Machine Learning for Automated Quality Assurance in Radiotherapy: A Proof of Principle using EPID Data Description

Issam El Naqa[1*], Jim Irrer[1], Tim A Ritter[2], John DeMarco[3], Hania Al-Hallaq[4], Jeremy Booth[5], Grace Kim[6], Ahmad Alkhatib[7], Richard Popple[8], Mario Perez[5], Karl Farrey[4], and Jean M. Moran[1]

[1]University of Michigan Department of Radiation Oncology, Ann Arbor, MI 48103, United States of America

[2]Virginia Commonwealth University Department of Radiation Oncology, Richmond, VA 23298, United States of America

[3]Department of Radiation Oncology, Cedars-Sinai Medical Center, Los Angeles, California 90048, United States of America

[4]University of Chicago Radiation and Cellular Oncology, Chicago, IL 60637, United States of America

[5]Royal North Shore Hospital, St Leonards, New South Wales 2065, Sydney, Australia

[6]University of California at San Diego, San Diego, CA 92093, United States of America

[7]Karmanos Cancer Institute, McLaren-Flint, Flint, Michigan 48532 United States of America

[8]University of Alabama at Birmingham, Birmingham, AL 35249, United States of America

*Corresponding author: email: ielnaqa@med.umich.edu.

Conflict of interest: Part of this work was supported by Varian Medical Systems (PI: JMM). Otherwise, the authors have no conflicts to disclose with regard to this publication.

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/MP.13433

Article Type: Research Article

Machine Learning for Automated Quality Assurance in Radiotherapy: A Proof of Principle using EPID Data Description

Issam El Naqa, Jim Irrer, Tim A Ritter, John DeMarco, Hania Al-Hallaq, Jeremy Booth, Grace Kim, Ahmad Alkhatib, Richard Popple, Mario Perez, Karl Farrey, and Jean M. Moran

Abstract

Purpose: Developing automated methods to identify task-driven quality assurance (QA) procedures is key towards increasing safety, efficacy, and efficiency. We investigate the use of machine learning (ML) methods for possible visualization, automation and targeting of QA, and assess its performance using multi-institutional data.

Methods: To enable automated analysis of QA data given its higher dimensional nature, we used nonlinear kernel mapping with Support Vector Data Description (SVDD) driven approaches. Instead of using labeled data as in typical Support Vector Machine (SVM) applications, which requires exhaustive annotation, we applied a clustering extension of SVDD, which identifies the minimal enclosing hypersphere in the feature space defined by a kernel function separating normal operations from possible failures (i.e., outliers). In our case, QA test data are mapped by a Gaussian kernel to a higher dimensional feature space and then the minimal enclosing sphere was identified. This sphere, when mapped back to the input data space along the principal components, can separate the data into several components, each enclosing a separate cluster of QA points that could be used to evaluate tolerance boundaries and test reliability. We evaluated this approach for gantry sag, radiation field shift, and MLC (multileaf collimator) offset data acquired using electronic portal imaging devices (EPID), as representative examples.

Results: Data from 8 LINACS and 7 institutions (n=119) were collected. A standardized EPID image of a phantom with fiducials provided deviation estimates between the radiation field and phantom center at 4 cardinal gantry angles. Deviation measurements in the horizontal direction (0°, 180°) were used to determine the gantry sag and deviations in the vertical direction (90°, 270°) were used to determine the field shift. These measurements were fed into the SVDD

clustering algorithm with varying hypersphere radii (Gaussian widths). For gantry sag analysis, two clusters were identified one of which contained 2.5% of the outliers and also exceeded the 1mm tolerance set by TG-142. In the case of field shifts, SVM clustering identified two distinct classes of measurements primarily driven by variations in the second principal component at 270°. Results from MLC analysis identified one outlier cluster (0.34%) along Leaf offset Constancy (LoC) axis that coincided with TG-142 limits.

Conclusion: Machine learning methods based on SVDD clustering are promising for developing automated QA tools and providing insights into their reliability and reproducibility.

Keywords: Machine learning, quality assurance, SVM, Linacs, higher dimension visualization.

I. Introduction

Cancer patients' safety and their treatment outcomes, despite rigorous regulations, may be compromised by rare but deadly errors that can occur during complex treatment planning and delivery of radiotherapy as highlighted by several editorials in national and international media reports in recent years [1]. Traditionally, quality assurance (QA) in radiotherapy follows the guidelines of national and international bodies such as the American Association of Physicists in Medicine (AAPM), American Society for Radiation Oncology (ASTRO), American College of Radiology (ACR), European Society for Radiotherapy and Oncology (ESTRO), and the International Atomic Energy Agency (IAEA). For instance, the AAPM and its widely used task group (TG) report TG-40 [2] has provided a comprehensive QA program for institutional radiation oncology practice. This report accounts for potential risks during the planning and delivery of high energy irradiation, harmonizing the treatment of patients and accommodating new advances in technology. TG-142 updated the requirements for advances in linear accelerator delivery technology[3]. A risk assessment and consensus evaluation of the critical requirements is presented in AAPM Medical Physics Practice Guideline 8a on linear accelerator QA [4]. Moreover, QA is a necessary process for credentialing institutions for multi-institutional radiotherapy clinical trials such as the ones carried out by the NRG Oncology consortium and

AAPM report TG113 [5-8]. While these QA guidelines have focused on monitoring all functional aspects of radiotherapy equipment, recent efforts have been geared towards identifying failures in workflow and processes. For instance, AAPM TG-100 has taken a risk-based approach using failure mode and effect analysis (FMEA) for designing QA protocols and prioritizing effort [9]. However, whether it is the traditional TG-40/142 or the new TG-100 guidelines, both approaches, as useful as they are, remain unfortunately subjective and are opinion-driven rather being data-driven; consequently, physicists are still left without an evidence-based answer to tailor a large number of laborious QA procedures to the associated failure risk. In the era of big data this limitation can be remedied [10,11].

Radiotherapy provides a fertile environment to harness the power of big data analytics, particularly in areas related to QA and safety [11-13]. Targeting of laborious QA tasks as needed has been recognized as a key component towards safer, more accurate and efficient radiotherapy administration [14]. However, traditional statistical methods cannot handle the challenges posed by radiotherapy big data, particularly the large class imbalance in navigating a great number of variables with a small sample size of relevant clinical data. This is further taken to the extreme in the case of QA, where the event rate is not only small but rare [15], due to improvements in software and hardware functionality and the tremendous efforts performed by the medical physicist. This issue constitutes a serious data analytics challenge.

Machine learning methods represent the computational vehicle for complex data analytics due to their ability to capture nonlinear and hidden patterns in the data, handle data imbalance, visualize higher dimensional space, and generalize to out-of-sample data [16]. Several studies have utilized different machine learning techniques for QA applications. These applications included automated error checkers of treatment plans using unsupervised learning such as k-means clustering [17] or supervised learning by neural networks [18], Bayesian networks [19], support vector machines (SVM) [20,21], and Poisson regression [22,23]. In addition, ML was applied to linear accelerator (Linac) machine QA such as supervised learning by neural networks of Linac beam symmetry [24] and multi-leaf collimator positional errors by random forest and Cubist methods [25].

In this work, we recognize that Linac machine QA processes in particular, are typically comprised of laborious tasks that are done based on prescriptive guidelines to monitor machines and equipment performance irrespective of the expected probability of failure risk. Methods based on process control charts have been proposed to assist in longitudinal monitoring of equipment function and separating random from systematic errors by defining action thresholds [14]. However, QA tests consist of multidimensional measures that exhibit complex and potentially nonlinear behavior among them. Thus, we hypothesize that the ability to visualize these tests in a higher dimensional space would allow for better identification of tolerance boundaries and assessment of the ability of these tests to detect failure risks. When applied to QA, it can potentially lead to a prioritized and targeted QA approach. Given the complex nature of radiotherapy QA processes and their redundancies, we will present an unsupervised machine learning tool to facilitate clustering and visualization of radiotherapy multidimensional test results. We highlight a method for estimating the tolerance boundaries and performance reliability of the tests by using the nontraditional Support Vector Data Description that does not require explicit training as typically practiced by SVM classification, and we evaluate its performance using multi-institutional data.

II. Materials and Methods

II.A Dataset

The dataset that will be used in this proposal is currently available institutionally and multi-institutionally through a consortium on Automated Quality Assurance (AQA) from 8 participating organizations. The consortium is focused on collecting comprehensive electronic portal imaging device (EPID) test results (Figure 1) from digital linear accelerators following TG-142 guidelines [26,27]. A dataset takes about 15 minutes to deliver and is subsequently submitted to the University of Michigan AQA database for analysis using an automated program (~1 minute to run) as described by Eckhause et al. [26]. In this study, the EPID images of a phantom constructed from Lucite and 2-mm diameter steel balls acting as fiducials are used to determine leaf and collimator positions relative to the fixed fiducials, which are localized in the image using a Canny edge detection algorithm [26]. The threshold for edge detection was adjusted until all the steel balls in the phantom were detected in the image. The location of the

phantom is defined as the location of the central ball bearing. Field edges were identified by averaging the in-field and out-of-field intensities on the images and the field position was determined from the location of these edges. Leaf edges were determined according to the leaf gap size; for large gaps, the peak intensity of the gradient parallel to the leaf was used, whereas for narrow gaps (e.g., picket fence test) the positions were calculated from the local peak in the intensity profile [26]. A total of 119 independent EPID measurements of several mechanical tests of the Linac (i.e., gantry sag, field shifts. leaf positions), taken at 1-4 week intervals by 7 institutions on 8 Varian TrueBeam accelerators, were analyzed.

In order to enable visualization and analysis of the EPID QA data in higher dimensions, we will investigate the use of nonlinear kernel mapping with Support Vector with Data Description (SVDD). SVM kernel methods have been proven to produce excellent classification rates by mapping relevant input features into higher dimensional space and building optimal hyperplanes to separate low from high risk categories by maximizing the separating margin between the classes (Figure 2). Successful application of SVM to medical applications has been demonstrated in many imaging and outcome modeling studies in radiation oncology [16]. However, instead of using labeled data, which would require exhaustive annotation, we will apply a cluster labeling extension of SVM using the SVDD algorithm [28,29].

II.B Data Description (SVDD) Clustering

The basic idea of SVDD is that input data (**x**) are mapped by a nonlinear kernel (e.g., Gaussian kernel) to the higher dimensional feature space, where one would search for the minimal enclosing hypersphere with a center a and radius R such that:

$$\min R^2 + C \sum_i \xi_i \quad (1a)$$

subject to:

$$\|\Phi(\mathbf{x}_i) - a\| \leq R^2 + \xi_i, \ \vee i, \quad (1b)$$

where ξ_i are slack variables to allow outliers in the dataset with a regularization parameter C and $\Phi(\cdot)$ is a nonlinear mapping function (Figure 2). Using a Lagrange multiplier approach, the following conditions can be attained:

$$\|\Phi(\mathbf{x}_i) - a\| < R^2 \longrightarrow \alpha_i = 0, \gamma_i = 0 \quad (2a)$$

$$\|\Phi(\mathbf{x}_i) - a\| = R^2 \longrightarrow 0 < \alpha_i < C, \gamma_i = 0 \quad (2b)$$

$$\|\Phi(\mathbf{x}_i) - a\| > R^2 \longrightarrow \alpha_i = C, \gamma_i > 0, \quad (2c)$$

where $\alpha_i \geq 0, \gamma_i \geq 0$ are Lagrange multipliers. Data points with $\alpha_i > 0$ are only needed to describe the mapping and are called support vectors (SVs), with points with $\alpha_i = C$ labelled as bounded SVs (BSVs). The solution can be obtained, as in other SVM approaches, using Quadratic Programing (QP) optimization techniques with a numerical complexity that depends on the underlying solver, which is generically between O(n^2) and O(n^3), where n is the number of training samples. Hence, the resulting hypersphere is given by:

$$a = \sum_i^{N_s} \alpha_i \Phi(\mathbf{x}_i) \quad (3a)$$

$$R^2(\mathbf{x}) = 1 - 2\sum_i^{N_s} \alpha_i K(\mathbf{x}_i, \mathbf{x}_k) + \sum_i^{N_s}\sum_j^{N_s} \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j)\,. \quad (3b)$$

For any $\mathbf{x}_k \in SV$, where N_s is the number of SVs and $K(\cdot,\cdot)$ is a kernel mapping representing the inner product: $K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)$. Typical kernel mapping is represented by Gaussian or radial basis functions (RBF):

$$K(\mathbf{x}_i, \mathbf{x}_j) = exp^{\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_k\|}{\sigma^2}\right)}\,, \quad (4)$$

where σ is the width of the RBF kernel. The hypersphere, when mapped back to the input data space, can separate the data into several components, each enclosing a separate related collection of points (a cluster of QA tests) labelled following efficient graph-based [29] or dynamical system equilibrium [30,31] algorithms for preserving the topological mapping characteristics as seen in Figure 3. The labeling approach we will be using is based on decomposing the data into several disjoint groups, where each group is represented by a stable equilibrium point (SEP) to which its members are assigned the same cluster label. An SEP represents the state when the clustering system reaches equilibrium, i.e., the eigenvalues of the Jacobian corresponding to Eq. (3b) are positive yielding a stable and topologically invariant solution [30].

II.C SVDD Clustering Application to QA

There are two parameters that control the behavior of the SVDD algorithm, namely: (1) the regularization parameter (C), which defines the soft margin boundaries and controls the number of SVs and (2) the width of the RBF (σ), which controls the number of clusters in the input space. These parameters and the resulting sphere radius (R) can be used to identify the accepted confidence levels in the QA test suite, analogous to control limits in control charts but providing the important advantage of visualization in higher dimensional space. In this case, C helps identify numerical outliers (failures) and controls possible overlap between the QA testing

clusters while σ controls the scale at which the data points are being probed (tolerance limits). Since the focus here is on visualization in higher dimensions, we fixed C=1 throughout the experiments while σ was varied between two categories: large width in which all points fit into a single cluster (k=1) and small width in which there are multiple clusters (k) $\geq$ 2. With this C = 1 setup, it also prevents BSVs (i.e., Equation 2c boundary condition). Moreover, the reported small width here corresponds to the largest possible experimental σ with k $\geq$ 2. For visualization in a 2D input space, dimensionality reduction by projection into principal component analysis (PCA) is used when the dimensions are greater than 2. The software tools used are based on extensions of MATLAB (Mathworks, Natick, MA, USA) for pattern recognition of data description [32] and efficient SVM cluster labeling [31]. The experiments were conducted on a 64-bit Windows 7 machine running an Intel Xeon-E5 processor with clock speed of 3.7 GHz and 32 GB of memory.

III. Results

III.A Gantry sag analysis by SVDD clustering

The gantry sags are primarily the result of gravitational torque. It is quantified as the difference in the field center with respect to the phantom center (central ball bearing on the EPID image), when the gantry is rotated from 0° to 180° using IEC standards [33]. Visualizing the EPID image as a matrix, the differences are estimated in the in-plane and cross-plane directions of the image and fed into the SVDD clustering algorithm. In Figure 4, we demonstrate the application of the proposed SVDD clustering algorithm to EPID-based measurements of gantry sag. Figure 4a (σ=0.5) shows a single cluster, while Figure 4b (σ=0.25) reveals two clusters, with "Cluster#2" here being the outliers' data considering the TG-142 recommended limits of gantry sag of 1 mm. Using Equation 3b, the RBF mapping with large and small widths (σ) corresponds to hyperradii (R) of 1.35 and 2.69, respectively. The calculations were performed in less than a second (i.e., on average 0.42±0.05s for σ=0.5 and 0.43±0.07s for σ=0.25). Interestingly, the members of this outlier cluster corresponded to different machines from different institutions. In this case, the percentage of outliers (Cluster#2) represents 2.5% while the TG-142 isotropic box is higher at 8.4%. Note that the PCAs here correspond to the 0° to 180° sag measurements, respectively, and demonstrate greater variation in the second principal component that corresponds to 180^{o}.

III.B Radiation field shift analysis by SVDD clustering

The shift in the radiation field is measured in the vertical direction and is defined as the difference between the radiation field positions with respect to the phantom averaged at gantry angles of 90° and 270° in the in-plane and cross-plane directions of the EPID image. In Figure 5, we demonstrate the application of the proposed SVDD clustering algorithm to EPID-based measurements of radiation field shift with Figure 5a showing a single cluster compared to Figure 5b revealing four clusters. The hyperradii corresponding to large and small RBF widths were R = 1.98 and 4.85, respectively. Three of the four clusters (Clusters #2-#4) were primarily inside the TG-142 recommended limits for radiation field shift of 1 mm. The outlier cases are estimated to be 2.5% while SVDD cluster analysis identified more outliers (2.5%) compared to the TG-142 limits (1.7%). Note that the PCAs here correspond to the 90° to 270° shift measurements, respectively. Again, the calculations were performed in less than a second (i.e., on average 0.45±0.1s for σ=0.5 and 0.42±0.06s for σ=0.25).

III.C MLC analysis by SVDD clustering

The multileaf collimator data included measurements for the Varian Millennium and high definition (HD) MLCs [26,27]. The Millennium MLCs consist of 120 leaves with the inner 40 leaves having widths of 0.5 cm and the outer leaves having widths of 1 cm. The HD MLCs consist of the inner 32 leaves having widths of 0.25 cm and the outer leaves having widths of 0.5 cm. Measurements of Leaf offset Constancy (LoC) and transmission were available for each leaf, for a total of 3486 points with the majority (83.5%) being from HD MLCs. The transmission measurements were adjusted from baseline on a per leaf basis following TG-142. Previous work by AQA consortium members demonstrated that the EPID-measured LoC is a comprehensive and efficient way to determine if the dosimetric leaf gap (DLG) is consistent with baseline[27]. The procedure for the EPID measurements were adapted from the LoSasso scheme for measuring DLG by using five fields: three sliding gap fields, a transmission field, and an open field [27]. The 3486 LoC and transmission data points were fed into the SVDD clustering algorithm. The TG-142 limit for leaf position repeatability of 1 mm was applied to the LoC and evaluated, and MLC transmission was assessed against a 0.5% allowable variation from baseline.

Figure 6 shows the clustering results for large (Figure 6a) and small (Figure 6b) RBF widths σ = 2 and 0.3 with corresponding hyperradii R = 0.35 and 1.12, respectively. Moreover, Figure 6b identified an outlier clusters (Clusters #2) along the LoC axis (i.e., principal component #1). Both the TG-142 and the SVDD estimates in this coincided with an outliers' rate of 0.34%. The calculation times here increased polynomially (i.e., on average 14.4±0.15s for σ=2 and 174.8±0.62s for σ=0.3).

IV. Discussion

We have presented a machine learning approach for visualization of QA data in higher dimensions and potentially for providing a mean for defining tolerance limits based on inherent data characteristics and detecting outliers. The approach was based on Support Vector Data Description (SVDD) with a clustering algorithm for analysis of QA data. As seen in the results, this method allows for visualization of higher dimensional QA test data and interpretation of non-isotropic boundary limits. The presented results were primarily qualitative, and the clusters are dependent on the selection of the RBF kernel width (sphere radii). Effects of the different QA tests on identifying failures could also be analyzed in this approach in a similar fashion to factor loading analysis, where the effect of including/excluding a test/parameter could be visualized in terms of separating annotated cases.

In this work, we have focused on applying SVDD as a visualization tool, to learn about the nature of the QA data, but it can subsequently be used as an effective outlier detector as presented in the results. For instance, when a deviation in gantry sag is detected, this can be reported to physics/engineering for maintenance and a decision made about the needed timing of maintenance and the clinical impact. [34]. In such a case the SVDD can be considered as having a one-class representation of normal Linac operations and the rest would be considered as outliers. In our case, we have heuristically determined the RBF width, as the largest one that would result in the number of clusters ≥ 2. We showed that a large width will result in one-cluster and the that there is a width that would yield ≥ 2, if outliers exist. To apply a more autonomous approach, one would consider assembling a training data with known normal

operations with or without known errors (i.e., failing data) . In such a task, a grid search is applied with cross-validation resampling to avoid overfitting in order to identify the hypersphere radius that would minimize the classification error in a similar fashion as supervised SVM training [35,36]. Moreover, the current application suggests batch processing of measurements. However, a strength of SVDD is that it can be also used as an online detector by applying incremental learning techniques [37,38], which would allow for efficient training and real-time monitoring in a similar fashion to control charts [39].

The importance of using measurements to evaluate leaf position reproducibility, such as with an EPID, rather than log files alone has previously been demonstrated by Agnew et al. [40]. A number of investigators have demonstrated the importance of the accuracy of MLC leaves on dosimetric accuracy of IMRT including when tolerances are considered. Others have noted that pre-treatment IMRT QA methods may be inadequate at identifying different types of delivery errors, especially when a gamma value is used that incorporates both distance and dose criteria [41,42]. The machine learning methods applied here for an evaluation of periodic QA permit a multi-dimensional evaluation of the results. The methods can also be used to identify dependencies of different QA results.

The current methodology shows promise in identifying the most sensitive QA parameters and quantifying the detection of outliers in a data-driven approach. In this context, SVDD can be used for visualization and failure monitoring. The RBF width and/or the hypersphere radius can be related to machine tolerances providing an anisotropic description of normal operations versus anomalies and a mean for estimating their likelihood of occurrence and detection, which can be subsequently used to rank the necessary frequency of QA tests. However, there are also limitations for using RBF kernels with sphere mapping, which performed adequately for the presented cases. However, other kernels/geometries or algorithms may be more appropriate in other instances. Moreover, in this work we simplified the representation of TG-142 by a bounding box, and the results are not intended to show preference but to provide a reference for comparison only, as supervised training with annotated data may be required to evaluate and establish definitive limits as discussed earlier. In addition, before applying principles such as those in TG 100, further data collection and analysis is required that incorporates a longer time

component and includes other events such as machine breakdowns and preventive maintenance on the linear accelerators. If dealing with large datasets for AQA applications becomes an issue in this context, faster training algorithms of SVDD are available and can be utilized [43,44]. There is a richness to such datasets because the same type of detector is used for all measurements. Since it is unlikely that a single institution can collect sufficient data over a few years, pooling data across institutions [26] may be required to create datasets of the size required to harness the power of machine learning. The application of machine learning extends beyond the traditional analysis of QA results, which focuses on whether or not a test limit was met or exceeded.

V. Conclusions

Machine learning methods based on SVDD clustering can be used as a promising tool for developing automated QA methods analysis and providing insights into the effectiveness, reliability, and reproducibility of such tests. Such methods offer an enhancement to the information that is typically available in an individual clinic and it is an area where collaboration and multi-institutional data can be valuable to establish a more efficient data-driven approach rather than an opinion-driven QA program in radiotherapy.

VI. Acknowledgements

Part of this work was presented at the AAPM Science Council Symposium, 2017. This work was supported in part by a grant from Varian Medical Systems.

VII. References

[1] BOGDANICH W. Radiation offers new cures, and ways to do harm The New York Times. New York, NY: The New York Times, 2010.

[2] Kutcher GJ, Coia L, Gillin M, et al. Comprehensive qa for radiation oncology: Report of aapm radiation therapy committee task group 40. Medical Physics 1994;21:581-618.

[3] Klein EE, Hanley J, Bayouth J, et al. Task group 142 report: Quality assurance of medical accelerators. Med Phys 2009;36:4197-4212.

[4] Smith K, Balter P, Duhon J, et al. Aapm medical physics practice guideline 8.A.: Linear accelerator performance tests. Journal of Applied Clinical Medical Physics 2017;18:23-39.

[5] Followill DS, Urie M, Galvin JM, et al. Credentialing for participation in clinical trials. Frontiers in Oncology 2012;2:198.

[6] Hartford AC, Galvin JM, Beyer DC, et al. American college of radiology (acr) and american society for radiation oncology (astro) practice guideline for intensity-modulated radiation therapy (imrt). American journal of clinical oncology 2012;35:612-617.

[7] Fitzgerald TJ, Bishop-Jodoin M, Bosch WR, et al. Future vision for the quality assurance of oncology clinical trials. Front Oncol 2013;3:31.

[8] Moran JM, Molineu A, Kruse JJ, et al. Executive summary of aapm report task group 113: Guidance for the physics aspects of clinical trials. Journal of Applied Clinical Medical Physics;0.

[9] 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. Medical Physics 2016;43:4209-4262.

[10] Ford EC Evans S. Incident learning in radiation oncology: A review. Med Phys 2018.

[11] Potters L, Ford E, Evans S, et al. A systems approach using big data to improve safety and quality in radiation oncology. International journal of radiation oncology, biology, physics 2016;95:885-889.

[12] El Naqa I. Biomedical informatics and panomics for evidence-based radiation therapy. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 2014;4:327-340.

[13] El Naqa I. The role of big data in radiation oncology: Challenges and potentials. In: Wang B, Li R and Perrizo W, eds. Big data analytics in bioinformatics and healthcare. Hershey, PA: IGI Global, 2014;pp. 163-185.

[14] Pawlicki T, Yoo S, Court LE, et al. Process control analysis of imrt qa: Implications for clinical trials. Physics in medicine and biology 2008;53:5193-5205.

[15] Ford EC Evans SB. Incident learning in radiation oncology: A review. Medical Physics 2018;45:e100-e119.

[16] El Naqa I, Li R Murphy MJ. Machine learning in radiation oncology: Theory and application. Switzerland: Springer International Publishing, 2015.

[17] Fatemeh A, David K, Jennifer GD, et al. Towards the development of an error checker for radiotherapy treatment plans: A preliminary study. Physics in Medicine & Biology 2007;52:6511.

[18] Willoughby TR, Starkschall G, Janjan NA, et al. Evaluation and scoring of radiotherapy treatment plans using an artificial neural network. International Journal of Radiation Oncology*Biology*Physics 1996;34:923-930.

[19] Alan MK, John HG, Eric CF, et al. Bayesian network models for error detection in radiotherapy plans. Physics in Medicine & Biology 2015;60:2735.

[20] Brown WE, Sung K, Aleman DM, et al. Guided undersampling classification for automated radiation therapy quality assurance of prostate cancer treatment. Medical Physics;0.

[21] El Naqa I. Detection and prediction of radiotherapy errors. In: El Naqa I, Li R and Murphy MJ, eds. Machine learning in radiation oncology: Theory and applications, vol. 1. Switzerland: Springer, 2015;pp. 237-241.

[22] Valdes G, Chan MF, Lim SB, et al. Imrt qa using machine learning: A multi- institutional validation. Journal of Applied Clinical Medical Physics 2017;18:279-284.

[23] Valdes G, Scheuermann R, Hung CY, et al. A mathematical framework for virtual imrt qa using machine learning. Medical Physics 2016;43:4323-4334.

[24] Li Q Chan MF. Predictive time- series modeling using artificial neural networks for linac beam symmetry: An empirical study. Annals of the New York Academy of Sciences 2017;1387:84-94.

[25] Joel NKC, Jong Min P, So-Yeon P, et al. A machine learning approach to the accurate prediction of multi-leaf collimator positional errors. Physics in Medicine & Biology 2016;61:2514.

[26] Eckhause T, Al-Hallaq H, Ritter T, et al. Automating linear accelerator quality assurance. Medical Physics 2015;42:6074-6083.

[27] Ritter TA, Schultz B, Barnes M, et al. Automated epid-based measurement of mlc leaf offset as a quality control tool. Biomedical Physics & Engineering Express 2018;4:027008.

[28] Tax DMJ Duin RPW. Support vector data description. Machine Learning 2004;54:45-66.

[29] Ben-Hur A, Horn D, Siegelmann HT, et al. Support vector clustering. J. Mach. Learn. Res. 2002;2:125-137.

[30] Jaewook L Daewon L. An improved cluster labeling method for support vector clustering. IEEE transactions on pattern analysis and machine intelligence 2005;27:461-464.

[31] Lee J Lee D. Dynamic characterization of cluster structures for robust and inductive support vector clustering. IEEE transactions on pattern analysis and machine intelligence 2006;28:1869-1874.

[32] Lei B, Xu G, Feng M, et al. Classification, parameter estimation and state estimation: An engineering approach using matlab. NJ, USA: John Wiley & Sons, 2017.

[33] Du W, Gao S, Wang X, et al. Quantifying the gantry sag on linear accelerators and introducing an mlc-based compensation strategy. Medical Physics 2012;39:2156-2162.

[34] Du W, Gao S, Wang X, et al. Quantifying the gantry sag on linear accelerators and introducing an mlc-based compensation strategy. Med Phys 2012;39:2156-2162.

[35] Theissler A Dear I. Autonomously International Journal of Computer and Information Engineering 2013;7:949-957.

[36] A.G. R, Abdulla MS S. A. Lightly trained support vector data description for novelty detection. Expert Syst. Appl. 2017;85:25-32.

[37] Oh JH, Naqa IE Yang Y. Online learning of relevance feedback from expert readers for mammogram retrieval Proceedings of the 43rd Asilomar conference on Signals, systems and computers. Pacific Grove, California, USA: IEEE Press, 2009; 17-21.

[38] Xie W, Uhlmann S, Kiranyaz S, et al. Incremental learning with support vector data description. 2014 22nd International Conference on Pattern Recognition. 20143904-3909.

[39] Sanghangthum T, Suriyapee S, Kim GY, et al. A method of setting limits for the purpose of quality assurance. Physics in medicine and biology 2013;58:7025-7037.

[40] Agnew A, Agnew CE, Grattan MW, et al. Monitoring daily mlc positional errors using trajectory log files and epid measurements for imrt and vmat deliveries. Physics in medicine and biology 2014;59:N49-63.

[41] Kruse JJ. On the insensitivity of single field planar dosimetry to imrt inaccuracies. Med Phys 2010;37:2516-2524.

[42] Steers JM Fraass BA. Imrt qa: Selecting gamma criteria based on error detection sensitivity. Medical Physics 2016;43:1982-1994.

[43] Chaudhuri A, Kakde D, Jahja M, et al. Sampling method for fast training of support vector data description. 2018 Annual Reliability and Maintainability Symposium (RAMS). 20181-7.

[44] Cao J, Zhang L, Wang B, et al. A fast gene selection method for multi-cancer classification using multiple support vector data description. Journal of Biomedical Informatics 2015;53:381-389.

FIGURE LEGENDS

Figure 1. (a) The QA phantom containing small spherical fiducials. The two pieces of plastic (upper left and lower right) create contrast for measuring image quality. (b) An EPID image of the QA phantom. The locations of the fiducials (marked with circles) are determined with automated analysis software (Reproduced from [26] with permission).

Figure 2. Kernel-based mapping from a lower dimensional space (X) to a higher dimensional space (Z) called the feature (Hilbert) space, where non-separable classes become linearly separable. In case of SVM, this mapping can be achieved using polynomials or radial basis functions to create higher order features from the input data. Samples on the borders constitute support vectors and they are represented by the most difficult cases to diagnose (Reproduced from [16]).

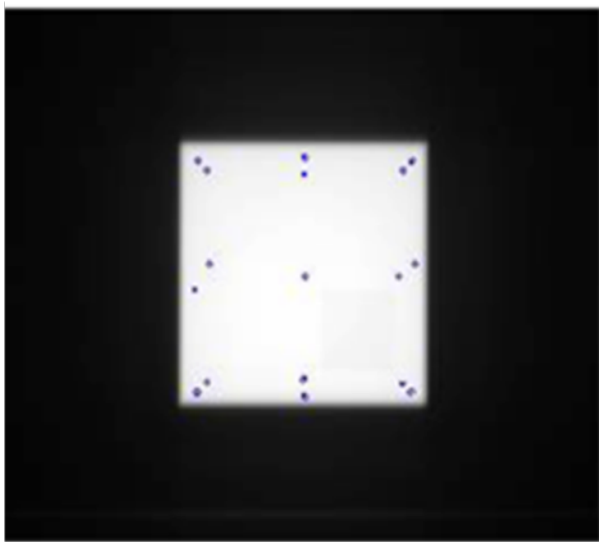
Figure 3. The main principle of the SVDD approach is that by first mapping input data from potentially different characteristics (e.g., normal Linac operation versus outliers) into a higher dimension and identifying the enclosing sphere (left), then re-mapping the sphere back into the data space, the data points can be divided efficiently into their corresponding clusters (right) [29].

Figure 4. Gantry sag analysis using SVDD clustering. The principal components 1 and 2 correspond to the 0° and 180° angles, respectively. (a) Using a large Gaussian kernel width (σ=0.5), it is noted that the cluster exceeds the bounds of the TG-142 recommendation (1 mm box) in the input data space. (b) Using a small Gaussian kernel width (σ=0.25), it is noted the presence of two clusters of measurements, with the smaller cluster representing the "true" outliers per the shape of data which is anisotropic in comparison to the TG-142 recommendation.

Figure 5. Radiation field shift analysis using SVDD clustering with principal components 1 and 2 corresponding to the lateral angles (90°, 270°) angles, respectively. (a) Using a large Gaussian kernel width (σ=0.5), the cluster encloses all measurements with the red circles showing the support vectors (boundary points). (b) Using a small Gaussian kernel width (σ=0.25), the presence of two distinct classes of measurements is noted primarily related to variations in the second principal component (270° measurements).

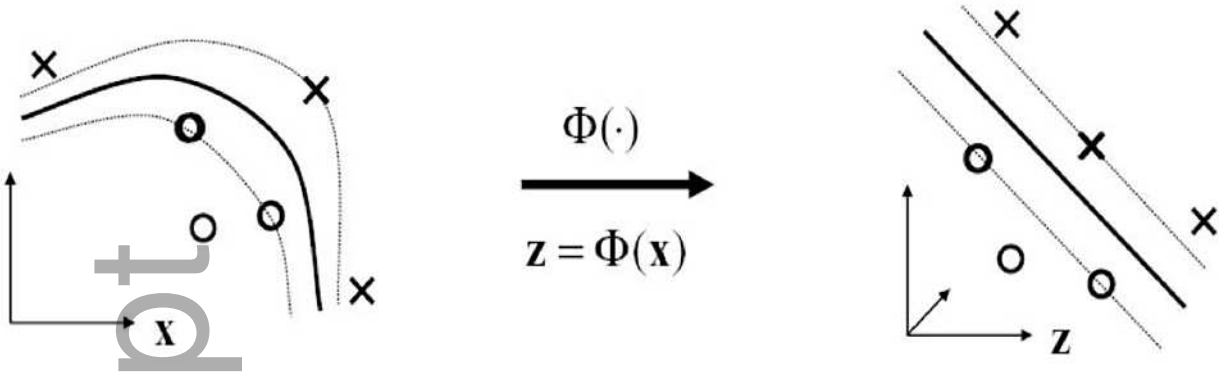
Figure 6. MLC shift analysis using SVDD clustering, the principal components 1 and 2 correspond to LoC and transmission respectively with the dashed rectangle representing TG-142 limits. (a) Using a large Gaussian kernel width (σ=2), it is noted that the cluster encloses all measurements, with the red circles showing the support vectors (boundary points). (b) Using a small Gaussian kernel width (σ=0.3), it is noted the presence of 2 regions (clusters) in the data in the LoC direction.

(a)

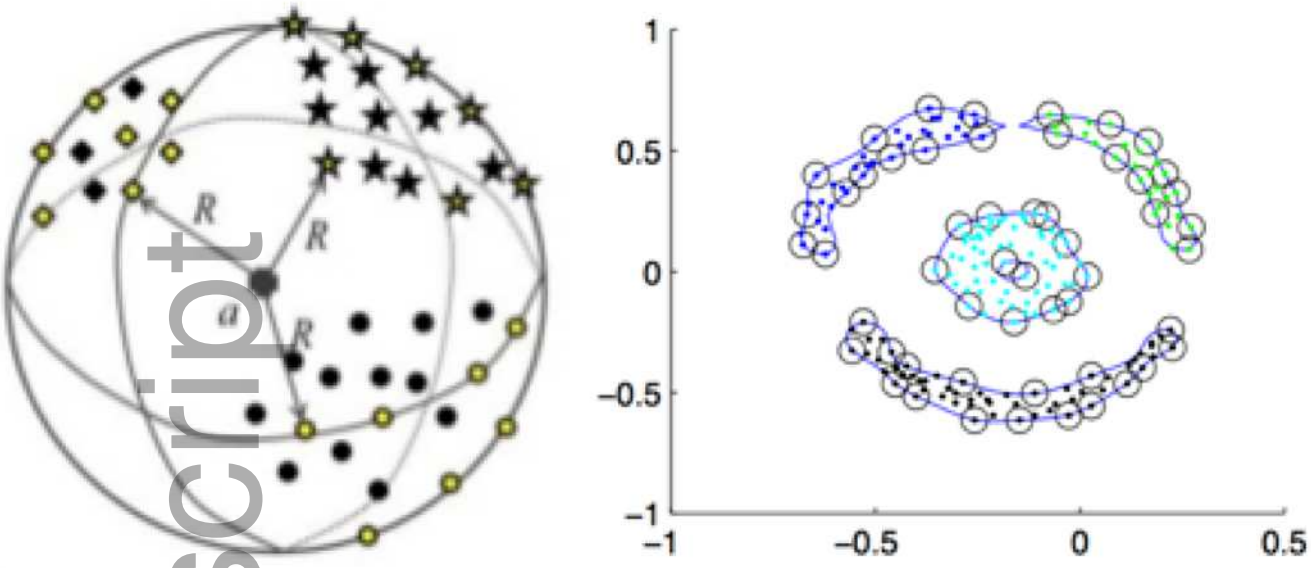


(b)

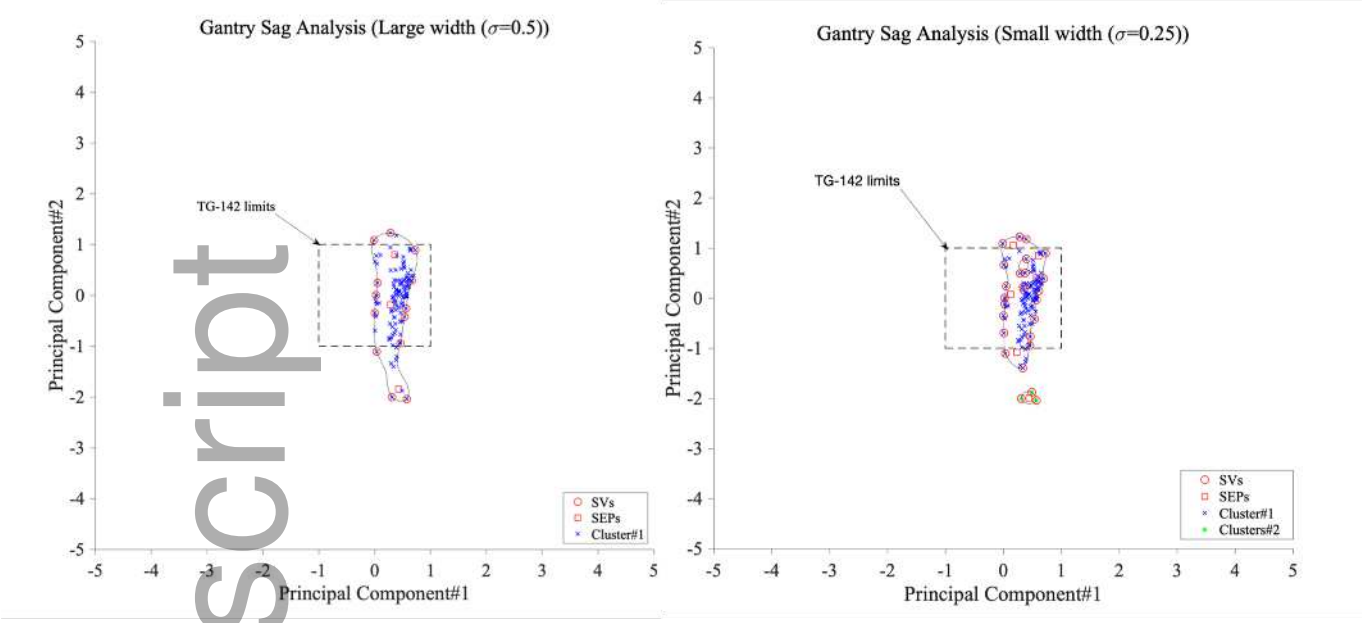
mp_13433_f1.tiff



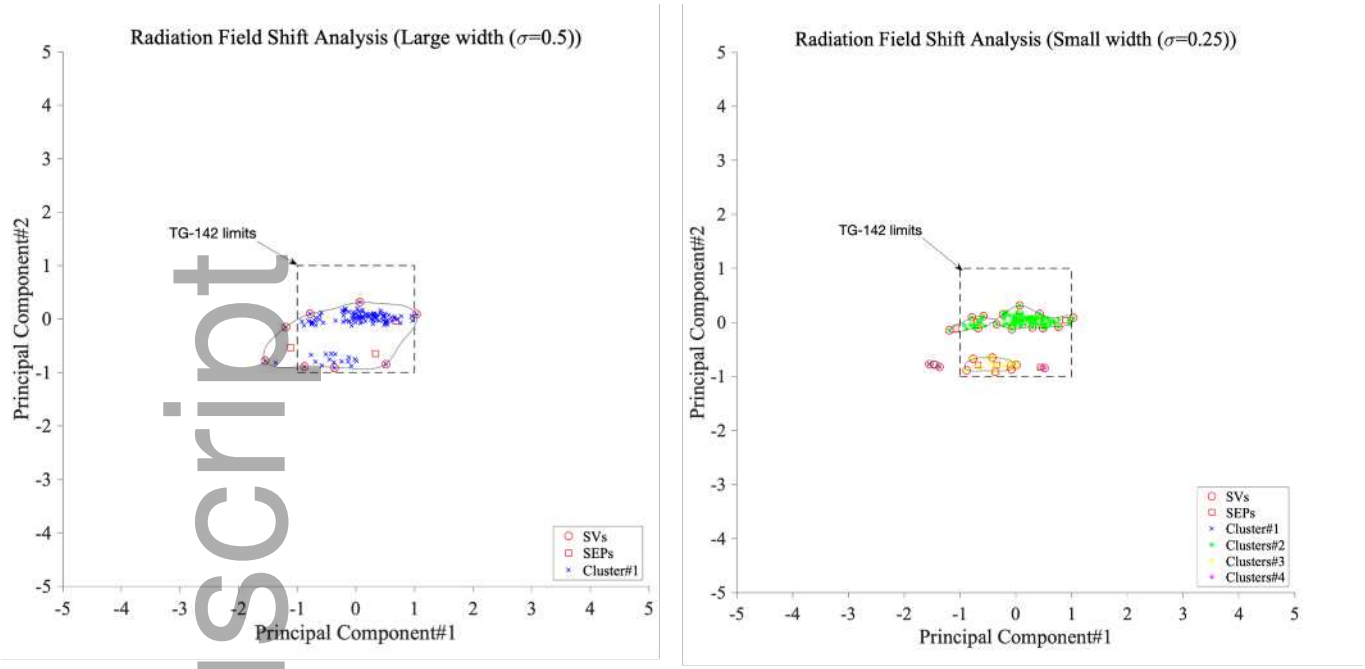
mp_13433_f2.tiff



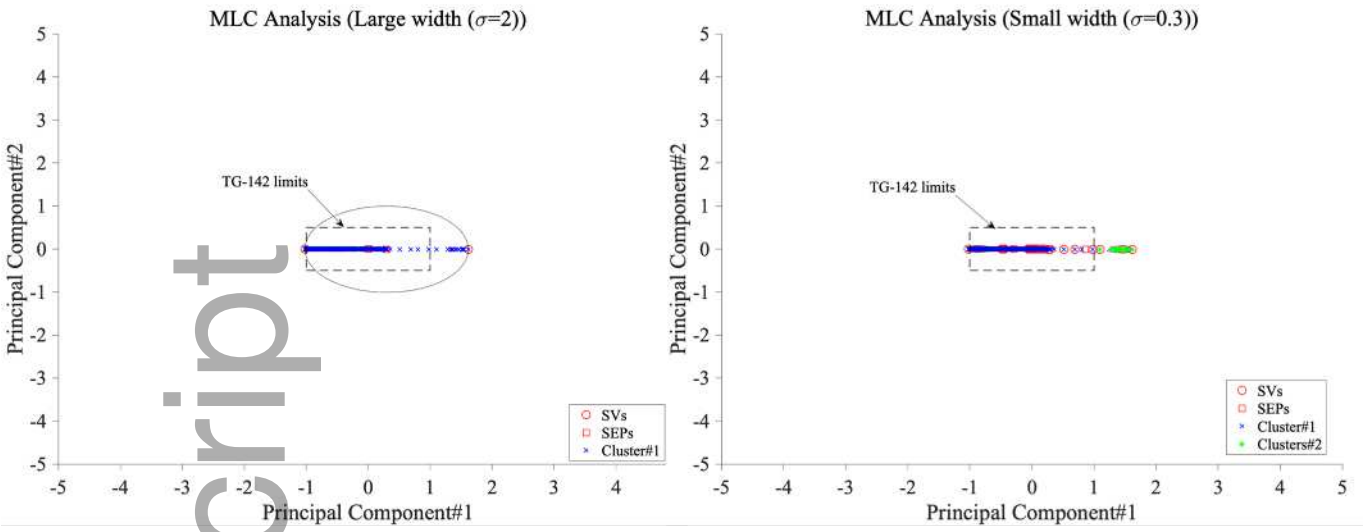
mp_13433_f3.tiff



mp_13433_f4.tiff



mp_13433_f5.tiff



mp_13433_f6.tiff