

UCSF

UC San Francisco Previously Published Works

Title

IMRT QA using machine learning: A multi-institutional validation

Permalink

<https://escholarship.org/uc/item/36f8n5t4>

Journal

Journal of Applied Clinical Medical Physics, 18(5)

ISSN

1526-9914

Authors

Valdes, Gilmer
Chan, Maria F
Lim, Seng Boh
et al.

Publication Date

2017-09-01

DOI

10.1002/acm2.12161

Peer reviewed

IMRT QA using machine learning: A multi-institutional validation

Gilmer Valdes^{1,3,a} | Maria F. Chan^{2,a} | Seng Boh Lim² | Ryan Scheuermann³ |
Joseph O. Deasy² | Timothy D. Solberg^{1,3}

¹Department of Radiation Oncology, University of California San Francisco Medical Center, San Francisco, CA, USA

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

³Department of Radiation Oncology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

Author to whom correspondence should be addressed. Gilmer Valdes

E-mail: gilmer.valdes@ucsf.edu;
Telephone: (415) 910-6105

Funding Information

NIH/NCI Cancer Center, Grant/Award Number: P30 CA008748

Abstract

Purpose: To validate a machine learning approach to Virtual intensity-modulated radiation therapy (IMRT) quality assurance (QA) for accurately predicting gamma passing rates using different measurement approaches at different institutions.

Methods: A Virtual IMRT QA framework was previously developed using a machine learning algorithm based on 498 IMRT plans, in which QA measurements were performed using diode-array detectors and a 3%local/3 mm with 10% threshold at Institution 1. An independent set of 139 IMRT measurements from a different institution, Institution 2, with QA data based on portal dosimetry using the same gamma index, was used to test the mathematical framework. Only pixels with $\geq 10\%$ of the maximum calibrated units (CU) or dose were included in the comparison. Plans were characterized by 90 different complexity metrics. A weighted poisson regression with Lasso regularization was trained to predict passing rates using the complexity metrics as input.

Results: The methodology predicted passing rates within 3% accuracy for all composite plans measured using diode-array detectors at Institution 1, and within 3.5% for 120 of 139 plans using portal dosimetry measurements performed on a per-beam basis at Institution 2. The remaining measurements (19) had large areas of low CU, where portal dosimetry has a larger disagreement with the calculated dose and as such, the failure was expected. These beams need further modeling in the treatment planning system to correct the under-response in low-dose regions. Important features selected by Lasso to predict gamma passing rates were as follows: complete irradiated area outline (CIAO), jaw position, fraction of MLC leaves with gaps smaller than 20 or 5 mm, fraction of area receiving less than 50% of the total CU, fraction of the area receiving dose from penumbra, weighted average irregularity factor, and duty cycle.

Conclusions: We have demonstrated that Virtual IMRT QA can predict passing rates using different measurement techniques and across multiple institutions. Prediction of QA passing rates can have profound implications on the current IMRT process.

^aFirst authors contribution

This article is presented in part at the 2016 annual AAPM meeting in Washington D.C.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2017 The Authors. *Journal of Applied Clinical Medical Physics* published by Wiley Periodicals, Inc. on behalf of American Association of Physicists in Medicine.

PACS
87.55.Qr

KEY WORDS

IMRT QA, machine learning, poisson regression, radiotherapy

1 | INTRODUCTION

Over 50% of cancer patients receive radiotherapy as partial or full cancer treatment, and radiotherapy is an increasingly complex process. Machine learning is a subfield of data science that focuses on designing algorithms that can learn from and make predictions on data. Machine learning applications in radiotherapy have emerged increasingly in recent years, with applications including predictive modeling of treatment outcome in radiation oncology,^{1–7} treatment optimization,^{8–11} error detection and prevention,^{12–15} and treatment machine quality assurance (QA).^{16–19} These machine learning techniques have provided physicians and physicists information for more effective and accurate treatment delivery as well as the ability to achieve personalized treatment.

To the best of our knowledge, however, little work with machine learning has been explored in the field of dosimetry and QA in clinical radiotherapy. It is common to perform patient-specific pretreatment verification prior to intensity-modulated radiation therapy (IMRT) delivery. This process is time consuming and not altogether instructive due to the myriad of sources that affect a passing result. In an earlier work, a machine learning algorithm, *Virtual IMRT QA*, was developed that can predict IMRT QA passing rates and identify underlying sources of errors not otherwise apparent.²⁰ The algorithm identified the correlation between the IMRT plan complexity metrics and gamma passing rates and was validated on a single planning/delivery platform. The objective of this study is to further validate the approach using a large, heterogeneous dataset using different QA measurement devices (diode-array detectors and portal dosimetry) on different models of treatment machines and at different institutions.

Identifying plans prone to QA failure allows physicists to concentrate resources in developing proactive approaches to QA and provides information on sources of errors needed to strategically improve the workflow of patient care as described in AAPM TG-100.²¹ Goals of this study are to provide a framework to establish universal standards and thresholds, intercompare results, safely and efficiently implement adaptive radiotherapy, and in the long term, eliminate failing QA altogether. This represents a fundamental paradigm change in the way in which QA is performed.

2 | MATERIALS AND METHODS

2.A | Methodologies and data collection

The *Virtual IMRT QA* framework was previously developed based on 498 IMRT plans, 416 and 82 using 6 and 15 MV, respectively, on TrueBeam and Clinac IX treatment units (Varian Medical Systems, Palo Alto, CA, USA), measured using a diode-array detector

(MapCHECK2, Sun Nuclear Corp) with a gamma criteria of 3%local/3 mm and 10% threshold. For each plan, parameters were extracted from the Eclipse treatment planning system (AAA, v11.0 – Varian Medical Systems, Palo Alto, CA, USA) using SQL queries. In order to further validate the framework, over 200 additional IMRT measurements (139 and 64 using 6 and 15 MV, respectively) based on Varian portal dosimetry (PD) on a Trilogy Linac at the Memorial Sloan Kettering Cancer Center (MSKCC) were tested using the same gamma criteria. The electronic portal imaging device (EPID) used in this study is a Varian aS1000 model with a pixel resolution of 0.392 mm and maximum image acquisition rate of 30 frames per second. The dark and flood field calibration along with a 10 × 10 cm, 100 MU at the source-to-detector distance of 105 cm were performed every day before QA to guarantee consistency of the measurements. Automatic registration was used to align the absolute dose distributions and analysis was performed using a gamma criteria 3%local/3 mm. For each IMRT beam, control point information was extracted from Eclipse v11.0 and different features that characterize the plans, Table 1, were

TABLE 1 Sample variables in Virtual IMRT QA modeling (continuous variables >90).

Number	Variable	Possible value
1	QA device	MapCHECK2, Portal dosimetry
2	Energy	6 MV, 15 MV
3	Machine type	TrueBeam, Trilogy, 211X, 21EX, 6EX
4	Collimator angle	Mean value averaged over all control points
5	CIAO area (i.e., 5→5 × 5)	<5, 5–10, 10–15, 15–20, 20–25, 25–30, >30
6	Jaw position	<5, 5–10, 10–15, 15–20, 20–25, >25
7	Small aperture score	Fraction of MLC gaps <2, 5, 10, 20 mm
8	Fraction of area receiving at least x% of CU	10, 20, 30, 40, 50
9	Irregularity factor	Fraction of area outside Radius = 5, 10, 20 cm
10	MLC leaf transmission	HD, M120-pre-2007, M120-post-2007
11	Perimeter	<10, 10–30, 30–50, 50–70, 70–90, 90–110, >110
12	Duty cycle (Total MU/Dose)	<2, 3–3, 3–4, 4–5, 5–6, >6
13	Modulation factor	Overall complexity (1, 2, 3)

calculated using Matlab Scripts, Matlab Inc, MA. Using all features and the corresponding passing rate, multivariate analysis was performed to identify the most important features that govern the passing rate. Poisson regression with Lasso regularization was trained using the new dataset acquired at MSKCC to learn the relation between the plan characteristics and each passing rate. Other details can be found in the original publication on the development of the Virtual IMRT QA method.²⁰ Figure 1 summarizes the entire workflow of the construction and validation of the predictive IMRT QA model.

For the portal dosimetry, only pixels with $\geq 10\%$ of the maximum calibrated units (CU) or dose were included in the comparison. Plans were characterized by 90 different complexity metrics. A weighted poisson regression with Lasso regularization was trained to predict passing rates using the complexity metrics as an input.

2.B | Weighted Poisson regression with Lasso regularization

Let $S = [(x_1, y_1), \dots, (x_n, y_n)]$ be a set of paired data, where y_i is the number of detectors (pixels) that fail IMRT QA and x_i is a vector of complexity metrics of length 90 + 1 (as x_1 is set to 1). Now let us assume that the number of measurement points that fail follows a Poisson distribution, as it is customary when counts are modeled:

$$p(y = y_i | D_i, fr(x_i)) = \frac{(D_i * fr(x_i))^{y_i} e^{-D_i * fr(x_i)}}{y_i!} \quad (1)$$

where D_i is the total number of detectors in the analysis and $fr(x_i)$ is the mean value of the failing rate of the plan i that depends on its complexity vector x_i .

We can model fr according to a Poisson regression as:

$$fr(x_i) = e^{\beta^T x_i} \quad (2)$$

where β is a constant vector the same size as x_i

Now, given the realization of the data S , let us find the most likely vector β . In order to obtain β , we use Bayes theorem:

$$p(\beta|S) = \frac{p(S|\beta) * p(\beta)}{p(S)} \quad (3)$$

where $p(\beta|S)$ is the posterior probability of β given S , $p(S|\beta)$ is the probability of obtaining S given β , $p(\beta)$ is the prior probability of β , and $p(S)$ is the probability of obtaining S regardless of β . We are interested in finding the β that maximizes the function $p(\beta|S)$, which is the same as:

$$argmax_{\beta} p(\beta|S) = argmax_{\beta} \frac{p(S|\beta) * p(\beta)}{p(S)} = argmax_{\beta} p(S|\beta) * p(\beta) \quad (4)$$

In eq. 4 we have taken into account that $p(S)$ does not depend on β and as such, it can be dropped from the optimization problem. Assuming all measurements (x_i, y_i) are conditionally independent given the model, the probability $p(S|\beta)$ can be written as:

$$p(S|\beta) = \prod_{i=1}^n \frac{(D_i * e^{\beta^T x_i})^{y_i} e^{-D_i * e^{\beta^T x_i}}}{y_i!} \quad (5)$$

And assuming a Laplace distribution with a mean of 0 and variance equal to $2\lambda^2$ for $p(\beta)$, as customary in Lasso regularization, we have

$$p(\beta) = \frac{1}{2\lambda} e^{-|\beta|/\lambda} \quad (6)$$

which results in:

$$argmax_{\beta} p(\beta|S) = argmax_{\beta} \left[\prod_{i=1}^n \frac{(D_i * e^{\beta^T x_i})^{y_i} e^{-D_i * e^{\beta^T x_i}}}{y_i!} * \frac{1}{2\lambda} e^{-|\beta|/\lambda} \right] \quad (7)$$

As maximizing $p(\beta|D)$ is equivalent to maximizing $\log(p(\beta|D))$, eq. 7 can be rewritten as:

$$argmax_{\beta} \log(p(\beta|S)) = argmax_{\beta} \log \left(\prod_{i=1}^n \frac{(D_i * e^{\beta^T x_i})^{y_i} e^{-D_i * e^{\beta^T x_i}}}{y_i!} * \frac{1}{2\lambda} e^{-|\beta|/\lambda} \right)$$

Applying the rules of logarithms and dropping the terms that do not depend on β results in:

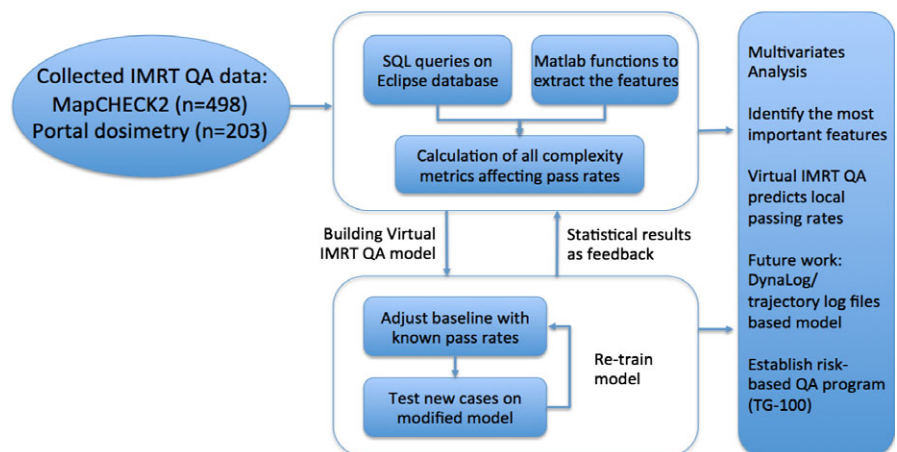


FIG. 1. The workflow of the validation of Virtual IMRT QA model.

$$\operatorname{argmin}_{\beta} \log(p(\beta|S)) = \operatorname{argmin}_{\beta} \left[- \sum_{i=1}^n w_i (fr_i \beta^T x_i - e^{\beta^T x_i}) + \lambda |\beta| \right] \quad (8)$$

where fr_i is the observed failing rate for plan i , $w_i = D_i/D_{max}$ is a weight factor for each observation proportional to the number of detectors in the measurement and D_{max} is a normalization constant. Equation 8 is a weighted Poisson regression problem with Lasso regularization where β^T can be obtained using the software package available at https://web.stanford.edu/~hastie/glmnet/glmnet_alpha.html.

Once β^T is obtained, this constant vector is used together with eq. 2 and the complexity metrics of each plan x_i to predict a specific plan's passing rates as:

$$fr_i = e^{\beta^T x_i} \quad (9)$$

3 | RESULTS

Equation 9 was used to construct histograms of measured versus predicted passing rates using a 3%/3 mm local gamma threshold (Figs. 2 and 3). This magnitude commonly called residual is the standard metric to evaluate the performance of regression algorithm in a similar way that Area Under the Curve is used to evaluate performance in classification algorithms.²² All composite plans measured using diode-array detectors were predicted within 3% accuracy², while passing rates for portal dosimetry on per-beam basis were predicted within <3.5% for 120 IMRT measurements delivered with 6 MV. The remaining measurements that used 6 MV (19) had large areas of low CU, where portal dosimetry exhibits poorer agreement with the calculated dose. This is due to the difference in response of an a-Si electronic portal imaging device to open and MLC transmitted radiation, and further modeling using an algorithm such as that presented by Vial et al. can improve the under-response in low-dose regions.^{23,24} These 19 IMRT beams with mean CU value of 0.127 ± 0.071 (range 0.022–0.254) have been excluded from the study pool. From the machine learning analysis, the important features selected by Lasso to predict gamma passing rates when the model was trained with MSKCC data were as follows: complete irradiated area outline (CIAO) area, jaw position, fraction of MLC leafs with gaps smaller than 20 or 5 mm, fraction of area receiving less than 50% of the total CU, fraction of the area receiving dose from penumbra, weighted average irregularity factor, and duty cycle among others. These features are the most likely feature to result in plans failing QA at different institutions that also use EPID dosimetry. Their specific quantitative contribution to the failing rate, however, will depend on the underlying model at each individual clinic. In addition, please note that all predictions are out of sample prediction, that is, the passing rates being predicted are not part of the data used to train the model.

Finally, a learning curve experiment was performed to estimate the number of plans needed in order for the model to converge to a solution. To evaluate confident intervals for the model, we

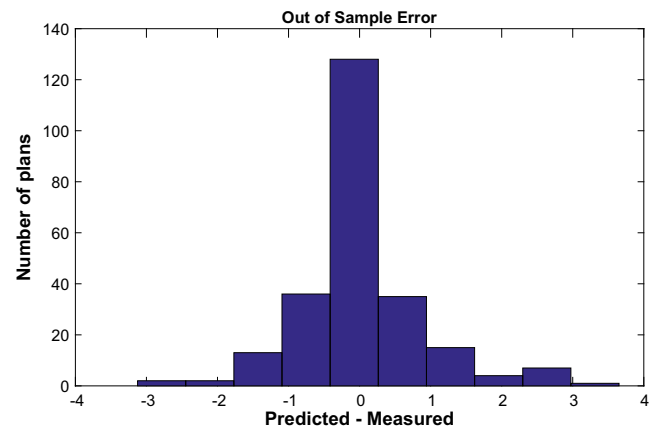


FIG. 2. Residual error for Clinac and TrueBeam Linacs measured using MapCHECK2 at Institution 1.

constructed models by varying the number of training samples and calculating the training and testing error. The process was repeated 10 times and confident intervals calculated. Figure 4 shows that after approximately 80 data patients, no further improvement is obtained on average in the testing error. This number is roughly half the number obtained by Valdes et al.²⁰ confirming that if the data are more homogeneous, that is, for Institution 2, only data from one Linac were analyzed, a smaller number of data points are needed to obtain a stable model.

4 | DISCUSSION

An algorithm using machine learning, *Virtual IMRT QA*, has been developed to correlate the characteristics of IMRT plan and delivery characteristics and the corresponding gamma passing rate. With some adjustments on the features in the original model, Poisson regression with Lasso regularization was trained to learn the relation between the plan characteristics and each passing rate for different measurement devices. The predictive model was validated for different QA devices and methodologies on seven Linacs across two institutions. With our validation results, *Virtual IMRT QA* can predict gamma passing rates accurately within 3.5% as maximum empirical error observed. In addition, our previous work showed that 200 patient-specific QA plans and results are required to train the virtual QA model to achieve an accurate prediction.²⁰ The required training QA data should be readily available assuming a measurement-based clinical patient-specific QA program is in place, and should not impose additional measurement burden on the practicing physicist. The potential benefit of this approach can be quite significant. For instance, *Virtual IMRT QA* could be run by the dosimetrist while planning. If an arbitrary threshold of 93.5% for *Virtual IMRT QA* is set, all plans that satisfy this threshold should pass IMRT QA with a passing rate higher than 90%. These plans could be further measured. However, those that have predicted passing rate smaller than 93.5% could be modified without the need to perform the QA

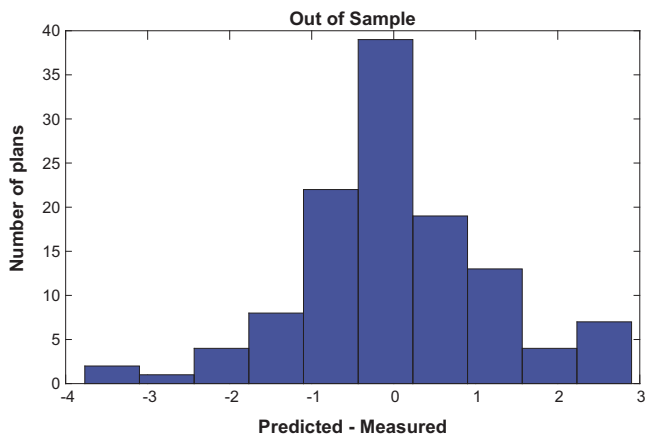


FIG. 3. Residual error for a Trilogy (6 MV) at using portal dosimetry at Institution 2. Note that the inherent Varian's Portal Dose Image Prediction algorithm assumes a radially symmetric response which is certainly different than the reality in 2D profiles of portal dosimetry.²³ This may add the additional uncertainty of this measurement.

potentially saving valuable time. In addition to single gamma results, this model provides insight into factors contributing to the resulting failure points by identifying relevant features. This allows physicists to quantitatively assess different risk factors-associated treatment plans. As TG-100 calls for a new risk-based QA program, this virtual QA model can be an invaluable tool to assist clinical physicists in their implementation. The virtual QA approach is not intended to replace measurement-based QA, but rather to complement the measurement-based program to provide a more comprehensive view. Through workflow improvement and risk analysis, the tool should enhance the overall QA and improve the safety of treatment delivery.

At present, the model is only capable of assessing fixed-beam IMRT planning/delivery; features critical to volumetric-modulated radiotherapy (VMAT) will be incorporated into future machine

learning models. This should be acknowledged as a limitation of the current method. Analysis of 3D detectors will require the collection of some of the key delivery information, such as gantry speed, MLC speed, and aperture size, and obtaining these parameters posts potential challenges. With the popularity of this treatment modality, further study in this direction will be a great asset to the community. In addition, the predictive model was trained to correlate the automatic registration of calculated and measured QA doses which has its pros (uncertainty due to phantom misalignment is removed) and cons (some mechanical errors producing a shift of dose are not detected in QA).

In this study, we have used the formalism as described by Valdes et al.²⁰ From eq. 9, however, it is clear that contributions from the different metrics were assumed to be linear (multiplication of a constant vector by the vector describing the plan characteristics), that is, no interaction terms between the different characteristics were considered. In addition, at the current stage, different models are needed for different combinations of delivery systems and energies. Ideally, one would like to incorporate all the data from an institution (delivery device, energy, QA devices, etc.) and input this data into a single mathematical formulation that will predict the gamma passing rate without the need for independent models which increases the data requirement. Even though beam data models for different Linacs/energies are different, they do share important characteristics and from an analytics point of view, it is inefficient to segregate the data by imposing practical constraints. We are currently developing a formulation that will take into account these limitations in order to obtain errors of local gamma passing rates within 2%. This number should be the ultimate goal of any quantitative analysis predicting gamma passing rate because as it has been reported by Nelms et al., some clinical relevant errors could create changes in gamma passing rate in the order of 2%.²⁵ One possible way to improve accuracy within the current framework would be to separate data by treatment site and build a model for each treatment site. As in the case of different models for different energies, this approach will increase

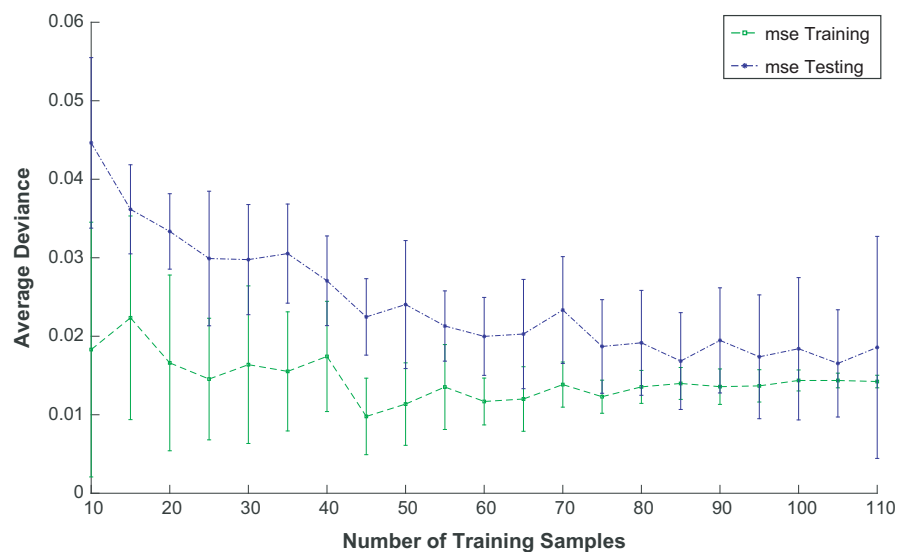


FIG. 4. Learning Curve. Testing and Training error versus number of data points used to build the model.

the need for more data. Finally, although we have decided to model here gamma passing rate with 3%/3 mm local analysis (for small targets 3%/2 mm might be more appropriate in the conventional QA), our methodology should not depend on the % dose or distance to agreement selected. That being said, at least in our dataset, stricter metric than 3%/3 mm results in having too much inherent noise for proper modeling.

5 | CONCLUSIONS

In this work with more extensive QA data, the validity of *Virtual IMRT QA* to accurately predict gamma passing rates, within 3.5% error, has been shown for different models of Linacs in different institutions, providing a strong validation of our IMRT QA predictive model. Compared to conventional measurement-based QA, the framework also provides significant insight into both machine and plan characteristics. Software-based, *Virtual IMRT QA* using machine learning has a unique position in the radiotherapy QA program and further provides a framework for a future integrated risk-based QA program such as that envisioned in AAPM TG-100.

ACKNOWLEDGMENTS

This research was funded in part through the NIH/NCI Cancer Center Support Grant P30 CA008748.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- Mihailidis D, Phillips MH. *Machine Learning in Radiation Oncology: Theory and Applications*. New York, NY: Springer; 2017.
- Zhang HH, D'Souza WD, Shi L, Meyer RR. Modeling plan-related clinical complications using machine learning tools in a multiplan IMRT framework. *Int J Radiat Oncol Biol Phys*. 2009;74:1617–1626.
- Valentini V, Dinapoli N, Damiani A. The future of predictive models in radiation oncology: from extensive data mining to reliable modeling of the results. *Future Oncol*. 2013;9:311–313.
- Robertson SP, Quon H, Kiess AP, et al. A data-mining framework for large scale analysis of dose-outcome relationships in a database of irradiated head and neck cancer patients. *Med Phys*. 2015;42:4329–4337.
- Kang J, Schwartz R, Flickinger J, Beriwal S. Machine learning approaches for predicting radiation therapy outcomes: a clinician's perspective. *Int J Radiat Oncol Biol Phys*. 2015;93:1127–1135.
- Oermann EK, Rubinsteyn A, Ding D, et al. Using a machine learning approach to predict outcomes after radiosurgery for cerebral arteriovenous malformations. *Sci Rep*. 2016;6:21161.
- Valdes G, Solberg TD, Heskell M, Ungar L, Simone CB II. Using machine learning to predict radiation pneumonitis in patients with stage I non-small cell lung cancer treated with stereotactic body radiation therapy. *Phys Med Biol*. 2016;61:6105.
- Zhang J, Ates O, Li A. Implementation of a machine learning-based automatic contour quality assurance tool for online adaptive radiation therapy of prostate cancer. *Int J Radiat Oncol Biol Phys*. 2016;96:E668.
- Schreibmann E, Fox T. Prior-knowledge treatment planning for volumetric arc therapy using feature-based database mining. *J Appl Clin Med Phys*. 2014;15:19–27.
- Price S, Golden B, Wasil E, Zhang HH. Data mining to aid beam angle selection for intensity-modulated radiation therapy. Paper presented at: Proceedings of the 5th ACM conference on bioinformatics, computational biology, and health informatics. 2014.
- Stanhope C, Wu QJ, Yuan L, et al. Utilizing knowledge from prior plans in the evaluation of quality assurance. *Phys Med Biol*. 2015;60:4873.
- Carlson JN, Park JM, Park S-Y, Park JI, Choi Y, Ye S-J. A machine learning approach to the accurate prediction of multi-leaf collimator positional errors. *Phys Med Biol*. 2016;61:2514.
- Tomatis S, Palumbo V, D'Agostino G, et al. EP-1329: data mining applied to a radiotherapy department: developing quality assurance tools for risk management. *Radiother Oncol*. 2015;115:S718.
- Bojchko C, Ford E. Quantifying the performance of in vivo portal dosimetry in detecting four types of treatment parameter variations. *Med Phys*. 2015;42:6912–6918.
- Ford EC, Terezakis S, Souranis A, Harris K, Gay H, Mutic S. Quality control quantification (QCQ): a tool to measure the value of quality control checks in radiation oncology. *Int J Radiat Oncol Biol Phys*. 2012;84:e263–e269.
- Li Q, Chan MF. Predictive time-series modeling using artificial neural networks for Linac beam symmetry: an empirical study. *Ann N Y Acad Sci*. 2017;1387:84–94.
- Chan MF, Li Q, Tang X, et al. Visual analysis of the daily QA results of photon and electron beams of a trilogy linac over a five-year period. *Int J Med Phys Clin Eng Radiat Oncol*. 2015;4:290.
- El Naqa I. SU-E-J-69: an anomaly detector for radiotherapy quality assurance using machine learning. *Med Phys*. 2011;38:3458–3458.
- Valdes G, Morin O, Valenciana Y, Kirby N, Pouliot J, Chuang C. Use of TrueBeam developer mode for imaging QA. *J Appl Clin Med Phys*. 2015;16:322–333.
- Valdes G, Scheuermann R, Hung C, Olszanski A, Bellerive M, Solberg T. A mathematical framework for virtual IMRT QA using machine learning. *Med Phys*. 2016;43:4323–4334.
- Huq MS, Fraass BA, Dunscombe PB, et al. The report of Task Group 100 of the AAPM: application of risk analysis methods to radiation therapy quality management. *Med Phys*. 2016;43:4209–4262.
- Hastie T, Friedman JH, Tibshirani R. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. Stanford, CA: Springer; 2009.
- Vial P, Greer PB, Hunt P, Oliver L, Baldock C. The impact of MLC transmitted radiation on EPID dosimetry for dynamic MLC beams. *Med Phys*. 2008;35:1267–1277.
- Hobson MA, Davis SD. Comparison between an in-house 1D profile correction method and a 2D correction provided in Varian's PDPC Package for improving the accuracy of portal dosimetry images. *J Appl Clin Med Phys*. 2015;16:43–50.
- Nelms BE, Chan MF, Jarry G, et al. Evaluating IMRT and VMAT dose accuracy: practical examples of failure to detect systematic errors when applying a commonly used metric and action levels. *Med Phys*. 2013;40:111722–111737.