IOP PUBLISHING PHYSICS IN MEDICINE AND BIOLOGY

Phys. Med. Biol. **54** (2009) S9–S30 doi:10.1088/0031-9155/54/18/S02

Predicting radiotherapy outcomes using statistical learning techniques*

Issam El Naqa[1], Jeffrey D Bradley[1], Patricia E Lindsay[2], Andrew J Hope[2] and Joseph O Deasy[1]

[1] Washington University, Saint Louis, MO, USA
[2] Department of Radiation Oncology, Princess Margaret Hospital, Toronto, ON, Canada

Received 4 February 2009, in final form 9 April 2009
Published 18 August 2009
Online at stacks.iop.org/PMB/54/S9

Abstract

Radiotherapy outcomes are determined by complex interactions between treatment, anatomical and patient-related variables. A common obstacle to building maximally predictive outcome models for clinical practice is the failure to capture potential complexity of heterogeneous variable interactions and applicability beyond institutional data. We describe a statistical learning methodology that can automatically screen for nonlinear relations among prognostic variables and generalize to unseen data before. In this work, several types of linear and nonlinear kernels to generate interaction terms and approximate the treatment-response function are evaluated. Examples of institutional datasets of esophagitis, pneumonitis and xerostomia endpoints were used. Furthermore, an independent RTOG dataset was used for 'generalizabilty' validation. We formulated the discrimination between risk groups as a supervised learning problem. The distribution of patient groups was initially analyzed using principle components analysis (PCA) to uncover potential nonlinear behavior. The performance of the different methods was evaluated using bivariate correlations and actuarial analysis. Over-fitting was controlled via cross-validation resampling. Our results suggest that a modified support vector machine (SVM) kernel method provided superior performance on leave-one-out testing compared to logistic regression and neural networks in cases where the data exhibited nonlinear behavior on PCA. For instance, in prediction of esophagitis and pneumonitis endpoints, which exhibited nonlinear behavior on PCA, the method provided 21% and 60% improvements, respectively. Furthermore, evaluation on the independent pneumonitis RTOG dataset demonstrated good generalizabilty beyond institutional data in contrast with other models. This indicates that the prediction of treatment response can be improved by utilizing nonlinear kernel methods for discovering important

* Part of this work was first presented at the Seventh International Conference on Machine Learning and Applications, San Diego, CA, USA, 11–13 December 2008.

0031-9155/09/180009+22$30.00

nonlinear interactions among model variables. These models have the capacity to predict on unseen data.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Advances in 3D treatment planning could potentially pave the way for individualized and patient-specific treatment planning decisions based on estimates of tumor local control probability or complication risk to surrounding normal tissues (Webb 1997). Accurate prediction of treatment outcomes would provide clinicians with better tools for informed decision making about patients expected benefits versus anticipated risk. Recently, there has been a burgeoning interest in using radiobiological models to rank patients' treatment plans in order to identify the 'optimal' or at least personalize the patient's plan (Brahme 1999, Deasy *et al* 2002). However, instead of using dose metrics only, other parameters associated with the radiobiological response can be incorporated into the treatment design process. For instance, these models could be used as a guide for treatment options (Armstrong *et al* 2005, Weinstein *et al* 2001). Alternatively, once a decision has been reached, these models could be included in an objective function, and the optimization problem driving the actual patient's treatment plan can be formulated in terms relevant to complication risk and tumor eradication (Moiseenko *et al* 2004, Brahme 1999).

Recent years have witnessed an extraordinary increase in patient-specific biological and clinical information due to the progress in genetics and imaging technology (Elshaikh *et al* 2006). Therefore, recent approaches have focused more on data-driven models, in which dosimetric metrics are mixed with other patient- or disease-based prognostic factors (Deasy and El Naqa 2007). This approach is motivated by recent reports of image-specific outcomes findings (de Crevoisier *et al* 2005, Hope *et al* 2005). In (de Crevoisier *et al* 2005) it was reported that rectal distension on the planning computed tomography (CT) scan is associated with an increased risk of biochemical and local failure in patients of prostate cancer when treated without daily image-guided localization of the prostate. Similarly, in (Hope *et al* 2005), it was found that tumor distance to the spinal cord was a significant predictor of failure in irradiated lung cancer patients. Moreover, biological markers were found to be predictive of biochemical failure in prostate cancer or lung injury in non-small cell lung cancer (NSCLC) post-radiotherapy treatment (post-RT) (Alan Pollack 2003, Chen *et al* 2005).

In standard modeling methods, model parameters could be chosen using traditional statistical techniques to define the abscissa of a logistic regression function, for instance (Blanco *et al* 2005, Bradley *et al* 2004, Levegrun *et al* 2001, Marks 2002, El Naqa *et al* 2006, Hope *et al* 2006). These methods, though useful, are incapable of handling potentially complex interactions, manifested as important nonlinear relationships between combinations of variables and resulting outcomes. Thus, limiting their predictive power and applicability in clinical practice. Handling such nonlinearities requires the active development of more sophisticated methods utilizing datamining approaches and advances in statistical learning theory (Hastie *et al* 2001).

Artificial intelligence techniques (e.g., neural networks and decision trees), which are able to emulate human intelligence by learning the surrounding environment from the given input data, have also been utilized because of their ability to detect nonlinear patterns in the data. In particular, neural networks were extensively investigated to model post-radiation treatment outcomes for cases of lung injury (Munley *et al* 1999, Su *et al* 2005) and biochemical failure and

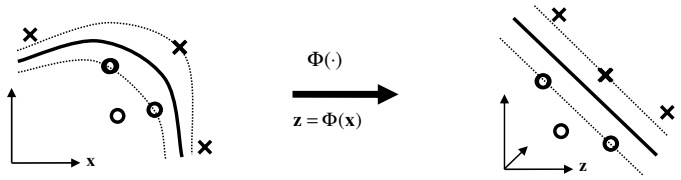

Figure 1. Kernel-based mapping from a lower-dimensional space (*X*) to a higher-dimensional space (*Z*) called the feature space, where non-separable classes become linearly separable. Already established linear theory could be used to estimate the separating hyperplane. Samples on the margin are denoted as support vectors and they define the prediction function, which could be implemented efficiently using the kernel trick.

rectal bleeding in prostate cancer (Gulliford *et al* 2004) However, these studies have mainly focused on using a single class of neural networks, namely feed-forward neural networks (FFNN) (Haykin 1999) with different types of activation functions. In our previous work (El Naqa *et al* 2005), we have shown that a different neural network architecture, referred to as generalized regression neural network (GRNN) (Specht 1991), outperforms classical neural networks. The major drawback using neural network methods is that they are based on greedy heuristic algorithms with no guarantee of global optimality or robustness, in addition to the extensive computational burden associated with them.

In this study, we focus on a branch of machine learning called kernel-based methods, and in particular an extension relying on support vector machine (SVM) methodology. This class of learning methods possesses the ability to adapt artificial intelligence with the avoidance of excessively fitting data incorrectly while maintaining the computational efficiency of the classical statistical methods (Vapnik 1998, Shawe-Taylor and Cristianini 2004). Kernel-based methods are able to incorporate prior knowledge, while handling missing information and uncertainty in the observed data. The basic philosophy is that with the aid of a certain projection or similarity measure (called the kernel) the data are implicitly embedded in a high-dimensional feature space, which allows computationally efficient and well-understood linear methods to be utilized, as illustrated in figure 1.

Kernel methods have been applied successfully in many diverse applications such as pattern recognition (Shawe-Taylor and Cristianini 2004, Vapnik 1998) and in computer-aided diagnosis in medical imaging, as in our previous work (El-Naqa *et al* 2002, 2004). In most, if not all of these applications the kernel-based methods outperformed the state-of-art technology or provided a competitive performance.

We therefore conjecture that kernel-based machine learning methods if properly utilized can help the outcome analyst gain a more insightful understanding of complex variable interactions that affect outcome and wider model applicability. This is partially due to the kernel's natural ability to identify patterns, variable interactions and higher-order relationships without the required guesswork of an analyst. In this paper, we test the hypothesis that kernel-based machine learning methods may improve upon outcomes models using institutional data and resampling methods. Furthermore, we evaluated its clinical applicability using an independent test data.

2. Background

2.1. Multi-metric modeling of radiotherapy outcomes

The approach we adopted for modeling outcomes follows an exploratory datamining-based approach. In the context of data-driven outcomes modeling, the observed treatment outcome

(e.g., normal tissue complication probability (NTCP) or tumor control probability (TCP)) is considered as the result of functional mapping of multiple dosimetric, clinical or biological input variables. Mathematically, this could be expressed as $f(\mathbf{x}; \mathbf{w}^*) : X \rightarrow Y$, where $x_i \in R^d$ are the input explanatory variables (dose–volume metrics, patient disease specific prognostic factors or biological markers) of length d and $y_i \in Y$ are the corresponding observed treatment outcome (TCP or NTCP), and $\mathbf{w}^*$ includes the optimal parameters of outcome model $f(\cdot)$ obtained by optimizing a certain objective criteria. In our previous work (Deasy and El Naqa 2007, El Naqa *et al* 2006), a logit transformation was used:

$$f(\mathbf{x}_i) = \frac{e^{g(\mathbf{x}_i)}}{1 + e^{g(\mathbf{x}_i)}}, \qquad i = 1, \ldots, n, \tag{1}$$

where n is the number of cases (patients), $\mathbf{x}_i$ is a vector of the input variable values used to predict $f(\mathbf{x}_i)$ for outcome y_i of the ith patient. The 'x-axis' summation $g(\mathbf{x}_i)$ is given by

$$g(\mathbf{x}_i) = \beta_o + \sum_{j=1}^{d} \beta_j x_{ij}, \qquad i = 1, \ldots, n, \quad j = 1, \ldots, d, \tag{2}$$

where d is the number of model variables, and the β's are the set of model coefficients determined by maximizing the probability that the data gave rise to the observations. However, a major weakness in using this formulation is that the model's capacity to learn out-of-sample data is limited as noted in our recent analysis of the 93–11 RTOG data (Bradley *et al* 2007). In addition, equation (2) requires the user's feedback to determine whether interaction terms or higher-order terms should be included in the model, making it a trial-and-error process. A solution to ameliorate these problems is offered by applying machine-learning methods as discussed in the next section.

2.2. *Kernel-based methods*

Kernel-based methods and its most prominent member, SVMs, are universal constructive learning procedures based on the statistical learning theory (Vapnik 1998).

2.2.1. Statistical learning. Learning is defined in this context as estimating dependences from data (Hastie *et al* 2001). There are two common types of learning: supervised and unsupervised. In this study, we will focus mainly on supervised learning where the endpoints of the treatments such as tumor control or toxicity grade are provided by experienced oncologists following RTOG or NCI criteria. Nevertheless, we will use unsupervised methods such as PCA to aid visualization of multivariate data and selection of kernel type as described later. For discrimination between patients who are at low risk versus patients who are at high risk of radiation therapy, the main idea of the kernel-based technique would be to separate these two classes with 'hyper-planes' that maximize the margin between them in the nonlinear feature space defined by implicit kernel mapping. The optimization problem is formulated as minimizing the following cost function,

$$L(\mathbf{w}, \xi) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{n} \xi_i, \tag{3}$$

subject to the constraint

$$y_i(\mathbf{w}^T \Phi(\mathbf{x}_i) + b) \geqslant 1 - \zeta_i \qquad i = 1, 2, \ldots, n,$$
$$\zeta_i \geqslant 0 \quad \text{for all } i$$

where $\mathbf{w}$ is a weighting vector, and $\Phi(\cdot)$ is a nonlinear mapping function. The ζ_i represents the tolerance error allowed for each sample being on the wrong side of the margin. Note that minimization of the first term in equation (3) increases the separation (improves generalizabilty) between the two classes, whereas, minimization of the second term (penalty term) improves fitting accuracy. The trade-off between complexity and fitting error is controlled by the regularization parameter C.

It stands to reason that such nonlinear formulation would suffer from the curse of dimensionality (i.e., the dimension of the problem becomes too large to solve) (Haykin 1999, Hastie *et al* 2001). However, computational efficiency is achieved from solving the dual optimization problem instead of equation (3), which is convex with a complexity that is dependent only on the number of samples (Vapnik 1998). The prediction function in this case is characterized only by a subset of the training data known as support vectors s_i:

$$f(\mathbf{x}) = \sum_{i=1}^{n_s} \alpha_i y_i K(\mathbf{s}_i, \mathbf{x}) + \alpha_0, \tag{4}$$

where n_s is the number of support vectors, α_i are the dual coefficients determined by quadratic programming, and $K(\cdot,\cdot)$ is the kernel function as discussed next.

2.2.2. Kernel construction. Kernels are the basic ingredient shared by all kernel-based methods. An admissible kernel should satisfy Mercer's positivity conditions since by definition they represent inner product functions (Schèolkopf and Smola 2002):

$$K(\mathbf{x}, \mathbf{x}') = \Phi(\mathbf{x})^T \Phi(\mathbf{x}'), \tag{5}$$

where the mapping Φ is implicit and need not to be defined.

2.3. Model variable selection

Any multivariate analysis often involves a large number of variables or features (Guyon and Elissee 2003). The main features that characterize the observations are usually unknown. Therefore, dimensionality reduction or subset selection aims to find the 'significant' set of features. Finding the best subset of features is definitely challenging, especially in the case of nonlinear models. The objective is to reduce the model complexity, decrease the computational burden and improve the generalizability on unseen data. The straightforward approach is to make an educated guess based on experience and domain knowledge; then, apply feature transformation (e.g., principle component analysis (PCA)) (Dawson *et al* 2005, Kennedy *et al* 1998), or sensitivity analysis by using organized search such as sequential forward selection or sequential backward selection or combination of both (Kennedy *et al* 1998). A recursive feature elimination (RFE) technique that is based on machine learning has also been suggested (Guyon *et al* 2002), in which the data set is initialized to contain the whole set, train the predictor (e.g., SVM classifier) on the data, rank the features according tocertain criteria (e.g., $\|\mathbf{w}\|$) and keep iterating by eliminating the lowest ranked one. In our previous work (El Naqa *et al* 2006), we used model order determination based on information theory and resampling techniques to select the significant variables.

3. Methods and materials

3.1. Kernel-based methods for modeling radiotherapy outcomes

However, another issue that arises with the generic SVM formulation in (3) is that the cost function treats samples from the two classes equally. However, this may not be the case in

many radiotherapy outcome cases, where event and non-event cases maybe imbalanced. One way to account for this issue is by assigning different penalty weights to the samples belonging in the different classes in equation (3) by decomposing C (El-Naqa *et al* 2004). This way, the penalty term is rewritten as

$$C^{+}\sum_{i\in Z^{+}}\xi_i + C^{-}\sum_{i\in Z^{-}}\xi_i, \tag{6}$$

where $\pm$ symbols represent the two classes. A higher penalty weight is assigned to the under-represented class.

Typically, used nonlinear kernels include

$$\begin{aligned}&\text{Polynomials:}\quad K(\mathbf{x},\mathbf{x}') = (\mathbf{x}^T\mathbf{x}' + c)^q\\ &\text{Radial basis function (RBF):}\quad K(\mathbf{x},\mathbf{x}') = \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{x}-\mathbf{x}'\|^2\right),\end{aligned} \tag{7}$$

where c is a constant, q is the order of the polynomial, and σ is the width of the radial basis functions (RBF). Note that the kernel in these cases acts as a similarity function between sample points in the feature space. Moreover, kernels enjoy closure properties, i.e., one can create admissible composite kernels by weighted addition and multiplication of elementary kernels. This flexibility allows for constructing a neural network by using the combination of sigmoidal kernels or one could choose a logistic regression equivalent kernel by replacing the hinge loss with the binomial deviance (Hastie *et al* 2001).

3.2. Visualization of higher-dimensional data

Prior to applying the kernel-based method, it is important to visualize the data distribution, as a screening test. This requires projecting the data into a lower-dimensional space. In this work, we chose the PCA approach due to its simplicity. In PCA analysis, the principal components (PCs) of a data matrix $\boldsymbol{X}$ (with zero mean) are given by (Härdle and Simar 2003)

$$PC = U^T X = \Sigma V^T, \tag{8}$$

where $U\Sigma V^T$ is the singular value decomposition of X. This is equivalent to transformation into a new coordinate system such that the greatest variance by any projection of the data would lie on the first coordinate (first PC), the second greatest variance on the second coordinate (second PC) and so on. For visualization purposes with the PCA, the heterogeneous variables were normalized using z-scoring (zero mean and unity variance).

The term 'Variance Explained,' used in PCA plots (cf figure 2), refers to the variance of the 'data model' about the mean prognostic input factor values. The 'data model' is formed as a linear combination of its principal components. Thus, if the PC representation of the data 'explains' the spread (variance) of the data about the full data mean, it would be expected that the PC representation captures enough information for modeling.

We used unsupervised PCA analysis to provide an indication about class separability; however, it should be cautioned that PCA is an indicator and is not necessarily optimized for this purpose as supervised linear discriminant analysis, for instance.

3.3. Evaluation and validation methods

To evaluate the performance of our classifiers, we used Matthew's correlation coefficient (MCC) (Matthews 1975) as a performance evaluation metric for classification, which is given by

$$\text{MCC} = \frac{\text{TP}\times\text{TN} - \text{FN}\times\text{FP}}{\sqrt{(\text{TN}+\text{FN})(\text{TP}+\text{FN})(\text{TN}+\text{FP})(\text{TP}+\text{FP})}}, \tag{9}$$

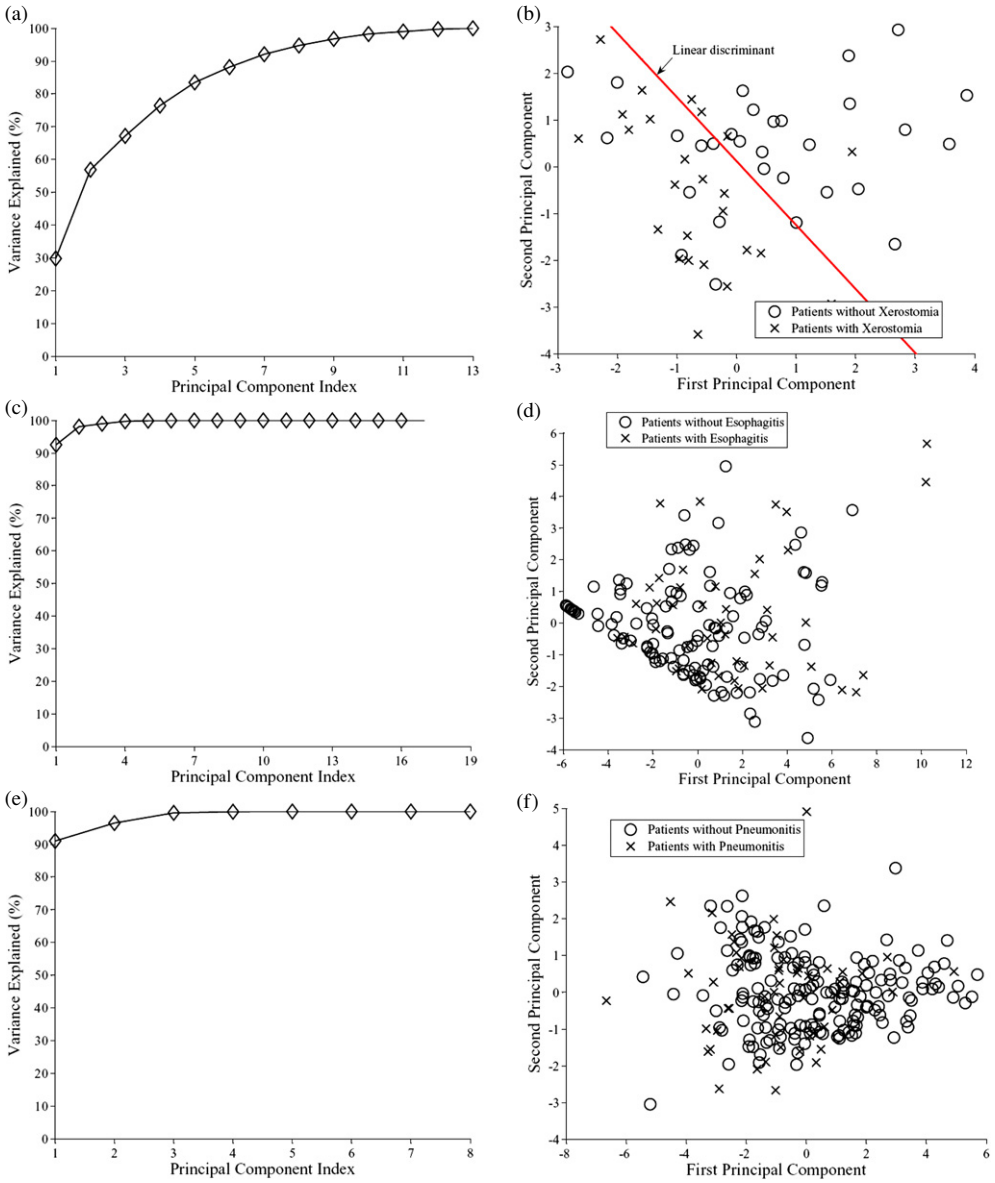

Figure 2. Visualization of higher-dimensional data by principle component analysis (PCA). The left column shows variation explanation versus principle component (PC) index. This represents the variance of the data model about the mean prognostic input factor values. The right column shows data projection into the first two components space. The different endpoints are (a), (b) xerostomia in head and neck, (c), (d) esophagitis and (e), (f) pneumonitis in lung cancer. Note the linear separation in the xerostomia case in contrast with the very high overlap in the pneumonitis case, which makes pneumonitis a better candidate for nonlinear kernel modeling.

where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives. An MCC value of 1 would indicate perfect classification, a value of –1 would indicate anti-classification, and a value close to zero would mean no correlation. For ranking evaluation, i.e., the ability of the classifier to measure prediction quality trends, we used Spearman's correlation, which provides a robust estimator of trend. This is a desirable property, particularly when ranking the quality of treatment plans for different patients.

We used resampling methods (leave-one-out cross-validation (LOO) and bootstrap) for model selection and performance comparison purposes. These methods provide statistically sound results when the available data set is limited (Good 2006). Application of these methods for radiotherapy outcome modeling is reviewed in our previous work (El Naqa *et al* 2006). Furthermore, we evaluated the generalizabilty of the model (i.e., applicability to unseen data) on an independent RTOG dataset.

3.4. Data sets

We used three datasets of patients as representative examples. The first set consisted of 55 head and neck cancer (HNC) patients who were evaluated by quantitative measurements of whole-mouth stimulated and unstimulated saliva flow prior to therapy and at 6 months follow-up post-RT (Blanco *et al* 2005). Xerostomia (dry mouth) was defined as occurring when the post-RT stimulated flow fell below 25% of the pretreatment value. Several mathematical models were used to fit the saliva data (Blanco *et al* 2005). The explored variables included relevant clinical variables (age, gender, race, performance status, chemo, stage, histology, tumor subsite, treatment technique, fraction size, total dose, etc) (Blanco *et al* 2005). (Blanco *et al* 2005, Chao *et al* 2001). In addition, we included a relatively successful model (Blanco *et al* 2005) that predicts the salivary function from each parotid gland to decrease according to and exponential function with an attenuation parameter equal 0.054 Gy^{-1}.

The next two data sets consisted of patients diagnosed with NSCLC treated with three-dimensional conformal radiation therapy (3D-CRT) with median doses around 70 Gy. One set consisted of 52 out of 219 patients diagnosed with post-radiation late pneumonitis (lung inflammation) (RTOG grade $\geqslant 3$) (Hope *et al* 2006). In the other set, 45 out of 166 patients were diagnosed with a highest acute esophagitis (esophagus inflammation) score greater than or equal to 3 according to the RTOG scale (Bradley *et al* 2004). An independent dataset from the RTOG 9311 after removing duplicate patients from our institution was used for evaluating generalizabilty to out-of-sample data. The RTOG 9311 is a radiation dose-escalation study largely based on the $V20$ (percent volume receiving more than 20 Gy) value. The endpoint of pneumonitis was tested (Bradley *et al* 2005). These data sets contain clinical, dosimetric and tumor location parameters. The clinical variables patient related information such as age, last follow-up date, status at follow-up, weight loss, gender, performance status and smoking history. In addition to tumor characteristics such as tumor histology, gross tumor volume (GTV) and tumor stage. The dosimetric variables radiation prescription dose, maximum dose (Gy), treatment time, fraction size (Gy), Vx values (normal lung tissue volume receiving more than x Gy), Dx values (minimum dose to the hottest x% of lung volume) and the use of either sequential or concurrent chemotherapy. The high-dose location effect within the lung by analyzing the center of mass (COM) of the GTV for each patient relative to lateral, anterior-to-posterior and superior-to-inferior dimensions of the lung bounding box (Bradley *et al* 2007).

The results presented here are only for demonstrating the use of our techniques and are not intended as formal clinical findings, which are presented elsewhere (Blanco *et al* 2005,

Bradley *et al* 2004, 2007, Hope *et al* 2006). Treatment planning data were de-archived and potential prognostic factors were extracted using CERR (Deasy *et al* 2003).

3.5. Kernel-based modeling of NTCP

As an example, we formulate the treatment outcome modeling as a binary-classification problem of post-treatment risk, i.e., patients who developed complications after treatment belong to class '+1' and patients who did not develop complications belong to class '−1'. Consequently, our objectives become to find the best predictor that separates the two classes. Note that designing a kernel-based SVM requires the determination of a regularization parameter 'C', which provides the best trade-off between machine complexity (defined in terms of the number of support vectors) and the tolerated classification error. Higher values of C indicate more complexity and more penalization of error. The input variables are normalized between 0 and 1 as a pre-processing step. In this work, we used our logistic regression-based modeling technique and contrasted this with SVM–RFE selection as discussed below.

In our simulations, we used LOO to select model parameters and measure the generalizability of the different classifiers. In which, all the data are used for training except for one left out for testing; the sample is permuted in a round-robin fashion. Matthew's correlation coefficient, averaged over cross-validation test samples, is used as the performance evaluation metric.

4. Experimental results

4.1. Data exploration and visualization

In the head and neck case, we modeled xerostomia using the ratio of pre- and 6 months post-RT stimulated salivary function, measured by whole mouth collection. The PCA analysis of figures 2(a), (b) shows that these data are almost linearly separable into high-risk and low-risk groups. Moreover, most of the data are explained by the exponential mean dose model alone, given that there is an uncertainty of 30% in the measurement of saliva. Hence, this type of data will not benefit from kernel-based approaches and could be modeled satisfactorily using conventional methods (El Naqa *et al* 2005).

In figures 2(c) and (d), we show the esophagitis data, where concurrent chemo is used in conjunction with the Vx (percentage esophagus volume that received at least x Gy). From PCA alone, one can notice that the overlap in the esophagitis case is less than that encountered in the pneumonitis case. As a result, a low-order polynomial kernel may be used to separate these two classes.

In figures 2(e) and (f), we analyzed the pneumonitis endpoint, with a pool of dosimetric variables that consisted only of Vx (i.e., the percentage volume of non-GTV lung that got at least x Gy) (Similar to Su *et al* (2005)). As Vx values are highly correlated (Hope *et al* 2006), dose was binned into eight groups ($V10$, $V20, \ldots$ and $V80$). Notice that more than 99% of the variation in the input data was explained by the first two components (figure 2(e)). Additionally, the overlap (figure 2(f)) between patients with and without radiation pneumonitis is very high, suggesting there is no linear classifier that can adequately separate these two classes. Similar overlap was also observed when other clinical and dosimetric variables were added emphasizing the nonlinear nature of this data.

4.2. Kernel-based modeling examples

4.2.1. NTCP modeling of xerostomia in head and neck cancer. In figure 3(a), we notice that the linear classifier with $C = 1$ (which indicates a preference for the simplest possible

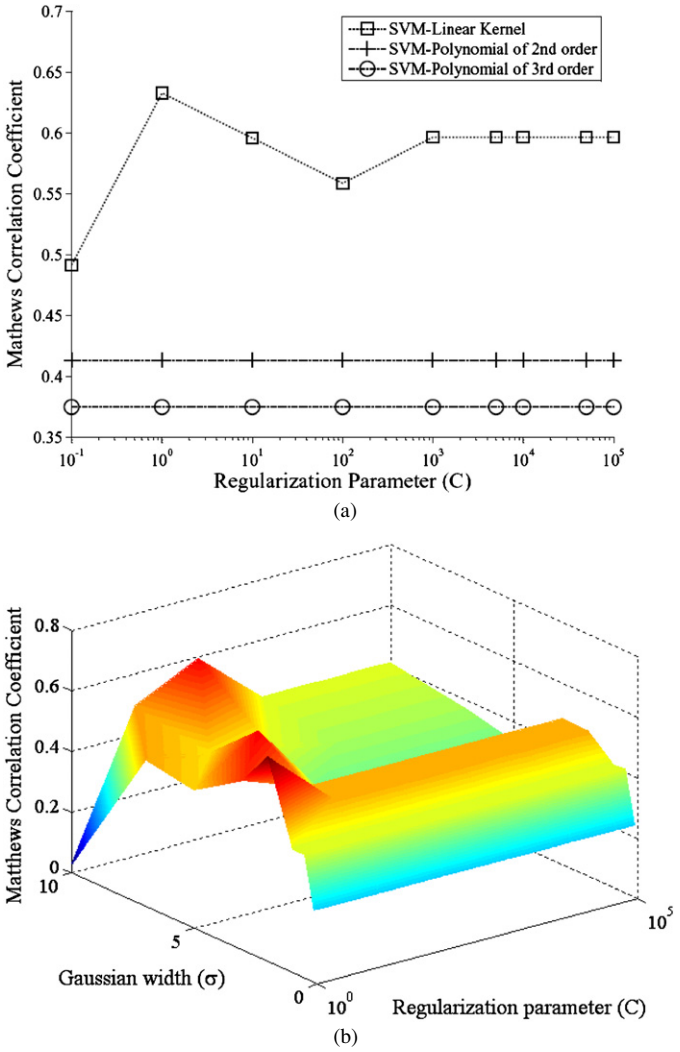

Figure 3. Kernel-based modeling of xerostomia using support vector machine (SVM) with different kernel types. All identified dosimetric and patient variables were included as inputs. (a) SVM using linear and polynomial kernels, (b) similar results of SVM using radial basis functions (RBF), but with a 3D plot is used because of the Gaussian width. The best prediction is obtained with a linear kernel. This means that there is little interaction between the effect of the best dosimetric indicators and other variables.

model) had better performance than any higher-order polynomial, with MCC = 0.64. Using the more complex RBFs also did not provide further improvement. This is an example where linear methods should be preferred in the modeling. In contrast, using a feed-forward neural network with seven neurons had an MCC = 0.3, while a GRNN yielded an MCC of 0.6 using a width of 1.75 (El Naqa *et al* 2005). These results are consistent with our multivariate logistic analysis where a single variable (exponential mean dose model) was sufficient to model the data indicating a sigmoidal dose effect only (Blanco *et al* 2005, Chao *et al* 2001). However, we believe that this case is not the norm but rather the exception in radiation therapy as demonstrated next.

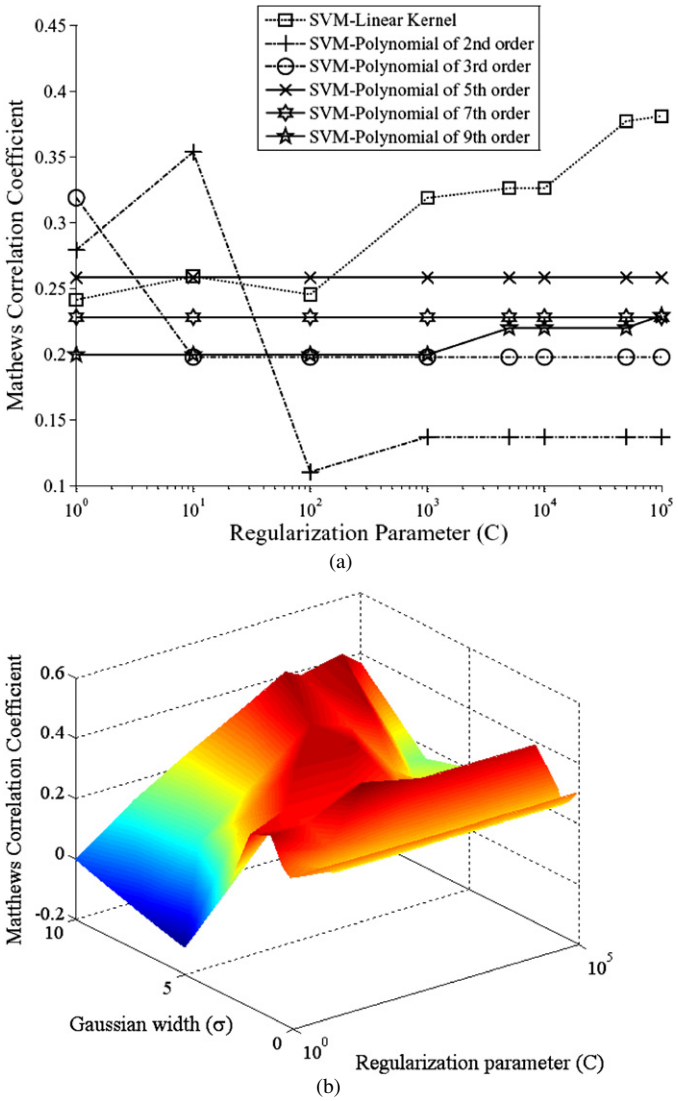

Figure 4. Kernel-based modeling of esophagitis in lung cancer using concurrent chemo and *Vx* with SVM of different kernel types (no pre-modeling variable selection was performed). (a) SVM using linear and polynomial kernels, (b) SVM results using a RBFs. The maximum predictive power is obtained with a radial basis kernel having ($\sigma = 1$, $C = 100$).

4.2.2. NTCP modeling of esophagitis in lung cancer. As shown in figure 4, the best overall performance was obtained using an SVM with a RBF of width 2 and $C = 100$ resulting in an MCC value of 0.43 (figure 4(b)). It is worth mentioning that this result is better by 21% than the generalizability on LOO cross-validation of the multi-metric logistic regression (MCC = 0.36).

4.2.3. NTCP modeling of pneumonitis in lung cancer. We will first consider dosimetric variables only to predict pneumonitis. Using the dosimetric variables, *Vx*, resulted in selecting *C* for linear and polynomial kernels as shown in figure 5(a). It is noted that the higher the

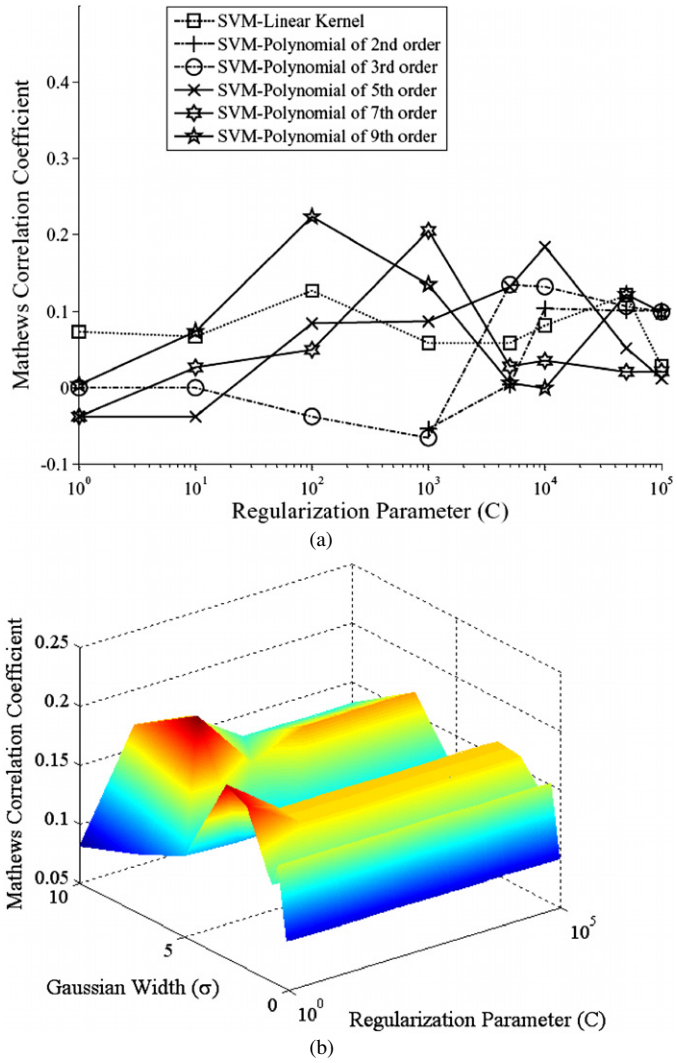

Figure 5. Kernel-based modeling of pneumonitis in lung cancer using *Vx* with SVM of different kernel types (no pre-modeling variable selection was performed). (a) SVM using linear and polynomial kernels, (b) similar results of SVM using RBFs. The best prediction is obtained with a polynomial kernel of the ninth order. In this case, there are significant variable interactions that can increase the predictive power, if detected and modeled.

polynomial order, the better the prediction accuracy. The best performance was obtained with polynomial kernel of order $p = 9$ and $C = 100$. In figure 5(b), we used an SVM with RBFs, where the width of the Gaussian needs to be determined alongside the regularization parameter yielding improved performance when $C = 100$ and a width of 5 were applied (MCC = 0.21). For comparison purposes (El Naqa *et al* 2005), an FFNN with three neurons had an MCC = 0.15. The training time of the FFNN is about one day on a 2.0 GHz Intel processor. Using the fast GRNN training (few seconds), we observed two peaks: one is at $\sigma = 3$, and the other one is at $\sigma = 5$, both having an MCC value of about 0.2. Multiple peaks are usually a sign that variable preference is not strong. However, we show how to improve the performance by including non-dosimetric variables next.

4.3. Variable selection effects in kernel-based methods

Variable selection plays a crucial role in the performance of any prediction method. (Guyon and Elissee 2003) In practice, however, the associations between the dosimetric, biological, clinical variables and the observed endpoints are unknown. The aim of the variable selection process is to find the best set of features that can improve the prediction power. In our previous work on multi-metric modeling (El Naqa *et al* 2006), we investigated methods based on stepwise forward selection and resampling methods to extract significant variables, and to demonstrate whether one variable set is robust versus a cohort of similarly performing models.

We now consider the potential effect of having a heterogeneous dataset that includes non-dosimetric variables in the pneumonitis prediction model. In this section, we explore the effect of variable selection over the entire variable pool on the prediction of pneumonitis in the lung using SVM–RBF as a classifier. A total pool of 58 variables were used, including clinical variables, dosimetric variables, such as Vx (lung volume getting at least x Gy), Dx (minimum dose to the hottest x% lung volume) and the relative location of the tumor within the lung. In figure 6, we show the top 30 selected variables using a recursive-feature-elimination SVM method, which was previously shown to be an excellent method for gene selection in microarray studies (Guyon *et al* 2002). We use variable pruning to account for multi-colinearity of correlated variables as shown in figure 6(a). In figure 6(b), we show the resulting SVM–RBF classifier using the top six variables (using a cutoff of 5% weighting score). The best MCC obtained was 0.22. In figure 7(a), we show the results of variable selection using our previous multi-metric approach based on model order selection and resampling with logistic regression (El Naqa *et al* 2006, Hope *et al* 2006). The model order was determined to be 3 with variables of $D35$, max dose and COM-SI (center of mass of tumor location in the superior–inferior direction) (Hope *et al* 2006). Figure 7(b) shows the evaluation results of applying the SVM methodology with RBF kernels using these selected variables. The resulting correlation (MCC = 0.34) on LOO testing data significantly improved our previously achieved multi-metric logistic regression by 46% as discussed below using a C^+:C^- ratio of 1.0 (El Naqa *et al* 2008). The basic interpretation of this improvement is that the SVM automatically identified and accounted for interactions between the model variables. Moreover, using the decomposed SVM approach, to account for data imbalance, with a C^+:C^- ratio of 1.5 (imposing higher penalty on missing pneumonitis events) yielded an MCC = 0.36 (60% improvement compared to multi-metric logistic regression and 14% compared to C^+:C^- ratio of 1.0). The ratio value in this case was selected by searching the C^+:C^- neighborhood from the value of 1–4.2, which is the number of samples to the number of pneumonitis events.

4.4. Comparison with previous work and validation on an independent dataset

For comparison purposes, previous reports used Spearman rank correlation (Rs) as a visible sign for improvement. For instance, when using the SVM classifier to predict radiation pneumonitis risk we achieved an Rs of 0.37 (or MCC = 0.36) compared to $V20$ which yields an Rs = 0.18 (or MCC = 0.20) or with the best logistic regression model resulting in an Rs = 0.22 using LOO evaluation. These results thus provide a 60% improvement in prediction power over our current best model. A comparison of risk prediction using proposed kernel-based approach versus conventional $V20$ and our previous three-term logistic regression model is shown in figure 8. In which the patients are sorted in an ascending order and divided into eight equal groups according to their risk prediction using logistic regression and kernel-based prediction as a reference in figures 8(a) and (b), respectively. The distinctive ability of the proposed approach to fit both low-risk and high-risk groups is demonstrated even in the case

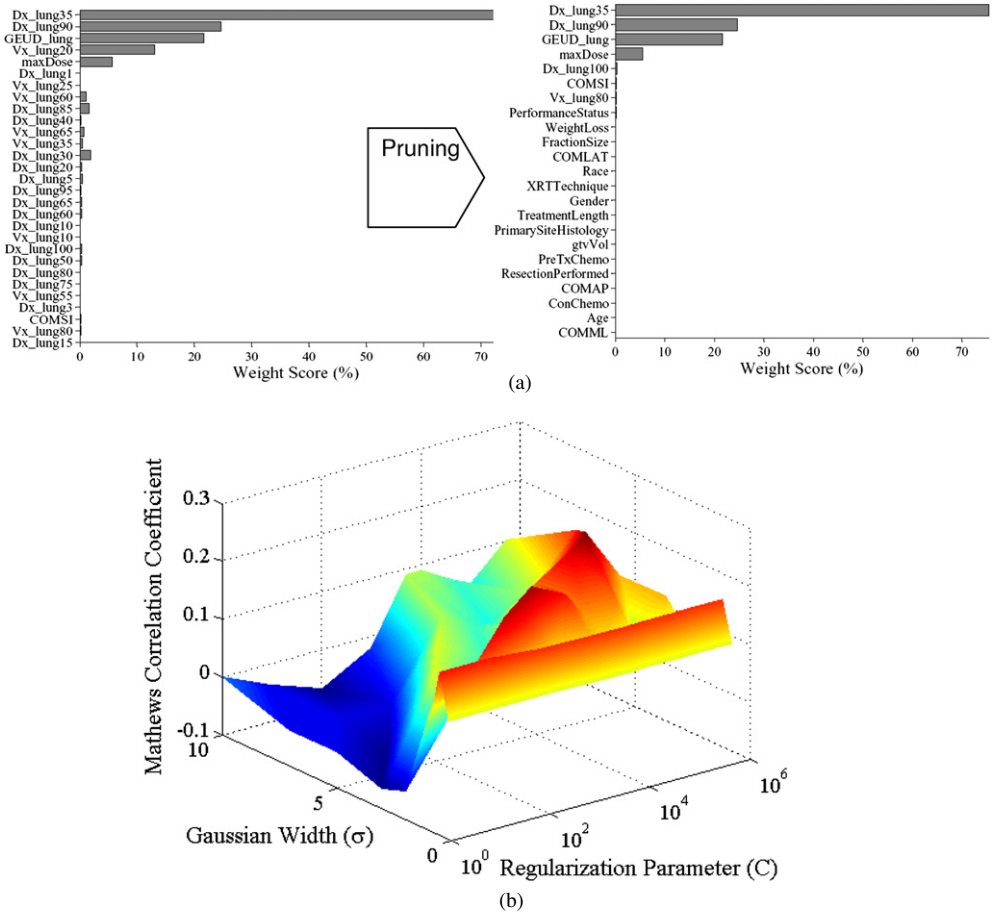

Figure 6. Pneumonitis with a pre-modeling variable selection using the recursive feature elimination (RFE) method. Variables were chosen from a pool of 58 dosimetric, positional and clinical variables. (a) The top 30 variables selected by SVM–RFE are shown before (left) and after (right) applying the pruning step to correct for multi-collinearity ($Rs = 0.75$), (b) The top six variables selected by the pruned SVM–RFE by applying a cutoff of 5% weighting score were used for modeling pneumonitis. An SVM–RBF classifier was tested on LOO testing data.

of imbalanced representation of events as in this case. Furthermore, this ability to discriminate between risk groups is demonstrated using complication-free survival plots using the Kaplan–Meier method in figure 9. Another important consideration for clinical applications is the generalizabilty to unseen data as discussed next.

For validation on independent dataset, we used the RTOG 93–11. We have previously reported that the best three-parameter multi-metric model from our institute resulted in poor performance when applied to the RTOG dataset ($Rs = 0.06$) (Bradley *et al* 2007). Using the SVM–RBF model trained solely on our institutional WUSTL (Washington University in St. Louis) data with the same three variables and applied blindly to the RTOG data resulted in $Rs = 0.16$ (or MCC = 0.15), a slightly better correlation nevertheless statistically significant ($p = 0.049$). It is noted that the application of variable selection to the RTOG data using logistic regression yielded $D15$ as a single variable model, while using SVM–RFE yielded

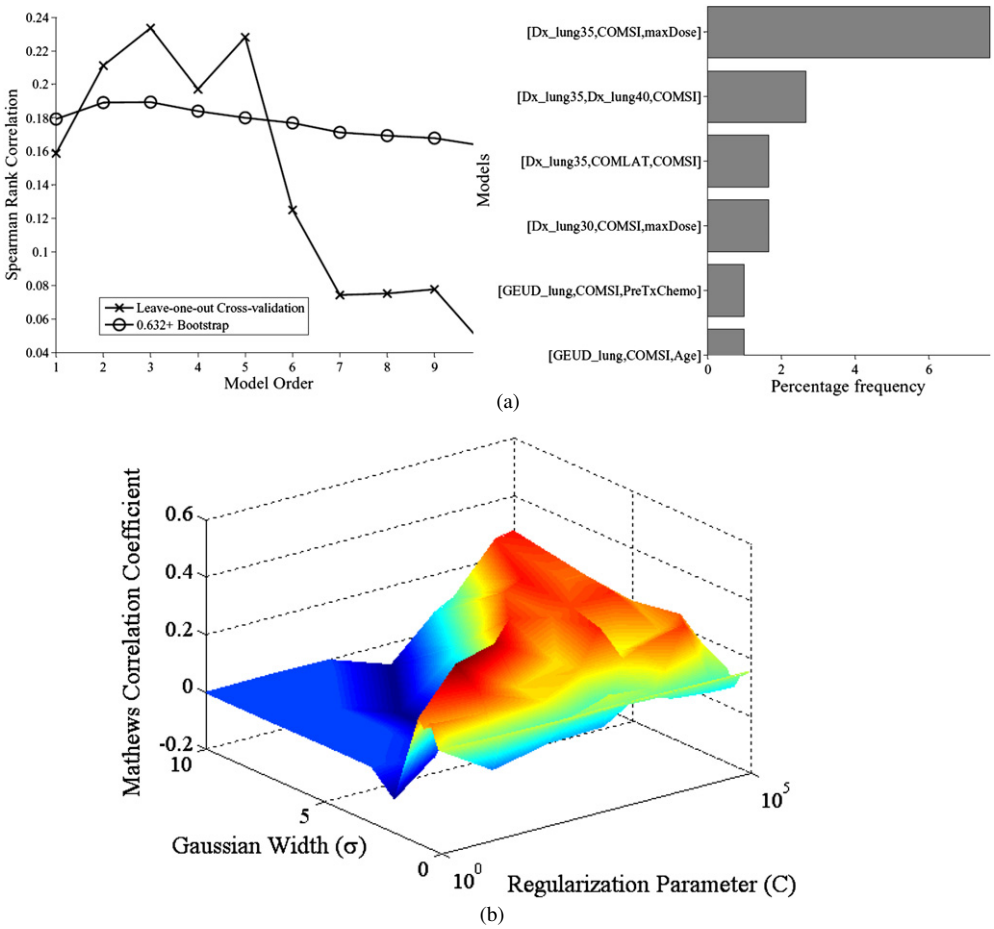

Figure 7. Pneumonitis with a pre-modeling variable selection using the multi-metric logistic regression approach. (a) The two steps variable selection of the model order (left) using resampling methods and the frequency of selected models (right). The best selected model consisted of three parameters (D35, COM-SI and maximum dose). (b) The results of applying the SVM methodology with RBF kernels using these selected variables on LOO testing data. Note the improved performance in this case compared to RFE variable selection.

*D*95, *D*25 and *D*15 as the top three variables in descending order of relevance. However, as mentioned earlier, variable selection plays an important role in learning performance. Therefore, we used mean lung dose (MLD) and dose center in the superior–inferior direction (COM-SI) as variables selected by the logistic regression analysis from the combined datasets (Bradley *et al* 2007). Then, we trained SVM solely on WUSTL data and applied the resulting learning machine to the RTOG data. The result yielded an $Rs = 0.31$ (or MCC $= 0.28$). This generalization ability is typically limited when using conventional methods.

In figure 10, we present the resulting kernel-based pneumonitis nonlinear prediction model as a function f of mean dose and dose center in the superior–inferior direction using an SVM–RBF according to equation (4) ($n_s = 128$, $\sigma_{\mathrm{RBF}} = 3$). Based on patient's characteristics (MLD and COM-SI), there are four possible regions for prediction based on the risk group and prediction confidence level: (1) region of low-risk patients with high confidence prediction

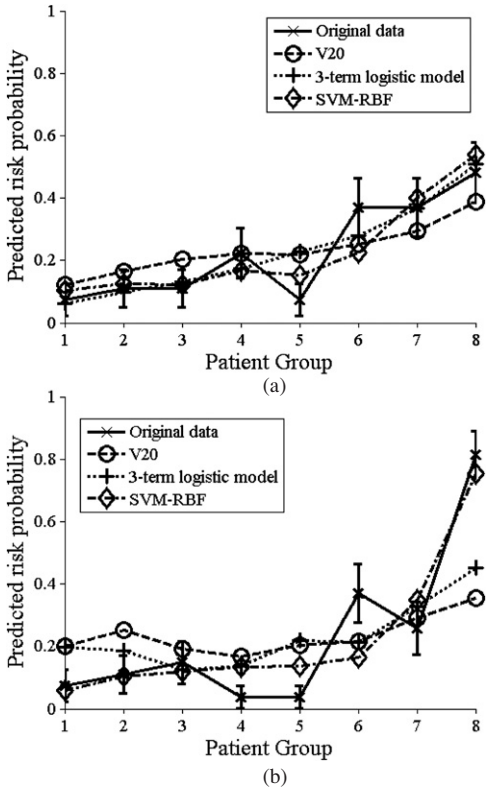

Figure 8. Risk plot comparison of different pneumonitis risk prediction models as a function of patients' binned equal groups. The SVM–RBF is compared to *V*20 and our previous best three-parameter logistic model. (a) Predictors binned by the three-parameter logistic model and (b) predictors binned by SVM–RBF. It is noted that prediction of low risk is quite similar, however, the SVM–RBF provides a significant superior performance in predicting high-risk patients.

level ($f \leqslant -1$), (2) region of low-risk patients with lower confidence prediction level ($-1 \leqslant f \leqslant 0$), (3) region of high-risk patients with lower confidence prediction level ($0 \leqslant f \leqslant 1$) and (4) region of high-risk patients with higher confidence prediction level ($f \geqslant 1$). These are translated into NTCP prediction probabilities using a sigmoidal function for illustration purposes. The lower confidence level group is patients whose characteristics lie within the margins for cases that are considered 'border-line' cases. The corresponding logistic regression plot is shown in figure 11 for comparison where the effect of linear versus nonlinear tessellation of the input space could be contrasted.

5. Discussion

Attempting to uncover predictive relationships in radiation oncology dose–volume analyses is hampered by many effects, of which we will briefly discuss four. First, effective modeling can only occur if the important features that determine outcomes are indeed included in the feature set ('*Variable set adequacy*'). Second, many of the variables (especially dosimetric variables) are highly correlated for most types of treatments ('*Variable correlations*'). Third, limitations in dataset size impose sampling fluctuations ('*Sample noise*'). A key consequence of *Variable*

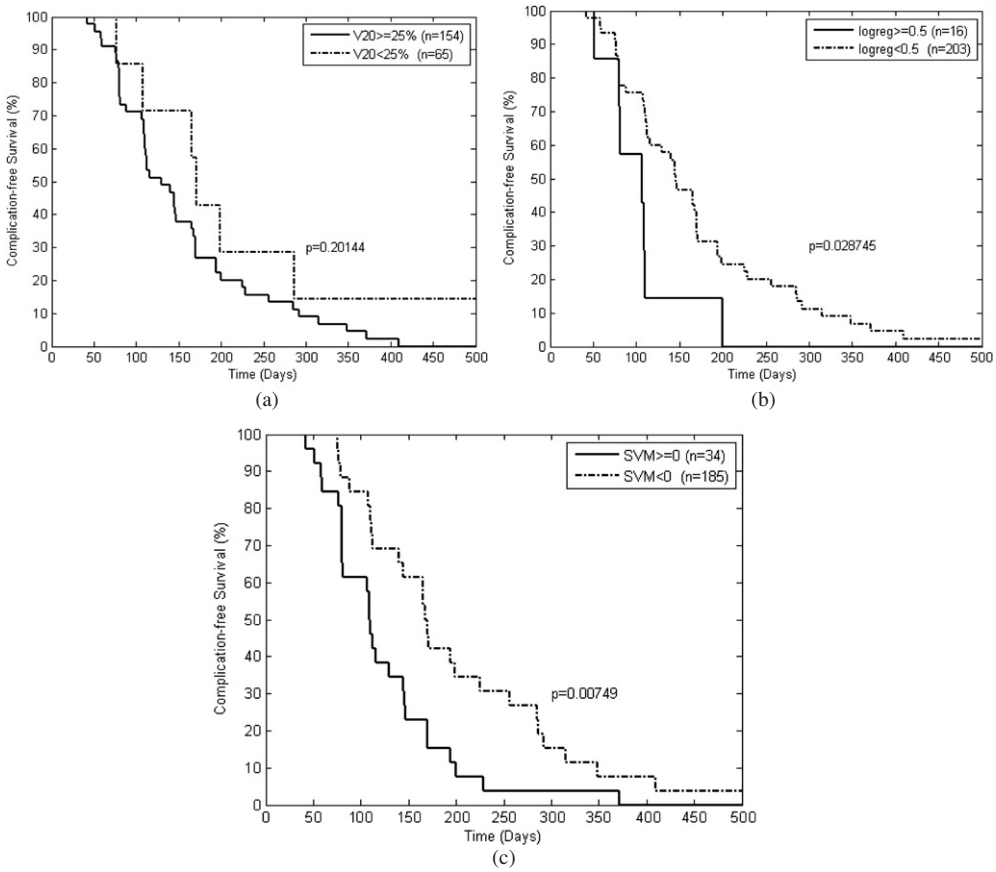

Figure 9. Complication free survival rate as a function of time after radiation therapy using the Kaplan–Meier and log-rank test. Comparison of the complication-free survival of RP, including discrimination between (a) the *V*20 metric with a cutoff of 25% (Bradley *et al* 2005), (b) multi-metric logistic regression with a cutoff of 0.5 using COM-SI, D35 and maximum dose, and (c) SVM with RBF kernel using COM-SI, D35, and maximum dose with patients with low risk defined in class −1 and patients at high risk in class +1. Note the highly significant *p*-value in the case of SVM reflecting a strong discrimination ability.

correlations and *Sample noise* together is that there is typically no such thing as 'the model': many candidate models may often have similar predictive or correlative power. This issue was partially addressed elsewhere via model building based on bootstrap resampling (Deasy and El Naqa 2007, El Naqa *et al* 2006). Last, the variables may interact in a complicated, nonlinear way to determine, or at least correlate with, the endpoint ('*Variable interactions*'). In our experience, for example, it appears that the risk model for xerostomia is relatively effective, whereas outcome models for pneumonitis are less so (though still useful for high- and low-risk plans). Therefore, by extending the modeling process to a nonlinear framework, including higher-order variable interactions and local variable averaging, we hope to better capture real classification effects.

The kernel approach presented here automates the search for higher-order interactions between input variables that may be relevant to risk classification and outcomes in addition to balancing generalizabilty versus fidelity to the data. This is accomplished through an implicit

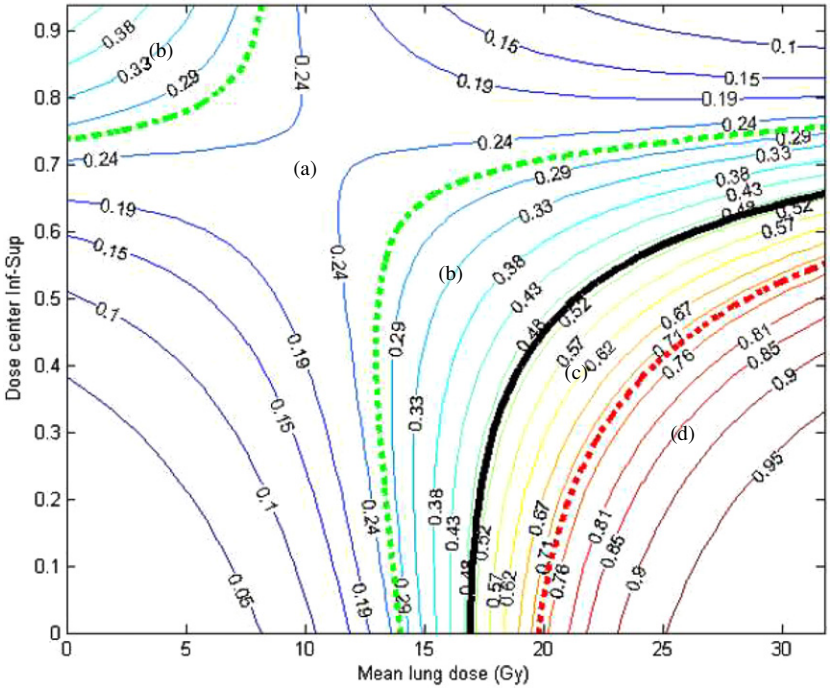

Figure 10. The kernel-based pneumonitis nonlinear prediction model ($f(x)$) as a function of mean dose and dose center in the superior–inferior direction. The model is obtained using the WUSTL data (training set) and was evaluated on the independent RTOG dataset (testing set). The plot shows four regions for risk prediction: (a) area of low-risk patients with high confidence prediction level, (b) area of low-risk patients with lower confidence prediction level, (c) area of high-risk patients with lower confidence prediction level and (d) area of high-risk patients with higher confidence prediction level. Note, patients within the 'margin' (cases (b) and (c)) represent intermediate-risk patients, which have intertwined characteristics that could fit belong to either group.

nonlinear mapping. The kernel/SVM approach maximizes the separation between events and non-events in feature space. Moreover, this approach has often been credited to yield better generalization with small datasets compared to the standard maximum likelihood approach (Schèolkopf and Smola 2002, Shawe-Taylor and Cristianini 2004).

Potential benefit from these methods can be predicted on the basis of PCA: if responses may be separated along a linear ridge in a PCA plot, then linear methods probably work well and nonlinear methods are unnecessary (cf figure 2). If there is no such linear ridge under PCA analysis, it is more likely that nonlinear features generated via kernel/SVM methods will improve the model prediction. For instance, no clear improvement was noticed in the case of linearly separable data (the xerostomia dataset) while using the kernel model versus traditional techniques. However, this should be evaluated on a case-by-case basis. For instance, if the two classes could be described by Gaussian distribution and have similar covariance matrices, a linear Bayesian classifier could be the optimal classifier.

In the current approach, variables were selected according to weights automatically generated by the kernel/SVM algorithm. There is a trade-off between the regularization parameter, C, and bias. When C is large, there are many non-zero variable terms, and the trend is toward overfitting the data, with increased noise in the predictions. When C is small,

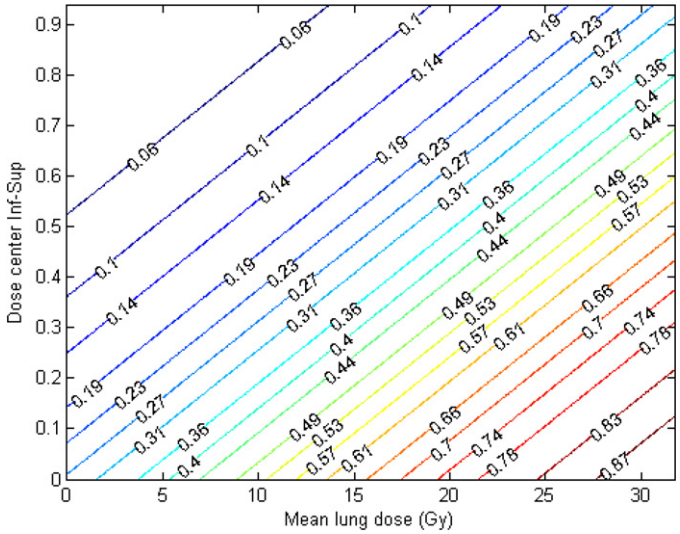

Figure 11. The logistic regression prediction model ($f(x)$) fitted to the WUSTL data as a function of mean dose and dose center in the superior–inferior direction.

many of the variables are dropped or reduced but the model may not capture real trends, thus introducing bias. We further introduced the concept of regularization decomposition, where risk groups could be weighted differently based on the available number of events. This approach seemed to improve the prediction power. Another advantage that is demonstrated on the independent multi-institutional dataset is the ability to perform well on an unseen data before, which is something not accounted for by other methods prone to over-fitting.

The plot in figure 10 could be used as a guideline in clinical practice for better prediction of pneumonitis risk based on this model. Estimates of the model could be used to stratify patients into different risk groups and therefore modulate treatment regimens accordingly. A favorable feature of this framework is that it highlights areas where the confidence level of prediction power is weak (inside the margin) versus areas of strong confidence level (outside the margin). Therefore, strengths and limitations of the model are provided as well to the analyst.

One of the main challenges of this framework is the selection of most relevant variables to include within the model. This is important clinically because it supports increased focus on potentially causative factors. Hence, future work will be dedicated to optimize the selection of the significant variables. As mentioned earlier, off-the-shelf techniques often fail to address the specificity of this application. Our previous selection method based on resampling and information theory seems to have produced better generalization results in comparison to SVM–RFE (El Naqa *et al* 2006). However, using nonlinear sensitivity analysis of kernel methods (Rakotomamonjy 2003) in conjunction with resampling techniques may provide better opportunities for robust selection of relevant variables.

Another issue that might have limited the predictive ability of the presented esophagitis or pneumonitis models ($Rs < 0.5$) is missing relevant variables that could be related to the genetics of the patient or underlying biology of the disease that are not currently captured in our existing clinical archives. To explore this issue, we are currently conducting a prospective pilot study in NSCLC patients that aims to build a comprehensive archive of clinical, physical

dosimetric variables and relevant biological markers. This could potentially compensate for the observed prediction gap (Spencer *et al* 2009). In this case, machine-learning algorithms are poised to play an increasingly important role in delineating such complex physical and biological interactions.

6. Conclusions

We have demonstrated an innovative approach for model exploration and building in radiotherapy based on nonlinear kernel-based statistical learning. The method was evaluated using resampling methods and validated on an independent dataset. The method efficiently and effectively handles high-dimensional space of potentially critical features. These methods are known to possess superior statistical power when learning from smaller sample sizes. For cases where nonlinear effects are important, this technique can significantly improve on the best results we achieved from the previous methods, by considering variable interactions and ability to generalize to unseen data. Future work will examine other aspects of nonlinear modeling for outcomes, such as incorporating prior information, adapting the kernel specifically to the expected response structure, this, as well as addressing the variable selection problem more comprehensively.

Acknowledgments

This work was supported in part by NIH grants K25 CA128809 and R01 CA85181.

References

Armstrong K, Weber B, Ubel P A, Peters N, Holmes J and Schwartz J S 2005 Individualized survival curves improve satisfaction with cancer risk management decisions in women with BRCA1/2 mutations *J. Clin. Oncol.* **23** 9319–28

Blanco A I, Chao K S, El Naqa I, Franklin G E, Zakarian K, Vicic M and Deasy J O 2005 Dose-volume modeling of salivary function in patients with head-and-neck cancer receiving radiotherapy *Int. J. Radiat. Oncol. Biol. Phys.* **62** 1055–69

Bradley J, Deasy J O, Bentzen S and El Naqa I 2004 Dosimetric correlates for acute esophagitis in patients treated with radiotherapy for lung carcinoma *Int. J. Radiat. Oncol. Biol. Phys.* **58** 1106–13

Bradley J, Graham M V, Winter K, Purdy J A, Komaki R, Roa W H, Ryu J K, Bosch W and Emami B 2005 Toxicity and outcome results of RTOG 9311: a phase I-II dose-escalation study using three-dimensional conformal radiotherapy in patients with inoperable non-small-cell lung carcinoma *Int. J. Radiat. Oncol. Biol. Phys.* **61** 318–28

Bradley J D, Hope A, El Naqa I, Apte A, Lindsay P E, Bosch W, Matthews J, Sause W, Graham M V and Deasy J O 2007 A nomogram to predict radiation pneumonitis, derived from a combined analysis of RTOG 9311 and institutional data *Int. J. Radiat. Oncol. Biol. Phys.* **69** 985–92

Brahme A 1999 Optimized radiation therapy based on radiobiological objectives *Semin. Radiat. Oncol.* **9** 35–47

Chao K S, Deasy J O, Markman J, Haynie J, Perez C A, Purdy J A and Low D A 2001 A prospective study of salivary function sparing in patients with head-and-neck cancers receiving intensity-modulated or three-dimensional radiation therapy: initial results *Int. J. Radiat. Oncol. Biol. Phys.* **49** 907–16

Chen Y, Hyrien O, Williams J, Okunieff P, Smudzin T and Rubin P 2005 Interleukin (IL)-1 A and IL-6: applications to the predictive diagnostic testing of radiation pneumonitis *Int. J. Radiat. Oncol. Biol. Phys.* **62** 260–6

Dawson L A, Biersack M, Lockwood G, Eisbruch A, Lawrence T S and Ten Haken R K 2005 Use of principal component analysis to evaluate the partial organ tolerance of normal tissues to radiation *Int. J. Radiat. Oncol. Biol. Phys.* **62** 829–37

Deasy J O, Blanco A I and Clark V H 2003 CERR: a computational environment for radiotherapy research *Med. Phys.* **30** 979–85

Deasy J O and El Naqa I 2007 *Radiation Oncology Advances* ed M Mehta and S Bentzen (New York: Springer)

Deasy J O, Niemierko A, Herbert D, Yan D, Jackson A, Ten Haken R K, Langer M and Sapareto S 2002 Methodological issues in radiation dose-volume outcome analyses: summary of a joint AAPM/NIH workshop *Med. Phys.* **29** 2109–27

de Crevoisier R, Tucker S L, Dong L, Mohan R, Cheung R, Cox J D and Kuban D A 2005 Increased risk of biochemical and local failure in patients with distended rectum on the planning CT for prostate cancer radiotherapy *Int. J. Radiat. Oncol. Biol. Phys.* **62** 965–73

El Naqa I, Bradley J and Deasy J 2005 *AAPM Symp. Proc. on Physical, Chemical and Biological Targeting in Radiation Oncology* ed M Mehta *et al* (Madison, WI: Medical Physics Publishing) pp 150–9

El Naqa I, Bradley J D and Deasy J O 2008 Nonlinear kernel-based approaches for predicting normal tissue toxicities *Machine Learning and Applications, 2008, ICMLA '08. 7th Int. Conf. on* pp 539–44

El Naqa I, Bradley J D, Lindsay P E, Blanco A I, Vicic M, Hope A J and Deasy J O 2006 Multi-variable modeling of radiotherapy outcomes including dose-volume and clinical factors *Int. J. Radiat. Oncol. Biol. Phys.* **64** 1275–86

El-Naqa I, Yang Y, Galatsanos N P, Nishikawa R M and Wernick M N 2004 A similarity learning approach to content-based image retrieval: application to digital mammography *IEEE Trans. Med. Imaging* **23** 1233–44

El-Naqa I, Yang Y, Wernick M N, Galatsanos N P and Nishikawa R M 2002 A support vector machine approach for detection of microcalcifications *IEEE Trans. Med. Imaging* **21** 1552–63

Elshaikh M, Ljungman M, Ten Haken R and Lichter A S 2006 Advances in radiation oncology *Annu. Rev. Med.* **57** 19–31

Good P I 2006 *Resampling Methods : A Practical Guide To Data Analysis* (Boston, MA: Birkhèauser)

Gulliford S L, Webb S, Rowbottom C G, Corne D W and Dearnaley D P 2004 Use of artificial neural networks to predict biological outcomes for patients receiving radical radiotherapy of the prostate *Radiother. Oncol.* **71** 3–12

Guyon I and Elissee A 2003 An Introduction to variable and feature selection *J. Mach. Learn. Res.* **3** 1157–82

Guyon I, Weston J, Barnhill S and Vapnik V 2002 Gene selection for cancer classification using support vector machines *Mach. Learn.* **46** 389–422

Härdle W and Simar L 2003 *Applied Multivariate Statistical Analysis* (Berlin: Springer)

Hastie T, Tibshirani R and Friedman J H 2001 *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (New York: Springer)

Haykin S 1999 *Neural Networks: A Comprehensive Foundation* (Englewood Cliffs, NJ: Prentice Hall)

Hope A J, Lindsay P E, El Naqa I, Bradley J D, Alaly J, Vicic M, Purdy J A and Deasy J O 2006 Radiation pneumonitis risk based on clinical, dosimetric, and location related factors *Int. J. Radiat. Oncol. Biol. Phys.* **65** 112–24

Hope A J, Lindsay P E, El Naqa I, Bradley J D, Vicic M and Deasy J O 2005 Clinical, dosimetric, and location-related factors to predict local control in non-small cell lung cancer *Astro 47th Ann. Meeting (Denver, CO)* p S231

Kennedy R, Lee Y, Van Roy B, Reed C D and Lippman R P 1998 *Solving Data Mining Problems Through Pattern Recognition* (Englewood Cliffs, NJ: Prentice Hall)

Levegrun S, Jackson A, Zelefsky M J, Skwarchuk M W, Venkatraman E S, Schlegel W, Fuks Z, Leibel S A and Ling C C 2001 Fitting tumor control probability models to biopsy outcome after three-dimensional conformal radiation therapy of prostate cancer: Pitfalls in deducing radiobiologic parameters for tumors from clinical data *Int. J. Radiat. Oncol. Biol. Phys.* **51** 1064–80

Marks L B 2002 Dosimetric predictors of radiation-induced lung injury *Int. J. Radiat. Oncol. Biol. Phys.* **54** 313–6

Matthews B W 1975 Comparison of the predicted and observed secondary structure of T4 phage lysozyme *Biochim. Biophys. Acta* **405** 442–51

Moiseenko V, Kron T and Van Dyk J 2004 Biologically-based treatment plan optimization: a systematic comparison of NTCP models for tomotherapy treatment plans *Proc. 14th Int. Conf. on the Use of Computers in Radiation Therapy (Seoul, Korea)*

Munley M T, Lo J Y, Sibley G S, Bentel G C, Anscher M S and Marks L B 1999 A neural network to predict symptomatic lung injury *Phys. Med. Biol.* **44** 2241–9

Pollack A, Cowen D, Troncoso P, Zagars G K, von Eschenbach A C, Meistrich M L and McDonnell T 2003 Molecular markers of outcome after radiotherapy in patients with prostate carcinoma: Ki-67, bcl-2, bax and bcl-x *Cancer* **97** 1630–8

Rakotomamonjy A 2003 Variable selection using SVM-based criteria *J. Mach. Learn. Res.* **3** 1357–70

Schèolkopf B and Smola A J 2002 *Learning With Kernels: Support Vector Machines, Regularization, Optimization, and Beyond* (Cambridge, MA: MIT Press)

Shawe-Taylor J and Cristianini N 2004 *Kernel Methods for Pattern Analysis* (Cambridge: Cambridge University Press)

Specht D F 1991 A general regression neural network *IEEE Trans. Neural Netw.* **2** 568–76

Spencer S J, Bonnin D Almiron, Deasy J O, Bradley J D and El Naqa I 2009 Bioinformatics methods for learning radiation-induced lung inflammation from heterogeneous retrospective and prospective data *J. Biomed. Biotechnol.* **2009** 892863

Su M, Miften M, Whiddon C, Sun X, Light K and Marks L 2005 An artificial neural network for predicting the incidence of radiation pneumonitis *Med. Phys.* **32** 318–25

Vapnik V 1998 *Statistical Learning Theory* (New York: Wiley)

Webb S 1997 *The Physics of Conformal Radiotherapy: Advances in Technology* (Bristol: Institute of Physics Publishing)

Weinstein M C, Toy E L, Sandberg E A, Neumann P J, Evans J S, Kuntz K M, Graham J D and Hammitt J K 2001 Modeling for health care and other policy decisions: uses, roles, and validity *Value Health* **4** 348–61