**Combination of deep learning and ensemble machine learning using intraoperative video images strongly predicts recovery of urinary continence after robot-assisted radical prostatectomy.**

#### Wataru Nakamura1, Makoto Sumitomo1, 2, Kenji Zennami1, Masashi Takenaka1, Manabu Ichino1, Kiyoshi Takahara1, Atsushi Teramoto3, Ryoichi Shiroki1

#### 1 Department of Urology School of Medicine, Fujita Health University

#### 2 Fujita Cancer Center, Fujita Health University

#### 3 Faculty of Radiological Technology, School of Medical Sciences, Fujita Health University

**Running title:** DL of video images to predict recovery of UI after RARP

(45 characters)

**Key Words:** deep learning, ensemble machine learning, video images, continence, RARP

**Correspondence to:** Makoto Sumitomo, M.D., Ph.D.

Department of Urology School of Medicine, Fujita Health University

1-98 Dengakugakubo, Kutsukake-cho, Toyoake, Aichi 470-1192 Japan

Tel: +81-562-93-9830

FAX: +81-562-93-7863

E-mail: [m-sumi@fujita-hu.ac.jp](mailto:m-sumi@fujita-hu.ac.jp)

**Support**

Not funded.

#### Abstract

#### Objectives: We recently reported that deep learning (DL) using pelvic magnetic resonance imaging is useful for predicting the severity of urinary incontinence (UI) after robot-assisted radical prostatectomy (RARP). However, our results were limited because the prediction accuracy was approximately 70%. We aimed to develop a more accurate prediction system that can be used to inform patients on recovery from UI after RARP using a DL model based on intraoperative video images.

#### Materials and Methods: This study included 101 patients with prostate cancer who underwent　RARP. Three snapshots showing the pelvic cavity (before bladder neck incision, just after prostate removal, and after vesicourethral anastomosis) from intraoperative video records, as well as preoperative and intraoperative covariates, were assessed. We evaluated the DL models plus simple or ensemble machine learning, and their sensitivity, specificity, and area under the receiver operating characteristic curve (AUC) were analyzed.

#### Results: Sixty-four and 37 patients demonstrated ‘early continence’ and ‘late continence’, respectively, at the 3-month follow-up. The combination of DL and simple machine learning using intraoperative video snapshots with clinicopathological parameters had a notably high performance (AUC, 0.683 to 0.749) for predicting early recovery from post-prostatectomy UI. Notably, the combination of DL and ensemble artificial neural network using intraoperative video snapshots had the highest performance (AUC, 0.882; sensitivity, 92.2%; specificity, 78.4%; overall accuracy, 85.3%) for predicting early recovery from post-prostatectomy UI. In contrast, DL and ensemble ML with clinicopathological parameters (Method 4) achieved no additive effects (AUC, 0.690 to 0.747) compared with DL and simple ML with clinicopathological parameters. Internal validation was performed on additional 30 cases with similar results.

#### Conclusions: Our results suggest that DL algorithms using intraoperative video images can be used to reliably inform patients regarding their recovery from UI after RARP. (287 words)

#### 1. Introduction

Artificial intelligence (AI) is a branch of computer science that uses algorithms to approximate human cognitive functions, such as problem solving, decision-making, object detection, and classification.1 An emerging machine learning (ML) technique, referred to as deep learning (DL),2 is gaining popularity in many areas of medical image analysis.3-5 Urology was one of the first fields in which AI was used for object detection, image classification, image segmentation, skill assessment, and outcome prediction for complex urologic procedures.6

Prostate cancer (PC) is a major disease that affects men worldwide, and robot-assisted radical prostatectomy (RARP) has become the standard surgery for localized PC. However, post-prostatectomy urinary incontinence (PPUI) severely impairs the quality of life of patients with PC and remains problematic. We recently reported that DL using pelvic magnetic resonance imaging (MRI) is useful for predicting the severity of urinary incontinence (UI) after robot-assisted radical prostatectomy (RARP).7 Our results from DL algorithms might help in the choice of treatment strategy, especially for patients with PC who wish to avoid prolonged UI after RARP. However, our results were limited because the prediction accuracy remained at approximately 70%.7 Recent studies suggest that the preservation of periprostatic structures by intraoperative surgical techniques such as the nerve-sparing (NS), bladder neck-preserving, and Retzius-sparing modalities are associated with early recovery from UI after RARP.8-10 Although surgical procedures can be expected to improve early recovery from UI, it is difficult to objectively assess the relationship between surgical procedures and the preservation of anatomic structures and whether they contribute to early recovery from UI after RARP.

We hypothesized that DL could objectively assess the relationship between surgical techniques and the preservation of anatomic structures and the relationship between these anatomical structures and early recovery from UI after RARP. In the present study, we aimed to develop a more accurate prediction system that can be used to inform patients regarding their recovery from UI after RARP using a DL model from intraoperative video images.

**2. Materials and Methods**

**2.1 Patient inclusion and exclusion criteria**

All experimental protocols were approved by the institutional review board (IRB) of Fujita Health University School of Medicine (IRB no. HM19-257). All methods were performed in accordance with the relevant local guidelines and regulations. This study was conducted with an explanation to the patients, and a website with additional information, including an opt-out option, was set up for the study. A database of 400 patients with PC was used in our recent study (August 2015 to July 2019).7 In 78 patients, surgical videos had been deleted or lost, and in 299 patients, video records could be viewed; however, snapshots necessary to perform the analysis could not be obtained (refer to Snapshots extraction from the intraoperative video). Thus, we adopted 101 patients whose intraoperative video records were available.

In addition, the video records of 30 additional patients who recently underwent surgery were prepared for internal validation.

**2.2 RARP procedure**

RARP was performed by nine surgeons using the da Vinci Si or Xi system (Intuitive Surgical, Inc., Sunnyvale, CA, USA). NS surgery was performed according to the clinical stage and risk criteria, and bladder neck preservation was routinely included. All patients underwent posterior and anterior reconstruction.

**2.3 Preoperative and intraoperative parameters and definition of continence**

Preoperative clinicopathological covariates, such as age, body mass index (BMI), neoadjuvant androgen deprivation therapy (NADT) history, membranous urethral length (MUL), prostate volume (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. Continence was evaluated using the Expanded Prostate Cancer Index Composite survey question: “How many pads per day did you usually use to control leakage during the last 4 weeks?” Patients who did not use pads with no urine leakage, or used 1 safety pad which was for less than 20 mL of urine within 3 months after RARP were categorized into the ‘early continence’ group, whereas others were categorized into the ‘late continence’ group.

**2.4 Snapshots extraction from the intraoperative video**

Three snapshots showing the pelvic cavity (before bladder neck incision, just after prostate removal, and after vesicourethral anastomosis) from the intraoperative video records. Snapshot extraction was performed in accordance with the following principles while considering reproducibility; 1. Anatomical structures near the pubic symphysis (prostatic apex) should be included in all images. 2. In "before bladder neck incision," the bladder neck and prostatic apex should be visible. 3. In the "just after prostate removal," bladder neck preservation, nerve preservation, and degree of bleeding should be visible. 4. In "after vesicourethral anastomosis," the anastomosis should be visible without tension on the urethra or bladder. Figure 1 shows the representation of the three snapshots extracted from the intraoperative video of the same patients.

**2.5 DL model**

First, the given images were input into a convolutional neural network (CNN), which is a DL technique2,11 that shows excellent ability for classifying images.12 CNN is applied in medical image processing and is widely used for lesion detection and differentiation and prognostication.7,13,14 We focused on DenseNet, which is a type of CNN.15 By tightly coupling the layers, information can be transmitted smoothly, even in a multilayered network, thereby improving the processing performance. DenseNet has several variations with different numbers of network layers; in this study, we used DenseNet169 (with 169 layers).

A common method for using CNNs is to input data and directly obtain the desired results, such as the classification output. In this study, three images were treated as input images; therefore, a simple DenseNet could not be used. CNNs are also used as feature extractors, where the convolutional layer of the CNN is responsible for extracting various features from the input images. CNNs trained on a large number of images are able to extract general-purpose features from images, and some studies have used these features for other purposes. In our previous study,7 a CNN was used to extract 4,096 features from a single image and predict whether urinary continence was good or poor using ML.

In this study, three images were input into DenseNet169, and 1920 features were obtained for each image. In all, 5760 features were used for prediction.

Because the number of these features was large compared with the number of samples, dimensional compression was necessary. Principal component analysis was used for dimensionality compression, and the data were compressed into 20 dimensional principal components. These data were the image features obtained from the CNN.

Using the image features and clinical information obtained as described above, an ML method was used to predict whether urinary continence was early or late. Naïve Bayes, support vector machine, random forest, and artificial neural networks (ANNs) were each used as the ML method. This method of prediction, using only one ML method, is called the single model.

We also introduced an ensemble model, in which the output of the above four ML methods (probability of UI) and 20 compressed features were input again to ML to predict UI.

The following six methods (Methods 1–6) were possible, depending on the variation in input information (images and clinical information) and network configuration (single and ensemble models). In this study, we compared the prediction performances of these methods as follows: Method 1, DL with three video images and simple ML; Method 2, DL with three video images and ensemble ML; Method 3, DL with 3 video images, addition of clinical factors, and simple ML; Method 4, DL with three video images, addition of clinical factors, and ensemble ML; Method 5, simple ML using clinicopathological factors; and Method 6, ensemble ML using clinicopathological factors.

The automated classification method for early recovery of urinary continence is summarized in Figure 2.

**2.6 Statistical analyses**

Statistical analyses were performed using EZR software (Saitama Medical Center, Jichi Medical University, Saitama, Japan).16 The Mann-Whitney test and chi-square test were used to compare the data between continent and incontinent patients. Multivariate logistic regression analyses were used to examine the variables associated with postoperative continence. Statistical significance was set as *p* < 0.05.

**3. Results**

The median age of the 101 patients at the time of RARP was 66 years, and the median body mass index (BMI) was 23.4 kg/m2. The median prostate-specific antigen level at PC diagnosis was 7.50 ng/mL, and the median prostate volume at PC diagnosis was 27.2 mL. The Gleason sum at diagnosis was 6 in 28 (28%), 7 in 52 (51%), and > 8 in 21 (21%) patients. The risk criteria were low in 17 (16.8%), intermediate in 49 (48.5%), and high in 35 (34.7%), and 27 (26.7%) patients received NADT. A total of 62 patients were operated on by expert surgeons, and 39 patients were operated on by non-expert surgeons under the supervision of experienced mentors as necessary. The median operative and console times were 160 and 119 min, respectively, and the median bleeding volume was 200 mL. NS surgery was performed in 80 patients (unilateral, n = 70; bilateral, n = 10). A histopathologically positive resection margin was observed in 19 patients (18.5%).

The characteristics of the early continence (64 patients) and poor (37 patients) continence groups at the 3-month follow-up are presented in Table 1. Significant differences were found between the groups with regard to the median BMI, Gleason sum, and MUL, as well as NADT history. Although no significant differences in intraoperative parameters were observed, there was a tendency for NS surgery to affect continence. Multivariate logistic regression analyses of continence status at the 3-month post-RARP follow-up showed that BMI, NADT history, MUL, and NS remained significant predictors of poor continence (Table 2); these findings were notably similar to our previous study.7

Figure 3A and Figure 3B shows the area under the receiver operating characteristic curve (AUC) and the accuracy of continence prediction using Methods 1 and 3. The results showed that the combination of DL and simple ML using intraoperative video snapshots with clinicopathological parameters (Methods 3) demonstrated a higher performance (AUC, 0.683 to 0.749) for predicting early recovery from PPUI (Fig. 3A and Supplementary Table 1), while intraoperative video snapshots alone (Method 1) achieved an AUC of 0.641 to 0.701 (Fig. 3B and Supplementary Table 1), suggesting that the combination of intraoperative video images with clinicopathological parameters showed additive effects for PPUI prediction. We then attempted a combination of DL and ensemble ML using intraoperative video images (Method 2 and 4). The results showed that the combination of DL and ensemble ANN using intraoperative video snapshots (Method 2) had the highest performance, with an AUC of 0.882 (sensitivity, 92.2%; specificity, 78.4%; overall accuracy, 85.3%) for predicting early recovery from PPUI (Fig. 3C and Supplementary Table 1). In contrast, DL and ensemble ML with clinicopathological parameters (Method 4) achieved no additive effects (AUC, 0.690 to 0.747) compared with DL and simple ML with clinicopathological parameters (Fig. 3D and Supplementary Table 1). The AUCs of the four ML algorithms according to the DL and ML methods are shown in Figure 4A. Ensemble ML involving clinicopathological parameters had a non-notable additive effect on performance compared with simple ML (Method 5 and 6 in Fig. 3). We finally performed an internal validation test using snapshot photographs extracted from surgical videos of 30 recently operated patients. Although the combination of DL and simple ML using intraoperative video snapshots (Methods 1) did not give excellent results, the combination of DL and ensemble ANN (Method 2) had the highest performance, with an AUC of 0.858 for predicting early recovery from PPUI (Fig. 4B). Our results suggest that ensemble ML has the potential to improve accuracy using information from intraoperative video images, but not with clinicopathological parameters, including intraoperative parameters.

**4. Discussion**

Notable effort has been expended to determine preoperative and intraoperative parameters17-19 that can help identify patients at risk of PPUI. Recent studies have demonstrated that superior surgeon performance is associated with superior continence outcomes after RARP,20-22 which has been proven by AI methods using semantic segmentation and automated performance status (APM). These findings strongly suggest that PPUI is influenced by surgeon (technical) and patient (anatomical) factors. In our previous study,7 we reported that the overall accuracy for predicting early recovery from PPUI was at most 70% using a DL model based on preoperative pelvic MRI. Hung et al. also reported that the overall accuracy was approximately 70% using the surgeon APM and clinical parameters.21 In the present study, our DL model combined with simple ML using intraoperative video images produced better results (AUC, approximately 0.75) for predicting early recovery from PPUI. Our results are meaningful because they suggest that the use of intraoperative video imaging has an additive effect on PPUI prediction compared to preoperative parameters alone. Notably, the DL model combined with the ensemble ANN showed over 80% accuracy in predicting early recovery from PPUI. To the best of our knowledge, this study is the first to report that a DL model using intraoperative video snapshots could provide better accuracy for early recovery from UI after RARP compared with previous reports on PPUI. Moreover, this is the first study to predict postoperative outcomes (including prognosis and complications) via DL using intraoperative video images rather than APM or semantic segmentation.

In the present study, we applied ensemble methods to classify PPUI status. **Ensemble methods** are ML techniques that combine several base models to produce an optimal predictive model, and they have been used in recent studies.23,24 Our results showed that ensemble ML has the potential to improve accuracy using intraoperative video information but not with clinicopathological parameters from our database. Our model allowed us to predict the risk of PPUI using preoperative and/or intraoperative factors using simple or ensemble ML. However, it remains unclear why the ANN produced superior prediction results in the present study. ANN was developed based on biological neurons of the human brain and is trained to generate an output outcome as a weighted combination of input variables.25,26 Support vector machine, random forest, and naïve bays are based on the principle of structural risk minimization and place the data into a multidimensional space to achieve classification with a hyperplane, which has distinct advantages in solving problems such as a small sample size and nonlinear or high-dimensional pattern types.27,28 Each approach has its advantages and disadvantages, and it is necessary to try different methods to seek a suitable model for classification. On the other hand, ANN aims to solve a variety of classification and pattern recognition problems.29 The main advantage of an ANN is that it can approximate any nonlinear mathematical function.27 It seems necessary to clarify the focus of DL in the intraoperative video images. DL is an AI technology that automates various processes. In the present study, multiple video snapshots made using gradient-weighted class activation mapping impossible; this is known as the black-box problem.30 One plausible scenario is that the ANN might focus on the anatomical relationships around the vesicourethral anastomosis, including the MUL. Although MUL is shown to be one of the strong preoperative predictors for PPUI in our study and other numerous reports 7,31-33, we doubt that MUL specificity alone can explain the prediction using the DL and ensemble ANN. The ANN method might discriminate on various factors other than MUL, such as "sphincter to urethral anastomosis ratio," "vesicourethral anastomosis location," "degree of preservation of puborectal muscle group," "degree of nerve preservation," and "degree of bleeding." In order to solve this problem, we planned an analysis involving DL prediction using higher resolution snapshots and photographs focusing on the "anastomotic area," "nerve-sparing area," and "peri-rectal area." However, due to image resolution issues, it was not easy to use these high-quality images to identify the focus of the AI method using ANN.

Our study was an evaluation of intraoperative images, not a study to "exclude unfit patients for surgery," as focused on in previous studies using preoperative information.7 However, our results indicate that both "anatomical" and "surgical factors (surgical techniques)" shown in the video are related to the risk of PPUI. Our results suggest that the information obtained from actual surgical images is more important for outcome prediction than surgical procedure and performance records transcribed from surgical records and databases. Identifying the hotspots in the black box of DL can help clarify the mechanism of PPUI from multiple perspectives, which may contribute to preoperative informed consent (to explain the prediction of PPUI) and education, including the improvement of surgical techniques. The most significant aspect of this study is its clarification of this point.

This study had several limitations. First, it was a retrospective study that was conducted at a single institute. Regarding this point, we obtained promising data as a result of internal validation. Further multi-institutional large-scale studies are warranted to externally validate the proposed algorithms. Second, PPUI was defined based on the use of a daily pad alone, although a standard PPUI definition has not been established.

**5. Conclusions**

Our findings seem useful for individualized counseling in clinical practice, using preoperative and intraoperative information regarding the probability of UI after RARP. The future development of methods to identify hotspots in intraoperative video information is expected, which will enable the use of this DL model as a tool for surgical navigation training to avoid prolonged UI after RARP.　(2899 words in Text）

**Conflict of interest**

None declared.

**Ethics statement**

This study was approved by the Institutional Review Boards of Fujita Health University Hospital.

**Data availability statement**

The data sets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

#### Author contributions

#### W.N. and M.S. had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. M.S. and A.T. conceived and designed the study. W.N., M.S., K.Z., M.T., M.I., K.T., and R.S. contributed to the data acquisition. W.N., M.S., and A.T. analyzed and interpreted the data. M.S. and A.T. performed the statistical analyses and drafted the manuscript, and all authors reviewed the manuscript. R.S. supervised the study.

#### ORCID:

Makoto Sumitomo  [https://orcid.org/0000-0003-2589-922X](https://orcid.org/0000-0003-2589-922X))

#### Abbreviations & Acronyms

#### AI, artificial intelligence; ANN, artificial neural network; APM, automated performance status; AUC, area under the receiver operating characteristic curve; BMI, body mass index; CNN, convolutional neural network; DL, deep learning; ML, machine learning; MRI, magnetic resonance imaging; NADT, neoadjuvant androgen deprivation therapy; NS, nerve-sparing; PC, prostate cancer; PPUI, post-prostatectomy urinary incontinence; RARP, robot-assisted radical prostatectomy; UI, urinary incontinence

**References**

1. Yu KH, Beam AL, Kohane IS. Artificial intelligence in healthcare. *Nat Biomed Eng.* 2018;2(10):719-731.

2. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* 2015;521(7553):436-444.

3. Kooi T, Litjens G, van Ginneken B, et al. Large scale deep learning for computer aided detection of mammographic lesions. *Med Image Anal.* 2017;35:303-312.

4. Cheng JZ, Ni D, Chou YH, et al. Computer-Aided Diagnosis with Deep Learning Architecture: Applications to Breast Lesions in US Images and Pulmonary Nodules in CT Scans. *Sci Rep.* 2016;6:24454.

5. Teramoto A, Tsukamoto T, Kiriyama Y, Fujita H. Automated Classification of Lung Cancer Types from Cytological Images Using Deep Convolutional Neural Networks. *Biomed Res Int.* 2017;2017:4067832.

6. Chang TC, Seufert C, Eminaga O, Shkolyar E, Hu JC, Liao JC. Current Trends in Artificial Intelligence Application for Endourology and Robotic Surgery. *Urol Clin North Am.* 2021;48(1):151-160.

7. Sumitomo M, Teramoto A, Toda R, et al. Deep learning using preoperative magnetic resonance imaging information to predict early recovery of urinary continence after robot-assisted radical prostatectomy. *Int J Urol.* 2020;27(10):922-928.

8. Egan J, Marhamati S, Carvalho FLF, et al. Retzius-sparing Robot-assisted Radical Prostatectomy Leads to Durable Improvement in Urinary Function and Quality of Life Versus Standard Robot-assisted Radical Prostatectomy Without Compromise on Oncologic Efficacy: Single-surgeon Series and Step-by-step Guide. *Eur Urol.* 2021;79(6):839-857.

9. Sood A, Grauer R, Jeong W, et al. Evaluating post radical prostatectomy mechanisms of early continence. *Prostate.* 2022;82(12):1186-1195.

10. Pedraza AM, Wagaskar V, Parekh S, Tewari A. Technical advances in nerve-sparing and continence preservation. *Curr Opin Urol.* 2022;32(2):204-210.

11. Fujita H. AI-based computer-aided diagnosis (AI-CAD): the latest review to read first. Radiological Physics and Technology. . 2020; doi:10.1007/s12194-019-00552-4.

12. Krizhevsky A SI, Hinton GE. . ImageNet classification with deep convolu- tional neural networks. . *Adv Neural Inf Process Syst* 2012(25):1106–1114. .

13. Teramoto A, Fujita H, Yamamuro O, Tamaki T. Automated detection of pulmonary nodules in PET/CT images: Ensemble false-positive reduction using a convolutional neural network technique. *Med Phys.* 2016;43(6):2821-2827.

14. Teramoto A, Kiriyama Y, Tsukamoto T, et al. Weakly supervised learning for classification of lung cytological images using attention-based multiple instance learning. *Sci Rep.* 2021;11(1):20317.

15. He KZ, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* 2016:pp. 770–778.

16. Kanda Y. Investigation of the freely available easy-to-use software 'EZR' for medical statistics. *Bone Marrow Transplant.* 2013;48(3):452-458.

17. Matsushita K, Kent MT, Vickers AJ, et al. Preoperative predictive model of recovery of urinary continence after radical prostatectomy. *BJU Int.* 2015;116(4):577-583.

18. Heesakkers J, Farag F, Bauer RM, Sandhu J, De Ridder D, Stenzl A. Pathophysiology and Contributing Factors in Postprostatectomy Incontinence: A Review. *Eur Urol.* 2017;71(6):936-944.

19. Averbeck MA, Woodhouse C, Comiter C, et al. Surgical treatment of post-prostatectomy stress urinary incontinence in adult men: Report from the 6th International Consultation on Incontinence. *Neurourol Urodyn.* 2019;38(1):398-406.

20. Hung AJ, Chen J, Ghodoussipour S, et al. A deep-learning model using automated performance metrics and clinical features to predict urinary continence recovery after robot-assisted radical prostatectomy. *BJU Int.* 2019;124(3):487-495.

21. Hung AJ, Ma R, Cen S, Nguyen JH, Lei X, Wagner C. Surgeon Automated Performance Metrics as Predictors of Early Urinary Continence Recovery After Robotic Radical Prostatectomy-A Prospective Bi-institutional Study. *Eur Urol Open Sci.* 2021;27:65-72.

22. Balvardi S, Kammili A, Hanson M, et al. The association between video-based assessment of intraoperative technical performance and patient outcomes: a systematic review. *Surg Endosc.* 2022.

23. Pathan S, Siddalingaswamy PC, Kumar P, Pai MMM, Ali T, Acharya UR. Novel ensemble of optimized CNN and dynamic selection techniques for accurate Covid-19 screening using chest CT images. *Comput Biol Med.* 2021;137:104835.

24. Ferreira JR, Armando Cardona Cardenas D, Moreno RA, de Fatima de Sa Rebelo M, Krieger JE, Antonio Gutierrez M. Multi-View Ensemble Convolutional Neural Network to Improve Classification of Pneumonia in Low Contrast Chest X-Ray Images. *Annu Int Conf IEEE Eng Med Biol Soc.* 2020;2020:1238-1241.

25. Bozorg-Haddad O, Aboutalebi M, Ashofteh PS, Loaiciga HA. Real-time reservoir operation using data mining techniques. *Environ Monit Assess.* 2018;190(10):594.

26. Rajan JR, Chelvan AC, Duela JS. Multi-Class Neural Networks to Predict Lung Cancer. *J Med Syst.* 2019;43(7):211.

27. Xiang Y, Sun Y, Liu Y, et al. Development and validation of a predictive model for the diagnosis of solid solitary pulmonary nodules using data mining methods. *J Thorac Dis.* 2019;11(3):950-958.

28. Wang W, Feng X, Duan X, et al. Establishment of two data mining models of lung cancer screening based on three gene promoter methylations combined with telomere damage. *Int J Biol Markers.* 2017;32(1):e141-e146.

29. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. *Comput Struct Biotechnol J.* 2015;13:8-17.

30. Zhang Y, Weng Y, Lund J. Applications of Explainable Artificial Intelligence in Diagnosis and Surgery. *Diagnostics (Basel).* 2022;12(2).

31. Paparel P, Akin O, Sandhu JS, et al. Recovery of urinary continence after radical prostatectomy: association with urethral length and urethral fibrosis measured by preoperative and postoperative endorectal magnetic resonance imaging. *Eur Urol.* 2009;55(3):629-637.

32. Hakimi AA, Faleck DM, Agalliu I, Rozenblit AM, Chernyak V, Ghavamian R. Preoperative and intraoperative measurements of urethral length as predictors of continence after robot-assisted radical prostatectomy. *J Endourol.* 2011;25(6):1025-1030.

33. Mungovan SF, Sandhu JS, Akin O, Smart NA, Graham PL, Patel MI. Preoperative Membranous Urethral Length Measurement and Continence Recovery Following Radical Prostatectomy: A Systematic Review and Meta-analysis. *Eur Urol.* 2017;71(3):368-378.

**Figure Legends**

**Figure 1. The representation of the three snapshots extracted from the intraoperative video of the same patients.**

Three snapshots showing the pelvic cavity (before bladder neck incision, just after prostate removal, and after vesicourethral anastomosis) from the intraoperative video records. Snapshot extraction was performed as described in Materials and Methods.

**Figure 2. Automated classification method of the early recovery of urinary continence using DL and simple ML or DL and ensemble ML.**

Intraoperative video images were input to pretrained VGG-16, and 20 kinds of features that contribute to classification were selected using information gain from 5760 output values of the last convolutional layer extracted as characteristic features. Subsequently, 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. Ensemble ML was performed instead of simple ML.

**Figure 3. ROC curves and accuracies on continence prediction using intraoperative video images and clinicopathological parameters analyzed by DL and simple ML or DL and ensemble ML.**

**A, B.** Intraoperative video images with (A) or without (B) clinicopathological parameters were analyzed by DL and simple ML as described in the legends of Figure 1 and ROC analyses were performed. The analyses were performed three times and the representatives were shown. **C, D.** Intraoperative video images without (C) or with (D) clinicopathological parameters were analyzed by DL and ensemble ML as described in the legends of Figure 1 and ROC analyses were performed. The analyses were performed three times and the representatives were shown.

**Figure 4. The results of AUCs and accuracies according to 6 kinds of Methods in the 101 patient data set and the internal validation set.**

Intraoperative video images with or without clinicopathological parameters were analyzed by DL and simple ML or ensemble ML as described as descried in the Materials and Methods. ROC analyses were performed and AUCs were calculated. A. AUCs according to 6 kinds of Methods in the 101 patient data set were shown. B. An internal validation test was performed using snapshot photographs extracted from surgical videos of 30 recently operated patients. AUCs and accuracies according to the Method 1 and the Method 2 were shown. Each analysis was performed three times with similar results and the representatives were shown.