MP19-02 Deep learning using preoperative MRI information to predict early recovery of urinary continence after robot-assisted radical prostatectomy.

Makoto Sumitomo[1, 2], Atsushi Teramoto[3], Ryo Toda[3], Naohiko Fukami[1], Kosuke Fukaya[1], Kenji Zennami[1, 2], Manabu Ichino[1], Kiyoshi Takahara[1], Mamoru Kusaka[1], Ryoichi Shiroki[1]

[1]Fujita Cancer Center, Fujita Health University , [2]Department of Urology School of Medicine, Fujita Health University, [3]Faculty of Radiological Technology, School of Medical Sciences, Fujita Health University

Introduction

Urinary incontinence (UI) remains a severe complication after robot-assisted radical prostatectomy (RARP), but there is no system to precisely predict the risk of post-prostatectomy UI (PPUI). We investigated whether deep learning (DL) model from magnetic resonance imaging (MRI) information is an accurate method to predict the risk of UI after RARP.

Patients and Methods

1. **Patients inclusion criteria:** We identified 561 patients who underwent RARP at Fujita Health University Hospital from August 2015 to July 2019. Patients who did not match the fat suppression condition and who have the image quality problem such as low resolution and blurring as several patients underwent MRI at different institutes were excluded before starting data analyses, leaving 400 patients eligible for analysis.
2. **RARP procedure:** RARP was done by nine surgeons with the da Vinci Si or Xi system (Intuitive Surgical, Inc., Sunnyvale, CA, USA). Nerve-sparing (NS) was basically conducted according to the clinical stage and risk criteria, and bladder neck preservation was included on a routine basis. All patients received posterior and anterior reconstructions.
3. **Continence definition:** Patients using 0 or 1 pad (for less than 20 g incontinence) /day within 3 months after RARP were categorized into the “good” group, whereas the other patients into the “bad” group.
4. **Preoperative and intraoperative parameters:** Preoperative clinicopathological covariates, such as age, BMI, NADT history, MUL, PV, continence status before RARP, serum prostate-specific antigen (PSA) level, Gleason score (GS sum), clinical stage, and risk criteria based on the risk stratification in the European Association of Urology guidelines, and intraoperative covariates, such as operator experience, total operation time, console time, with or without NS, and bleeding volume, were assessed. We considered surgeons with more than 50 cases of RARP experience as experts, whereas the others were non-expert.
5. **MRI:** DICOM data of preoperative T2-weighted MRI without fat suppression were collected. We selected one imaging slice showing the maximal diameter of the prostate from the axial and coronal planes and that showing the membranous urethra from the sagittal plane per patient. MUL, defined as the distance from the prostatic apex to the level of the urethra at the penile bulb, was measured in a blinded manner by two experts.
6. **DL model:** Automated classification method of the early recovery of urinary continence is shown in Figure 1. MR images were input to pretrained VGG-16, and 4096 output values of the last convolutional layer were extracted as characteristic features. Thirty types of features that contribute to classification were selected using information gain, used for feature selection in multivariate analysis and machine learning. The selected image features and preoperative and intraoperative parameters were given to a plurality of ML algorithms to distinguish between good and bad urinary incontinence.
7. **nvestigation of DL models using Grad-CAM:** DL (CNN) is a black box, and determining the employed image features based only on the judgment result is difficult. Selvaraju et al. proposed the gradient-weighted class activation mapping (Grad-CAM) to produce visual explanations of decisions made by CNN-based models.
8. **Statistical analyses:** Statistical analyses were performed with EZR (Saitama Medical Center, Jichi Medical University, Saitama, Japan), which is a graphical user interface for R (The R Foundation for Statistical Computing, Vienna, Austria) [18]. The Mann-Whitney test and chi-square test were employed to compare the data between continent and incontinent patients. Multivariate logistic regression analyses were used to examine variables associated with postoperative continence. A p-value of less than 0.05 was considered significant.

Conclusions

We showed that DL algorithm using MRI could improve the accuracy for predicting the risk of PPUI than ML using conventional clinicopathological parameters. DL predictions may help in allocating treatment strategies for PC patients who dislike prolonged UI after PARP.

Discussion

1. This study is the first to report that a DL model using MRI could provide better information than simple ML using conventional clinicopathological parameters such as BMI and MUL.

2. Our results provide some advantages to compare with results form past studies because of the following reasons: First, data from the DL algorithms are not affected by human judgment and can be more objective and reliable than direct MRI measurement performed and confirmed blindly by humans. Second, our results provided better results on AUC and specificity, which increased over 15% and 20% better on DL using MRI, respectively, than on those using conventional clinicopathological parameters including MUL and BMI.

3. Our preliminary results by Grad-CAM methods may highlight the importance of continuous physical training for improving early continence recovery.

4. Our prediction model is rudimentary in the present version in which only one axial MRI slice was used. In the future, the development of the method using multiple slices will be desired.

5. There may remain biases from many surgeons with different robotic experience and skills for evaluating intraoperative parameters. Our results showing that the accuracy of UI recovery prediction remained around 70% suggest the possibility that approximately 30% of PPUI may be caused by other factors, including intraoperative parameters.

References

[1] Di Pierro GB, et al. Eur Urol. 2011;59:1-6.
[2] Ficarra V, et al. Eur Urol. 2012;62:405-17.
[3] Krambeck AE, et al. BJU Int. 2009;103:448-53.
[4] Averbeck MA, et al. Neurourol Urodyn. 2019;38:398-406.
[5] Heesakkers J, et al. Eur Urol. 2017;71:936-44.
[6] Shao IH, et al. J Int Med Res. 2018;46:421-9.
[7] LeCun Y, et al. Nature. 2015;521:436-44.
[8] Kooi T et al. Med Image Anal. 2017;35:303-12.
[9] Cheng JZ et al. Sci Rep. 2016;6:24454.
[10] Teramoto A, et al. Biomed Res Int. 2017;2017:4067832.
[11] von Bodman C, et al. J Urol. 2012;187:945-50.
[12] Krizhevsky A SI, et al. Adv Neural Inf Process Syst 2012:1106–14. .
[13] Fujita H. et al. 2020; doi:10.1007/s12194-019-00552-4.
[14] Simonyan K ZA, 2015. arXiv:1409.1556.
[15] Deng J, et al. 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, . 2009, pp. 248-255.
[16] Kullback S LR. 1951;22(1):79- 86. doi:10.1214/aoms/1177729694.
[17] Selvaraju RR CM, ICCV. 2017. p. 618–26. https://doi.org/10. 1109/ICCV.2017.74. .
[18] Kanda Y. Bone Marrow Transplant. 2013;48:452-8.

Figure 1. Automated classification method of the early recovery of urinary continence

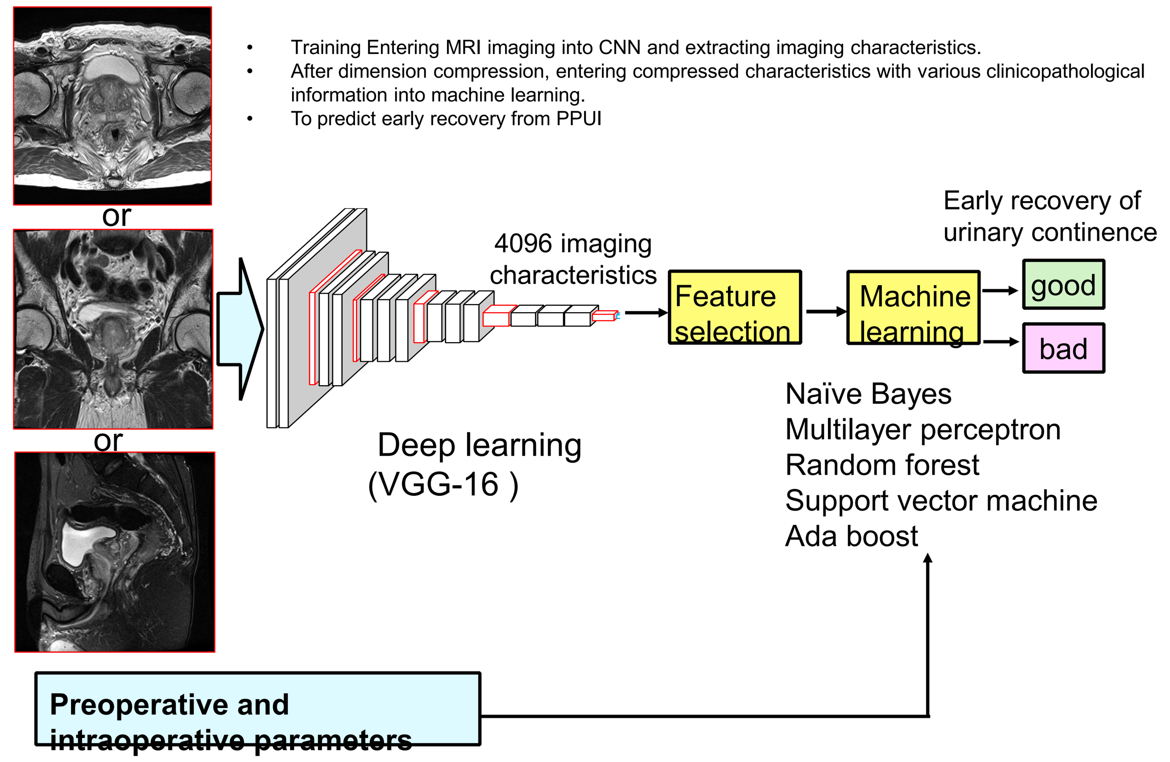

Figure 2. Kaplan-Meier curve showing PPUI recovery in 400 patients

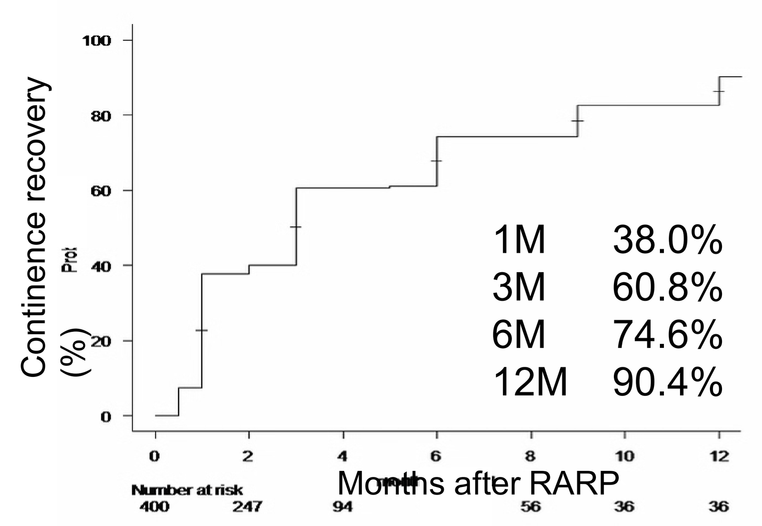

Table 1. Comparison of characteristics between good and bad groups on urinary continence 3 months after RARP.

	Good group (n = 246)	Bad group (n = 154)	*p*-value
Median Age, years (IQR)	67 (63.0-71.0)	66 (62.3-70.8)	0.603
Median BMI, kg/m^2 (IQR)	22.9 (21.4-24.2)	23.6 (21.8-25.7)	< 0.001
NADT no	175	93	0.029
yes	71	61	
PV, cm3 (IQR)	31.0 (24.0-40.8)	32.3 (24.5-44.8)	0.469
MUL, mm (IQR)	13 (12-15)	12 (11-14)	0.001
Incontinent before RARP no	240	149	0.756
yes	6	5	
PSA, ng/ml (IQR)	7.6 (5.5-11.0)	8.14 (6.0-11.9)	0.213
GS sum	7 (7-8)	7 (7-7)	0.008
T stage ≦2b	197	112	0.111
≧2c	49	42	
Risk criteria low	34	16	0.032
intermediate	134	69	
high	78	69	
Operated by expert surgeons	165	105	0.827
by non-expert surgeons	81	49	
Operation time	165 (143-208)	165 (144-197)	0.902
Console time	120 (100-149)	119 (99-150)	0.570
Bleeding volume	200 (100-300)	200 (101.5-300)	0.217
Nerve sparing not done	72	46	0.268
unilateral	146	98	
bilateral	28	10	

Table 2. Logistic regression analyses of predictive factors including intraoperative factors on UI 3 months after RARP.

	Odds ratio	95% CI	*p*-value
Age	0.995	0.959-1.030	0.807
BMI	1.150	1.060-1.250	<0.001
NADT	1.220	0.666-2.220	0.533
PV	1.010	0.997-1.020	0.127
MUL	0.850	0.762-0.947	0.003
Continence status before RARP	1.570	0.450-5.500	0.478
PSA	0.992	0.970-1.010	0.439
GS sum	1.310	0.918-1.860	0.137
T stage ≧2c (vs ≦2b)	1.090	0.616-1.920	0.771
Risk criteria high (vs low or intermediate)	1.230	0.616-2.440	0.562
Non-expert surgeon	1.300	0.753-2.240	0.347
Operation time	1.010	0.990-1.020	0.485
Console time	0.991	0.974-1.010	0.272
Bleeding volume	1.000	0.999-1.000	0.699
Nerve sparing	0.997	0.664-1.500	0.987

Table 3. Preliminary results of AUC and accuracies on continence prediction by DL algorithms according to three MRI planes in 30 patients.

Axial plane

Model	AUC	Precision
AdaBoost	0.683 ± 0.000	0.687 ± 0.000
Naive Bayes	0.750 ± 0.000	0.680 ± 0.000
Neural Network	0.596 ± 0.000	0.605 ± 0.000
Random Forest	0.531 ± 0.073	0.543 ± 0.049
SVM	0.352 ± 0.115	0.520 ± 0.000

Naive Bayes

		Predicted	
		Good	Bad
Actual	Good	9	3
	Bad	5	8

Overall accuracy = 68.7%

Coronal plane

Model	AUC	Precision
AdaBoost	0.420 ± 0.000	0.518 ± 0.000
Naive Bayes	0.545 ± 0.000	0.553 ± 0.000
Neural Network	0.375 ± 0.000	0.410 ± 0.000
Random Forest	0.489 ± 0.151	0.464 ± 0.075
SVM	0.316 ± 0.046	0.322 ± 0.000

Naive Bayes

		Predicted	
		Good	Bad
Actual	Good	5	3
	Bad	6	5

Overall accuracy = 55.3%

Sagittal plane

Model	AUC	Precision
AdaBoost	0.071 ± 0.000	0.062 ± 0.000
Naive Bayes	0.449 ± 0.000	0.500 ± 0.000
Neural Network	0.510 ± 0.000	0.578 ± 0.000
Random Forest	0.274 ± 0.083	0.265 ± 0.061
SVM	0.200 ± 0.059	0.000 ± 0.000

Neural Network

		Predicted	
		Good	Bad
Actual	Good	3	4
	Bad	2	5

Table 4. Results of AUC and accuracies on continence prediction by DL algorithms using the axial MRI plane information and preoperative clinicopathological parameters in 400 patients.

MRI Axial T2WI only

Model	Overall accuracy	Sensitivity	Specificity	AUC
Naive Bayes	68.5	71.1	64.3	0.758
Random Forest	67.8	82.1	44.8	0.689
SVM	65.8	87.0	31.8	0.653
ANN	59.3	66.7	47.4	0.605
AdaBoost	58.5	67.5	44.2	0.558

Naive Bayes

		Predicted class	
		Good	Bad
Actual class	Good	175	71
	Bad	55	99

Overall accuracy	68.5
Sensitivity	71.1
Specificity	64.3

Preoperative clinicopathological parameters only

Model	Overall accuracy	Sensitivity	Specificity	AUC
Naive Bayes	60.8	71.1	44.2	0.622
Random Forest	58.5	74.1	33.1	0.606
SVM	61.0	87.0	19.5	0.594
ANN	58.5	67.9	43.5	0.580
AdaBoost	53.0	63.0	37.0	0.500

Naive Bayes

		Predicted class	
		Good	Bad
Actual class	Good	175	71
	Bad	86	68

Overall accuracy	60.8
Sensitivity	71.1
Specificity	44.2

MRI Axial T2WI + Preoperative clinicopathological parameters

Model	Overall accuracy	Sensitivity	Specificity	AUC
Naive Bayes	70.3	72.0	67.5	0.775
Random Forest	67.8	82.1	44.8	0.707
SVM	65.3	84.1	35.1	0.681
ANN	61.8	69.9	48.7	0.647
AdaBoost	55.3	61.8	44.8	0.533

Naive Bayes

		Predicted class	
		Good	Bad
Actual class	Good	177	69
	Bad	50	104

Overall accuracy	70.3
Sensitivity	72.0
Specificity	67.5

Grad CAM (Gradient-weighted Class Activation Mapping) [arXiv, 2017]

- A robust and parsimonious approach for indoor navigation using DCNN
- A potential capability in space feature learning and recognition
- DCNN based approach to look into the visual similarity and visual distinctiveness of interior space

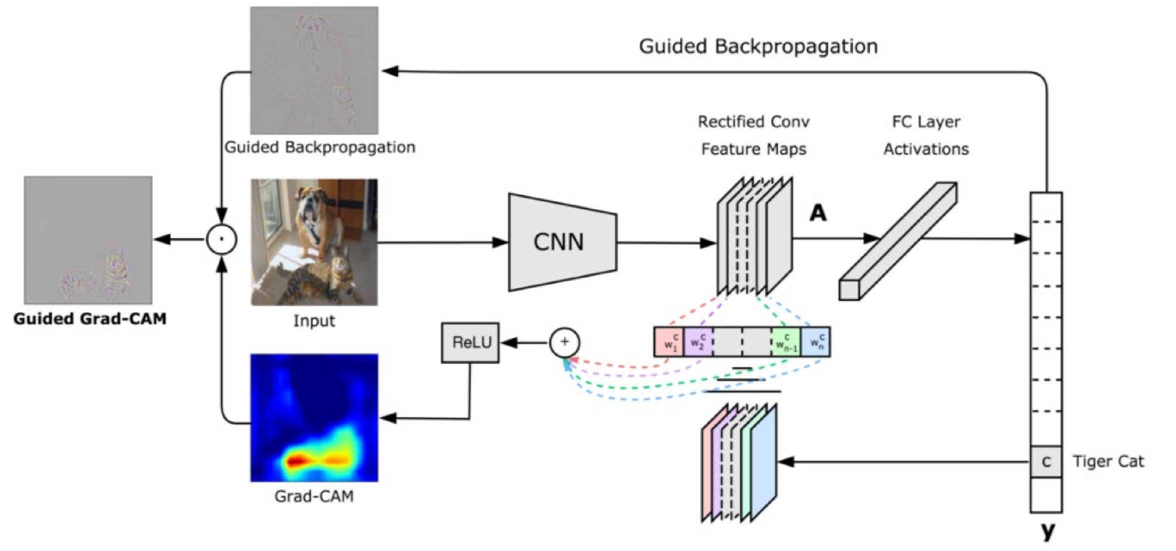

Figure 3a. Visualizing of activation maps by DL algorithms in MRI axial planes of patients with ‘good’ postoperative continence.

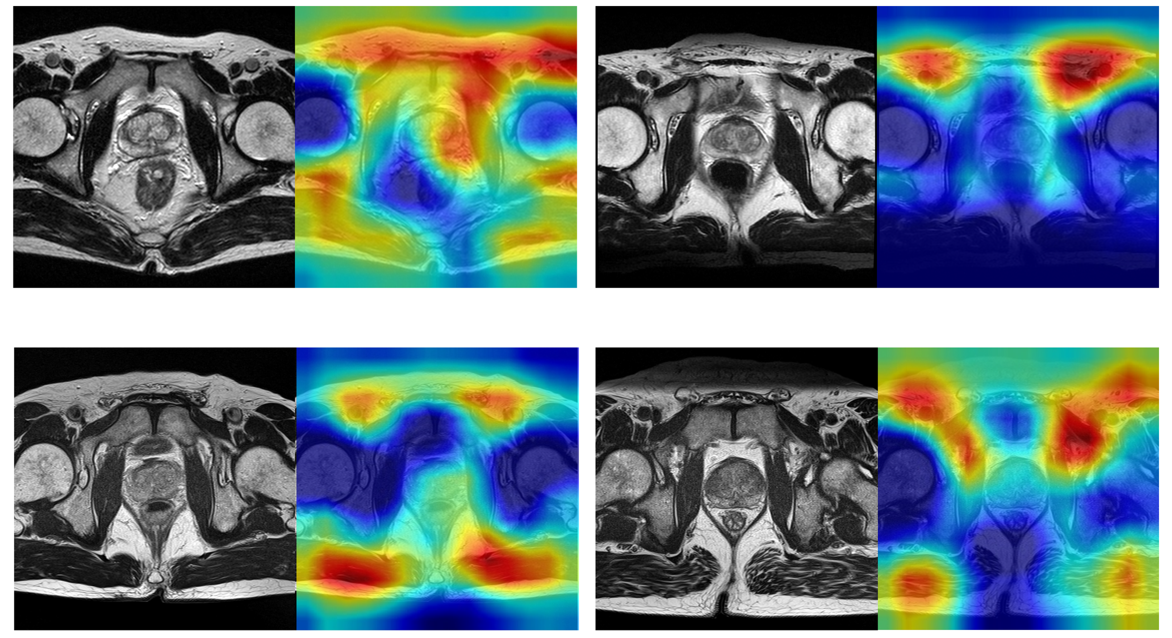

Figure 3b. Visualizing of activation maps by DL algorithms in MRI axial planes of patients with ‘bad’ postoperative continence.