\; \Re; \qquad (8)$$

in other words, a kernel function takes in input two points, *x* and *x'*, and returns a real number representing, in some sense, their similarity.

The basic idea behind the kernel function consists of mapping, via a suitable transformation, the original training data into a higher dimensional feature space, where it is possible to determine a separating hyperplane maximizing the margin. This corresponds to the construction of a non linear separating surface in the original input space. By using a kernel matrix, it is possible to compute the separating hyperplane without explicitly carrying out the transformation into the feature space.

In general, the choice of the kernel function should reflect prior knowledge about the problem at hand. In fact, the kernel function defines, to some extent, the similarity measure that respects the process generating the specific patterns.

In the medical domain, in most of the cases, it could be very difficult to have a quite useful prior knowledge of the specific problem. In this sense, the obvious approach is to test the performance of several kernels with different characteristics.

Under this respect, in our experiments we chose the following kernel functions:

- Inhomogeneous Polynomial:

$$K(x,x') \;=\; (x^T x' + 1)^d, d \geq 1. \qquad (9)$$

- Gaussian Radial Basis Function:

$$K(x,x') \;=\; \exp(-\|x - x'\|^2 / 2\sigma^2), \sigma > 0. \qquad (10)$$

- Laplacian Radial Basis Function:

$$K(x,x') \;=\; \exp(-\|x - x'\| / \sigma), \sigma > 0. \qquad (11)$$

Broadly speaking, polynomial kernels transform the original data into a product feature space of all products of *d* entries (polynomial classifiers [9]); Gaussian and Laplacian kernels are, typically, suitable kernels for Gaussian and Laplacian processes with variance σ (radial basis function classifiers [10]), whereas. For more details on kernel functions definition and properties, the interest reader is referred to [11].

4. EXPERIMENTAL TESTING AND VALIDATION

In this section, we will describe how we have faced the relevant medical decision problem considered in this paper. In particular, we will illustrate the testing and validation of the several kernels, previously described, embedded into the Support Vector Machine model.

For the testing and validation of the proposed classifiers, we used the well-known *10-fold cross validation procedure* [12]. All the computational experiments have been carried out on the workstation S.G.I. Origin 2000, under Unix

operating system. The relevant software has been developed in Gnu C standard language and for solving the model (1) and (5) we used the state-of-the-art solver CPLEX 6.5 [13].

With the aim to have a better control of the misclassified cases (in terms of false positive and false negative), we assigned different values to the tuning parameter C with respect to the relevant class: a value C^+ for penalizing the error on the "Good" cases (false negative), whereas a value C^- for penalizing the error on the "Poor" cases (false positive). Since it is quite evident that a false positive error could have much more serious consequences on the patient, we stronger penalized the false positive error by assigned increasing values to C^- with respect to C^+. In particular, the numerical experiments have been carried out with the following pairs of values respectively for C^+ and C^-:

$$\begin{pmatrix} C^+ \\ C^- \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 10 \end{pmatrix}, \begin{pmatrix} 5 \\ 20 \end{pmatrix}. \qquad (12)$$

For each kernel, the relevant parameter have been varied randomly in a given range of values, as follows:

- Inhomogeneous Polynomial: $d = 2, 3, 4$.
- Gaussian Radial Basis Function: $\sigma = 1, 1.2, 1.5, 1.8, 2, 3, 5$.
- Laplacian Radial Basis Function: $\sigma = 4, 5, 6, 7, 8, 9, 10$.

We have compared the performance of the mentioned kernels, with the behavior of the classifier based on the linear Support Vector Machine model (1), with the same values of the tuning parameters C^+ and C^-. As a consequence of the several setting of the different parameters, we have run a total of *54* instances of the basic classification problem.

The overall numerical results are synthesized in the Table 1, where we have reported, for each classifiers, the best and worst values of the testing correctness (averages over ten folds), obtained by varying the tuning parameters C^+ and C^-, and the relevant kernel parameter (*8* instances of the total *54* instances considered).

Table 1. 10-Fold Cross Validation Testing Correctness Results.

Kernel	Test Performance	
	Best	**Worst**
Linear	**78.8 %** $C^+ = 1, C^- = 10$	77.5 % $C^+ = 1, C^- = 1$
Polynomial	**82.5 %** $C^+ = 5, C^- = 20$ $d = 3$	**80.0 %** $C^+ = 1, C^- = 1$ $d = 3$
Gaussian	**77.5 %** $C^+ = 1, C^- = 1$ $\sigma = 1$	**65.0 %** $C^+ = 1, C^- = 10$ $\sigma = 1$
Laplacian	**76.3 %** $C^+ = 1, C^- = 1$ $\sigma = 5$	**63.2 %** $C^+ = 1, C^- = 10$ $\sigma = 7$

On the basis of the numerical results and their comparisons, we make the following remarks.

In terms of generalization property, the performance of the proposed classifiers ranges over the interval [**63.2%**, **82.5%**]. In fact, the best performing classifier (**82.5%** test set correctness) has been developed by considering $C^+ = 5$, $C^- = 20$ and the Polynomial kernel with $d = 3$. On the other hand, the worst performance (**63.2%** test set correctness) has been obtained by the classifier based on $C^+ = 1$, $C^- = 10$ and Laplacian kernel with $\sigma = 7$.

The performance of the proposed kernels, for the Nonlinear Support Vector Machine, can be considered quite satisfactory, especially in the case of the Polynomial kernel, which is clearly the more stable and robust, demonstrating that this kernel function is potentially able to capture the "knowledge" hidden into the clinical data set.

On the other hand, it is worth while to point out also the good behavior of the Linear Support Vector Machine.

Finally, the choice to set different values to the tuning parameters C^+ and C^- seems to play an effective role in the cases of Linear and Polinomial kernel, while for the Gaussian and Laplacian kernels the best test performance has been obtained by setting the same value to C^+ and C^-.

It is worth while to observe that, in terms of best test performance, the obtained figures are comparable to those reported in [1] and [2], even though the structure of the relevant decision making problems are quite different.

In particular, the potential practical value of our approach could be established by the fact that, by only considering clinical features from biopsy samples of the tumor and, more important, without any evaluation of the axillary metastasis, we were able to get performance comparable and, in some cases, better than that typically obtained by the standard TNM staging systems (compare the discussion in [2]).

However, some limitations of the proposed approach should be detailed. First, the number and

type of patients considered for the medical knowledge base could be not suitable in order to fully satisfy the clinical validation of the approach. Second, the performance of the learning method used could be impaired by the choice of the relevant parameters (the value of C^+ and C^-, the values of the kernel parameters). These limitations will be considered in a future work.

5. CONCLUSION

In this paper we have proposed the solution of a difficult medical decision making problem - the prognosis evaluation of breast cancer patients - by using new and very promising learning methodologies, based on embedding suitable Kernels into a Support Vector Machine framework.

By developing a medical knowledge base for the related domain, we were able to devise well reliable inference engines on the basis of Kernel based Support Vector Machine learning algorithms, which showed good generalization properties.

The overall experimentation give evidence of the very interesting potentialities of Kernel based Support Vector Machine learning algorithms for approaching, in a new way, many decision making problems in medical domains.

6. REFERENCES

[1]   Y. J. Lee, O. L. Mangasarian, W. H. Wolberg, Survival-Time Classification of Breast Cancer Patients, *Computational Optimization and Applications* 25 (2003), pp. 151-166.

[2]   M. De Laurentiis, S. De Placido, A. R. Bianco, G. M. Clark, P. M. Ravdin, A Prognostic Model that makes quantitative estimates of the Probability of Relapse for Breast Cancer Patients, *Clinical Cancer Research* 5 (1999). pp. 4133-4139.

[3]   W. H. Wolberg, W. N. Street, D. M. Heisey, O. L. Mangasarian, Computer-derived Nuclear Grade and Breast Cancer Prognosis, *Anal. Quant. Cytol. Histol.* 17 (4) (1995), pp. 257-264.

[4]   M. Ferno, Prognostic factors in breast cancer: a brief review, *Anticancer Research* 18 (3C) (1998), pp. 2167-2171.

[5]   V. N. Vapnik, *Statistical Learning Theory*, John Wiley & Sons, New York, 1998.

[6]   V. Cherkassky, F. Mulier, *Learning from Data - Concepts, Theory and Methods*, John Wiley & Sons, New York, 1998.

[7]   N. Cristianini, J. Shawe-Taylor, *An Introduction to Support Vector Machines*, Cambridge University Press, Cambridge, 2000.

[8]   O.L. Mangasarian, Generalized Support Vector Machines, *Advanced in Large Margin Classifiers*, A.J. Smola, P. Barlett, B. Scholkopf, D. Schuurmans, eds., MIT Press, Cambridge, Massachusetts, 2000, pp. 135-146.

[9]   J. Schurmann, *Pattern Classification: a unified view of statistical and neural approaches*, John Wiley & Sons, New York, 1996.

[10]   M.A. Aizerman, E'.M. Braverman, L.I. Rozonoér, Theoretical foundations of the Potential Function method in Pattern Recognition Learning, *Automation and Remote Control* 25 (1964) pp. 821-837.

[11]   B. Scholkopf, A. J. Smola, *Learning with Kernels*, MIT Press, Cambridge, 2002.

[12]   M. Stone, Cross-validation choices and assessment of statistical predictions, *Journal of the Royal Statistical Society*, Series B, 36 (1974) pp. 111-147.

[13]   CPLEX, *ILOG CPLEX 6.5: User's Manual*, CPLEX Optimization, Inc., Incline Village, NV, 1999.

*Prof. **Domenico Conforti**, Associate Professor in Operations Research, D.E.I.S., Università della Calabria (Italy). He graduated in Electrical Engineering and Computer Science at the same University and completed his background in models and methods of numerical optimization at the University of Wisconsin, Madison (USA). His current research activities regards mainly the development of decision support systems for medical decision making and the design of optimization models for health care delivery.*

*Dr. **Domenico Costanzo**, Adjunct Professor in Computer Science at the Università "Magna Graecia", Catanzaro (Italy), Faculty of Medicine. He graduated in Computer Science at the University of Pisa. His research activity is mainly devoted to the development and application of advanced information technologies to medicine and biology.*

*Dr. **Rosita Guido**, Teaching Assistant at D.E.I.S., Università della Calabria (Italy). She graduated in Management and Industrial Engineering at the same University. She is currently PhD student in Optimization and Operations Research, working on the development of optimization models and methods for machine learning applications, with particular emphasis on Kernel methods and Support Vector Machine models.*