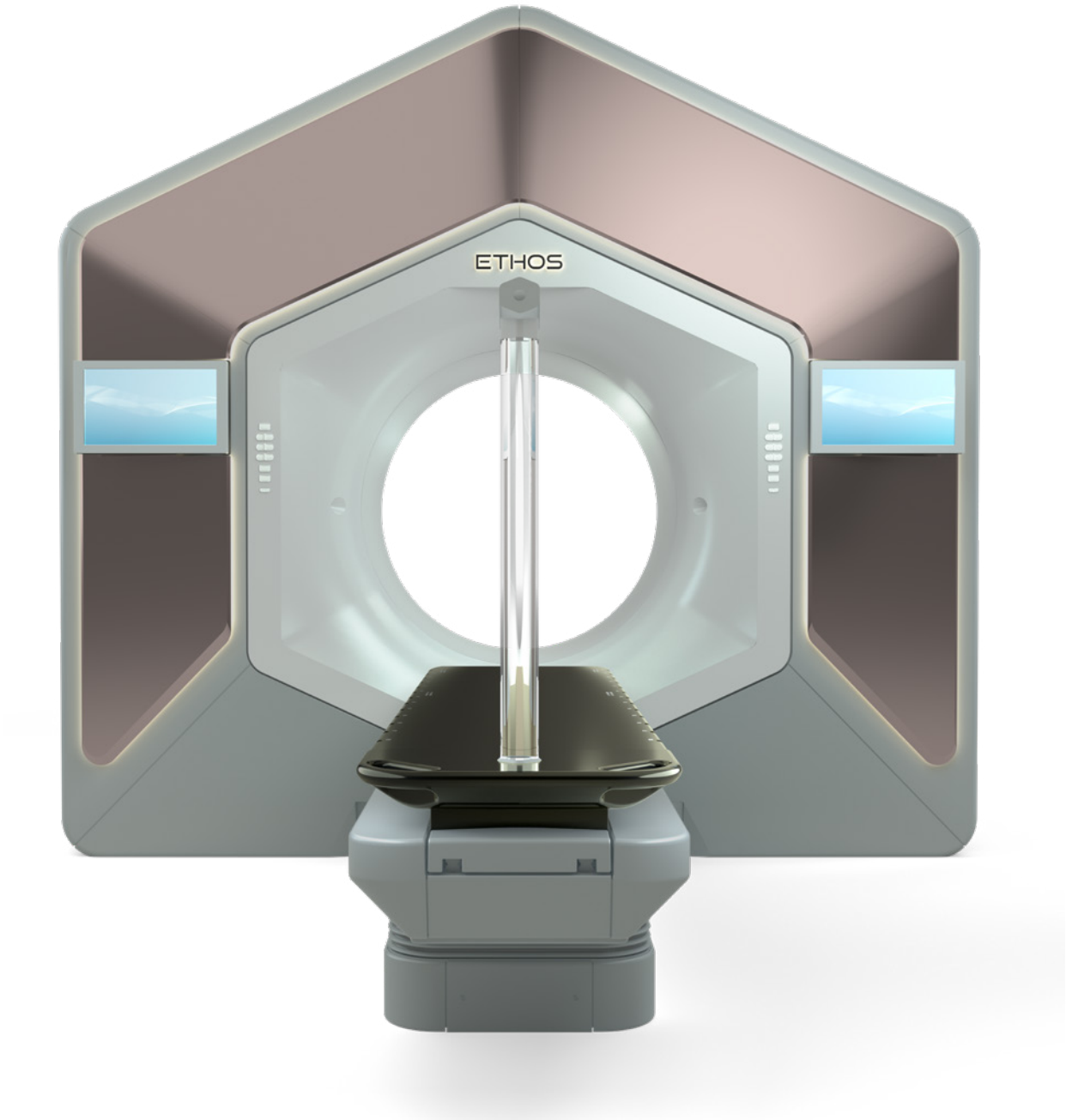

Ethos™ therapy AI

Technical Brief

Intelligent software drives Ethos™ therapy throughout the solution. Ethos treatment planning and Ethos treatment management incorporate machine learning, artificial intelligence, and intelligent algorithms. From the first treatment plan in Ethos treatment management to the last plan delivered on the Ethos radiotherapy system. We describe our approaches, algorithms, and machine learning-based tools in the following sections.

varian | ETHOS™

Table of Contents

Ethos therapy deep learning

Approach

During the adaptive workflow on the Ethos radiotherapy system, an artificial intelligence based (AI) algorithm is used to contour normal organ structures which influence the daily shape of the target on the newly acquired kV-CBCT. These "influencer" structures are those structures that are in the closest proximity to the target and have the biggest impact on the shape of common clinical target structures. The AI models and algorithms used to segment the influencer structures are based on a convolutional neural network. The neural network models used in the influencer segmentation process are static and do not continuously learn based on user input. This ensures the stability and performance of the algorithms over time.

Neural networks

A neural network is a collection of connected units or nodes called "artificial neurons", which behave in the same way as biological neurons. They have an input layer, an output layer (prediction) and one or more hidden layers. The depth of the network depends on the number of hidden layers. Deep neural networks are neural networks with multiple hidden layers. Deep learning convolutional neural networks (CNNs) make the explicit assumption that the inputs are images, which allows for the incorporation of certain properties into their architecture. CNNs are best for solving problems related to image recognition, object detection, and other computer vision applications. A typical CNN can be viewed as a sequence of layers that transform an image volume into an output volume.

Varian's in-house developed and trained deep learning model for Ethos therapy works with TensorFlow, CUDA and cuDNN libraries, and processes images on different resolution levels that are interconnected. Ethos uses full image deep convolutional neural networks with proprietary architectures that share many similarities with U-Net and DenseNet, which is widely used in image segmentation tasks. The network itself takes the full 3D iCBCT as an input and returns the same size of segmentation as an output during inference.

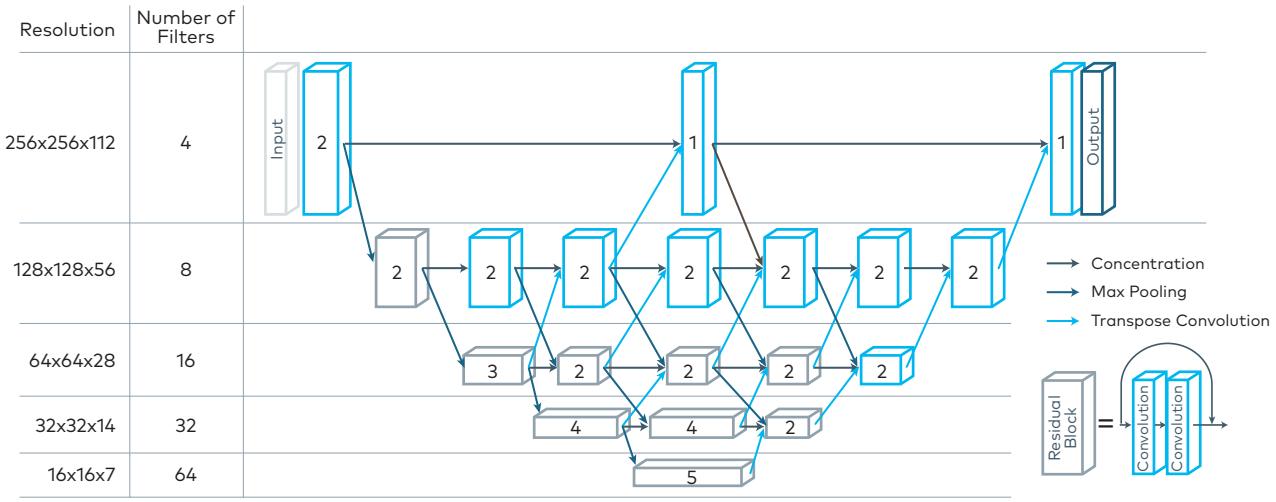

Figure 1 depicts an example of Varian's deep convolution neural network architecture.

*Note: For further information on Convolutional Neural Network layers refer to the Ethos Algorithms Reference Guide.

Deep neural network model production

The neural network training was performed in a supervised learning setting using images and ground truth contours of several hundred patients. Data was acquired from multiple clinics from the Americas, Europe, Australia and Asia. Images of the training set were selected to represent a realistic spectrum of anatomical variety and typical image artifacts. Ground truth contours were created by human anatomy experts as part of the algorithm development. A single set of contours was produced for each training image.

Training was done for each model in a supervised learning setting and involves three separate datasets:

1. Training dataset
    a. The training dataset is used to fit the model. This process involves utilizing a large set of consistently contoured data which is used to perform the actual training of the neural networks.
2. Validation dataset
    a. The validation dataset is used to provide an unbiased evaluation of a model fit on the training dataset while tuning the model with hyperparameters.
3. Test dataset
    a. The test dataset is a smaller set of data used to provide an evaluation of a final model fit.

A neural network is trained using the classical backward-error-propagation algorithm. An error is computed at the output and distributed throughout the network layers. The gradient descent optimization algorithm uses back propagation to adjust the weight of neurons by calculating the gradient of the cost function. A cost function is a measure of how good a neural network performs with respect to the given training sample and the expected output. The cost function is typically expressed as a difference or distance between the predicted value and the actual value. It can be estimated by iteratively running the model to compare estimated predictions against the ground truth.

Hyperparameters in deep learning models

Hyperparameters are settings that can be tuned to control the behavior of a machine learning algorithm. Conceptually, they can be considered orthogonal to the learning model itself in the sense that, although they live outside of the models, there is a direct relationship between them.

Examples of hyperparameters:

- Learning rate
    — The learning rate quantifies the learning progress of a model in a way that can be used to optimize its capacity.
- Number of hidden units
    — The number of hidden units is key to regulating the representation capacity of the model.
- Convolution kernel width
    — In CNNs, the kernel width influences the number of parameters in a model which, in turn, influences its capacity.

Hyperparameters may be tuned using two basic approaches: manual or automatic selection. Both approaches are technically viable, but the decision of which type of approach to use typically represents a trade-off. The decision is related to how deep of an understanding is required of machine learning models to select hyperparameters manually versus the high computation costs required for automatic selection algorithms. During training of Ethos therapy deep learning models hyper search is used for random weight initialization, as well as, defining the loss function and the layer order.

Post-processing of networks

Segmentation output of the networks is passed through a post-processing module. Smaller segments or dislocated segments are excluded from the results. Since the resolution used for calculation is typically larger than the image resolution, the final contours are smoothed.

Intelligent optimization engine

Within Ethos therapy, the treatment planning is highly automated to allow the user to focus on the clinical aspects of the patient's therapy. In order to automate the plan generation, we introduce an algorithm which orchestrates the plan optimization, the Intelligent Optimization Engine. The Intelligent Optimization Engine aims to perform all the actions necessary to generate high quality dose distributions which meet the clinical expectations for the plan, and ensure the plan is dosimetrically accurate. It sets up the optimization problem for the Photon Optimization algorithm, then controls and monitors the progress.

The Intelligent Optimization Engine is used in dose preview generation (which provides a fast, optimized dose distribution to check for potential clinical trade-offs), and the automated planning algorithm (which produces IMRT and VMAT plans for a given set of inputs). In both cases, the Intelligent Optimization Engine (IOE) works as follows:

Pre-processing: Translation of goals to objective functions

For each plan type to be generated (IMRT or VMAT), and for the Dose Preview, the IOE performs translation of the clinical goals to objective functions for the Photon Optimizer and creates Q-functions to monitor the progress of the optimization. Since the clinical goals from Ethos treatment management have an enforced syntax, and the overlap handling is well described, this translation step is straightforward. The Q-functions are described below.

Pre-processing: Overlap handling and objective setting

Prior to initiating the optimization and plan generation, a pre-processing step is performed by the IOE. In this step, the system examines the contoured organs, planning organ-at-risk volumes (PRV), and targets provided from Ethos treatment management, and assesses for conflicts and overlaps between targets and organs, and between targets with different dose levels.

A common overlap situation which needs resolution prior to the plan generation is when a target overlaps with an organ or a PRV and the user has specified goals for the target and the organ which cannot be physically met.

This overlap handling step needs guidance as to which goals for targets, organs, or PRVs have priority (which goal should "win") if there is a conflict. This goal hiearchy is taken initially from the priority group defined in the Ethos treatment management RT Intent module by the physician and provided to Ethos treatment planning and dose preview. The system reads the goals, determines how to resolve the overlaps, and then creates objectives for the Eclipse Photon Optimizer (PO) based on the clinical goal. In Dose Preview, the physician can fine-tune the order within the priority groups.

prior to authorizing the RT Intent.

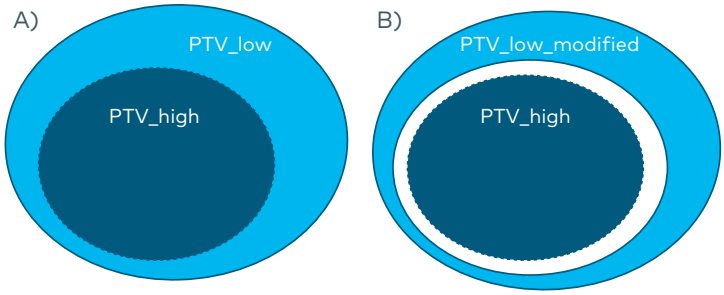

Fig 2. Conflict from multiple overlapping targets. A). High dose target (PTV_high) overlapping with lower dose target (PTV_low). B) User specifies a goal for maximum dose in PTV_low which is in direct conflict with the minimum dose goal for PTV_high. The system crops the PTV_low from PTV_high (with some margin for dosimetric fall-off) and then applies a maximum dose objective which follows the needed form for the Eclipse PO algorithm (created from the input maximum dose goal) to the remainder of PTV_low.

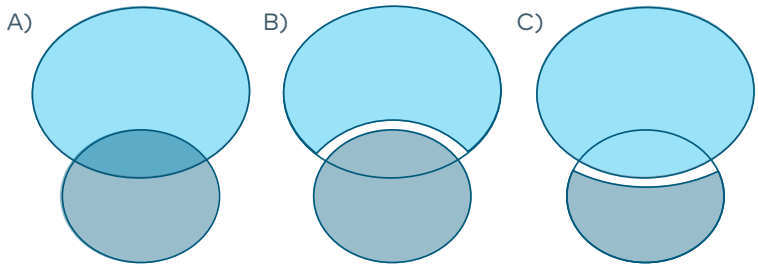

Fig 3. Conflict from organ (blue) overlapping with target (red). A). Target in shaded red with overlapping organ in shaded blue. B). Organ goal has higher priority than target goal, i.e., brainstem organ has a goal for maximum dose that is lower than the desired target dose. The target is cropped from organ including a margin for dosimetric falloff. C) Target goal exists with higher priority than organ goal, the organ is cropped from the target, again with a margin for dosimetric falloff. In both cases, objectives for the Eclipse PO algorithm are created by translating the input goals into the correct objective functions.

Figure 2 and Figure 3 show examples of overlaps and how they are resolved.

Optimization progress monitoring: Q-functions

The IOE establishes a set of piecewise continuous "quality" functions (Q-functions) based on prototype functions for target lower dose goals (goals which specify the minimum dose desired for a target), target upper dose goals (goals which specify the tolerated maximum dose to the target), and organ upper dose goals similar in nature to the target upper goals, but with a different prototype function. These functions are created from their prototypes, and "stretched" in vertical or horizontal directions depending on their relative priority and the input goal value in dose or volume.

Figure 4 provides an example of the functions for a set of 3 goals, one for each goal type: target upper, target lower, and organ upper. The Q-functions have a simple role in the plan generation similar to the automatic lower dose optimizer (ALDO) feature implemented in v15.5 Eclipse. Once Q-functions are established, the system starts the optimization and then interrogates the achieved values for each goal at specified intervals (number of iterations) and then uses the associated Q-function to

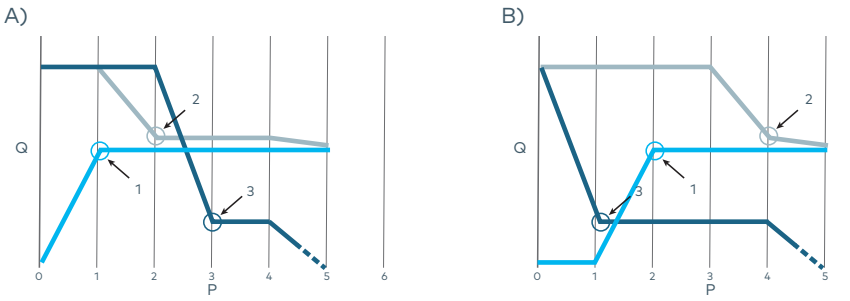

Fig 4. Intelligent Optimization Engine Q-functions (goal functions). In this example, the vertical axis in each panel is a quality measure for a goal. A goal meets its quality measure if the achieved value is the same or better than the goal value. The horizontal axis is a depiction of the relative priority of the set of goals. A). Line 1 is a target lower dose goal function with highest priority, line 2 a target upper dose function with second highest priority, and line 3 an organ upper dose function with third highest priority. In each, the circled point is the goal value in dose. Note that organ goal functions have a component which reaches toward very low or zero dose. B) Same goal functions but user has decided line 3 (organ upper) is highest priority, and line 2 (target upper) is lowest priority.

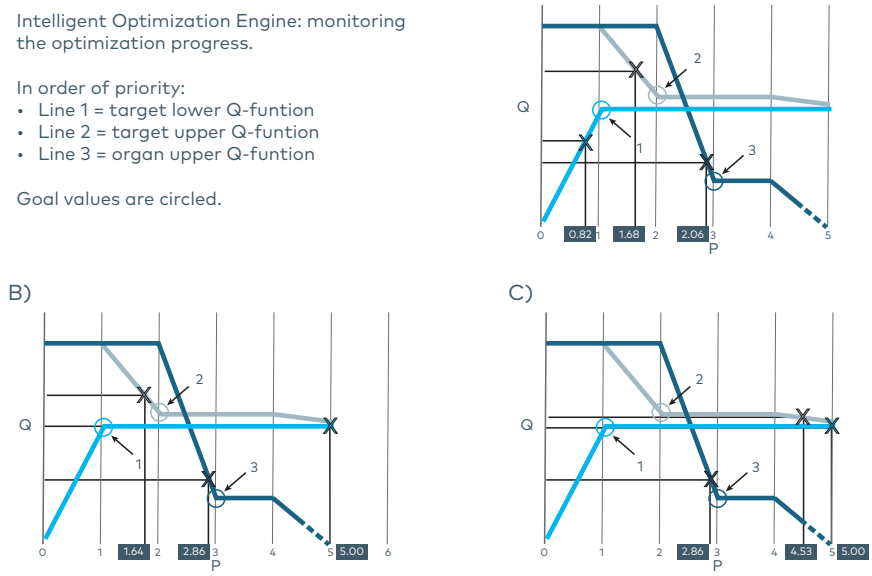

Fig 5. IOE Example with three goals. Optimization is started with fixed weightings for all objective functions. (A) After x iterations, the dose distribution is interrogated, and P values obtained. IOE elevates cost for objective function with lowest P (target lower dose goal=0.82; red line). (B) After another x iterations, dose distribution is interrogated again. Target lower dose goal is met and P=5.00. IOE elevates cost for the goal with P=1.68 (target upper). (C) After another x iterations, dose distribution is interrogated a third time. Target upper goal is now met with P evaluated to 4.53. IOE elevates cost for goal with P=2.68 (organ upper). Optimization continues until the collection of P is maximal and cannot further be improved.

determine an achieved "P" value for the goal. The goal of the IOE is to maximize the collection of P values in a given optimization. Figure 5 illustrates how the Q-functions are utilized to monitor the progress of the optimization.

The shape and location of the goal functions along the priority axis causes the optimization to progress similarly to how a human planner will work. Unmet goals with higher priority receive attention before unmet goals of lower priority, and when goals are met, additional effort is placed to reduce the dose to organs where possible. The example in Figure 5 could be performed with the goal functions from panel B of Figure 4. In that case, the highest priority goal is the organ upper and would be achieved first, prior to focusing attention to the lower priority goals. Since this is goal-based optimization, the system does not stop

when one goal is unmet; instead, the IOE detects the condition and re-baselines the goal function to higher (for organ or target upper goals) or lower (target lower) values.

RapidPlan™ knowledge-based planning

RapidPlan is an machine learning tool that can potentially introduce greater quality and efficiency into treatment planning though the power of historical patient data. The machine learning models within RapidPlan take inputs from dosimetric and geometric parameters of all plans included in its training set. As an output, the models can predict the dose volume histograms for modeled structures, as well as, generate the optimization objectives needed to drive DVHs to those predictions when used within Eclipse. [Eclipse Algorithm Reference Guide P1020505-002-B].

Organ Partitioning

Organ volume partitioning is performed on each structure of every plan included in the training set and in application of a specific model to a new clinical case. The beam geometry is used to create the partitions, as the beam eye view (BEV) from each field or control point is necessary to determine if the structure will receive any radiation dose at all. Combining the information obtained during partitioning is the mechanism which permits the ability to predict dose volume histograms (DVH) for modeled structures. As shown in Figure 6, the organ partitions are:

- Out-of-field region – the region of the structure that receives only scattered radiation dose
- Leaf transmission region – the region where the structure is always covered by leaves from a multileaf collimator (MLC)
- Overlap of the organ with the target (or union of all targets)
- The in-field region – the region that that is distal or proximal to the target in the beam eye view (BEV) and is not one the aforementioned regions. It represents the greatest contribution of dose to the modeled structure.

Partition modeling

Every case in the training set undergoes the partitioning and the average and standard deviations of the dose in the out of field partitions, leaf transmission partitions, and target overlap partitions are extracted. The in-field partition receives different treatment. This partition uses a supervised regression model of machine learning to infer characteristics which permit prediction of dose for this region. Combined with the result of the other 3 partitions, the entire DVH can be predicted.

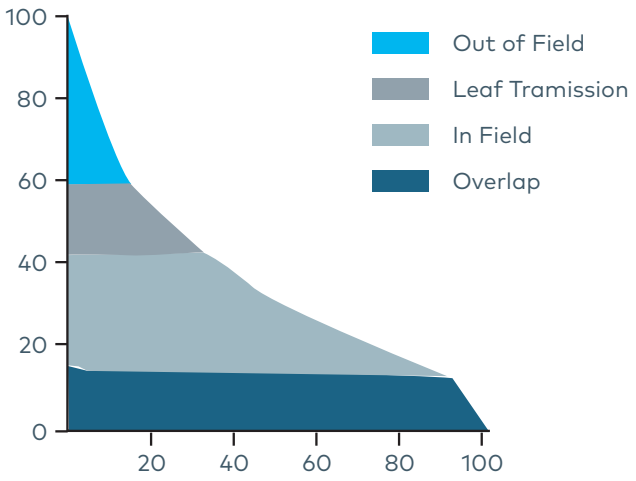

Fig 6. Organ volume partitions used in the machine learning DVH models. Each partition contributes to the sum total DVH as shown.

Geometry Expected Dose

In order to extract information which connects the geometry of the patient to the observed radiation dose, we utilize the concept of Geometry Expected Dose (GED). The GED is a score for each voxel within the treatment volume based solely on basic photon beam characteristics and the relationship of the structure with the radiation fields.

GED includes understanding the following properties [Eclipse Algorithm Reference Guide P1020505-002-B]:

- Field geometry
- Photon behavior
- Target geometry and dose levels
- Heuristics about which kind of beam arrangements lead to sparing of normal tissue

The GED can be calculated very quickly and only requires the field geometry, planning CT, and the structure geometry on the planning CT.

Supervised Regression

Once the GED for a given case is calculated, we can tabulate the GED volume histogram for the in-field partitions of modeled structures. These are considered to be highly correlated to the dose volume histograms for the same structures. This assumption is an observation based on the fact that the geometry relation (proximity) to a target highly influences how well a structure can be spared in a given treatment plan. Over a population of similar treatment plans, or treatment plans from a similar anatomical site, the DVH of the in-field partition and the GED volume histogram will be highly correlated.

To extract the correlations, RapidPlan uses principle component analysis applied to both the collection of in-field DVH and GED volume histograms in a training data set. The coefficient obtained from the principle components can be arranged and analyzed through regression models to extract the correlation from a given GED volume histogram to an observed dose volume histogram [Eclipse Algorithm Reference Guide P1020505-002-B].

For any case for which a DVH prediction is desired, the dose volume histogram for the in-field partition is predicted from this regression obtained from the training data set.

Compatibility with Ethos therapy

Any RapidPlan DVH Estimation model can be imported and applied to a RT Intent within Ethos treatment management. If attached to an RT Intent, the DVH estimates for modeled structures are shown in the Dose Preview and in the Plan Review work area. Additionally, the lower border of the DVH estimation band is used to derive a line objective which is applied during the plan generation process for both the initial planning and adaptive planning workflows in Ethos treatment planning.

Because there is not a known priority order for the line objective derived from the RapidPlan model, we cannot utilize the Intelligent Optimization Engine to effectively monitor or modify the strength of this line objective. The cost function derived from the line objective is added to the overall optimization, but at a level low enough not to overwhelm the objectives that the IOE determines, places, and monitors from the input goals and priority rankings. As such, its primary use in Ethos therapy is as a quality monitor. If the Ethos treatment planning Dose Preview or candidate plans from automated plan generation cannot achieve a result within the DVH predictions, additional goals may need to be added, the order of goals may need to be changed, beam geometry changed, or the case may not be suitable for automated planning.

Varian is your partner in this important enterprise because beating cancer takes a team. As always, we are committed to being by your side every step of the way while you make a vital difference in your patients' lives.

Intended Use Summary
Varian Medical Systems' linear accelerators are intended to provide stereotactic radiosurgery and precision radiotherapy for lesions, tumors, and conditions anywhere in the body where radiation treatment is indicated.

Important Safety Information
Radiation treatments may cause side effects that can vary depending on the part of the body being treated. The most frequent ones are typically temporary and may include, but are not limited to, irritation to the respiratory, digestive, urinary or reproductive systems, fatigue, nausea, skin irritation, and hair loss. In some patients, they can be severe. Treatment sessions may vary in complexity and time. Radiation treatment is not appropriate for all cancers.

varian

varian.com

USA, Corporate Headquarters and Manufacturer
Varian Medical Systems
Palo Alto, CA
Tel: 650.424.5700
800.544.4636

Authorized Representative in the EU
Varian Medical Systems
Nederland B.V.
Houten, The Netherlands
customer.relations@varian.com

Headquarters Europe, Eastern Europe, Middle & Near East, India, Africa
Varian Medical Systems
International AG
Stienhausen, Switzerland
Tel: 41.41.749.8844

Asia Pacific Headquarters
Varian Medical Systems
Pacific, Inc.
Kowloon, Hong Kong
Tel: 852.2724.2836

Australasian Headquarters
Varian Medical Systems
Australasia Pty Ltd.
Sydney, Australia
Tel: 61.2.9485.0111

Latin American Headquarters
Varian Medical Systems
Brasil Ltda.
São Paulo, Brasil
Tel: 55.11.3457.2655

510(k) pending—Not available for sale in all jurisdictions.

Varian Medical Systems as a medical device manufacturer cannot and does not recommend specific treatment approaches.
Specifications subject to change without notice.
Not all features or products are available in all markets and are subject to change.

RAD10690 Approved for distribution in the United States and Europe. 09/2019