sciendo
Polish Journal of Medical Physics and Engineering
The Journal of Polish Society of Medical Physics
ISSN 1898-0309, doi: 10.2478/pjmpe-2018-0023
December 2018
Vol 24, Issue 4

Scientific Paper

Gantry angle classification with a fluence map in intensity-modulated radiotherapy for prostate cases using machine learning

Hideharu MIURA[1,2,a], Shuichi OZAWA[1,2], Tsubasa ENOSAKI[1], Masahiro HAYATA[1], Kiyoshi YAMADA[1], Yasushi NAGATA[1,2]
[1]*Hiroshima High-Precision Radiotherapy Cancer Center, 3-2-2, Futabanosato, Higashi-ku Hiroshima 732-0057, Japan*
[2]*Department of Radiation Oncology, Institute of Biomedical & Health Science, Hiroshima University, 1-2-3, Kasumi, Minami-ku, Hiroshima, 734-8551, Japan*
[a]*E-mail address: miura@hiprac.jp*

(received 1 February 2018; revised 10 July and 27 October 2018; accepted 31 October 2018)

Abstract

We investigated the gantry-angle classifier performance with a fluence map using three machine-learning algorithms, and compared it with human performance. Eighty prostate cases were investigated using a seven-field-intensity modulated radiotherapy treatment (IMRT) plan with beam angles of 0°, 50°, 100°, 155°, 205°, 260°, and 310°. The k-nearest neighbor (k-NN), logistic regression (LR), and support vector machine (SVM) algorithms were used. In the observer test, three radiotherapists assessed the gantry angle classification in a blind manner. The precision and recall rates were calculated for the machine learning and observer test. The average precision rate of the k-NN and LR algorithms were 94.8% and 97.9%, respectively. The average recall rate of the k-NN and LR algorithms were 94.3% and 97.9%, respectively. The SVM had 100% precision and recall rates. The gantry angles of 0°, 155°, and 205° had an accuracy of 100% in all algorithms. In the observer test, average precision and recall rates were 82.6% and 82.6%, respectively. All observers could easily classify the gantry angles of 0°, 155°, and 205° with a high degree of accuracy. Misclassifications occurred in gantry angles of 50°, 100°, 260°, and 310°. Machine learning could better classify gantry angles for prostate IMRT than human beings. In particular, the SVM algorithm had a perfect classification of 100%.

Key words: classification; machine learning; fluence map; IMRT.

Introduction

Machine learning is an interdisciplinary field that combines computer science and mathematics to develop models to deliver maximum prediction accuracy. Findings on machine learning for medical physics fields have been reported [1-3]. For instance, Zhu *et al*. developed a planning quality quantitative evaluation tool using a machine-learning approach [2]. Carlson *et al*. used machine-learning techniques to train models to predict discrepancies between planned and delivered movements of multileaf collimators (MLCs), assessed the accuracy of the model predictions, and examined the effects of these errors on quality assurance (QA) procedures and dosimetry [3]. A new assistance tool called knowledge-based planning (for radiation treatment planning), was developed and released for clinical use. Knowledge-based planning is a promising technique that has been demonstrated to improve plan quality and increase planning efficiency [4,5]. Medical physicists need to facilitate the introduction of this technology to the radiotherapy field [6].

Intensity modulated radiotherapy treatment (IMRT) for prostate cancer can improve target coverage, and reduce the organ at risk (OAR) dose relative to a three dimensional conformal radiotherapy (3D-CRT) [7]. An IMRT fluence map is modulated by various optimization parameters during the treatment planning process. In practice, OAR must be avoided using a large number of intensity modulations. To verify the delivered treatment methods using a fluence map, patient-specific QA is partly performed using two-dimensional detector arrays (such as an electronic portal imaging device), prior to treatment [8,9]. A standard beam arrangement and constraint template are used, as the position relationship among the bladder, rectum, and prostate is in the same geometry in each patient. Therefore, the fluence map is quite similar, even if the patients are different. We questioned whether a gantry angle could be classified using machine learning. Image classification using a machine-learning algorithm can identify which category new images belong to according to a training dataset that contains images whose category is known. If a fluence map could be classified using machine learning, it could be used to detect possible inappropriate delivery of the linear accelerator system or for QA purposes.

In this study, we investigated the gantry-angle classifier performance using a machine-learning algorithm and compared it with human performance.

Materials and Methods

Treatment-planning process

The analysis included data for 80 prostate cases. A seven-field coplanar treatment plan (with beam angles of 0°, 50°, 100°, 155°, 205°, 260°, and 310°) was generated by a 6-MV X-ray beam using a Vero4DRT system (Mitsubishi Heavy Industries, Ltd., Hiroshima, Japan, and BrainLAB, Feldkirchen, Germany) with a 5 mm-wide MLC. The properties of the Vero4DRT system are described elsewhere [10]. A mean dose of 74 or 78 Gy was prescribed for the planning target volume (PTV) in 37 or 39 fractions for all patients. The treatment plans were created using an iPlan® RT v. 4.5.3 treatment planning system (TPS) (BrainLAB, Feldkichen, Germany), and the Monte Carlo dose calculation algorithm was used with a spatial resolution of 2.0 mm and a mean variance of 1%. The fluence map of each gantry angle was exported from the TPS (**Figure 1**). The fluence map has a spatial resolution of 1.0 mm and 151 × 151 pixels. Five hundred and sixty fluence maps were used in this study.

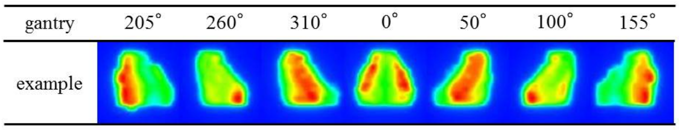


Figure 1. Typical per-field fluence map for each gantry angle for prostate IMRT.

Machine learning

We randomly split our dataset into 420 fluence maps for the training set, and 140 for the test set. All images for machine learning were scaled to a size of 50 × 50 pixels. Down sampling is commonly used to reduce the size of the data to process. The training set represents the set of pixel data and their respective angles that we input to our machine-learning model to learn. For classification in the computer, various classifiers are available that possess different characteristics and features. Three machine-learning algorithms were used in this study: k-nearest neighbor (k-NN), logistic regression (LR), and support vector machine (SVM). All hyper-parameters were tuned using the grid-search cross-validation method to identify the best algorithm. Grid search is the simplest method to train an algorithm. A fivefold cross validation was used to obtain higher cross-validation accuracy. For the k-NN algorithm, k in k-NN is the number of instances taken into account for the determination of affinity with classes. We searched a few possible candidates for k (in the order 1–20) and determined the optimal k = 3. For the LR algorithm, the parameter C is the inverse of regularization strength and is an important factor. We searched a few possible candidates for C (0.001, 0.01, 0.1, 1, 10, and 100) and determined the optimal C = 1.0. For the SVM algorithm, a kernel function was chosen to create the model. The four basic kernels are linear, polynomial, radius basis function, and sigmoid. We used the linear kernel in the SVM algorithm. The test set represents the set of pixel data and their respective angles used to evaluate our model predictions. All scientific computing tasks were performed using Python v.3.6.0 (http://www.python.org). Statistical modeling was performed using SciKit-Learn v.0.18 (http://scikit-learn.org).

We calculated the precision and recall rates of the gantry-angle classifier performance using the following equations:

$$\text{Precision} = \frac{TP}{TP+FP} \quad \text{Eq. 1}$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad \text{Eq. 2}$$

where TP, FN, FP, and TN are true positive, false negative, false positive, and true negative, respectively. The time-measurement function in Python was used to measure the computation time spent for the test data with each machine learning iteration.

Observer test

For the observer test, we used the same total of 140 fluence maps as the test set for machine learning. Three radiotherapists that use the Vero4DRT system in a clinical study independently assessed the gantry angle classification in a blind manner. Informed consent was obtained from all observers. Each observer was given training time and we measured the time spent reading the images. The gantry angle, which was classified by the observers, was written in the answer column. After the gantry angle classification in the fluence map was presented to the three observers, we calculated the precision and recall rates of the gantry angle classifier performance using the above equations.

Results

Table 1 shows the precision and recall rates for the machine learning and observer test. The confusion matrix results for the machine learning and observer test are shown in **Figure 2**. With regard to machine learning, the average precision and recall rates of the k-NN algorithm were 94.8% and 94.3%, respectively. The average precision and recall rates of the LR algorithm were 97.9% and 97.9%, respectively. The SVM algorithm had a perfect classification rate of 100%. The precision rates for the k-NN and LR algorithms ranged from 83.3% to 100% and from 95.2% to 100%, respectively. The recall rates of the k-NN and LR algorithms were from 80.0% to 100% and from 95.0% to 100%, respectively. The gantry angles of 0°, 155°, and 205° displayed an accuracy of 100% in all algorithms.

Figure 3 shows an example of the incorrect classifier predictions using machine learning. Most machine-learning algorithm errors were caused by a misclassification in gantry angles of 50°, 100°, 260°, and 310°. The predicted times per a set of 140 fluence maps for the k-NN, LR, and SVM algorithms were 0.093, 0.001, and 0.027 s, respectively.

With regard to the observer test, the average precision and recall rates were 82.6% and 82.6%, respectively. The precision rate by the observers ranged from 64.2% to 98.4%, respectively. The average recall rate by the observers ranged from 56.7% to 100%, respectively. All observers easily classified the gantry angles of 0°, 155°, and 205° in our test set with high accuracy. Similar to machine learning, most observer errors came from the misclassification of gantry angles of 50°, 100°, 260°, and 310°. The average time for the observers was 776 s.

Table 1. Precision and recall rates for each angle using k-NN, LR, and SVM algorithms and observer test.

gantry angle (°)	k-NN		LR		SVM		observer	
	precision (%)	recall (%)	precision (%)	recall (%)	precision (%)	recall (%)	precision (%)	recall (%)
0	95.2	100.0	100.0	100.0	100.0	100.0	98.4	100.0
50	83.3	100.0	95.2	100.0	100.0	100.0	64.2	56.7
100	100.0	80.0	100.0	95.0	100.0	100.0	64.7	73.3
155	100.0	100.0	95.2	100.0	100.0	100.0	98.3	96.7
205	100.0	100.0	100.0	100.0	100.0	100.0	96.7	98.3
260	94.7	90.0	100.0	95.0	100.0	100.0	79.0	81.7
310	90.0	90.0	95.0	95.0	100.0	100.0	76.8	71.7
average	94.8	94.3	97.9	97.9	100.0	100.0	82.6	82.6

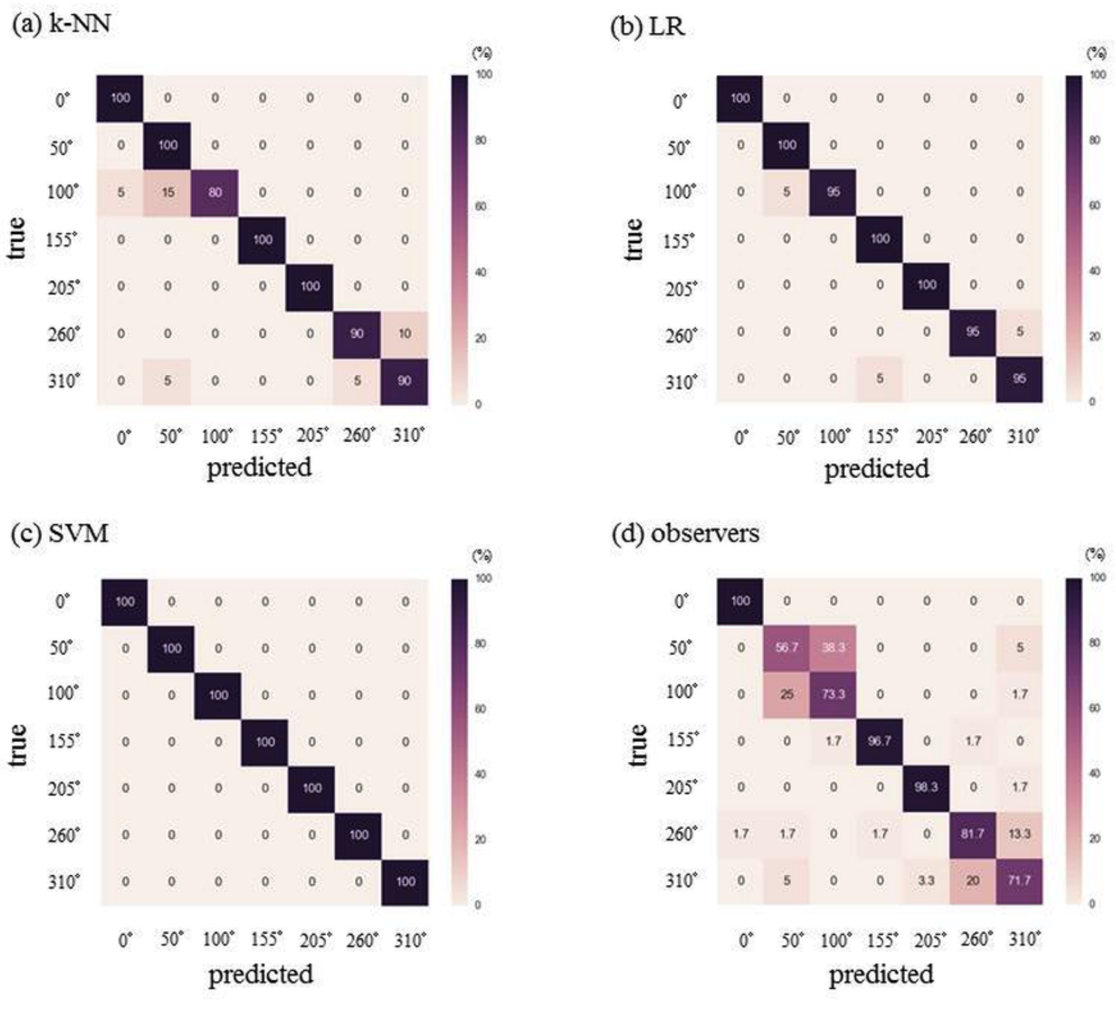
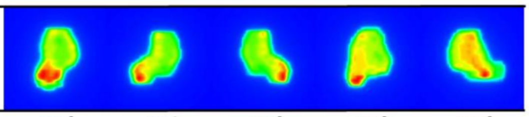


Figure 2. Confusion matrix for (a) k-NN, (b) LR, (c) SVM algorithms, and (d) observers for each gantry angle. The column consists of the true labels, and the predicted labels are shown in the rows.



True	100°	100°	260°	100°	310°
Predicted	50°	50°	310°	50°	260°
Misclassification algorithm	LR	LR K-NN	LR K-NN	LR	K-NN

Figure 3. Incorrect fluence map test set predictions using k-NN, LR, and SVM algorithms. The k-NN and LR algorithms caused the misclassification in gantry angles of 50°, 100°, 260°, and 310°. The SVM algorithm could correctly classify all gantry angles.

Discussions

We reported the gantry-angle classification performance with fluence maps using three machine-learning algorithms. Furthermore, we compared the results of the observer test with those using machine learning. The performance results of the three machine-learning algorithms all differed from one another. The kNN algorithm is simplest instance based learning method used to classify objects based on their closest training examples in the feature space. LR algorithm focuses on maximizing the probability of the data. SVM algorithm tries to find the widest possible separating margin of the data. All machine-learning algorithm results in this study significantly outperformed the observer test results. Moreover, the observer test achieved a poor average recall rate of 83.0%, while the best performing machine-learning approach, SVM, achieved perfect 100% performance. The gantry angle with a fluence map for prostate IMRT could be easily classified by SVM machine learning. In the observer test, fluence maps with 0°, 155°, and 205° gantry angles were easily classified because modulation is produce dose distribution sparing the rectum. In the other gantry angles, the target or OAR shapes, volumes, patient sizes, and optimization weight conditions were more difficult to classify in terms of visual appearance. Some fluence maps could not even be classified by the k-NN and LR machine-learning algorithms. A large number of the machine learning misclassifications were due to mistakes at 50° for 100° and 260° for 310° (and vice versa), as these fluence maps are very similar (**Figure 3**) and depend on rectum, bladder, and prostate sizes. Thus, the fluence map pattern varies slightly vary from patient to patient. We should emphasize that the classification of these gantry angles is a very difficult and time consuming task for human beings. When compared to the human observer prediction time, the k-NN, LR, and SVM algorithms had nearly instantaneous prediction times, which were primarily dependent on the computer execution environment.

The SVM algorithm perfectly classified the gantry angle for prostate IMRT in this study. The SVM algorithm is widely used and well known in the machine-learning field. Several authors have used the SVM algorithm to predict radiation pneumonitis after chemotherapy [11], local control after lung stereotactic body radiotherapy [12], and chemoradiosensitivity in esophageal cancer [13]. Machine learning plays an essential role in medical image analysis and computer-aided diagnosis because accurately representing the lesions and organs in medical images may be to complex to understand using only a simple equation [14].

A question arises on how to introduce our results to the medical physics fields. In the current system, the IMRT plan is created by a medical physicist. Semi-automated planning algorithms have been recently used to improve the overall plan quality and consistency, and to decrease the time required for planning [15]. Machine create treatment plans may be able to use a similarity fluence map to enhance the reliability of treatment planning.

A limitation of the current study is that we used only k-NN, LR, and SVM algorithms for machine-learning. SciKit-Learn has additional sets of statistical learning algorithms. For future work, we hope to improve machine learning accuracy using more sophisticated classifiers.

Conclusion

We have investigated the gantry angle classification performance with a fluence map using three machine learning algorithms, and compared their performance with that of human beings. The SVM algorithm achieved the best correct recognition rate of 100%, followed by the LR and k-NN algorithms with a near 95% accuracy. The precision and recall rates by machine learning in all gantry angles were higher than those by the observers. This study shows that machine learning can better classify gantry angles with a fluence map than human beings.

References

[1] Garapati SS, Hadjiiski L, Cha KH, *et al*. Urinary bladder cancer staging in CT urography using machine learning. Med Phys. 2017;44(11):5814-5823.

[2] Zhu X, Ge Y, Li T, *et al*. A planning quality evaluation tool for prostate adaptive IMRT based on machine learning. Med Phys. 2011;38(2):719-726.

[3] Carlson JN, Park JM, Park SY, *et al*. A machine learning approach to the accurate prediction of multi-leaf collimator positional errors. Phys Med Biol. 2016;61(6):2514-2531.

[4] Zhu X, Ge Y, Li T, *et al*. A planning quality evaluation tool for prostate adaptive IMRT based on machine learning. Med Phys. 2011;38(2):719-26.

[5] Chanyavanich V, Das SK, Lee WR, *et al*. Knowledge based IMRT treatment planning for prostate cancer. Med Phys. 2011;38(5):2515-2522.

[6] Wang G, Kalra M, Orton CG. Machine learning will transform radiology significantly within the next 5 years. Med Phys. 2017;44(6):2041-2044.

[7] Michalski JM, Yan Y, Watkins-Bruner D, *et al*. Preliminary toxicity analysis of 3-dimensional conformal radiation therapy versus intensity modulated radiation therapy on the high-dose arm of the Radiation Therapy Oncology Group 0126 prostate cancer trial. Int J Radiat Oncol Biol Phys. 2013;87(5):932-938.

[8] Jursinic PA, Nelms BE. A 2-D diode array and analysis software for verification of intensity modulated radiation therapy delivery. Med Phys. 2003;30(5):870-879.

[9] van Zijtveld M, Dirkx ML, de Boer HC, *et al*. Dosimetric pre-treatment verification of IMRT using an EPID; clinical experience. Radiother Oncol. 2006;81(2):168-175.

[10] Kamino Y, Takayama K, Kokubo M, *et al*. Development of a four-dimensional image-guided radiotherapy system with a gimbaled X-ray head. Int J Radiat Oncol Biol Phys. 2006;66(1):271-278.

[11] Chen S, Zhou S, Yin FF, *et al*. Investigation of the support vector machine algorithm to predict lung radiation-induced pneumonitis. Med Phys. 2007;34(1):3808-3814.

[12] Klement RJ, Allgäuer M, Appold S, *et al*. Support vector machine-based prediction of local tumor control after stereotactic body radiation therapy for early-stage non-small cell lung cancer. Int J Radiat Oncol Biol Phys. 2014;88(3):732-738.