

Learning the Relationship between Patient Geometry and Beam Intensity in Breast Intensity-Modulated Radiotherapy

Renzhi Lu¹, Richard J. Radke^{1*}, Linda Hong², Chen-Shou Chui²,
Jianping Xiong², Ellen Yorke², and Andrew Jackson²

¹Electrical, Computer, and Systems Engineering Department,
Rensselaer Polytechnic Institute, Troy, NY 12180

²Department of Medical Physics, Memorial Sloan-Kettering Cancer Center,
New York, NY 10021

lur@rpi.edu, rjradke@ecse.rpi.edu, hongl@mskcc.org, chuic@mskcc.org
xiongj@mskcc.org, yorkee@mskcc.org, jacksona@mskcc.org

* Please address correspondence to Richard Radke.

Abstract

Intensity modulated radiotherapy (IMRT) has become an effective tool for cancer treatment with radiation. However, even expert radiation planners still need to spend a substantial amount of time adjusting IMRT optimization parameters in order to get a clinically acceptable plan. We demonstrate that the relationship between patient geometry and radiation intensity distributions can be automatically inferred using a variety of machine learning techniques in the case of two-field breast IMRT. Our experiments show that given a small number of human-expert-generated clinically acceptable plans, the machine learning predictions produce equally acceptable plans in a matter of seconds. The machine learning approach has the potential for greater benefits in sites where the IMRT planning process is more challenging or tedious.

I. INTRODUCTION

An important recent advance in cancer treatment with radiation is the introduction of a new technology known as *intensity-modulated radiotherapy (IMRT)* [23], [25]. This computer-controlled method of delivering radiation can precisely irradiate a target with complex 3-D shape while simultaneously protecting normal radiation-sensitive tissues. This is accomplished by delivering nonuniform patterns of radiation to the tumor area along beams oriented at several different angles, using a device called a multileaf collimator (MLC) attached to a linear accelerator. Doctors and radiation physicists use the shapes of sensitive tissues (e.g. obtained from computed tomography (CT) scans) to determine a set of beam intensities that meets the clinical objectives.

The IMRT problem is generally solved by dividing each radiation beam into subcomponent pencil beams (or “beamlets”), and applying numerical optimization algorithms to determine the beamlet intensities I^* such that the resultant radiation dose distribution $D(I^*)$ best matches the requirements specified by the physician. The clinical objectives of planning are encapsulated by an objective function F , which assigns a numerical score to each plan. Mathematically,

$$I^* = \arg \min_I F(D(I)). \quad (1)$$

A basic difficulty is the formulation of F . IMRT planners typically have to make compromises between competing clinical objectives, e.g., delivering as high and as uniform a dose as possible to the planning target volume (PTV), while sparing organs at risk (OARs) and normal tissues as much as possible. However, the compromise desired in any given case is not easy to specify in terms of the parameters (e.g., weights, dose constraints) defining the objective function F . In current implementations of IMRT,

prior knowledge of these parameters is not available, and planners can spend a substantial amount of time adjusting parameters in order to get a clinically acceptable plan (which is often a subjective decision). The bottleneck is not so much the efficiency of the numerical optimization procedures themselves as the manual trial-and-error process of adjusting parameters in the objective function. Circumventing or at least minimizing this procedure would save many person-hours of effort.

This paper evaluates several machine learning algorithms for determining a clinically acceptable set of field intensities for two-field breast IMRT directly from the patient's geometry and the plan objectives, thus entirely circumventing the use of the objective function F and dose calculation D in (1). The goal is to “learn by example”, since the breast is a case where clinical objectives are standard and many patients have been treated. The basic framework of our approach is shown in Figure 1. We treat the IMRT process as an unknown system, and build an input-output model to simulate the same system through machine learning. The raw input for each patient consists of (1) CT imagery of the breast to be treated, and (2) several parameters pertaining to radiation delivery, described below. The output consists of the radiation profile (e.g. a 256 x 256 matrix) for each of the two beams used to treat the patient. The learning process requires less than a minute of offline training using a small set of previously planned patients, and the prediction process takes a matter of seconds. We show that the resulting field intensities differ from an expert's plan for the same patient by less than 3%, and verify that the machine-predicted plans are equally clinically acceptable.

This result indicates a new, potentially fruitful approach to IMRT planning. While a breast IMRT plan might currently take a human planner half an hour or more to complete, a prostate or head and neck IMRT plan typically takes several hours and many back-and-forth objective function adjustments, even for an expert. If machine learning can reveal the underlying relationships between patient geometry and field intensities in more challenging cases (by either directly predicting a clinically acceptable plan, or providing a better initial guess for the optimization than the planner could), planning time could be greatly reduced.

We note that since the machine learning algorithms are trained on clinically acceptable plans generated by an expert planner, we should not expect them to produce plans that are better than the expert's. This is not a shortcoming of the process, but the goal of the technique. We assume that the expert's plans are generated according to the clinical protocol deemed by the treatment center to be most effective at treating a given cancer. If the underlying protocol of the training plans changes, the machine learning prediction will change with it. This approach also has the benefit of producing planner- or treatment-center-specific predictions. For example, one treatment center's protocol may be more conservative than another's in

planning dose to a certain region. The machine learning system will produce correspondingly different plans for the same input geometry depending on which center's training plans were used.

In related work, Hunt et al. [19] summarized the influence of optimization parameters on the dose distribution, and specified a procedure for changes to be made in these parameters given specific deficits in clinical plans. However, this method only suggested the direction of change, and was actually a heuristic implemented by hand. Similarly, Barbieri et al. [4] searched for the best optimization parameters via structured grid searches, using a recipe based on observations about which parameters are likely to affect the solution. This method only applies to planning for prostate radiotherapy. Xing et al. [35] suggested an iterative algorithm to determine the optimal weights automatically; this method requires prior knowledge such as the optimal dose-volume histogram.

Efforts have recently been made towards multiobjective (MO) optimization, first introduced in [13], which results in a representative set of Pareto efficient solutions for various conflicting optimization objectives. Neither the trial-and-error process described above nor prior knowledge about the parameters is required, but planners still need to select the "best" solution from the available set. Bortfeld et al. [5] found the Pareto-efficient space of solutions for MO optimization, and developed an interface for the planner to examine the way the tradeoffs change when the optimization parameters are altered. Other recent approaches are described in [27], [36]. One drawback to the MO approach is speed, since it is very time-consuming to search for all possible solutions. Furthermore, given a set of solutions, the problem of choosing the best one from a clinical and patient-specific perspective still remains [3]. A decision-making process that incorporates clinical reasoning was introduced in [24], but it only works when the trade-offs between different plans are very large.

The above optimization methods all suffer from the very large number of degrees of freedom to be determined. Alber et al. [1] studied the eigenvalues of the second order Hessian matrix of the IMRT cost function, which is determined by patient geometry and the objectives of the optimization problem. They found that the number of degrees of freedom that make a difference to the objective function (i.e. dominant modes) is much smaller than the total number of beamlet intensities. For fixed patient geometry and beam settings, these dominant modes achieve sensible clinical goals such as moving intensity edges in regions where target and sensitive normal tissues overlap, or sharpening field gradients at the edge of the target. That study analyzed only one patient under different beam settings; the problem of obtaining dominant modes for a significant population of different patients using such a method may be difficult. Furthermore, while such an approach reveals a qualitative relationship between the intensity profile and patient's geometry, it does not easily allow for quantitative prediction.

The paper is organized as follows. In the next section, we review the clinical practice of breast IMRT at our partner institution, Memorial Sloan-Kettering Cancer Center (MSKCC). In Section III, we describe our approach to feature selection, which reduces the dimensionality of the problem and makes it computationally tractable. In Section IV, we describe how these features are extracted from real image data. In Section V, we introduce several global and local machine learning algorithms that we apply to the problem. Section VI presents results and discussion, and Section VII concludes with ideas for future work.

II. IMRT FOR THE BREAST

Breast cancer is the most common female cancer in the United States; according to the American Cancer Society, over 200,000 new cases of invasive breast cancer are projected for 2005. External beam radiation therapy following breast-conserving surgery (lumpectomy) has been demonstrated to be an effective treatment for early-stage breast cancer, providing both good cosmesis and local control equivalent to mastectomy [26]. It is estimated that breast patients account for 25-30% of the patients at a given clinic [34]. The goal of breast radiation therapy is to give a uniform dose to the affected breast while protecting such adjacent normal structures as the lung, heart and contralateral breast. While considerable normal organ protection is provided by the geometry of the “tangential” beam arrangement shown in Figure 2, dose uniformity in the breast is more difficult to achieve. However, uniformity is important, since the tumor may recur in low dose regions (cold spots) while high dose regions (hot spots) may cause local fibrotic complications.

Intensity modulated tangential beams have been used for whole breast treatment since 1999. IMRT has been found to improve dose homogeneity within the breast volume, to reduce the dose to the contralateral breast and ipsilateral lung, and to reduce heart dose for patients treated to the left breast [18], [20], [29]. In a 2004 national survey by the American Association of Physicists in Medicine, 40% of the responding clinics reported using IMRT for breast cancer treatments [2]. At Memorial Sloan-Kettering Cancer Center (MSKCC), an IMRT planning method referred to as “simplified IMRT”, or sIMRT, was implemented clinically in July 2001 [8], [11], [17]. This method requires fewer user-drawn anatomical contours and optimization constraints than full-blown IMRT, but provides equivalent dose homogeneity and contralateral breast protection. It has been used at MSKCC to treat more than 750 patients in either the supine [11] or prone [17] position; over 269 of these were treated in 2004. Although sIMRT greatly reduces the work of the treatment planner, variability in the patient’s anatomy and the user’s skill level often results in several optimization trials to obtain the clinical plan. For the training set used in this

study, the average number of trial optimizations was between 3 and 4, with a range from 1-9. Each run requires manual adjustment of objective function parameters and visual evaluation of the corresponding dose distribution. Below, we review the protocol used at MSKCC, emphasizing that it is unknown to the machine learning algorithms described subsequently. A good medical physics reference that goes into more detail on some of the terms below is [21].

The interaction of the beam and the patient is represented in the Beam’s Eye View (BEV) coordinate system, which allows the clinician to look down the radiation beam and see the target from the perspective of the beam source. The major step in sIMRT determines the intensity value for each beamlet through a direct calculation, followed by back-and-forth adjustments for certain parameters. The breast is treated with a pair of oppositely directed beams from a medical linear accelerator (linac). One beam enters from the medial side and one from the lateral, as shown in Figure 2. The beams encompass the entire breast and provide an additional margin beyond the patient’s surface (called the “skin flash”) to allow for breathing and other types of patient motion. The primary clinical goal is to deliver a uniform dose to the entire breast. As described in [9], in an intensity modulated field with a given intensity distribution $I(u, v)$, the dose to a point whose BEV coordinates are (x, y) and whose equivalent depth¹ is d is:

$$D(x, y, d) = TMR(d, W \times H) \times pOCR(x, y, d) \times \left(\frac{SAD}{f} \right)^2 \times \left[\frac{\int \int \int I(u, v) K(x - u, y - v, d - w) dudvdw}{\int \int \int K(x - u, y - v, d - w) dudvdw} \right]. \quad (2)$$

Here, W and H are the radiation field width and height, respectively. TMR is the tissue maximum ratio function, and $pOCR$ is the primary off-center ratio function, which are determined from look-up tables [21]. SAD and f are the distance from the radiation source to the isocenter² (here fixed to 1000 mm), and the distance from the radiation source to the point (x, y, d) , respectively. The term in brackets accounts for the effects of beamlet scattering interactions, described by a kernel K .³

From (2), the dose D and intensity I are clearly correlated, but the beamlet convolution implies that the dose to any point is influenced by the whole distribution I . In performing the optimization for sIMRT,

¹The equivalent depth is the aggregate “distance” the pencil beam has traveled, weighted for the approximate electron densities in the tissue it passed through; see Section IV.

²The isocenter is the point of intersection of the axis around which the radiation source rotates and the axis of rotation of the beam’s collimator. As part of normal clinical procedure, the patient is positioned so that isocenter is inside the treated breast (thus inside the PTV). In our study, the isocenter is mapped to (0,0) in BEV coordinates.

³TMR and pOCR also depend on the beam energy and the particular linac used for treatment (see Section VI). A patient will always have both tangent beams treated on the same linac and almost always with the same energy.

several further simplifications are made. First, for the points along each pencil beam, it is assumed that $D(x, y, d)$ is only influenced by the corresponding pencil beam intensity $I(x, y)$ (i.e. the scattering effect of neighboring pencil beams is neglected), which simplifies the term in brackets to $I(x, y)$. Second, instead of solving for the intensity $I(x, y)$ so that every point (x, y, d) has a uniform dose $D_{prescribe}$, only certain points located at the intersection of each pencil beam and a 2D surface roughly bisecting the breast are considered (see Figure 3). To determine the surface, lines are drawn parallel to the posterior edge of the beam on each transverse CT slice. The midpoints of the line segments that intersect the breast volume form the 2D surface. When the patient is treated with two tangential beams from the medial and lateral tangents, the posterior edges of the fields are coplanar, which leads to the desirable property that both beams share the same set of midpoints. Accordingly, each pencil beam should deliver half of the prescribed uniform dose to the midpoint:

$$I(x, y) = \frac{D_{prescribe}}{2 \cdot TMR(d, W \times H) \cdot pOCR(x, y, d) \cdot \left(\frac{SAD}{f}\right)^2} \quad (3)$$

Here, d and f take their values at the midpoint corresponding to the BEV coordinate (x, y) . After the intensity distribution has been calculated according to (3), a full dose calculation is performed which includes the previously omitted scattered radiation, and the resulting dose distribution is evaluated. The final intensity output $I(x, y)$ used for the clinical sIMRT beams typically differs somewhat from (3) due to subsequent back-and-forth parameter adjustment resulting from evaluation of the full dose distribution and several clinical rules of thumb used at MSKCC [8]. One rule requires that the intensity in the region near the edge of the field increase by approximately 5% to compensate for the penumbra effect (i.e. that the edge of an irradiated volume receives less than the full dose due to lack of scattering contributions from blocked radiation). Another rule requires adjustment so that the dose to the tip of the breast (the apex dose) is between 102% and 105%.

III. FEATURE SELECTION AND DIMENSIONALITY REDUCTION

Since the input for each patient is composed of about 100 512 x 512 CT images, and the output intensity profile for each beam is a 256 x 256 matrix, the dimensions of both the input and output are extremely high. At the moment, the number of clinical training samples we have access to for this study is limited to the order of tens or hundreds. If the raw CT voxels were directly used as a high-dimensional feature vector, it would be very difficult to use so few training samples to capture the major variations in the input/output data, or to predict the output for a new input. Hence, we require special steps to reduce the dimensionality of both the input and output.

One common technique to reduce the dimensionality of data is principal component analysis (PCA) [15], which captures the dominant modes of variation of a dataset. A new data sample can be projected onto the subspace spanned by the first several principal modes, resulting in a low-dimensional feature vector. However, applying PCA to three-dimensional volumes from different patients is a difficult problem, since it generally requires finding corresponding points between different datasets. While several solutions have been presented in the computer vision literature (e.g. [16], [31]), this correspondence problem is time-consuming to solve, and may have no “right answer”. It is similarly difficult to directly compare field intensity matrices from patients with different geometries and clinical objectives.

Our approach here is to convert the 3D treatment volume into appropriate 2D maps in the BEV coordinate system, and reduce the dimensionality by independently considering the input and output features for each beamlet. In this case, even if we possess a relatively small number of patients with which to train the algorithm, we actually have a large number (i.e. more than 500 per patient) of beamlets—more than enough for the purposes of machine learning. Hence, we divide the plan for each training patient into pencil beams. All the pencil beams from different patients are trained together to learn the relationship between the input features and the output feature. Prediction of the intensity profiles for an unknown patient is accomplished by merging the predicted pencil beams into an intensity matrix.

The key to an effective learning algorithm is an appropriate feature space, such that the input features are representative and sufficient to capture all the factors that affect the intensity result. From (3), the output (i.e. intensity) for each pencil beam is related to TMR , $pOCR$ and the source-to-point distance f , all defined for the corresponding midpoint in the treatment volume. Furthermore, TMR is a function of the depth d and the equivalent square field size $L = \frac{2W \times H}{W+H}$, while $pOCR$ is a function of d and the radial distance from center $r = \sqrt{x^2 + y^2}$ [21]. Hence, we choose d, L, r and f as input features for each pencil beam. To allow the algorithm to learn the clinical rules of thumb that do not obey (3), we also include the distances to the field edge in the x and y directions (e_x, e_y), which are related to the penumbra effect, and the distance to the skin flash region s , which is related to the apex adjustment. Since the penumbra effect/apex adjustment only takes effect near the field edge/skin respectively, we keep these features constant elsewhere. Specifically, we set s to be the minimum value of the actual distance and $4mm$, and $e_{x/y}$ to be the minimum value of the actual distance and $6mm$. Thus, the learning problem is to estimate a one-dimensional output (intensity) from each seven-dimensional input $(d, L, r, f, e_x, e_y, s)$. This is a substantial dimensionality reduction over a non-pencil-beam method that would treat the entire CT image as input and the entire intensity field as output. We describe how these features are extracted in the following section.

As far as our clinical input data is concerned, there are two additional special issues. One is that in each plan, the values at several rows of the intensity matrix have already been averaged for radiation delivery with a multileaf collimator, since the pencil beams are typically calculated on a finer grid than the leaf width. This averaging means that the general rule relating input to output doesn't hold for certain positions, which needs to be taken into account in both training and prediction (see Section VI). The other issue is that the intensity profiles have been normalized for each beam pair, so that the maximum intensity in the plan is 1000. This normalization factor varies among patients. To be able to directly compare the intensities from different plans in machine learning and prediction, we use the known normalization factors to obtain unnormalized absolute intensity values. The final predicted output is then normalized at the end of the process, so that it can be compared to other plans for the same patient. We note that the training and testing patients in our study have a wide range of ages and breast sizes (see Section VI) characteristic of the general population of breast patients treated at MSKCC.

IV. FEATURE EXTRACTION

Among the seven input features, the BEV depth d and BEV midpoint distance f are of great importance. We extract these from each raw CT volume using the following steps.

- 1) Read the CT data from the file. To avoid processing every voxel in the patient's body contour, we determine the treatment area based on the isocenter, field width/height, gantry angle and couch angle (which can all be obtained from the accompanying plan file). The treatment area for each CT slice is bounded by the breast contour and its intersection with the posterior edge of the beam.
- 2) Convert the Hounsfield Units (HU) in each image into the corresponding relative electron density (ED). The transform is linear in three stages, according to [30].

$$ED = \begin{cases} 1.0 + HU/1000 & HU \leq 0 \\ 1.0 + 0.0881/1000 \times HU & 0 < HU \leq 100 \\ 1.088 + (HU - 100) \times 0.9973/1800 & HU > 100 \end{cases} \quad (4)$$

- 3) Build 3D models of the breast and pencil beams [32]. A grid of pencil beams is created with a spacing of 2 mm by 2 mm measured at the isocenter distance. Each pencil beam is drawn from the radiation source to one grid point in the BEV plane within the field range (see Figure 4).
- 4) Trace along each pencil beam to determine its intersection with the midpoint surface described above. The midpoint is defined as the sample point that evenly bisects a line segment through the point parallel to the posterior beam edge (see Figure 3). The BEV depth, or equivalent depth d , for the midpoint is obtained by summing the incremental path length multiplied by the relative

electron density of each voxel the pencil beam passes through within the treatment volume. For those midpoints in the build-up region near the apex (i.e. $0 < d < d_{max}$), d is set to be d_{max} .⁴ For the pencil beams that do not intersect the breast, we set d to zero. The BEV depth maps for the medial beam and the lateral beam of one patient are illustrated in Figure 5.

- 5) Similarly, obtain the BEV midpoint distance map, defined as the distance from the source to the pencil beam midpoint determined above. For the pencil beams that do not intersect the breast, we set the BEV midpoint distance to zero. The BEV distance maps for the medial beam and the lateral beam of one patient are illustrated in Figure 6. Most of the values are nearly, but not exactly equal to, 1000 mm, which means that the treatment volume is not symmetric with respect to the plane perpendicular to the beam central axis, and that the midpoints do not lie on a single plane.

Figure 7 shows the output intensity distributions for the patient in Figures 5 and 6. Looking at these three figures, the relationship between the intensity and the input features is unclear. We can improve the situation by noting that only a small rectangular region in the 256x256 matrix contains intensity values greater than zero, which corresponds to the field region. We only consider this valid area for training and learning purposes. Furthermore, within the valid rectangular region, in some cases corners of the field are blocked to avoid irradiation of the arms or other regions outside the breast. Finally, for the pencil beams that do not intersect the treatment volume (which can be determined from the BEV midpoint distance map), the intensity is set to be nearly constant. This forms the “skin flash” region, which is also excluded from training since there is no patient geometry along the pencil beam to correlate with the intensity. Neglecting the blocking area and skin flash region results in a zoomed-in version of the output as shown in Figure 8.

As we can see from Figures 5, 6 and 8, the intensity distributions are visually correlated with the patient’s geometry, in the sense that the 2D shape of the BEV distance map resembles the 2D shape of the intensity, with deviations in intensity related to the BEV depth map. For example, in the area overlaying the projection of the lung volume onto the BEV plane (the left side of Figure 5a), the intensities are slightly reduced to compensate for the small equivalent depth (since lung is less dense than breast tissue).

In conclusion, we know that a relationship exists between the intensity and geometry constraints that is approximated by (3), and selected our input features accordingly. In the next stage, we present several machine learning algorithms to quantitatively discover this relationship.

⁴ d_{max} is the depth that receives the maximum dose. Generally, as the depth d increases, the effect of radiation reaches its maximum at d_{max} , and then attenuates.

V. MACHINE LEARNING ALGORITHMS

The methods we investigated to learn the relationship between the input and output are all based on regression analysis. Patterns inferred from the training data are used to evaluate the function at previously unobserved input points. The assumption is that both the training data and test data arise from the same underlying statistical distribution.

There are two general types of solutions for regression [7]. The first is global estimation, in which the same estimated model function can be applied to any point in the input space. The parametric form of the function is specified, and the parameters that minimize a certain global cost function are estimated. The second type of regression is based on local risk minimization. The value of the function at a new point is estimated based only on the values of the output for nearby input points from the training set. There are tradeoffs in both types of methods. A global method is able to abstract from data when the underlying function is smooth and the functional form is easy to hypothesize. Prediction for new inputs is nearly instantaneous. However, updating the model is expensive, since new data must be incorporated into the model by re-computing all the parameters. Local methods are nonparametric techniques without any assumptions on the functional form, which make them more flexible. Newly observed data can be easily incorporated into the model without expensive re-computation. The disadvantage is that all of the training data must be retained to evaluate the function at new points, and the search procedure required for prediction is usually time-consuming. In the following, we describe two global and two local approaches that we took to modeling the geometry/intensity relationship in breast IMRT.

A. Global Approaches

1) *Parametric Regression:* Since we already know a rough functional form for the input-output relationship, a natural approach is nonlinear parametric regression [6]. Certain terms in (3) can be modeled based on empirical observations; for example, according to [28],

$$TMR(d, L) \approx (Ade^{-Bd})(1 - e^{-kL}) + Ee^{-Fd}, \quad (5)$$

where d, L refer to the equivalent depth and equivalent square field size respectively, and A, B, k, E, F are unknown parameters. However, it is unnecessary and impractical to use this form directly in the nonlinear model for intensity prediction, since the number of unknown parameters is high, and the TMR term is only one part of the model. Instead, based on (3), we assume that all terms, (i.e., TMR , $pOCR$, the inverse-square factor, penumbra compensation and apex adjustments) appear as multiplicative factors in the model. Furthermore, from (5), we hypothesize that the equivalent depth d and equivalent square field

size L contribute exponentially to the model. Finally, from [8], [10], we assume that the other five features (i.e., source-to-point distance f , radial distance from center r , distance from edge e_x , e_y and distance from skin-flash area s) appear in the model as power terms. Applying a logarithmic transformation, we get the following simplified equation:

$$\begin{aligned} \log(I) = & \beta_0 + \beta_1 d + \beta_2 L + \beta_3 \log(f) + \beta_4 \log(1 + r) + \\ & \beta_5 \log(1 + e_x) + \beta_6 \log(1 + e_y) + \beta_7 \log(1 + s) \end{aligned} \quad (6)$$

The coefficients β_0, \dots, β_7 can be estimated using a least-squares fit. However, we note that from the point of view of evaluating algorithms that can generalize to other sites, parametric regression is less desirable than an algorithm that assumes nothing about the functional form of the input-output relationship. We explore three such methods below.

2) *Support Vector Regression*: Support vector regression (SVR) [33] is a powerful tool for nonlinear function approximation in high-dimensional input spaces. SVR requires no prior knowledge of a parametric functional form, and is generally more robust to noise than least-squares regression.

Given a training set of N samples (\mathbf{x}_i, y_i) with a p -dimensional input vector \mathbf{x}_i and scalar output y_i , we want to approximate the unknown nonlinear transfer function with precision ϵ . The nonlinearity is achieved by mapping the input space X into a new feature space F , and then estimating the relationship in F by linear regression. The functional form in SVR is:

$$y = f(\mathbf{x}) = \langle \Phi(\mathbf{x}), \boldsymbol{\beta} \rangle + \gamma, \quad (7)$$

where $\langle \cdot, \cdot \rangle$ is an inner product on F , $\gamma, \boldsymbol{\beta}$ are regression coefficients, and $\Phi(\cdot) : X \rightarrow F$ denotes a nonlinear mapping represented implicitly by

$$\langle \Phi(\mathbf{x}), \Phi(\mathbf{x}_i) \rangle = K(\mathbf{x}, \mathbf{x}_i), \quad (8)$$

where K is a known kernel. The SVR solution only requires us to be able to compute inner products with Φ , not the knowledge of Φ itself. In our case, we use a Gaussian function centered at \mathbf{x}_i of width σ :

$$K(\mathbf{x}, \mathbf{x}_i) = \exp\left(-\|\mathbf{x} - \mathbf{x}_i\|^2 / 2\sigma^2\right). \quad (9)$$

We want to approximately solve the linear regression problem (7); the solution is approximate in the sense that small errors (i.e. those less than ϵ) are considered to be negligible and set to 0. The number of free parameters in the function approximation scheme is equal to the number of support vectors, which can be obtained by defining the width of the tolerance band ϵ . Hence, the number of free parameters is

directly related to the approximation accuracy and does not depend on the dimensionality of the input space.

The tolerance bandwidth ϵ and the width of the Gaussian kernel σ must be determined in advance; however, determining the best set of parameters is not a solved problem and can be computationally expensive. In our experiments, we used SVMTorch [12], a tool designed for large scale regression, and chose these parameters through cross-validation.

B. Local Approaches

Since there are thousands of pencil beams available for training, there is enough data to make a local regression approach feasible. That is, for a new feature vector, it is likely that we have seen similar feature vectors in the training set, and can predict the output based on these nearest neighbors.

1) *K-nearest Neighbors*: The k -nearest neighbor rule [15] for classifying an unknown input is to assign it the output label associated with the majority of the k nearest training samples in the input space. Since the output is a continuous variable in our regression problem, the predicted output is some statistic (e.g. the mean or median value) of the outputs of the k nearest training samples. In our results below, we used $k = 5$ and the mean value.

2) *Barycentric Interpolation*: Instead of using a simple average of the nearest neighbors (neglecting how close the testing point may be to one of the training samples), barycentric interpolation uses a weighted average of the neighbors' values to make a more reasonable estimate. This approach assumes that the feature points have been triangulated, e.g. using the well-known Delaunay triangulation [14]. For the input point where we want to estimate the function, we find the enclosing Delaunay triangle (if the feature space is d -dimensional, then $d + 1$ points form the "triangle"). The input feature vector is then written in barycentric coordinates:

$$x = \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_{d+1} x_{d+1}, \quad (10)$$

where $0 \leq \alpha_i \leq 1$ and $\sum_{i=1}^{d+1} \alpha_i = 1$. After finding the α_i , we apply the same linear combination to the output:

$$\hat{y} = \alpha_1 y_1 + \alpha_2 y_2 + \dots + \alpha_{d+1} y_{d+1}. \quad (11)$$

The barycentric method may fail for some testing points, since an enclosing triangle for a new sample point may not always exist. We can simply use the nearest-neighbor method for these points.

VI. EXPERIMENTAL RESULTS

A. Experimental Design

We obtained clinical sMRT plans (lateral and medial tangential beams) for 22 breast cancer patients from MSKCC, all planned by the same experienced expert. These samples correspond to all breast cancer patients treated in the supine position with sMRT at MSKCC over a three-month period in 2004. The patients were divided into three groups based on machine type and beam energy. Group 1 consisted of 11 patients treated with 6MV X-rays on a Varian Clinac 2100EX linear accelerator, Group 2 consisted of 8 patients treated with 6MV X-rays on a Varian Clinac 600C linear accelerator, and Group 3 consisted of 3 patients treated with 15MV X-rays on a Varian Clinac 2100EX linear accelerator. The ages and breast sizes (defined as the breast width at the posterior border of the tangent field) of the patients are shown in Figure 9, and are characteristic of the general population of breast patients treated at MSKCC. In each case, from the planning CT volumes, beam settings and intensity profiles, both the input and output features defined in Section III were extracted for all the pencil beams. Before machine learning, we scaled each input feature of the training data to lie in $[-1, 1]$, to ensure that features with larger values do not unduly affect the solution.

We note that the machine type and beam energy affect the output intensity via the TMR and $pOCR$ terms [9]. One possible approach is to include the beam energy E and machine type T as two additional input features. However, due to the small amount of data and the fact that each additional variable has only two discrete values in the experiments, we found it preferable to sort the patients into 3 groups according to the beam energy and machine type, and learn in each group separately. Again, we emphasize that while there are a relatively small number of patients in each group, the number of training *beamlets* available for machine learning is in the thousands and thus sufficiently large.

Since the intensity profiles we used for training were already averaged for delivery by a multileaf collimator with a 1 cm leaf width, we downsampled the image and field intensity data by 5 in the X direction of Figure 8 to ensure that consistent data was supplied to the learning procedures. Predicted field intensities can be compared to the original plans by extending the downsampled prediction to the neighboring 4 grid lines, as shown in Figure 10.

We designed three sets of experiments to test the performance of the learning algorithms:

- 1) An “intra-patient” experiment, in which training and testing are limited to one patient at a time. The intensity of each pencil beam is predicted using a model trained by only the other pencil beams for the same patient. The motivation is to confirm that the intensities within one patient are indeed

consistent and predictable, and to demonstrate that the model learned for one patient cannot be applied to another, so the subsequent experiments are necessary.

- 2) An “inter-patient” experiment, in which the training samples from different patients are gathered together. For each testing case, the training data set is composed of all the other patients in the same group. The motivation is to confirm the main hypothesis of the paper: that the intensities for a new patient can be well predicted based only on other examples of the input/output relationship.
- 3) Another “inter-patient” experiment using a smaller set of training data. For Group 1 (6MV, 2100EX), we used 5 patients for training, and 6 patients for testing. For Group 2 (6MV, 600C), we used 4 patients for training and 4 patients for testing. For Group 3 (15MV, 2100EX), we report the results of 3 leave-one-out tests, each time using 2 patients for training and the remaining 1 patient for testing. The motivation is to determine whether results similar to the previous experiment can be obtained with a smaller representative set of training data (which makes the local methods run much faster and indicates that our amount of training data is sufficient).

In each experiment, the predicted result is compared to ground truth (i.e. the expert’s sIMRT plan). We evaluated two error measures based on the entire intensity distribution for each patient:

$$MPE = \frac{1}{N} \sum_{i=1}^N \frac{(I_i^{predicted} - I_i^{true})}{I_i^{true}} \times 100\% \quad (12)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \frac{|I_i^{predicted} - I_i^{true}|}{I_i^{true}} \times 100\% \quad (13)$$

Here, I_i^{true} is the true intensity value for the i th pencil beam, $I_i^{predicted}$ is the predicted output value for the i th pencil beam, and N is the number of pencil beams for the patient being evaluated. The mean percent error (MPE) can be positive or negative and indicates whether the prediction is biased from the ground truth. The mean absolute percent error (MAPE) is always positive and disallows cancellation of errors. These are standard measures for predictor performance [22]. Since, as discussed in Section III, the skin-flash and blocking regions were excluded from both training and prediction, errors in these regions do not contribute to (12)-(13). Each data point below represents the average between medial beam and lateral beam prediction for each patient. A sample comparison of the average times needed for feature selection, training, and prediction in the four algorithms is shown in Table I, which are all on the order of a few seconds.

| | PR method | SV method | 5-NN method | Barycentric method |
|-------------------|-----------|-----------|-------------|--------------------|
| Feature selection | 15 | 15 | 15 | 15 |
| Training | 3 | 5 | 6 | 10 |
| Prediction | < 1 | < 1 | 4 | 7 |

TABLE I

AVERAGE TIMES, IN SECONDS, FOR FEATURE SELECTION AND TRAINING (USING FIVE PLANS), AND PREDICTION OF ONE PLAN, IMPLEMENTED ON A PENTIUM 4 1.4 GHZ, 1024MB RAM PC.

B. Error analysis

Figure 11 illustrates the MAPE for the first (intra-patient) experiment using the two global methods and two local methods. The mean MAPE across all patients for parametric regression was 1.74%, while the mean MAPE for support vector regression was 1.46%. The mean MPE across all patients was 0.05% for parametric regression, and 0.02% for support vector regression, indicating virtually no bias. Thus, both global models fit the intra-patient data very well, which means that the intensity is quite predictable from the patient's geometry. Similar results were obtained from the local methods. The mean MAPE was 1.78% for 5-nearest neighbor regression and 1.69% for barycentric regression, and the mean MPE was 0.06% and 0.02%, respectively. We note that global methods are preferable in this case, since they give almost the same prediction error while requiring much less computation. However, the intra-patient model we learn is not particularly useful since it cannot be applied to another patient with different field size, machine type or beam energy. For example, using the parametric regression model learned from Patient 1 to predict the plan for the remaining patients in Group 1 results in an average MAPE of 8.9% and an average MPE of 3.1%. Using the local methods would lead to even higher error rates.

Figure 12 illustrates the MAPE for the second (inter-patient) experiment using the two global methods and two local methods, where all the other patients in the same group were used for training. The mean MAPE across all patients for parametric regression was 2.15%, while the mean MAPE for support vector regression was 2.07%. The mean MPE across all patients was 0.23% for parametric regression, and -0.16% for support vector regression. Table II breaks down the mean MAPE and MPE for each group and method. The errors are only slightly worse than in the intra-patient experiment, showing that the models are able to generalize to different patients and plans. We also note that Group 3 has relatively smaller prediction error, and that the field sizes for those 3 patients were almost the same.

For the other 2 groups, those testing patients whose field size is most different than the patients in the training set usually had the largest error. Thus, field size seems to be a major feature that accounts for the inter-patient difference. As for the local methods, the mean MAPE across all patients was 2.87% for 5-nearest neighbor regression and 2.56% for barycentric regression, while the mean MPE was 0.52% and 0.41%, respectively. Compared to Figure 11, the performance of the local methods is relatively worse than that of the global methods, indicating that the former are less capable of generalizing the inter-patient difference. The normalization procedure described in Section III is also an important factor in this experiment; without it, the errors increase by roughly a factor of 2. We also note that we did not observe any correlations between the error and the patient’s age or breast size.

| | | Experiment 2 | | | | Experiment 3 | | | |
|---------|---------|--------------|-------|-------|-------|--------------|-------|-------|-------|
| | | PR | SV | 5-NN | Bary | PR | SV | 5-NN | Bary |
| Group 1 | MAPE(%) | 2.25 | 2.14 | 3.00 | 2.68 | 2.31 | 2.52 | 2.86 | 2.60 |
| | MPE(%) | 0.24 | -0.16 | -0.28 | -0.12 | 0.11 | 0.86 | -0.30 | -0.22 |
| Group 2 | MAPE(%) | 2.22 | 2.07 | 2.92 | 2.68 | 2.35 | 2.25 | 2.92 | 2.77 |
| | MPE (%) | 0.27 | -0.14 | 0.38 | 0.27 | 0.68 | -0.80 | 0.56 | 0.33 |
| Group 3 | MAPE(%) | 1.59 | 1.80 | 2.12 | 1.72 | 1.59 | 1.80 | 2.12 | 1.72 |
| | MPE (%) | -0.25 | 0.29 | 0.24 | 0.13 | -0.25 | 0.29 | 0.24 | 0.13 |

TABLE II

PREDICTION RESULTS FOR THE FOUR METHODS IN EXPERIMENTS 2 AND 3. THE LEARNING ALGORITHMS ARE: PR (PARAMETRIC REGRESSION), SV (SUPPORT VECTOR REGRESSION), 5-NN (FIVE NEAREST NEIGHBOR INTERPOLATION), AND BARY (BARYCENTRIC INTERPOLATION).

The third (inter-patient) experiment uses a smaller set of training data. The training cases were deliberately selected to cover the range of field sizes. Figure 13 illustrates the MAPE for the 13 testing patients using each method, and Table II breaks down the mean MAPE and MPE for each group and method. The overall prediction results are slightly worse than, but quite comparable to, those in the first inter-patient experiment, which means that training based on fewer samples that are representative of the expected variation is feasible. Among the global methods, support vector regression has the larger drop in prediction performance. We suspect that this method is more sensitive to the number of different field sizes in the training set. Among the local methods, barycentric interpolation based on Delaunay triangulation is superior to that based on nearest neighbors. The overall performance of global methods

outweighs that of the local methods in Group 1 and Group 2, but not in Group 3, where the field sizes for training and testing are nearly the same. In light of the results from Figures 11 and 12, it seems that the field sizes of the testing data and the training data must be nearly identical for local methods to perform as well as a global method.

C. Comparison of dose distributions

Another natural way to evaluate the algorithms is to analyze the actual dose distribution imposed by the predicted intensity profiles, since the goal of sIMRT is to achieve uniform dose distributions within the breast. We did this evaluation for the two global methods in the second experiment, to determine if the numerical errors produced significant clinical differences. Each plan is normalized to give 100% dose at a point on the lung/chest-wall interface.

Table III reports the performance of parametric regression and support vector regression in the second experiment with respect to the generated dose distributions, giving the minimum, maximum, and mean dose to the planning target volume (PTV)⁵ for each plan, averaged over all patients. We also report the mean D95 (minimum dose to hottest 95% of the PTV), D05 (minimum dose to hottest 5% of the PTV) and V95 (volume receiving 95% of prescribed dose) values, which are commonly used in medical physics to evaluate the quality of a plan [23]. The differences of the D95, D05, and V95 values between each predicted plan and the expert’s plan were judged to be of no practical significance by the planner. Figure 14 compares the expert, parametric regression, and support vector regression dose distributions on the transverse, coronal, and sagittal slices for one typical patient. MSKCC planners agreed that the dose distributions of the machine-learning predictions matched very well with the expert plans and were clinically acceptable.

VII. DISCUSSION AND CONCLUSIONS

The machine learning approach presents a promising direction for clinical breast IMRT planning, for which practical application on a large scale has proven challenging [34]. While the clinical sIMRT technique requires 3 to 4 repeat optimizations on average for each patient, our approach needs only one initial training phase, and can predict each new plan in a few seconds. There is no need to re-train the model unless the treatment protocol of the institute changes. The problem of high dimensionality in the

⁵See [18] for a formal definition of the PTV in this case- essentially all the palpable breast tissue plus a 2 cm margin posterior, sup and inf.

| | sIMRT ground truth | PR method 2 | SV method 2 | PR method 3 | SV method 3 |
|--------------|--------------------|-------------|-------------|-------------|-------------|
| max. dose(%) | 108.5±2.3 | 108.1±2.6 | 108.7±3.2 | 108.3±2.3 | 108.7±3.0 |
| min. dose(%) | 81.4 ±7.2 | 81.5 ±6.9 | 81.8 ±6.9 | 81.8 ±6.8 | 82.0 ±6.7 |
| mean dose(%) | 101.5±1.2 | 101.6±1.2 | 101.8±2.3 | 101.5±2.3 | 101.5±2.3 |
| D_{05} (%) | 97.1±1.6 | 97.0±1.3 | 97.1±1.7 | 97.3±1.3 | 97.5±1.6 |
| D_{95} (%) | 104.6±1.9 | 104.9±2.0 | 105.3±3.0 | 105.2±1.8 | 105.2±2.7 |
| V_{95} (%) | 97.5±2.0 | 97.7±1.6 | 97.6±1.9 | 97.9±1.5 | 97.9±1.8 |

TABLE III

SUMMARY OF DOSE DISTRIBUTION RESULTS FOR ALL PATIENTS (MEAN \pm STANDARD DEVIATION). THE LEARNING ALGORITHMS ARE: PR 2 (PARAMETRIC REGRESSION, LEAVE-ONE-OUT), SV 2 (SUPPORT VECTOR REGRESSION, LEAVE-ONE-OUT), PR 3 (PARAMETRIC REGRESSION USING FEWER TRAINING SAMPLES), AND SV 3 (SUPPORT VECTOR REGRESSION USING FEWER TRAINING SAMPLES).

input and output was addressed by learning the intensity relationship on a beamlet basis. We were able to predict intensity profiles directly from the patient’s geometry to within a few percent of an expert’s plan in a matter of seconds. The numerical differences from the expert’s plan had a negligible effect on the quality of the resulting dose distributions. Thus, we can improve the efficiency of breast planning, while preserving the dose homogeneity and critical structure protection already achieved by the clinical protocol. This approach also makes it possible, given appropriate training data, to predict different plans for the same patient according to different planners’ subjective criteria.

In the breast, parametric regression performs well in most cases, since the functional form is roughly pre-determined by (3). However, due to their nonparametric nature, we suggest that support vector regression and local regression methods are promising for more complex learning tasks in IMRT, where parametric functional relationships between the input and output are far from obvious. Support vector regression seems to be a good choice when the available training datasets are representative and plentiful enough. If the variation in training plans is limited (e.g. only a few field sizes or machine types), local regression may be the better choice. It seems feasible to group a large number of existing plans into several training databases, each of which has the same beam energy/machine type and similar field sizes, and apply local regression to predict an unknown plan using the most suitable training set.

The next step in our research program is to extend the machine learning approaches described here to prostate IMRT, which is much more complicated and difficult to optimize. The number of repeat

optimizations for prostate IMRT plans is typically between 5 and 30, and the optimizations themselves are more time-consuming. At MSKCC, five equally-spaced beams are frequently used for radiation, and interaction between beams must be explicitly considered in planning. The input features we use for machine learning must change to reflect the multiple-beam configuration and the more complex objective function. One approach to dimensionality reduction in this case is to predict a set of parameters defining the IMRT objective function that will result in an acceptable set of beam intensities when the function is optimized. While the manual planning/validation process may not (and should not) be entirely circumvented, the resulting parameters/plans may serve as better initial estimates for IMRT, thus reducing the number of trial optimizations and shortening the overall planning time.

One shortcoming of machine learning approaches in general is that there must be sufficient training data to enable accurate modeling and prediction. Despite the limited number of patients in the experiment, we were able to obtain good results due to the large number of training beamlets. As mentioned above, different mechanisms of dimensionality reduction need to be explored for other sites to make machine learning feasible. Our initial experiments with the prostate site using a similar study size are promising. We also note that machine learning prediction will be unreliable for patients that violate the modeling assumption, e.g. the patient size/shape is grossly atypical, or the clinical compromise between target/normal tissues changes. However, such patients are probably best planned manually to begin with. Our primary concern is to reduce the tedium of planning a “typical” patient, without compromising the quality of the treatment.

ACKNOWLEDGMENTS

This work was supported by the National Cancer Institute under grant 5P01CA59017-13, and CenSSIS, the NSF Center for Subsurface Sensing and Imaging Systems, under the award EEC-9986821.

REFERENCES

- [1] M. Alber, G. Meedt, and F. Nusslin. On the degeneracy of the IMRT optimization problem. *Medical Physics*, 29:2584–2589, 2002.
- [2] American Association of Physicists in Medicine. 2004 Salary Survey and Profile of Radiation Oncology Departments. <http://www.aapm.org/AAPMUtilities/download.asp?file=AAPM-Salary04.pdf>, 2004.
- [3] H. I. Amols and C. C. Ling. EUD but not QED. *International Journal of Radiation Oncology Biology Physics*, 52:1–2, 2002.
- [4] J. Barbieri, M. Chan, J. Mechalakos, D. Cann, K. Schupak, and C. Burman. A parameter optimization algorithm for intensity-modulated radiotherapy prostate treatment planning. *Journal of Applied Clinical Medical Physics*, 3:227–234, 2002.

- [5] T. Bortfeld, K.-H. Kufer, M. Monz, A. Trofimov, and A. Niemierko. Problems with current IMRT prescription practices and planning systems [abstract],. *Medical Physics*, 31:1761, 2004.
- [6] S. Chatterjee. *Regression Analysis by Examples*. Wiley Series, 2000.
- [7] V. Cherkassky and F. Mulier. *Learning from Data*. Wiley-Interscience, 1998.
- [8] C.-S. Chui, L. Hong, and M. Hunt. A simplified intensity modulated radiation therapy technique for the breast. *Medical Physics*, 29:522–529, 2002.
- [9] C.-S. Chui, T. LoSasso, and S. Spirou. Dose calculation for photon beams with intensity modulation generated by dynamic jaw of multileaf collimators. *Medical Physics*, 21:1231–1244, 1994.
- [10] C.-S. Chui and R. Mohan. Off-center ratios for three-dimensional dose calculations. *Medical Physics*, 13:409–412, 1986.
- [11] C.-S. Chui, L. Hong and B. McCormick. Intensity-modulated radiotherapy technique for three-field breast treatment. *International Journal of Radiation Oncology Biology Physics*, 62: 1217–1223, 2005.
- [12] R. Collobert and S. Bengio. SVM Torch: Support vector machines for large-scale regression problems. *Journal of Machine Learning Research*, pages 143–160, 2001.
- [13] C. Cotrutz, M. Lahanas, C. Kappas, and D. Baltas. A multiobjective gradient based dose optimization algorithm for conformal radiotherapy. *Physics in Medicine and Biology*, 46:2161–2175, 2001.
- [14] M. de Berg, O. Schwarzkopf, M. van Kreveld, and M. Overmars. *Computational Geometry: Algorithms and Applications*. Springer-Verlag, 2000.
- [15] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley-Interscience, 2001.
- [16] D. Freedman, R. Radke, Y. Jeong, T. Zhang, D. M. Lovelock, and G.T.Y. Chen. Model-based segmentation of medical imagery by matching distributions. *IEEE Transactions on Medical Imaging*, 24(3):281–292, March 2005.
- [17] K. A. Goodman, L. Hong, R. Wagman, M. A. Hunt and B. McCormick. Dosimetric analysis of a simplified intensity modulation technique for prone breast radiotherapy. *Int. J. Radiat Oncol Biol Phys*, 60(1):95–102, September 2004.
- [18] L. Hong, M. Hunt, C.-S. Chui, S. Spirou, K. Forster, H. Lee, J. Yahalom, G.J. Kutcher, and B. McCormick. Intensity-modulated tangential beam irradiation of the intact breast. *Int. J. Radiat Oncol Biol Phys*, 44:1155–1164, 1999.
- [19] M. A. Hunt, C.-Y. Hsiung, S. V. Spirou, C.-S. Chui, H. I. Amols, and C. C. Ling. Evaluation of concave dose distributions created using an inverse planning system. *International Journal of Radiation Oncology Biology Physics*, 54:953–62, 2002.
- [20] L.L. Kestin, M.B. Sharpe, R.C. Frazier, F.A. Vicini, D. Yan, R.C. Matter, A.A. Martinez, and J.W. Wong. Intensity modulation to improve dose uniformity with tangential breast radiotherapy: initial clinical experience. *Int. J. Radiat Oncol Biol Phys*, 48:1559–1568, 2000.
- [21] F. M. Khan. *The Physics of Radiation Therapy*. Lippincott Williams and Wilkins, 2003.
- [22] M. Kutner, C. Nachtsheim, and J. Neter. *Applied Linear Regression Models*. McGraw-Hill, 2004.
- [23] C. C. Ling et al. *A Practical Guide to Intensity-Modulated Radiation Therapy*. Medical Physics Publishing, 2004.
- [24] J. Meyer, M. H. Phillips, P. S. Cho, I. Kalet, and J. N. Doctor. Application of influence diagrams to prostate intensity-modulated radiation therapy plan selection. *Physics in Medicine and Biology*, 49:1637–1653, 2004.
- [25] J. R. Palta and T. R. Mackie, editors. *Intensity-Modulated Radiation Therapy: The State of the Art*. Medical Physics Publishing, 2003.
- [26] C. A. Perez, M. E. Taylor, J. Bradley, D. Mansur, and M. M. Sanchez-Aragon. Breast: Stage T1 and T2 tumors. In *Principles and Practice of Radiation Oncology*, 4th edition, C. A. Perez, L. W. Brady, E. C. Halperin, and R. K. Schmidt-Ullrich, Lippincott Williams & Wilkins, Philadelphia, 2004.

- [27] H. E. Romeijn, J. F. Dempsey, and J. G. Li. A unifying framework for multi-criteria fluence map optimization models. *Physics in Medicine and Biology*, 49:1991–2013, 2004.
- [28] M. Schell. Empirical equation for tissue-maximum ratios/scatter maximum ratios for indirectly ionizing radiotherapy beams. *Medical Physics*, 6:65–67, 1979.
- [29] E.A. Strom. Breast IMRT: new tools leading to new vision. *Int. J. Radiat Oncol Biol Phys*, 54:1297–1298, 2002.
- [30] S. J. Thomas. Relative electron density calibrations of CT scanners for radiotherapy treatment planning. *The British Journal of Radiology*, 72:781–786, 1999.
- [31] A. Tsai, A. Yezzi, W. Wells, C. Tempany, D. Tucker, A. Fan, E. Grimson, and A. Willisky. A shape based approach to curve evolution for segmentation of medical imagery. *IEEE Trans. Medical Imaging*, 22(2), February 2003.
- [32] B. van Asselen, C. P. Raaijmakers, and P. Hofman. An improved breast irradiation technique using three-dimensional geometrical information and intensity modulation. *Radiotherapy and Oncology*, 58:341–347, 2001.
- [33] V. Vapnic. *The Nature of Statistical Learning Theory*. Springer, 1995.
- [34] F.A. Vicini, M. Sharpe, L. Kestin, A. Martinez, C.K. Mitchell, M.F. Wallace, R. Matter, and J. Wong. Optimizing breast cancer treatment efficacy with intensity-modulated radiotherapy. *International Journal of Radiation Oncology Biology Physics*, 54:1336-1344, 2002.
- [35] L. Xing, J. Li, S. Donaldson, Q. Le, and A. Boyer. Optimization of importance factors in inverse planning. *Physics in Medicine and Biology*, 44:2525–2536, 1999.
- [36] Y. Yu. Multiobjective decision theory for computational optimization in radiation therapy. *Medical Physics*, 24:1445–1454, 1997.

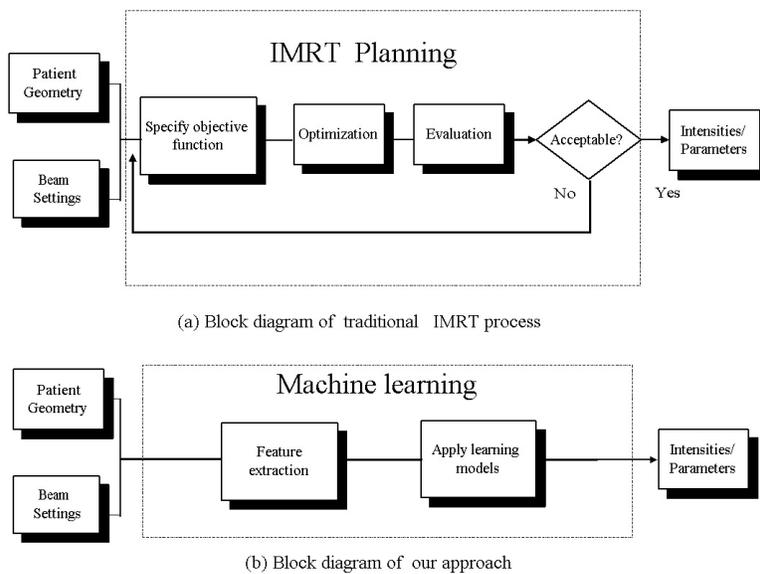


Fig. 1. The basic framework of the machine learning approach. The standard IMRT planning process is shown in (a). Our goal is to develop a machine learning algorithm (b) that can reproduce the results of the IMRT planning process based on many training examples of input/output pairs.

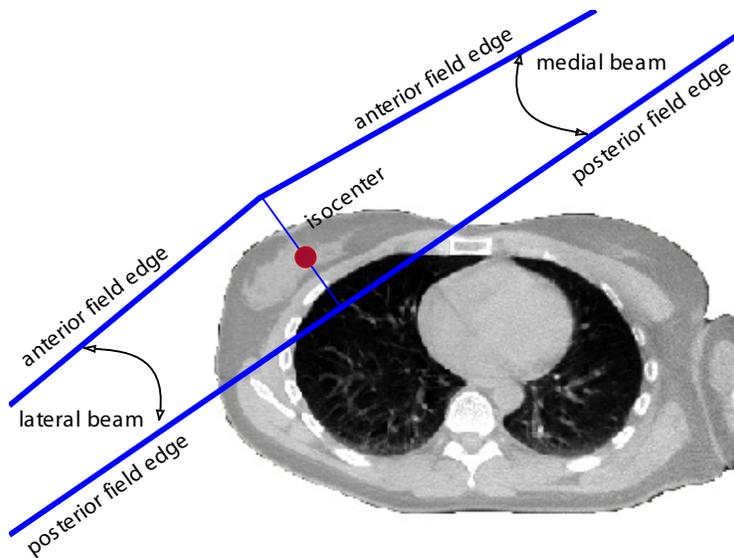


Fig. 2. In breast sIMRT, the breast is treated with a pair of oppositely directed beams from a medical linear accelerator. One beam enters from the medial side, one from the lateral.

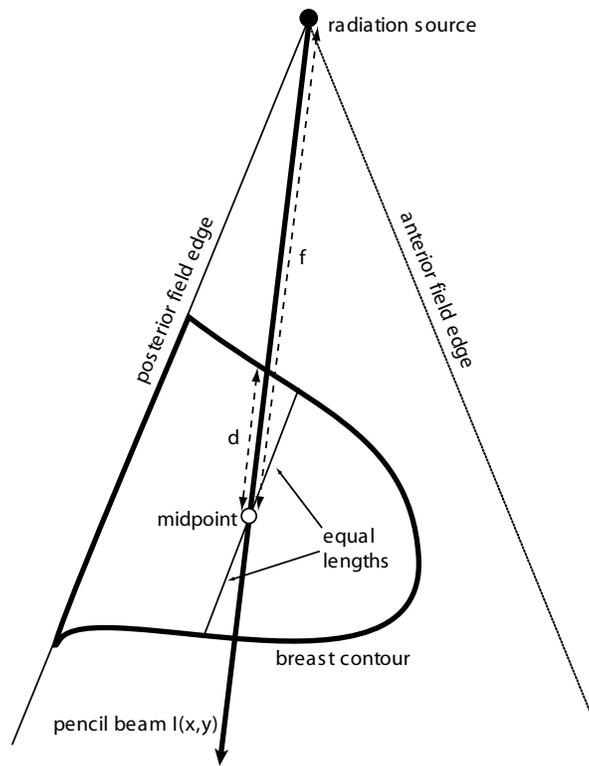


Fig. 3. The midpoint for each pencil beam is determined as the sample point that evenly bisects a line segment through the point parallel to the posterior beam edge.

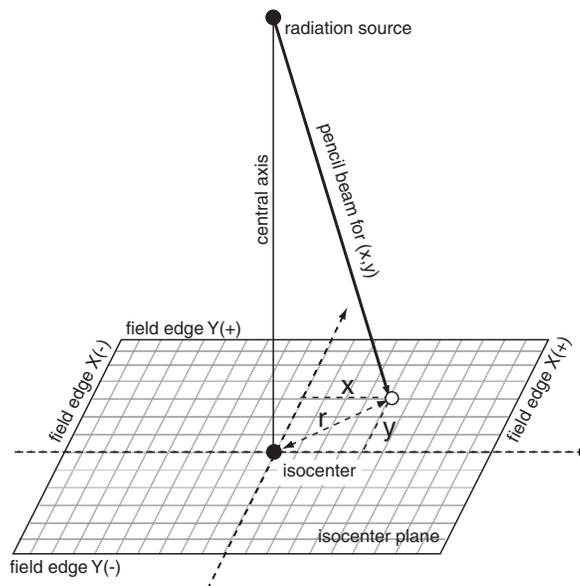


Fig. 4. Illustrates the relationship between the radiation source, pencil beams, isocenter, field edges and BEV plane.

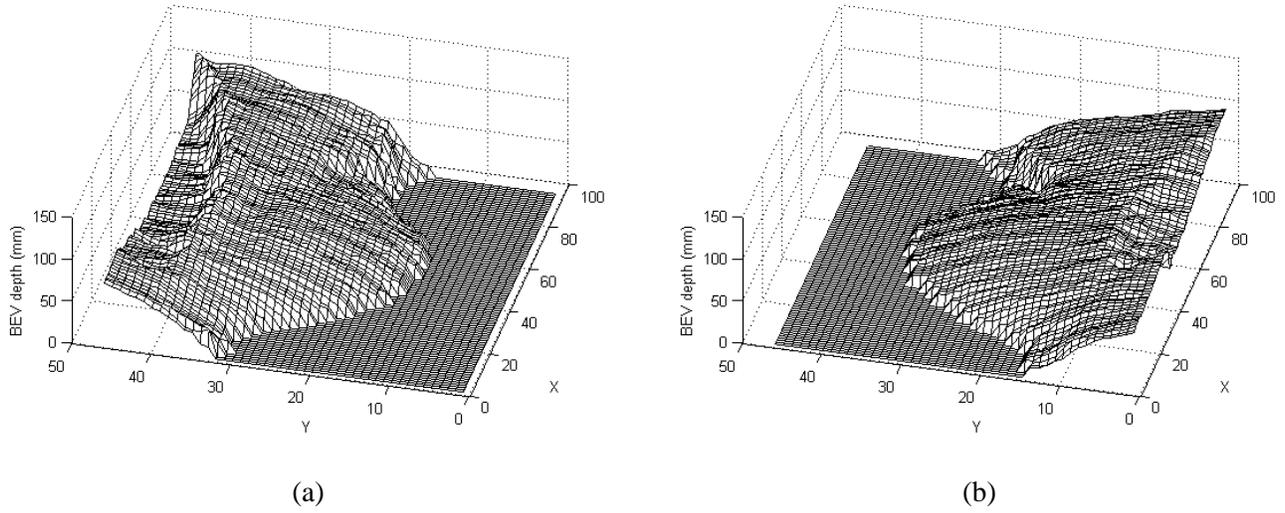


Fig. 5. BEV depth map for (left) medial beam and (right) lateral beam of Patient 7.

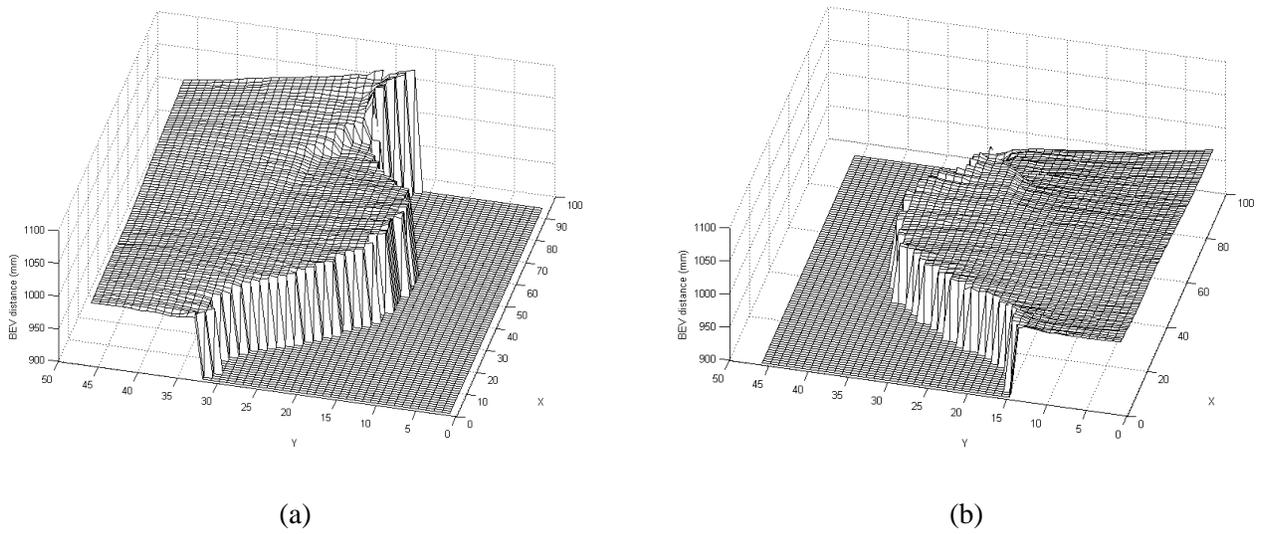


Fig. 6. BEV midpoint distance map for (left) medial beam and (right) lateral beam of Patient 7.

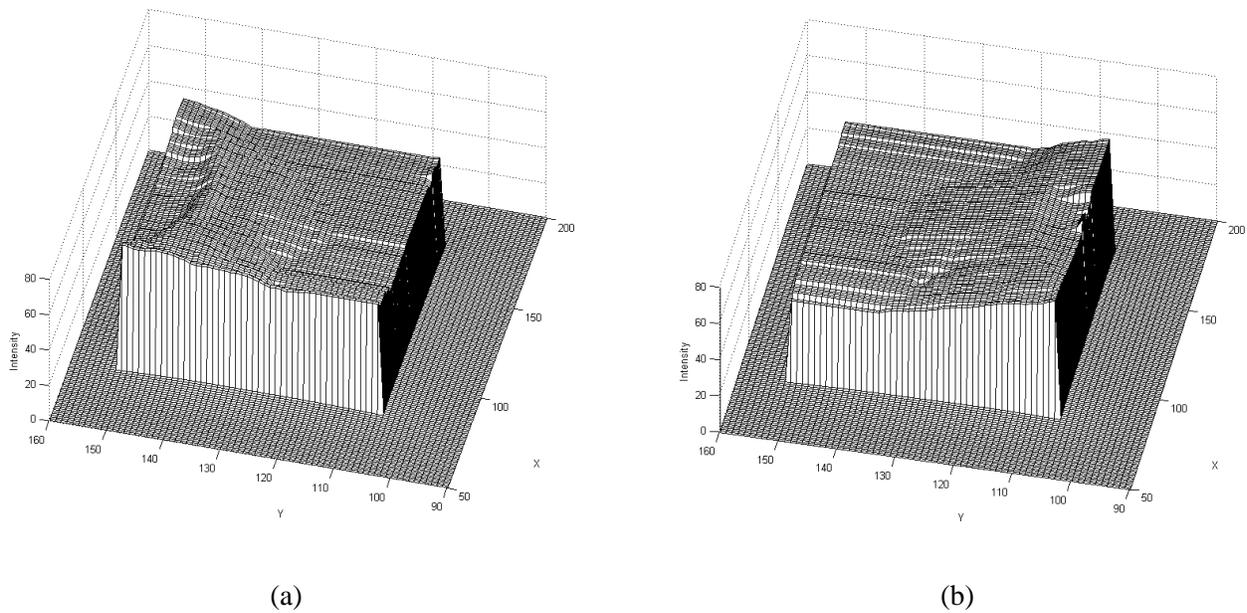


Fig. 7. Output intensity profiles for (left) medial beam and (right) lateral beam of Patient 7.

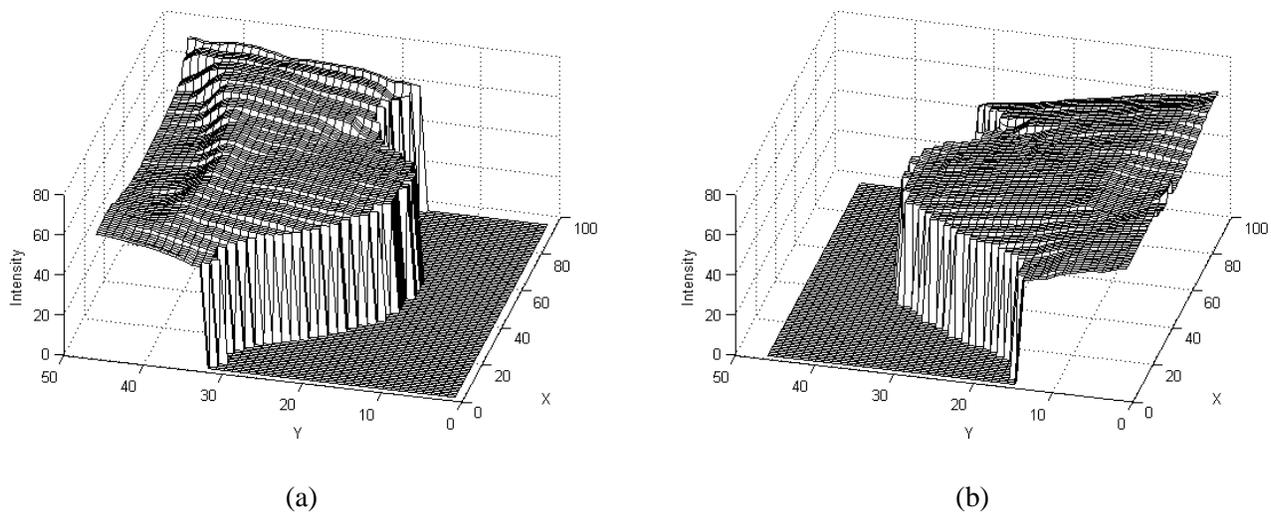


Fig. 8. Output intensity profiles for (left) medial beam and (right) lateral beam of Patient 7 after zooming and cropping. These are visually correlated with the BEV depth and distance maps in Figures 5 and 6.

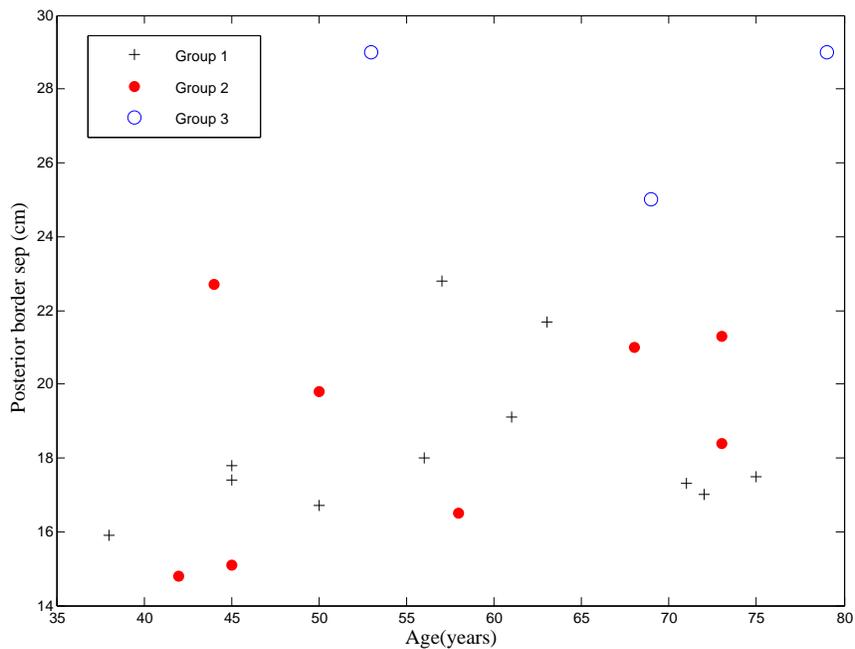


Fig. 9. Scatterplot of age (in years) and breast size (posterior border separation, in cm) of the patients in the experimental dataset. Patients in different machine learning groups are indicated with different symbols.

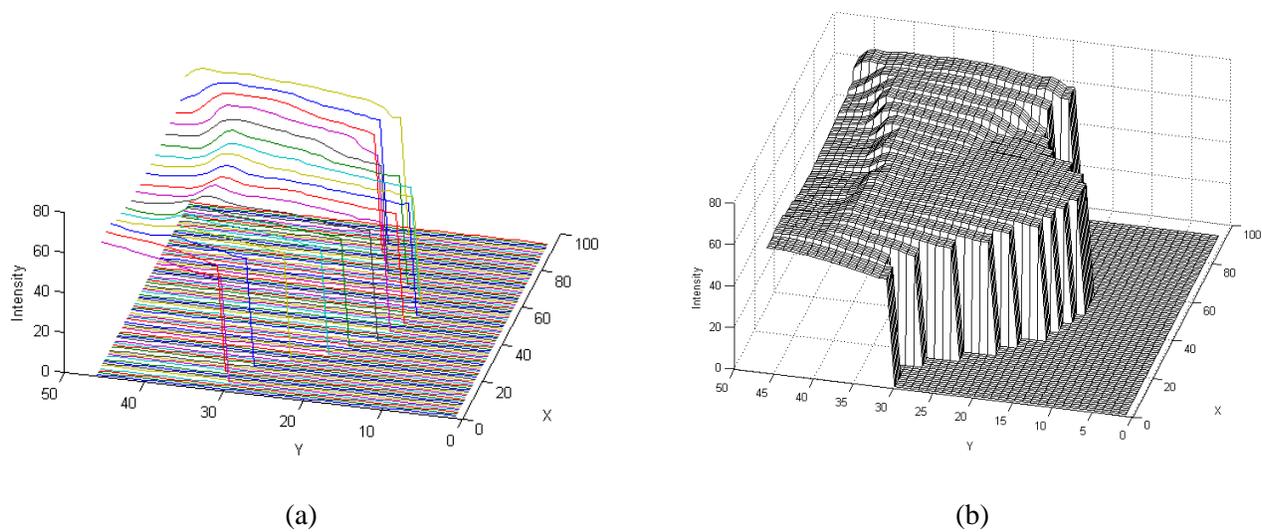


Fig. 10. Prediction for Patient 7. Left: Prediction for the intensities in the downsampled grid. Right: Intensity extension to produce full plan.

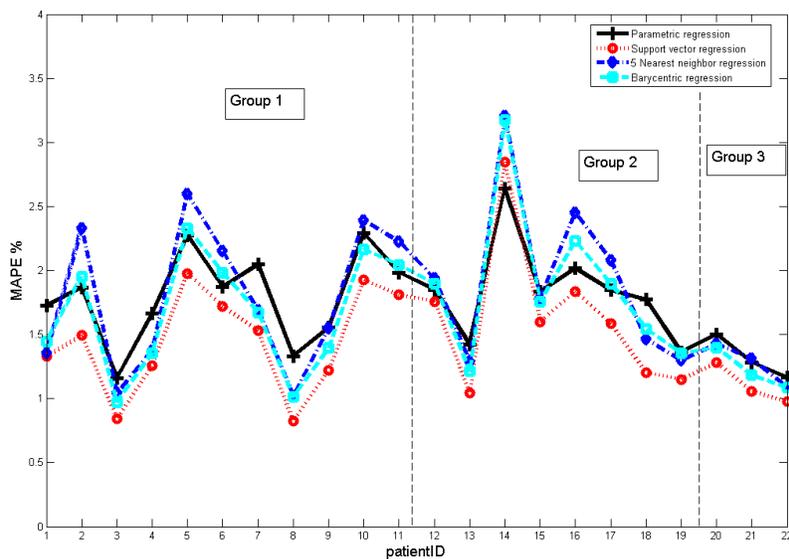


Fig. 11. MAPE for the intra-patient experiment. The solid and dotted lines are the parametric regression and support vector regression results, respectively. The dash-dot and dashed lines represent 5-nearest neighbor and barycentric interpolation, respectively.

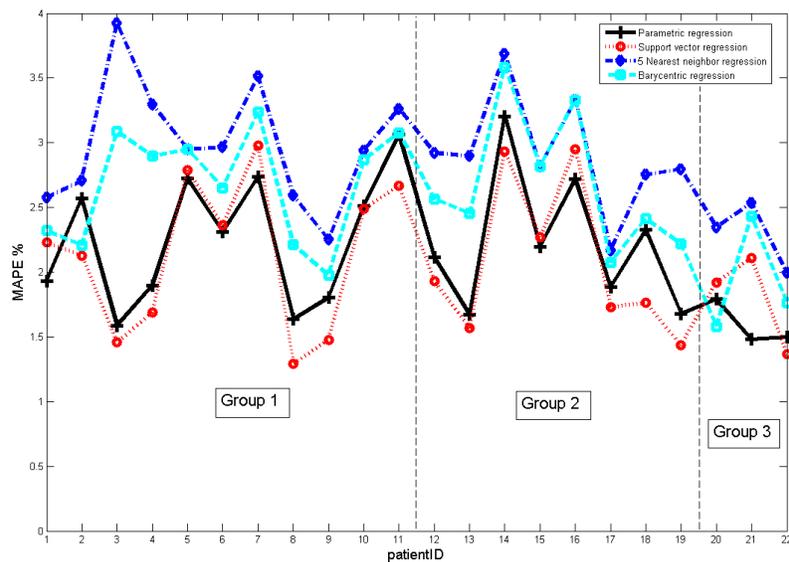


Fig. 12. MAPE for the first inter-patient experiment, where all the other patients with the same beam energy are used for training. The line style corresponding to each learning method is the same as in Figure 11.

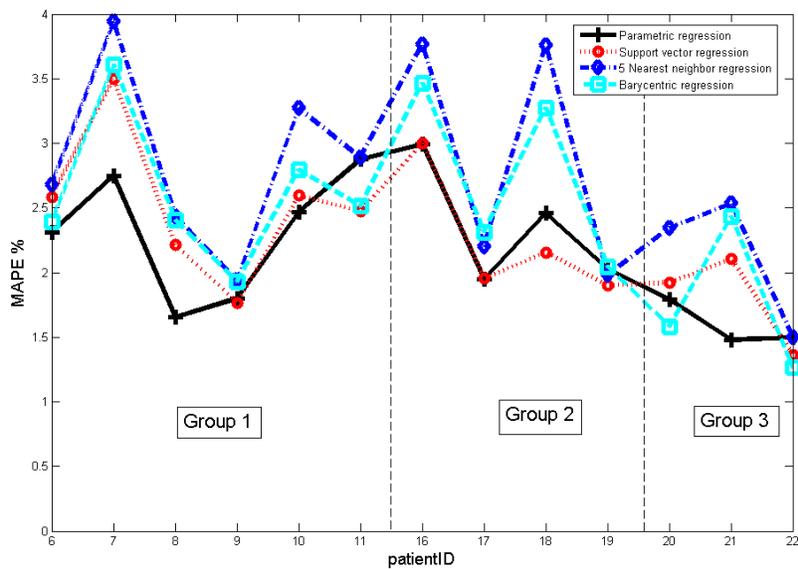


Fig. 13. MAPE for the second inter-patient experiment. In Group 1, patients 1 to 5 are used for training, and patients 6 to 11 for testing. In Group 2, patients 12 to 15 are used for training, and patients 16 to 19 for testing. In Group 3, training and testing are the same as in the first inter-patient experiment. The line style corresponding to each learning method is the same as in Figure 11.

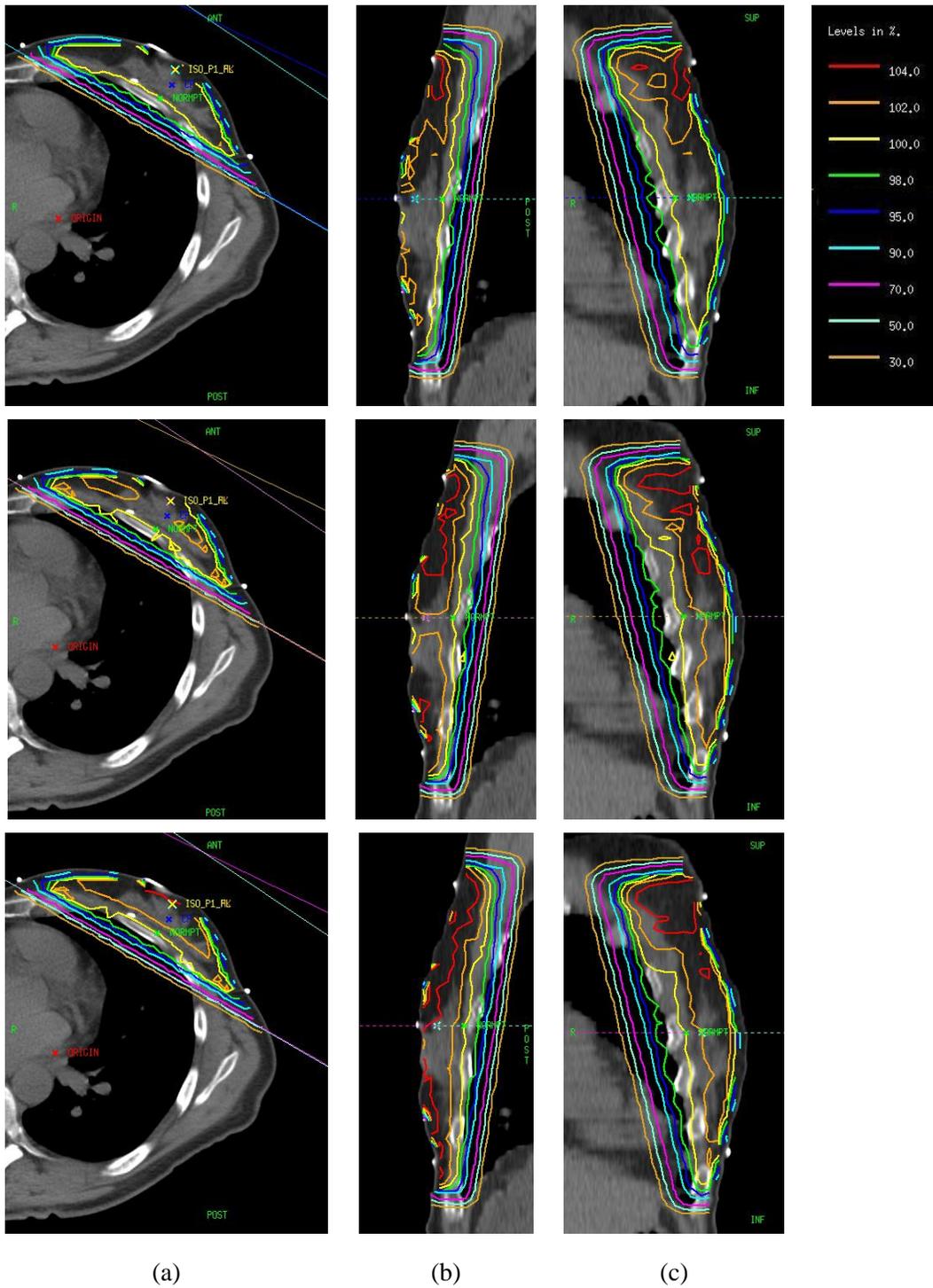


Fig. 14. Comparison of isodose distributions for leave-one-out experiment, patient 2, on the (a) transverse, (b) sagittal, and (c) coronal planes. The upper row is an expert's sIMRT plan; the middle row is the prediction result using parametric nonlinear regression; the lower row is the prediction result using support vector regression. Plans were normalized to be 100 at a specified point on the lung chest-wall interface (labeled NORMPT on the yellow contour).