Journal of Radiotherapy in Practice

Journal of Radiotherapy in Practice (2017)
16, 161–170 © Cambridge University Press 2017
doi:10.1017/S1460396916000583

Original Article

Predictive modelling analysis for development of a radiotherapy decision support system in prostate cancer: a preliminary study

Kwang Hyeon Kim[1,2], Suk Lee[2], Jang Bo Shim[2], Kyung Hwan Chang[5], Yuanjie Cao[6], Suk Woo Choi[3], Se Hyeong Jeon[4], Dae Sik Yang[2], Won Sup Yoon[2], Young Je Park[2], Chul Yong Kim[2]

[1]*Department of Biomedical Science, Graduate School of Korea University,* [2]*Department of Radiation Oncology, College of Medicine, Korea University,* [3]*Asan Medical Center, Seoul, South Korea,* [4]*Department of Radiation Oncology, Tianjin Medical University Cancer Institute and Hospital, Tianjin, China,* [5]*CQURE Healthcare, Seoul,* [6]*Innotems, Daejeon, South Korea*

(Received 27 October 2016; revised 23 October 2016; accepted 23 October 2016; first published online 31 January 2017)

Abstract

Purpose: The aim of this study is to develop predictive models to predict organ at risk (OAR) complication level, classification of OAR dose-volume and combination of this function with our in-house developed treatment decision support system.

Materials and methods: We analysed the support vector machine and decision tree algorithm for predicting OAR complication level and toxicity in order to integrate this function into our in-house radiation treatment planning decision support system. A total of 12 TomoTherapy™ treatment plans for prostate cancer were established, and a hundred modelled plans were generated to analyse the toxicity prediction for bladder and rectum.

Results: The toxicity prediction algorithm analysis showed 91·0% accuracy in the training process. A scatter plot for bladder and rectum was obtained by 100 modelled plans and classification result derived. OAR complication level was analysed and risk factor for 25% bladder and 50% rectum was detected by decision tree. Therefore, it was shown that complication prediction of patients using big data-based clinical information is possible.

Conclusion: We verified the accuracy of the tested algorithm using prostate cancer cases. Side effects can be minimised by applying this predictive modelling algorithm with the planning decision support system for patient-specific radiotherapy planning.

Keywords: predictive modelling; prostate cancer; radiation treatment planning decision support program (PDSS); radiation treatment planning (RTP) system; toxicity

Correspondence to: Suk Lee, PhD, Department of Radiation Oncology, Korea University Medical Center, 126-1, Anamdong, Seongbukgu, 02841 Seoul, Korea. Tel: +82-2-920-5519, Fax: +82-2-927-1419, E-mail: sukmp@korea.ac.kr

INTRODUCTION

A number of treatment plans are generated for each patient in order to establish the optimal radiation treatment plan. The final treatment plan is selected by applying a quantitative analysis method by determining the delineation shape of the planning target volume, organ at risk (OAR) and a qualitative analysis method based on the dose volume histogram (DVH).

However, there is no guarantee that the treatment plan selected by this analysis and evaluation will not cause radiotherapy side effects in the patient. Therefore, if the radiation oncologists and medical physicists consider historical clinical data on complications with typical treatment plan factors, such as DVH and OAR dose constraint range, they can establish the optimal treatment plan minimising OAR concerns by suppressing the normal tissue complication probability (NTCP) and increasing the tumour control probability (TCP).

Current advances in diagnostic and therapeutic technologies are under research and development through innovative tools that combine oncology, diagnostics, genetics and computer science to improve the quality of life of patients after treatment. Thus, the clinical decision support system is also being researched using clinical big data with application of prediction modelling using a machine learning algorithm. The role of the prediction model in the radiation treatment decision support system is to maximise tumour control and minimise side effects after treatment and to determine whether the plan offers acceptable dose risk by applying classification and regression methods to dose-volume data using an existing clinical database.[1]

Table 1. *Characteristics of the patients with prostate cancer*

Patients	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Clinical diagnosis						Prostate cancer						
PD (Gy)	77	77	77	77	77	77	77	77	77	77	77	77
FD (Gy)	2·2	2·2	2·2	2·2	2·2	2·2	2·2	2·2	2·2	2·2	2·2	2·2
Fraction	35	35	35	35	35	35	35	35	35	35	35	35
Age	73	75	79	73	79	68	75	69	76	57	64	78
TNM stage	T1c	T3b	T2c	T1c	T3a	T3a	T3b	T1c	T1c	T2b	T3a	T3b
	N0	N0	N0	N0	N0	N0	N0	N0	N0	N0	N0	N0
	M0	M0	M0	M0	M0	M0	M0	M0	M0	M0	M0	M0
Histological diagnosis						Adenocarcinoma						
OP	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Chemotherapy	Y	Y	Y	Y	Y	Y	Y	N	N	Y	Y	Y
Weight (kg)	60·50	95·60	80·00	75·00	65·45	80·70	78·00	81·00	69·10	52·65	72·20	97·95

Note: To establish original treatment plans for patients (n = 12).
Abbreviations: PD, prescription dose; FD, fractional dose; OP, operation; TNM, tumour-node-metastasis.

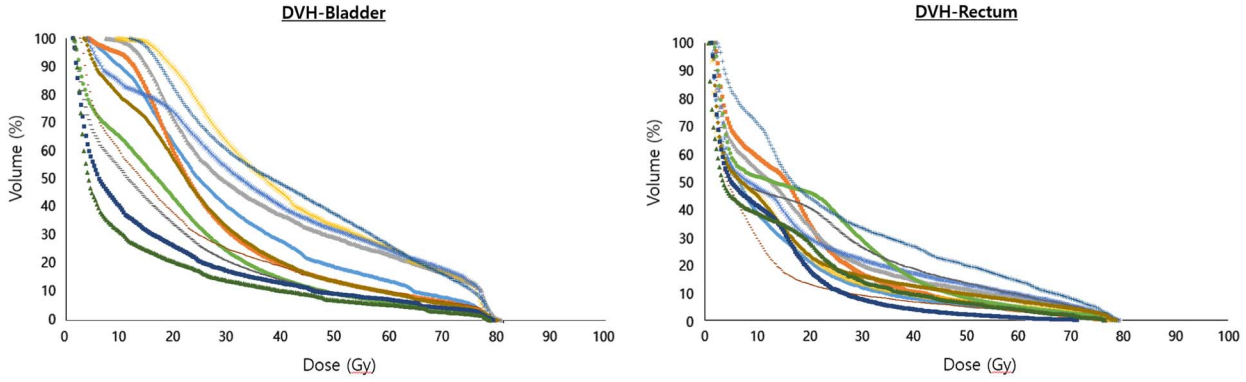

Figure 1. Established dose volume histogram (DVH) of bladder and rectum for 12 patients with prostate cancer (n = 12).

https://doi.org/10.1017/S1460396916000583 Published online by Cambridge University Press

Table 2. *Dose-volume constraints for bladder and rectum as organs at risk*

	Prostate			
Volume	25% bladder	50% bladder	25% rectum	50% rectum
Dose (Gy)	29–30	25–26	27–28	24–25
Maximum dose (Gy)	<80			

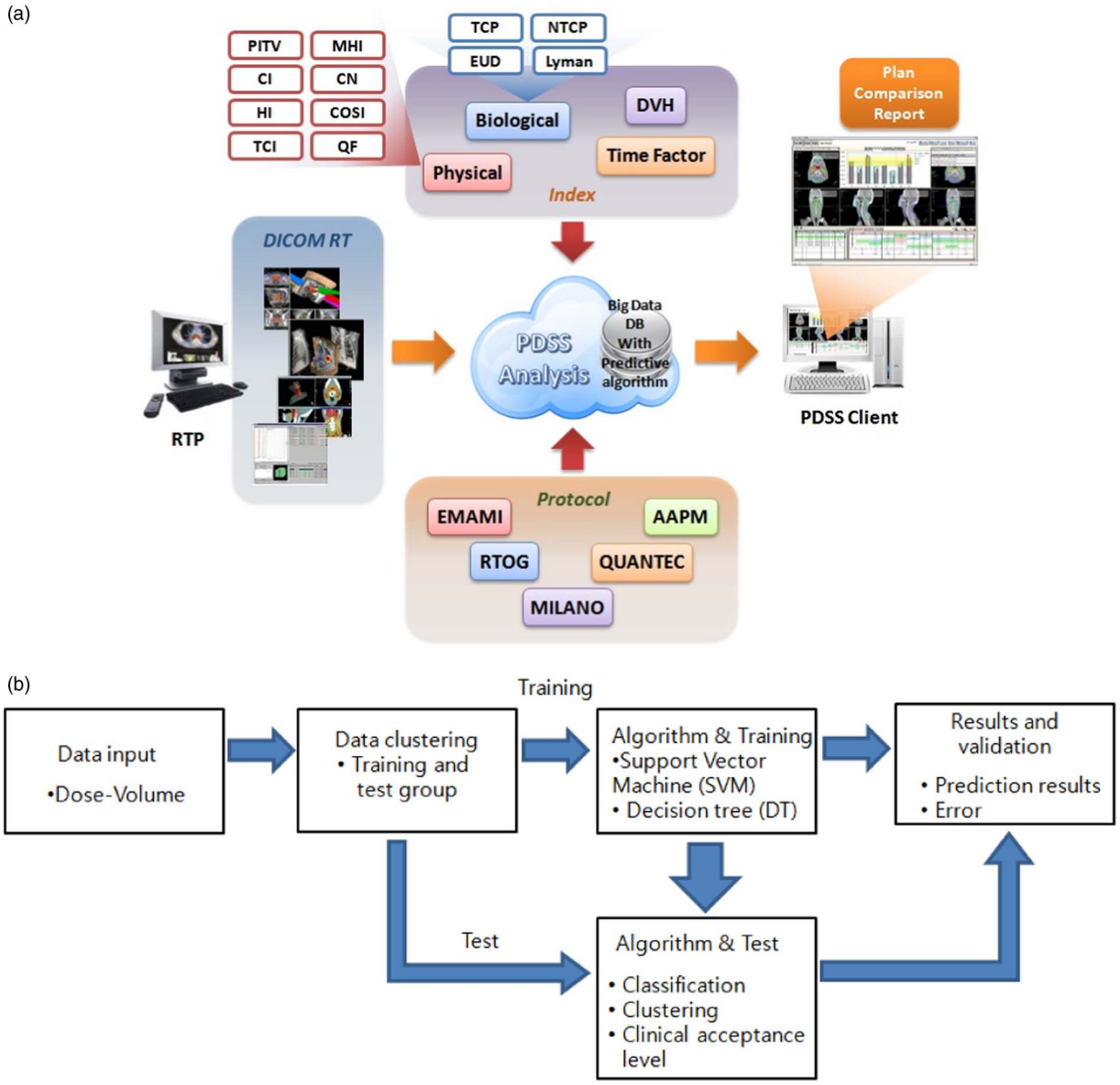

Figure 2. Model of a radiation treatment planning decision support system (a) and its predictive modelling flow chart for the support vector machine algorithm (b) of this study.
Abbreviations: PITV, prescription isodose to target volume; CI, conformity index; HI, homogeneity index; TCI, target coverage index; MHI, modified homogeneity index; CN, conformity number; COSI, critical organ scoring index; QF, quality factor

https://doi.org/10.1017/S1460396916000583 Published online by Cambridge University Press

Table 3. Classification of bladder and rectum complications by 100 modelled plans for the support vector machine algorithm (n = 100)

Modelled plan	Dose (Gy)		Complication	Modelled plan	Dose (Gy)		Complication
	Bladder	Rectum			Bladder	Rectum	
1	80·0377	77·9819	NC	51	80·3914	78·3167	NC
2	79·6505	78·9365	NC	52	79·3221	79·8448	NC
3	76·1811	79·2330	NC	53	78·0480	78·1449	NC
4	77·6200	78·9067	NC	54	80·1392	79·7010	NC
5	77·4846	78·8788	NC	55	77·2004	79·8897	NC
6	78·5789	79·4210	NC	56	79·0379	78·6822	NC
7	78·8271	79·6142	C	57	79·6613	79·5449	NC
8	77·1987	78·7761	C	58	79·6883	78·4031	NC
9	78·4139	79·2300	NC	59	76·4638	78·9204	NC
10	76·9016	79·7037	C	60	78·1089	78·6489	NC
11	80·1479	78·6796	NC	61	77·3774	78·2692	NC
12	76·4127	79·0802	NC	62	78·3918	79·6447	NC
13	78·6748	79·3314	C	63	77·8309	78·2732	NC
14	80·8180	78·0694	C	64	76·1092	79·0914	NC
15	80·1506	78·1815	NC	65	80·4119	79·9812	NC
16	80·5519	79·1045	C	66	80·7564	79·3584	NC
17	79·8819	78·8390	NC	67	76·5465	79·8955	NC
18	76·8203	78·8539	NC	68	79·8080	79·7441	NC
19	80·2812	78·9812	C	69	76·6590	78·4122	NC
20	77·9303	78·0680	C	70	76·1768	79·6625	NC
21	79·8707	79·7474	NC	71	76·2645	78·8370	NC
22	80·0537	78·3673	C	72	78·1316	78·5324	NC
23	77·8346	78·3078	NC	73	78·8725	79·9957	NC
24	78·0072	78·5484	NC	74	76·3180	79·9355	NC
25	78·3074	79·0777	NC	75	80·3172	79·7227	NC
26	77·1998	79·9182	NC	76	78·6868	78·5264	NC
27	77·8006	79·3755	NC	77	76·9282	79·6660	NC
28	80·9858	79·6930	NC	78	80·4061	78·4209	NC
29	80·4397	79·2423	NC	79	79·9404	79·9549	NC
30	80·5549	78·8034	NC	80	79·7061	79·2188	NC
31	78·3315	79·7445	NC	81	79·9936	78·4416	NC
32	77·8646	78·8211	NC	82	78·8844	79·4065	NC
33	76·7901	79·9276	NC	83	78·4796	79·5722	NC
34	76·7348	78·5325	NC	84	77·2741	79·1082	NC
35	77·6375	79·7284	NC	85	76·3125	79·3722	NC
36	76·7663	79·4000	NC	86	77·0770	78·5326	NC
37	76·3330	78·5323	NC	87	78·6952	79·6816	NC
38	79·4022	79·3970	NC	88	80·2204	78·0121	NC
39	78·1475	79·3029	NC	89	77·6894	79·8281	NC
40	79·2097	78·1327	NC	90	77·4124	78·4549	NC
41	79·8559	79·6310	NC	91	78·3786	78·6659	NC
42	78·1151	78·8874	NC	92	78·1064	78·1041	NC
43	80·2623	78·5010	NC	93	78·1939	79·6165	NC
44	80·1614	78·6939	NC	94	76·5212	79·1112	NC
45	76·1559	79·3378	NC	95	77·2312	79·4407	NC
46	77·3598	78·1836	NC	96	77·7528	79·4644	NC
47	77·4777	79·1634	NC	97	80·7835	78·4271	NC
48	80·8127	79·6918	NC	98	80·5662	79·4653	NC
49	76·7704	78·5975	NC	99	77·2284	79·5797	NC
50	79·3743	79·9104	NC	100	79·6497	79·2562	NC

Abbreviation: NC, non-complication.

With this aim, Zhang et al. recently studied complication prediction of radiation therapy using machine learning in the field of radiation therapy.[2] In addition, Guidi et al. are also making progress in studies to predict criticality in which machine learning algorithms target particular cases, such as patients with head and neck cancer.[3]

https://doi.org/10.1017/S1460396916000583 Published online by Cambridge University Press

However, no known published studies have integrated machine learning algorithms and dosimetrical and biological index analysis functions into a radiation treatment planning decision support system to determine the optimal plan.

Table 4. Complication prediction for bladder and rectum using 20 representative plans for the decision tree algorithm (n = 20)

Plan	25% bladder	50% bladder	25% rectum	50% rectum	Complication
1	42·5103	24·6570	17·7323	6·5689	NC
2	34·6096	22·5537	24·0404	15·4760	NC
3	56·6986	28·8961	25·0143	12·5988	NC
4	60·6693	36·7794	18·6604	6·6963	NC
5	59·8209	32·3455	25·1762	8·1739	NC
6	29·4472	16·9597	32·3111	12·7246	NC
7	61·2788	38·0032	41·8223	16·5369	C
8	30·1537	13·7383	11·3235	4·1326	NC
9	25·6640	11·3204	31·2221	6·6586	NC
10	35·3631	22·2015	18·8032	7·0208	NC
11	21·1052	6·1848	17·3420	5·0307	NC
12	14·7914	4·2515	21·4121	3·3405	NC
13	59·3810	59·0841	50·2707	50·6603	C
14	58·5683	42·6852	59·5842	55·8519	C
15	55·0244	54·8645	55·2236	55·3067	NC
16	56·3760	47·7500	54·0928	41·2431	NC
17	58·9306	49·8144	50·9266	55·7675	NC
18	44·3577	54·6872	46·4401	53·3388	C
19	51·3241	42·5009	44·9372	56·5772	C
20	45·1958	42·4967	50·8894	56·3337	NC

Abbreviation: NC, non-complication; C, complication.

Cao et al. performed integrated analysis studies of dosimetrical and biological index data using prostate cancer cases.[4,5] In addition, big data analysis studies of prostate cancer using machine learning approaches have been performed.[6–8] According to Çınar et al., prostate cancer is currently the most common type of cancer in men except lung cancer.[6] Therefore, we used prostate cancer cases as the model system.

The aim of this study was to develop a predictive model solution that includes the functions of support vector machine and decision tree algorithm to predict OAR complication level and suitable classification of OAR dose-volume values and to combine this function with an in-house developed treatment decision support system in a preliminary study.

MATERIAL AND METHODS

Patient group

The target population was 12 male patients with adenocarcinoma of the prostate, for whom 12 treatment plans had been established. The patient characteristics are as follows: average age, 72 years; average weight, 75·68 kg; tumour-node-metastasis (TNM) stage, T1c–T3b, N0 and M0 (Table 1). The treatment planning system used was TomoTherapy® (Accuray Incorporated, Sunnyvale, CA, USA).

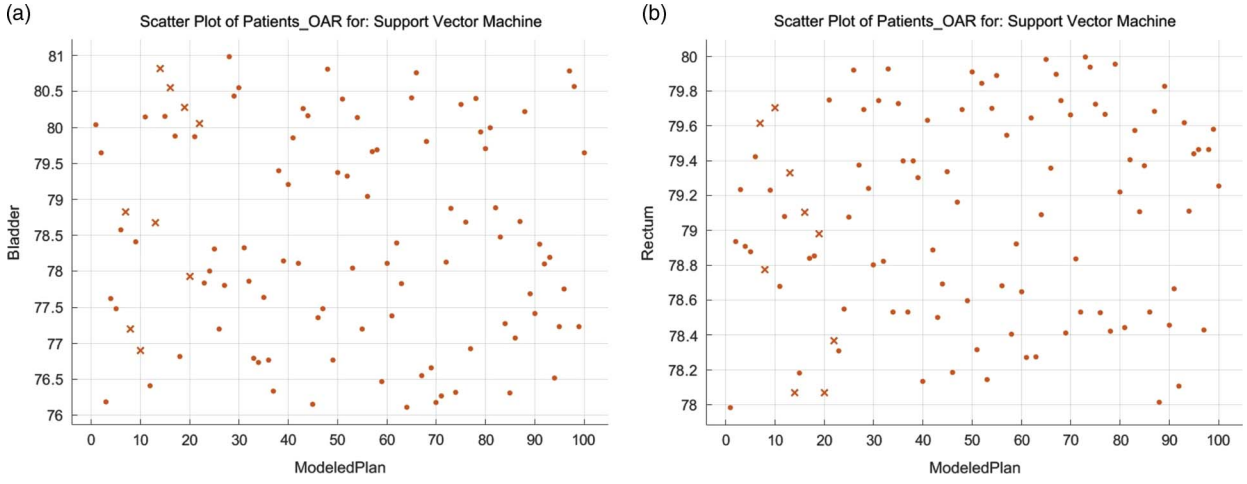

Figure 3. Scatter plot of patient organs at risk (OARs) with the support vector machine for 100 modelled plans.
Note: red dot (·): correctly classified as NC; red cross (x): misclassified as NC. (a) Bladder scatter plot with modeled plans; (b) rectum scatter plot with modeled plans.
Abbreviation: NC, non-complication.

https://doi.org/10.1017/S1460396916000583 Published online by Cambridge University Press

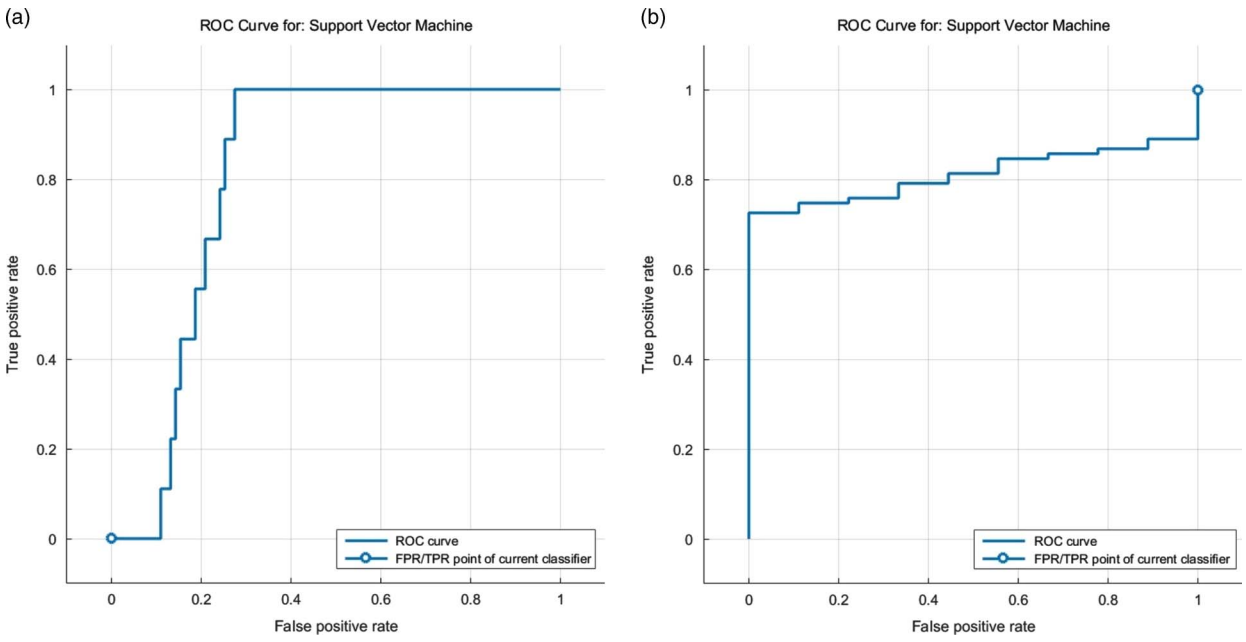

Figure 4. Receiver operating characteristic (ROC) of the classifier with the support vector machine.
Note: Area under curve (AUC) = 0·8107, (a) = positive class for complication, (b) = positive class for non-complication.

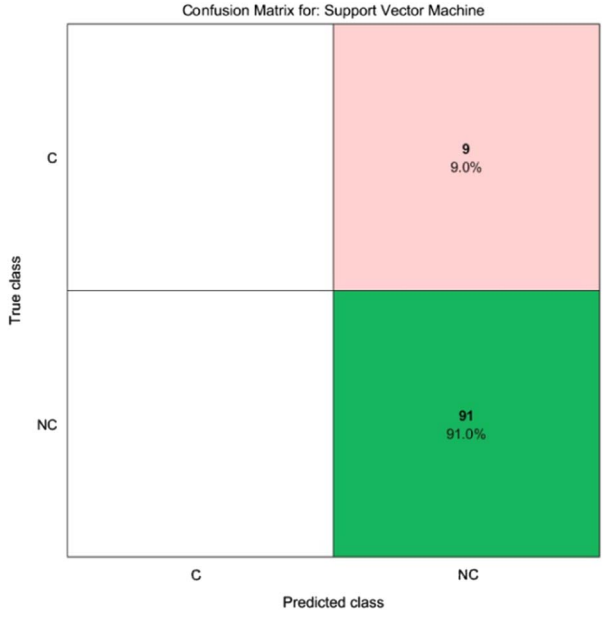

Figure 5. Confusion matrix for support vector machine analysis and the predicted class for non-complication (NC) with 91·0% accuracy.

The DVHs for the OARs of bladder and rectum are shown in Figure 1.

We developed an in-house planning decision support programme to input DVH information from the treatment plan and integrated the results of this study into our system. As an example, we used dose-volume data of the OAR to predict complications as a constraint value (Table 2).

Predictive modelling using machine learning algorithm

The machine learning algorithm for the radiation treatment planning decision support system can be used in the prediction of complications in OARs exposed to radiation as the peripheral target during radiation therapy.[9] That is, the predictive modelling algorithm calculates the results using comprehensive data in accordance with the state of the indicator characteristic of the patients and treatment plans.

Accordingly, there is a need to verify the results of late toxicity through a decision tree model or dose-volume data analysis based on current knowledge and historical clinical outcomes.

A prediction model can be applied using index data such as age, TNM stage, gender, prescribed dose, tumour control probability and survival rate.[10] In addition, the support vector machine (SVM) algorithm can be applied to classify the different OAR dose constraints.[2] The DVH of the patients during radiation therapy is a significant predictive indicator.[10]

https://doi.org/10.1017/S1460396916000583 Published online by Cambridge University Press

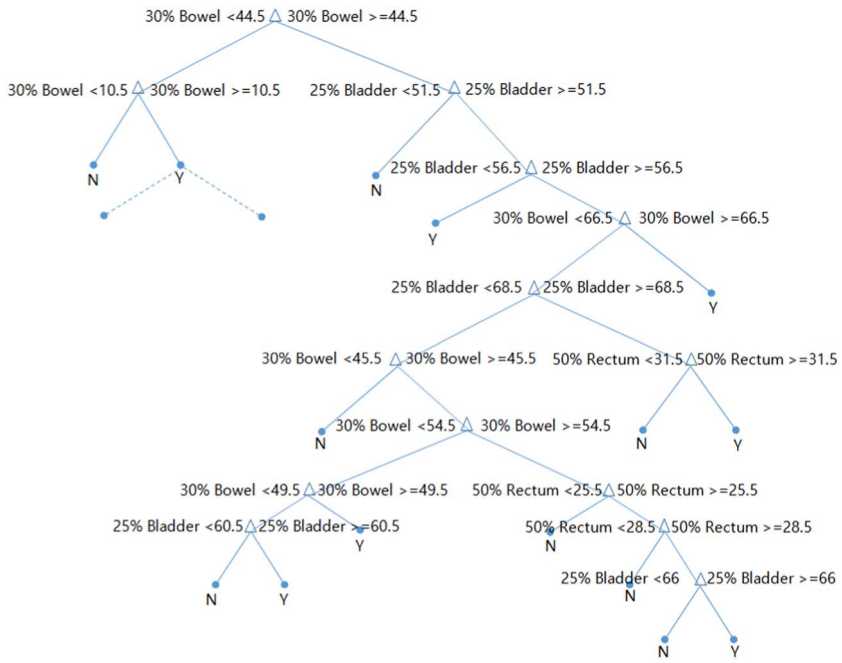

Figure 6. Decision tree for grade 2 rectal complication classification for 100 plans with 25% bladder, 50% rectum, 30% bowel.

Therefore, we used the DVH data of patients as input parameters for the application of clinical big data, and machine learning techniques were used in the SVM and decision tree as described previously for complication prediction.[2]

Support vector machine

The algorithm is developed to select the best classifier to separate two groups by drawing a perpendicular line between groups in the hyperplane. In the case of the nonlinear model, the kernel method is used to distinguish the linear machine.[9]

A hyperplane is defined as the set of all points $x \in R_{\text{dimension}}$ that satisfy $h(x) = 0$, where $h(x)$ is the function of the hyperplane, as follows in equation (1) in d dimensions:[7,8]

$$h(x) = w^T x + b \tag{1}$$

In this study, we modelled the SVM algorithm using dose-volume input in test and training models as shown in the flow chart in Figure 2. Figure 2a shows the entire analysis system from the treatment planning data, including quantitative analysis for homogeneity index, conformity index and conformation number and dosimetrical indices, TCP, NTCP of biological indices in addition to a big data-based prediction algorithm, to the results. The predictive algorithm component is further defined as in Figure 2b, which describes how dose-volume data for every patient is used as the input and the training and test processes involved in the SVM for classification to achieve an accurate final outcome.

A total of 100 model plans were generated based on 12 treatment plans for analysis of the support vector machine algorithm (Table 3).

Decision tree

A decision tree requires that critical decision points be selected for outgoing confirmation based on specific conditions by selecting a final value with these conditions. This can be formalised by a simple pattern and is an algorithm that can be programmed using machine learning.

The decision point $X_j \leq v$ divides the input data space, R, into two sections: R_Y and R_N. The division of R into R_Y and R_N also derives a binary section of the corresponding input data point D_{Input}. This means that a division point of the form $X_j \leq v$ derives the data into sections, as in equations (2) and (3).

$$D_Y = \{x | \, x \in D_{\text{Input}}, x_j \leq v\} \tag{2}$$

$$D_N = \{x \, | \, x \in D_{\text{Input}}, x_j > v\} \tag{3}$$

https://doi.org/10.1017/S1460396916000583 Published online by Cambridge University Press

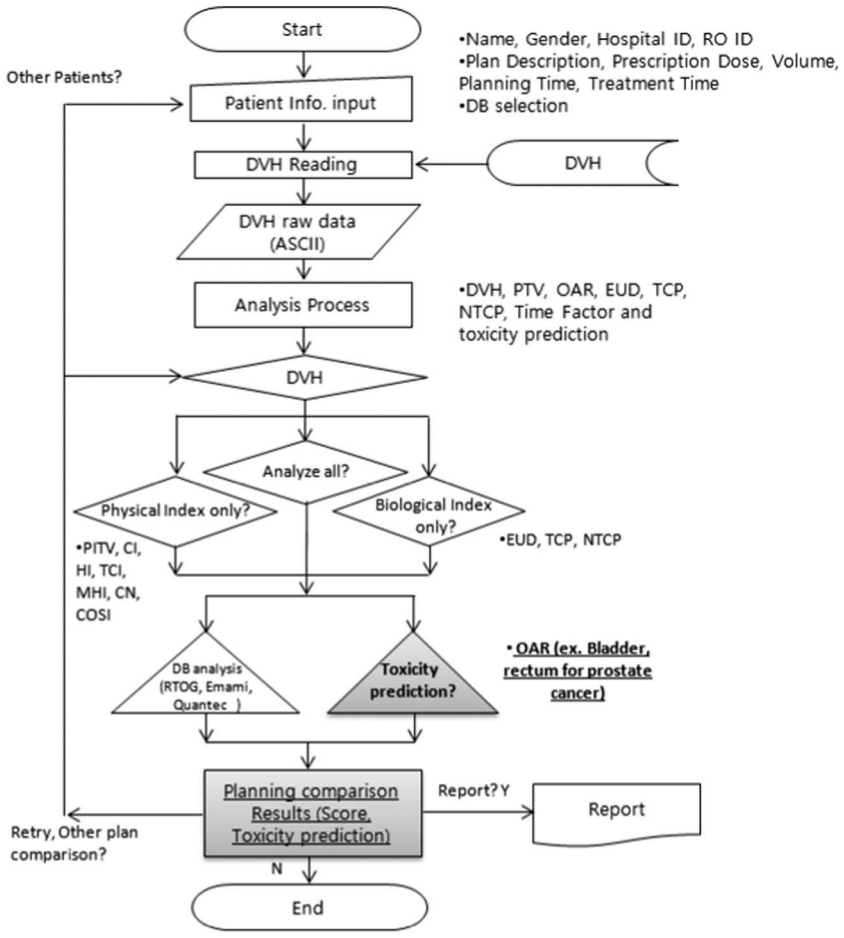

Figure 7. Integrated flow chart of toxicity prediction, dosimetric biological index analysis and overall factors for SMART^{RT}.
Abbreviation: DVH, dose volume histogram; PTV, planning target volume; OAR, organ at risk; TCP, tumour control probability; NTCP, normal tissue complication probability; PITV, prescription isodose to target volume; CI, conformity index; HI, homogeneity index; TCI, target coverage index; MHI, modified homogeneity index; CN, conformity number; COSI, critical organ scoring index; RO, radiation oncology; DB, database; RTOG, radiation therapy oncology group; EUD, equivalent uniform dose.

where D_Y is the group of data points lying in region R_Y, and D_N is the group of input points lying in R_N.[8]

To analyse the decision tree algorithm, we calculated an additional eight plans based on the 12 original treatment plans, expanding the analysis to 20 plans. Table 4 shows the doses (Gy) of 25% bladder, 50% bladder, 25% rectum and 50% rectum using this prediction.

RESULTS

The results of analysis with the machine learning algorithm showed 91·0% accuracy after the training process with respect to 100 modelled plans using SVM.

In addition, the OAR complication analysis showed possible classification of potential risk factors as complication (C) and non-complication (NC) relative to 25% bladder, 50% rectum and 30% bowel using the decision tree. Therefore, we could combine a programme including this machine learning algorithm and our in-house developed planning decision support system to allow complication predictions for patients based on clinical big data.

Predictive modelling analysis results

SVM

Figure 3 shows the results of classification analysis for bladder and rectum with respect to the 100-model plan. Quadratic SVM analysis correctly separated NC cases: red dots in Figure 3

https://doi.org/10.1017/S1460396916000583 Published online by Cambridge University Press

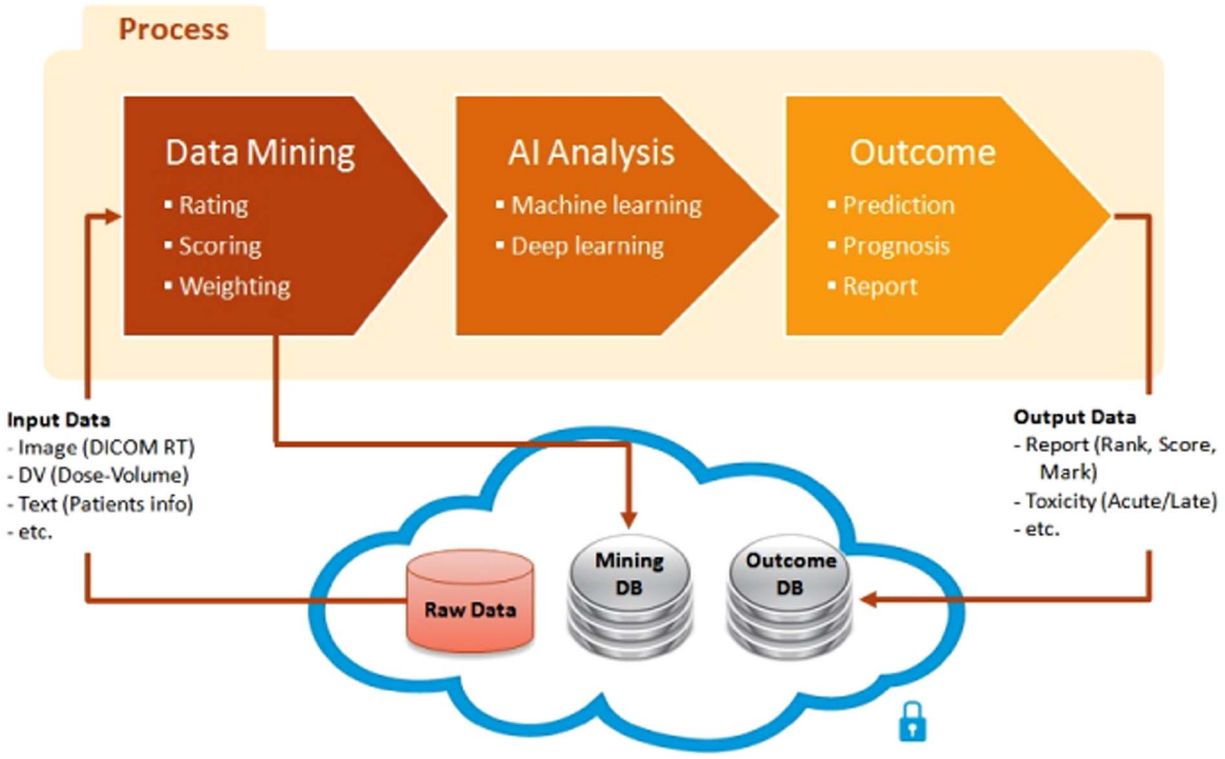

Figure 8. The artificial intelligence (AI)-based integrated clinical decision support system of this study. Abbreviation: DICOM RT, Digital Imaging and Communications in Medicine Radiation Therapy.

indicate correct classification, and red crosses show misclassified NC. The true positive rate and false positive rate were obtained to demonstrate the performance of the SVM classifier for the analysis and showed an area under curve of 0·8107 (Figure 4).

In addition, confusion matrix analysis was performed to calculate the error matrix, showing a 91% rate of accuracy for the predicted class and true class (Figure 5).

Decision tree

As a result of the decision tree analysis, complication prediction was possibly based on the dose of 25% bladder, 50% rectum and 30% bowel. When the radiation oncologists and medical physicists decide the final treatment plans before radiation therapy, the dose constraint for every OAR makes it complicated to determine an optimal plan; thus, a method considering these complex factors would be a useful analytical tool to predict complications (Figure 6).

Integration with SMART^{RT}

SMART^{RT} is an in-house radiation treatment planning decision support system (PDSS) that was developed to give a final scoring scheme that included DVH information for the patient from the treatment plan and functions with dosimetrical and biological index analysis through the overall quality factor result. However, if the toxicity prediction function is added into the SMART^{RT} programme using clinical big data and comprehensive clinical side effects could be linked to solve complication prediction, we might be able to achieve the optimal patient-specific PDSS (Figure 7).

DISCUSSION AND CONCLUSION

To improve the quality of life of the patient after treatment, more accurate patient treatment plans are needed in the field of radiation oncology. This should allow more accurate prognosis of patient outcomes after treatment. This issue is being addressed by planning decision support system

https://doi.org/10.1017/S1460396916000583 Published online by Cambridge University Press

research using fundamental DVH analysis, as well as dosimetrical and biological indices with TCP and NTCP.[4,11–13]

The treatment plan has to be compared more accurately according to the optimum patient-specific PDSS, which includes a predictive model-based function and will represent a significant breakthrough in patient care through machine learning research that can be linked to clinical big data (Figure 7). We present the total artificial intelligence-based integrated clinical decision support system of this study in Figure 8. This system includes an intelligent clinical decision support algorithm with machine learning and deep learning as the artificial intelligence system using clinical big data that could be further expanded.

Machine learning analysis-based studies with clinical cases in radiation oncology are being researched.[14–18] Therefore, it seems likely that more patient cases and multi-institutional studies will be compiled to increase the amount of training data and provide more accurate results. This will be the foundation for the development of optimal patient-specific PDSS for prognosis.

Acknowledgement

This work was supported by a Korea University grant.

Conflicts of Interest

None.

References

1. Meldolesi E, van Soest J, Damiani A et al. Standardized data collection to build rediction models in oncology: a prototype for rectal cancer. Future Oncol 2016; 12 (1): 119–136.
2. Zhang H H, D'Souza W D, Shi L, Meyer R R. Modeling plan-related clinical complications using machine learning tools in a multiplan IMRT framework. Int J Radiat Oncol Biol Phys 2009; 74 (5): 1617–1626.
3. Guidi G, Maffei N, Vecchi C et al. A support vector machine tool for adaptive tomotherapy treatments: prediction of head and neck patients criticalities. Phys Med 2015; 31 (5): 442–451.
4. Cao Y J, Lee S, Chang K H et al. Patient performance-based plan parameter optimization for prostate cancer in tomotherapy. Med Dosim 2015; 40 (4): 285–289.
5. Cao Y J, Lee S, Chang K H et al. Optimized planning target volume margin in helical tomotherapy for prostate cancer: is there a preferred method? J Kor Phys 2015; 67 (1): 26–32.
6. Çınar M, Engin M, Engin E Z, Ateşçi Y Z. Early prostate cancer diagnosis by using artificial neural networks and support vector machines. Exp Sys App 2009; 36: 6357–6361.
7. De Bari B, Vallati M, Gatta R et al. Could machine learning improve the prediction of pelvic nodal status of prostate cancer patients? Preliminary results of a pilot study. Cancer Invest 2015; 33 (6): 232–240.
8. Mohammed J Z, Wagner Meira J R. Data Mining and Analysis. England: Cambridge University Press, 2014.
9. El Naqa I, Li R, Murphy M J. Machine Learning in Radiation Oncology: Theory and Applications, 2nd edition. Switzerland: Springer, 2015.
10. Lambin P, van Stiphout R G, Starmans M H et al. Predicting outcomes in radiation oncology – multifactorial decision support systems. Nat Rev Clin Oncol 2015; 10 (1): 27–40.
11. Sanchez-Nieto B, Nahum A E. BIOPLAN: software for the biological evaluation of radiation therapy. Med Dosim 2000; 25 (2): 71–76.
12. Pinter C, Lasso A, Wang A, Jaffray D, Fichtinger G. SlicerRT: radiation therapy research toolkit for 3D Slicer. Med Phys 2012; 39 (10): 6332–6338.
13. Bentzen S M, Constine L S, Deasy J O et al. Quantitative analyses of normal tissue effects in the clinic (QUANTEC): an introduction to the scientific issues. Int J Radiat Oncol Biol Phys 2010; 76 (3): S3–S9.
14. El Naqa I, Bradley J D, PE L, Hope A J, Deasy J O. Predicting radiotherapy outcomes using statistical learning techniques. Phys Med Biol 2009; 54 (18): S9.
15. 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 (5): 1127–1135.
16. Bibault J E, Giraud P, Burgun A. Big data and machine learning in radiation oncology: state of the art and future prospects. Cancer Lett 2016; 16: 30346–30349.
17. Trifiletti D M, Showalter T N. Big data and comparative effectiveness research in radiation oncology: synergy and accelerated discovery. Front Oncol 2015; 5: 274.
18. Coates J, Souhami L, El Naqa I. Big data analytics for prostate radiotherapy. Front Oncol 2016; 6: 149.