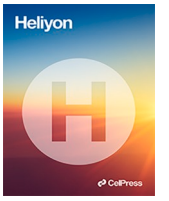
Heliyon 9 (2023) e20337

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



An artificial intelligence method for predicting postoperative urinary incontinence based on multiple anatomic parameters of MRI

Jiakun Li [a,b], Xuemeng Fan [a,b,1], Tong Tang [b,c], Erman Wu [b], Dongyue Wang [d], Hui Zong [b], Xianghong Zhou [a], Yifan Li [a], Chichen Zhang [a], Yihang Zhang [a], Rongrong Wu [b], Cong Wu [b], Lu Yang [a,**], Bairong Shen [b,*]

[a] *Department of Urology, West China Hospital, Sichuan University, Chengdu, China*
[b] *Institutes for Systems Genetics, Frontiers Science Center for Disease-related Molecular Network, West China Hospital, Sichuan University, Chengdu, China*
[c] *Department of Computer Science and Information Technologies, Elviña Campus, University of A Coruña, A Coruña, Spain*
[d] *Department of Ophthalmology, West China Hospital, Sichuan University, Chengdu, China*

A R T I C L E I N F O

Keywords:
Explainable deep neural network model
Prediction
Urinary continence
Robotic-assisted radical prostatectomy
Clinical practice

A B S T R A C T

Background: Deep learning methods are increasingly applied in the medical field; however, their lack of interpretability remains a challenge. Captum is a tool that can be used to interpret neural network models by computing feature importance weights. Although Captum is an interpretable model, it is rarely used to study medical problems, and there is a scarcity of data regarding MRI anatomical measurements for patients with prostate cancer after undergoing Robotic-Assisted Radical Prostatectomy (RARP). Consequently, predictive models for continence that use multiple types of anatomical MRI measurements are limited.

Methods: We explored the energy efficiency of deep learning models for predicting continence by analyzing MRI measurements. We analyzed and compared various statistical models and provided reference examples for the clinical application of interpretable deep-learning models. Patients who underwent RARP at our institution between July 2019 and December 2020 were included in this study. A series of clinical MRI anatomical measurements from these patients was used to discover continence features, and their impact on continence was primarily evaluated using a series of statistical methods and computational models.

Results: Age and six other anatomical measurements were identified as the top seven features of continence by the proposed model UINet7 with an accuracy of 0.97, and the first four of these features were also found by primary statistical analysis.

Conclusions: This study fills the gaps in the in-depth investigation of continence features after RARP due to the limitations of clinical data and applicable models. We provide a pioneering example of the application of deep-learning models to clinical problems. The interpretability analysis of deep learning models has the potential for clinical applications.

* Corresponding author. Institutes for Systems Genetics, Frontiers Science Center for Disease-related Molecular Network, West China Hospital, Sichuan University, Chengdu, Sichuan, China.
** Corresponding author. Department of Urology, Institute of Urology, West China Hospital, Sichuan University, Chengdu, Sichuan, China.
E-mail addresses: wycleflue@163.com (L. Yang), bairong.shen@scu.edu.cn (B. Shen).
[1] The authors have contributed equally to this work.

https://doi.org/10.1016/j.heliyon.2023.e20337
Received 15 June 2023; Received in revised form 12 September 2023; Accepted 19 September 2023
Available online 20 September 2023

1. Introduction

Prostate cancer is the second most prevalent cancer among males globally [1,2]. In over half of the countries (105/185), prostate cancer is the most frequently diagnosed cancer among men [3]. According to guidelines and expert consensus, patients with localized prostate cancer can achieve a high five-year survival rate with prompt treatment [4–6]. Therefore, the current guidelines for prostate cancer treatment not only aim to prolong survival, but also improve the patient's quality of life by ensuring a good prognosis [6,7].

Radical prostatectomy (RP) is the predominant surgical treatment for localized prostate cancer [8]. It can effectively improve survival and tumor-specific survival rate [4]. Numerous studies have indicated that urinary incontinence (UI) is the primary complication of radical prostatectomy [9,10]. The postoperative urinary incontinence rate ranges from 6% to 20% [11], significantly reducing the patients' quality of life [12–14]. Thus, it is crucial to study the factors that correlate with postoperative UI. Predicting postoperative UI in patients before surgery is beneficial. It is important for medical teams to prevent and manage potential urinary incontinence [6,7,15].

Previous studies have shown that a patient's anatomical structure is associated with postoperative UI [16,17]. They linked membranous urethral length (MUL) on preoperative MRI to postoperative UI [18–21]. However, this study requires further validation. To the best of our knowledge, most existing studies have only focused on one or two preoperative MRI anatomical measurements, without comprehensively assessing how a patient's anatomical structure affects postoperative urinary continence (UC) [18–22]. However, patients with prostate cancer have complex anatomical structures in the pelvis, which could be a potential factor related to postoperative UI and the quality of life [17,23]. Recent studies investigating outcomes of UC have used statistical approaches limited to univariate or multiple regression analyses [21,23]. Although these statistical methods are classic, the selection of variables requires subjective human judgement, which could lead to omissions and selection biases in comparison to the actual situation. Thus, they are not intelligent [24]. Therefore, it is necessary to identify better techniques for thoroughly examining and confirming the correlation between UI and other body structures based on preoperative MRI assessments.

Deep learning methods have been reported more frequently [25]. In 2019, Andrew et al. [26] utilized the DeepSurv deep learning model to predict UI recovery in patients who underwent robot-assisted radical prostatectomy (RARP) by using automatic performance metrics and clinical characteris. In 2020, Makoto et al. [27] also employed a deep learning model that achieved an AUC curve of 77.5% to predict the recovery of UI in RARP patients. Deep learning models have been used to predict UI following radical prostatectomy surgery or approaches; however, they do not consider the influence of MUL or other MRI parameters on UC following radical prostatectomy and their significance.

Although deep-learning models can usually predict or judge the results of clinical studies, they lack interpretability [28]. Researchers often focus only on the prediction results of the model and neglect to understand the neural network calculation process because of the typically large parameter scale of neural networks. At present, few studies have analyzed the impact of the parameters on the model from the perspective of explaining neural networks. However, the measurement of these crucial parameters can significantly affect the model and has a direct connection with practical clinical concerns [29]. Therefore, it is necessary to explore the interpretability of deep-learning assessments of UI features after radical prostatectomy.

In 2019, Facebook launched the Captum Library, an open-source model interpretation tool offering a range of attribution algorithms. Researchers can use them to understand the weight or size of input features and model structure on the model output, as well as to explain the model's operation mechanism [30] and identify the key factors that influence the model results. This inspired us to determine what is related to post-radical prostatectomy UI. To date, only a few studies have used the captum to interpret deep learning models in clinical research.

In this study, we employed deep learning and Captum to examine the influence of MUL and other preoperative MRI parameters on postoperative UI. Captum provided timely recommendations for interventions and improved the quality of life of patients after radical prostatectomy, while allowing them to understand the model's workings.

2. Methods

Ethical approval

Patient data were approved by the Clinical Ethics Committee of West China Hospital, Sichuan University (Chengdu, China; No. 2019-869). The hospital's informatics department controlled a secure data server, where patient records were de-identified and saved. In addition, the patients' private information was removed from the dataset. Original patient identifiers were replaced by random numbers. The data and MRIs are anonymous. The study team was trained in patient data security and privacy policies before accessing the data. Privacy rights of the human subjects were observed throughout the study.

2.1. Patients

The inclusion criteria included patients who underwent RARP at the West China Hospital of Sichuan University between July 2019 and December 2020 were included. The exclusion criteria were as follows: (1) missing UI assessment information, (2) missing baseline information, (3) patients without a preoperative MRI examination in the hospital, and (4) Patients with UI who were placing catheters and could not be evaluated.

2.2. Data collection

2.2.1. Clinicopathological characteristics

(1) The demographic indicators included the patient's ID, date of birth, age, height, weight, and BMI; (2) the clinicopathological characteristics included diagnosis, operation method, operation time, Charlson score, pathological TNM stage, and Gleason score.

2.2.2. Evaluation of UC in patients

UC was assessed at a random time between 1 month and 12 months postoperatively. According to the corresponding questions of UI score dimension in EPIC-26 scale [31] (q-27, question 3: how many sanitary pads or adult diapers do you need every day to control urine dripping), evaluate the patient's UC. This study defined UC as 'zero piece of sanitary pad or adult diaper per day' used by postoperative patients [31].

2.2.3. Acquisition and measurement of MRI parameters before operation

GE, USA's 3.0T equipment was used to perform MRI examinations in this study. Under the demonstration and guidance of imaging physicians, all preoperative MRI parameters were measured by seven blinded observers, including the authors who were blinded. All preoperative MRI parameters were measured using T2 weighted images. The definitions/measurement methods for the specific measurement parameters are listed in Table 1.

2.3. Statistical analysis

This study analyzed the correlations between preoperative MRI parameters and UC outcomes. To compare the baseline characteristics of the two groups of patients with or without UC recovery, the Kruskal-Wallis rank-sum test was used to reveal differences in continuous variables, and Fisher's exact probability test was used to reveal the differences in categorical variables.

Table 1
MRI measurement names and their definitions.

Item	Feature	Abbreviation	Definition
Sagittal plane	Membranous urethral length	MUL	The distance from the top of prostate to the bottom of the bulbar urethra
	Membranous urethral Angle	MUA	An obtuse angle obtained by crossing a horizontal line under the pubic symphysis and a straight line parallel to the membranous urethra
	Prostate length	PL	The maximum height of the prostate measured from the base of the prostate to the apex in any plane
	Pubic symphysis-prostate apex length	PAL	Refers to the distance between the sagittal line of the prostatic crest and the prostatic crest line, and the vertical distance between the horizontal extension line of the suprapubic crest line and the horizontal extension line of the prostate apex.
Axial plane	Thickness of left obturator internus muscle	TLOIM	The maximum thickness of left obturator muscle measured on any plane
	Thickness of right obturator internus muscle	TROIM	The maximum thickness of right obturator muscle measured on any plane
	Prostate width	PW	The maximum prostate width measured at the same level as the maximum width
	Prostate height	PH	The maximum prostate height measured at the same level as the maximum width.
	Urethral width	UW	Defined as the maximum diameter of the urethra.
	Distance of outer of the levator ani muscle	DOLAM	Defined as the distance from the outer edge of the levator ani muscle measured at the same level as the distance from the levator ani muscle
	Distance of inner of the levator ani muscle	DILAM	Defined as the narrowest distance from the inner edge of levator ani muscle to the urethra below the tip tail edge of the prostate
	Transverse membranous urethral thickness	TMUT	Measure the transverse maximum thickness of membranous urethra at the penis root level
	Anterior and posterior membranous urethral thickness	APMUT	Measure the maximum longitudinal thickness of membranous urethra at the level of penis root
	Left anal sphincter thickness	LST	The maximum thickness of the left anal sphincter measured in any plane
	Right anal sphincter thickness	RST	The maximum thickness of the right anal sphincter measured in any plane
	Urethral wall thickness	UWT	The widest part of the membranous urethra in the axial sequence before entering the prostate.
Coronal plane	Thickness of left levator ani muscle	TLLAM	The measured maximum thickness of left levator ani muscle.
	Thickness of right levator ani muscle	TRLAM	The measured maximum thickness of right levator ani muscle.
Calculated value	The thickness of levator ani muscle in axial plane	TLAM	The difference of DOLAM and DILAM divided by 2
	The cross section surface area of membranous urethra	CSAMU	Calculated based on the formula: (TMUT * 1/2) * (APMUT * 1/2)*π
	Prostate volume	PV	Calculated based on the following formula: PL * PH * PW *π/6
	Membranous urethral volume	MUV	Calculated by the following formula: CSAMU * MUL

Least Absolute Shrinkage and Selection Operator (LASSO) regression analyses were used to select potential predictors by measuring the predictability of preoperative MRI parameters in UC. After 10-fold cross-validation, the tuning parameter (λ) was tested, and the preoperative MRI parameters were selected when the lambda value achieved the minimum mean cross-validated error.

Univariate Logistic regression analysis was used to analyze the preoperative MRI anatomical parameters screened by LASSO regression, as well as the clinicopathological indicators of the patients. The results are presented as odds ratios (ORs), 95% confidence intervals, and p-values.

All statistical analyses were performed using R language and EmpowerStats statistical analysis software. Statistical significance was set than 0.05. significant.

2.4. Model construction

This study included clinical data related to UI from 91 patients. Because the number of samples in the data was not balanced, we used weighted sampling to address class imbalance. We defined the weight of each class as 1/total number of classes.

We built a neural network consisting of input, hidden, and output layers. We then set the number of nodes in these layers to 35, 256, and 2. In addition, the neural network used cross-entropy loss as the loss function and AdamW as the optimizer. We set the initial learning rate of the network to 0.01 and then decayed the learning rate using cosine annealing. After training the neural network 10,000 times, we obtained a neural network with an accuracy rate close to 100%.

Captum was used to explore the neural network model and its integrated gradient method was used to determine the impact of each feature on the predictive results of the neural network. Captum calculated the contribution of feature α to the results, like $C_\alpha = [C_1^\alpha, C_2^\alpha, \cdots, C_N^\alpha]$, where C_i^α represented the individual contribution value of the i-th sample under the feature α and N was the number of samples. We defined the total contribution of feature α as $TC_\alpha = \sum_{i=1}^{N} |C_i|/N$, which represented the average of all samples.

Then, features with an absolute value greater than 0.05 were selected as key features that might affect UC. Finally, we identified the features that were critical for UC. These features are used to construct a deep neural network model. The neural network consists of an input layer, four hidden layers, and an output layer. The numbers of nodes in the layers were 35, 256, 128, 128, 64, and 2.

The model was used to perform 100 accuracy comparisons with machine learning methods, such as K Nearest Neighbors (KNN), Logistic Regression, Support Vector Machine (SVM), Random Forest, and XGboost (random seeds can be found in Supplementary Table S1). For the neural networks, we set random number seeds to ensure that the training samples were fully utilized, and for

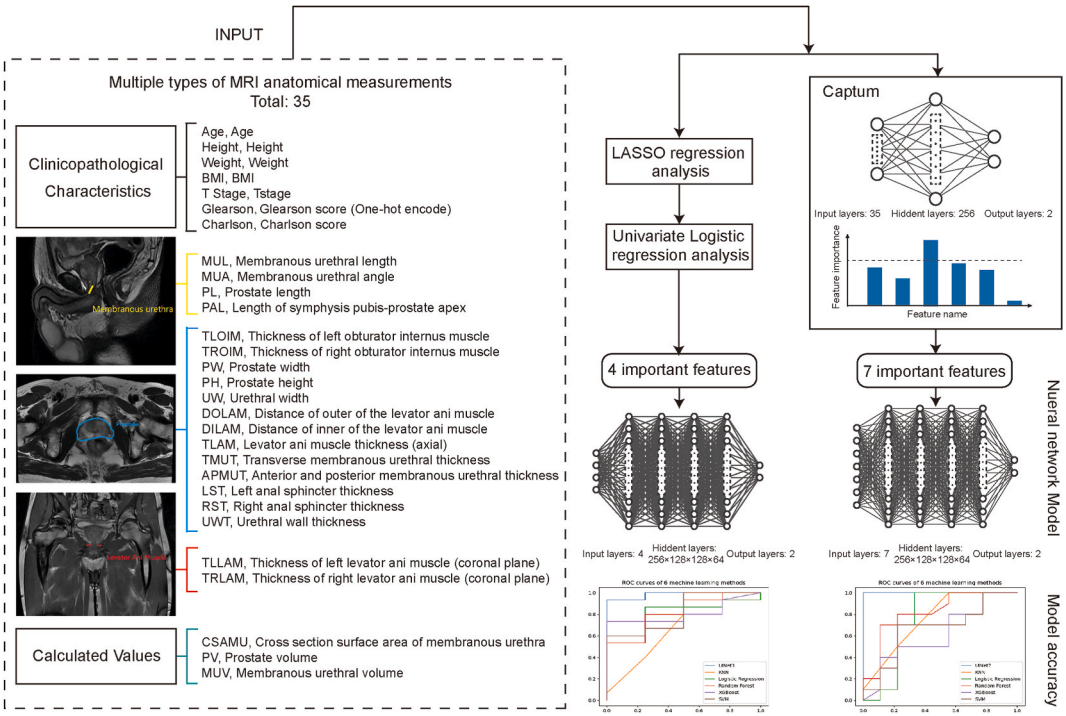


Fig. 1. Overview of study design and the architecture of explainable neural network model. A total of 35 features (one named Gleason score were encoded by one hot), including clinicopathological characters, anatomical measurements, and calculated values were collected and used in feature selection. Two methods were performed in feature selection. One was the traditional method, which included LASSO regression and univariate logistic regression. And the other was called Captum, which is a novel algorithm used to explain neural network models. Both methods were used to screen out key features. Then these key features were fed into neural network models as input. Finally, ROCs and AUCs were used to compare the performance of neural network models to other models.

machine learning, a three-fold cross-validation method was adopted to enhance data availability. The accuracy of the model was measured using the average accuracy over 100 tests and a 95% confidence interval (with a 95% confidence interval). Differences in accuracy between the models were compared using t-tests.

3. Result

In this study, we presented a model for predicting UI after RARP using inputs selected from different types of MRI anatomical measurements There are two methods for feature selection: LASSO regression and univariate Logistic regression; and Captum, a novel algorithm for interpreting neural network models. The key features selected by these two methods were used to construct neural network models. Fig. 1 shows that the neural network models outperformed the other five machine learning models in predicting UC.

3.1. Clinicopathological characteristics and univariate analysis

A total of 91 patients with prostate cancer who underwent RARP were enrolled in this study. In total, 58 patients were identified without UI and 33 patients with UI were identified during the follow-up period. There were significant differences (p-value <0.05) in age, MUL, and PAL between the two groups of patients with and without UI, but no significant differences in other aspects. Details of the baseline data and MRI parameters are presented in Table 2.

Table 2
Clinicopathological characteristics and MRI parameter information of patients.

Information	Total	Urinary incontinence	Urinary continence	P-value
Number of patients	91	33	58	
Age (year)	67.20 ± 7.55	69.70 ± 5.11	65.78 ± 8.35	0.008*
Height (cm)	166.92 ± 6.23	166.09 ± 7.63	167.40 ± 5.30	0.335
Weight (kg)	67.34 ± 10.17	67.63 ± 10.99	67.17 ± 9.76	0.655
BMI (km/m^2)	24.11 ± 2.88	24.41 ± 2.60	23.94 ± 3.04	0.656
Charlson score	2.57 ± 1.36	2.55 ± 0.94	2.59 ± 1.56	0.495
MUL (mm)	15.12 ± 3.63	13.49 ± 3.23	16.04 ± 3.55	<0.001*
TLOIM (mm)	18.45 ± 3.38	18.18 ± 3.76	18.61 ± 3.16	0.65
TROIM (mm)	18.63 ± 3.40	17.91 ± 3.45	19.03 ± 3.32	0.22
PL (mm)	38.89 ± 7.89	38.93 ± 6.79	38.87 ± 8.51	0.901
PW (mm)	44.43 ± 7.85	43.70 ± 8.51	44.85 ± 7.49	0.324
PH (mm)	33.59 ± 7.50	33.22 ± 7.08	33.80 ± 7.78	0.754
MUA (°)	1318.07 ± 137.35	1325.65 ± 140.01	1313.76 ± 136.86	0.751
PAL (mm)	30.16 ± 6.36	32.88 ± 6.25	28.62 ± 5.94	0.004*
UW (mm)	10.07 ± 2.85	9.73 ± 2.84	10.26 ± 2.86	0.478
DOLAM (mm)	37.55 ± 4.95	38.55 ± 5.87	36.98 ± 4.29	0.349
OILAM (mm)	18.28 ± 4.79	19.76 ± 6.36	17.44 ± 3.38	0.084
TMUT (mm)	11.67 ± 2.80	11.44 ± 2.56	11.80 ± 2.94	0.723
APMUT (mm)	11.61 ± 2.36	11.55 ± 2.36	11.65 ± 2.38	0.928
LST (mm)	6.92 ± 1.93	6.94 ± 2.03	6.91 ± 1.89	0.722
RST (mm)	6.63 ± 1.94	6.58 ± 2.31	6.67 ± 1.73	0.485
TLLAM (mm)	9.98 ± 2.48	9.95 ± 2.92	10.00 ± 2.21	0.612
TRLAM (mm)	9.88 ± 2.48	10.06 ± 3.14	9.78 ± 2.03	0.83
UWT (mm)	13.55 ± 2.78	13.92 ± 2.97	13.34 ± 2.66	0.56
TLAM (mm)	9.64 ± 2.08	9.40 ± 2.15	9.77 ± 2.04	0.552
CSAMU (mm^2)	110.55 ± 46.89	107.79 ± 41.84	112.12 ± 49.82	0.911
PV (mm^3)	31792.70 ± 15092.17	30697.46 ± 14593.36	32415.85 ± 15459.40	0.368
MUV (mm^3)	1672.06 ± 777.73	1499.28 ± 598.69	1770.37 ± 852.47	0.238
T stage				0.858
1	1 (1.10%)	1 (3.03%)	0 (0.00%)	
2	21 (23.08%)	9 (27.27%)	12 (20.69%)	
2a	15 (16.48%)	5 (15.15%)	10 (17.24%)	
2b	20 (21.98%)	6 (18.18%)	14 (24.14%)	
2c	13 (14.29%)	5 (15.15%)	8 (13.79%)	
3	1 (1.10%)	0 (0.00%)	1 (1.72%)	
3a	7 (7.69%)	2 (6.06%)	5 (8.62%)	
3b	7 (7.69%)	2 (6.06%)	5 (8.62%)	
4	6 (6.59%)	3 (9.09%)	3 (5.17%)	
Gleason score				0.682
3 + 3 = 6	11 (12.09%)	5 (15.15%)	6 (10.34%)	
3 + 4 = 7	29 (31.87%)	9 (27.27%)	20 (34.48%)	
4 + 3 = 7	25 (27.47%)	7 (21.21%)	18 (31.03%)	
3 + 5 = 8	1 (1.10%)	0 (0.00%)	1 (1.72%)	
4 + 4 = 8	11 (12.09%)	6 (18.18%)	5 (8.62%)	
4 + 5 = 9	12 (13.19%)	5 (15.15%)	7 (12.07%)	
5 + 4 = 9	2 (2.20%)	1 (3.03%)	1 (1.72%)	

The data in the table are: mean ± standard deviation/number of cases (%)* <0.05.

LASSO regression analysis was used for the preliminary calculation of preoperative MRI parameters to screen for potential predictors of UC. The minimum λ value of the average cross-validation error was selected as TROIM, PAL, MUL, and DILAM.

The scores were calculated using the following formula: 0.02035 × TROIM (mm) + 0.12496 × MUL (mm) × 0.0599 × PAL (mm) × 0.05918 × DILAM (mm). (Supplementary Fig. S1).

Univariate analysis was performed to analyze the relationship between the selected MRI parameters, patient baseline indicators, and UC. The analysis revealed that age (OR: 0.93, 95% CI: 0.877–0.99, *p*-value: 0.0200), MUL (OR: 1.28, 95% CI: 1.09–1.50, *p*-value: 0.0024), PAL (OR: 0.89, 95% CI: 0.82–0.96, *p*-value: 0.0032), and DILAM (OR: 0.89, 95% CI: 0.80–1.00, *p*-value: 0.0421) were potentially associated with UI (Supplementary Table S2).

3.2. Discovered feature

Based on the follow-up results, we divided patients into two groups according to the criteria of "zero use of sanitary pad or adult diaper per day" after the operation: UI (label = 0) and UC (label = 1). The AUC of the one-hidden-layer neural network model was close to 100% (see Methods). We used Captum, which is based on a neural network, to assess the importance of these features in the model. Fig. 2 shows the results obtained from the capture. We selected age, MUL, PAL, UW, DILAM, APMUT, and MUV as the key features when the threshold was set to 0.05. Among these features, lasso regression and univariate analysis demonstrated that age, MUL, PAL, and DILAM were substantially associated with UI.

Seven Captum features were chosen to reconstruct UINet7, a 5-layer neural network. The datasets were divided into a training and test sets at a 4:1 ratio, and the UINet7's prediction accuracy was 0.97 (95% CI: 0.90–1.00) and AUC was 0.98. (95% CI: 0.93–1.00). Fig. 3(A) shows the approaches used for comparison with UINet7: K-Nearest Neighbors (KNN), Logistic Regression, Random Forest, Extreme Gradient Boosting (XGBoost), and Support Vector Machine (SVM). UINet7 is better than the other machine learning methods.

Using six techniques (Deep Neural Network, KNN, Logistic Regression, Random Forest, XGBoost, and SVM), we compared the four characteristics discovered by the univariate analysis with the seven features selected by Captum. The results are shown in Fig. 3(B), Table 3, Fig. 4(A) and (B). Other machine learning models exhibited little improvement in addition to the four-feature model's accuracy and AUC value, which were significantly reduced when compared to the seven-feature model. This indicating that Captum was more suitable than univariate analysis for identifying the key features of neural networks.

Although the UINet7's prediction accuracy (0.97) was substantially higher than that of UINet4, it required healthcare staff to measure more than twice the number of features. However, UINet4's prediction accuracy may still approach 0.81, even when only four features were measured. This satisfies the requirements of clinical therapy in terms of diagnosis and treatment. Therefore, we suggest employing the UINet4 model for applications that do not require precision. The UINet7 model should be applied to medical situations requiring high precision.

4. Discussion

With the advancements in artificial intelligence, the field of medicine has entered an era of enhanced precision and intelligence [32]. However, most existing clinical studies still rely on traditional regression analysis as the primary statistical method for data processing and fail to fully exploit the potential of artificial intelligence. In terms of data analysis and modeling, machine learning and neural networks have unique advantages and characteristics. In this investigation, we used postoperative UI in patients with prostate cancer as a practical example to explore these methods. To conduct a comprehensive comparison, six algorithms–Deep Neural Network, KNN, Logistic Regression, Random Forest, XGBoost, and SVM– were used to analyze and compare their performances. It is

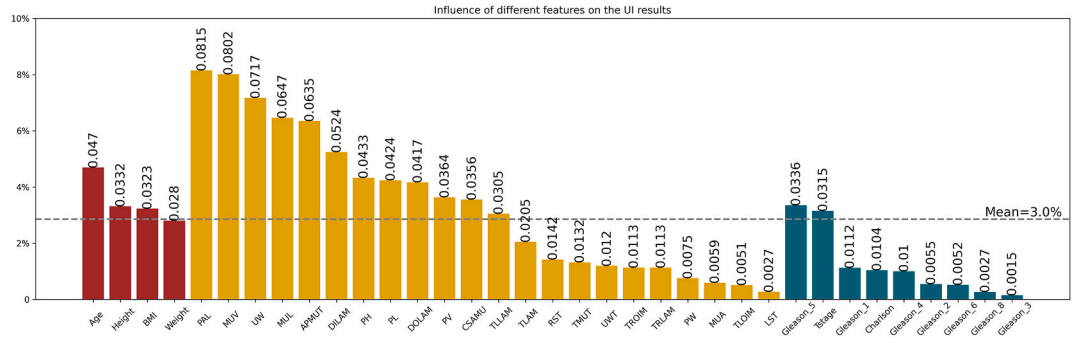
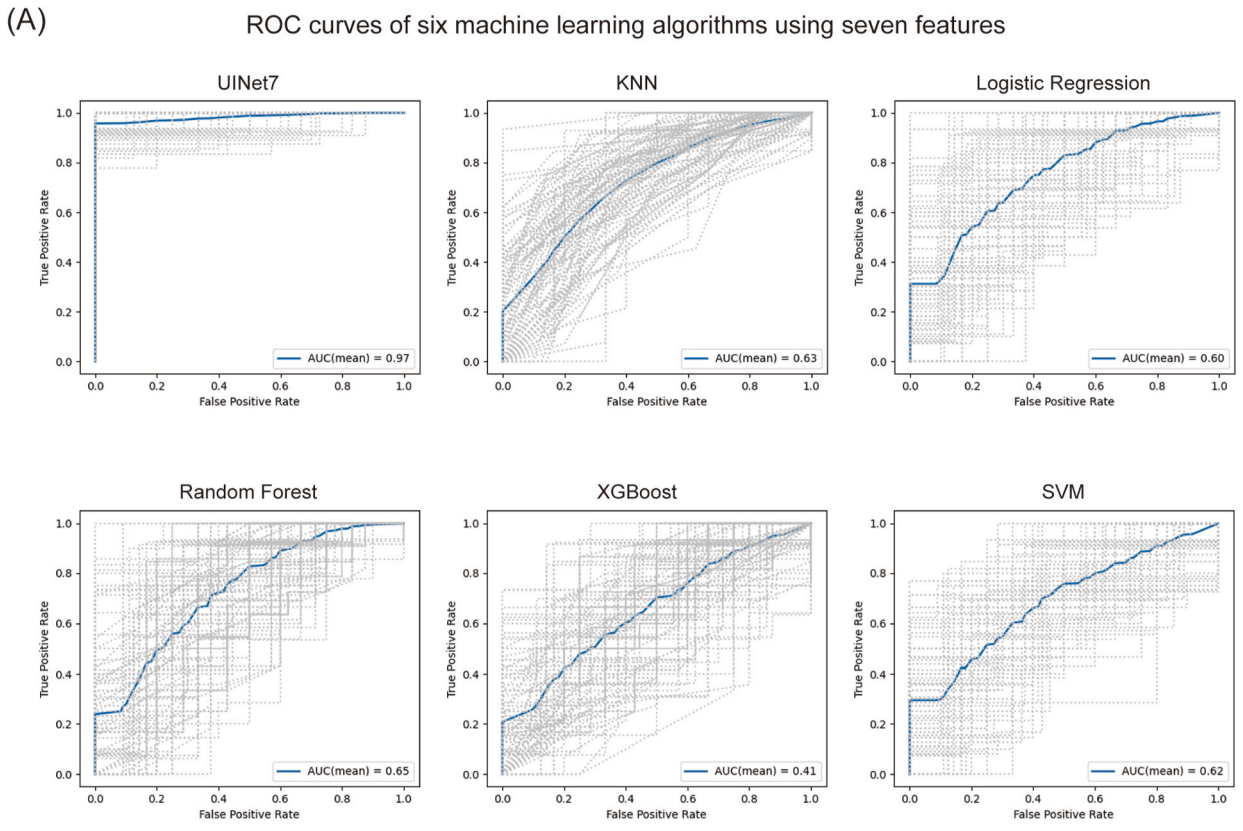
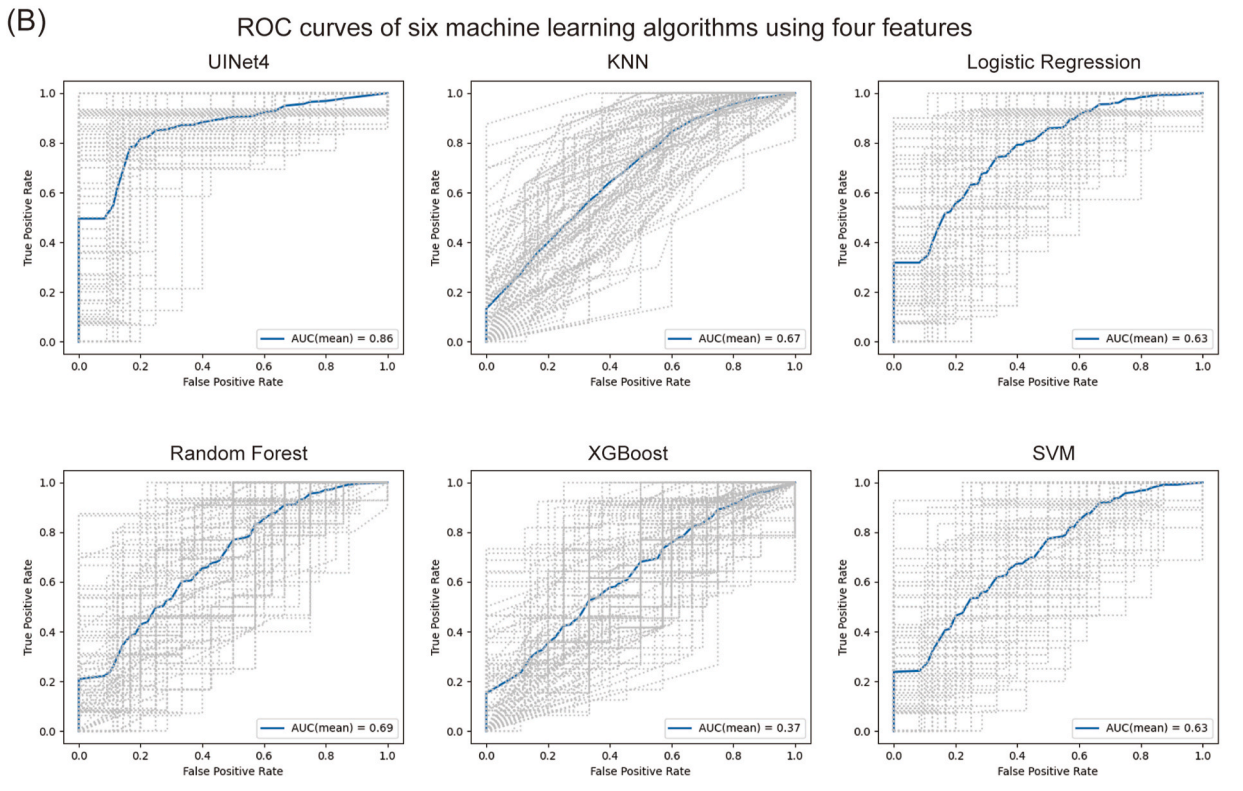


Fig. 2. The contribution of each features as evaluated by Captum. The abscissa signifies the abbreviated designations of the features, while the ordinate delineates the proportional contribution of each individual feature towards the predictive outcome, represented as percentages on the vertical scale. These contributions encapsulate their respective influence on the incidence of urinary incontinence subsequent to radical prostatectomy. The numerical notations on each bar mirror the precise values as derived from Captum, subject to an upper limit of 1. The colour differentiation of each bar indicates distinct classifications of data: red symbolises the baseline patient information, yellow represents anatomical measurements ascertained from MRI, and green designates patient pathology data. A dotted horizontal line superimposed on the graph signifies the mean value of the influence of all parameters on postoperative urinary incontinence.





(caption on next page)

Fig. 3. Comparisons of the ROC results of different models. The figure shows the ROC of UINet7, KNN, Logistic Regression, Random Forest, XGBoost, and SVM models, respectively. A total of 100 different tests were performed, the gray curve represents the ROC curve obtained for each test, and the blue curve represents the average of 100 curves.

Table 3
The performances of six machine learning models when using four features obtained by univariate analysis and seven features selected by Captum respectively.

Method	UINet4		UINet7	
	Mean ACC (95% CI)	Mean AUC (95% CI)	Mean ACC (95% CI)	Mean AUC (95% CI)
Neural Network	0.86, CI:0.72–1.00	0.85, CI:0.67–1.00	0.97, CI:0.90–1.00	0.98, CI:0.93–1.00
KNN	0.67, CI:0.45–0.90	0.68, CI:0.44–0.92	0.63, CI:0.38–0.89	0.73, CI:0.52–0.93
Logistic Regression	0.63, CI:0.37–0.90	0.77, CI:0.54–0.99	0.59, CI:0.31–0.89	0.74, CI:0.51–0.98
Random Forest	0.69, CI:0.47–0.91	0.69, CI:0.44–0.94	0.65, CI:0.40–0.90	0.73, CI:0.49–0.96
XGboost	0.37, CI:0.10–0.64	0.63, CI:0.38–0.88	0.41, CI:0.13–0.69	0.65, CI:0.39–0.90
SVM	0.63, CI:0.38–0.89	0.70, CI:0.46–0.94	0.62, CI:0.30–0.94	0.68, CI:0.48–0.88

KNN═K-Nearest Neighbors; XGBoost = Extreme Gradient Boosting; SVM═Support Vector Machine.

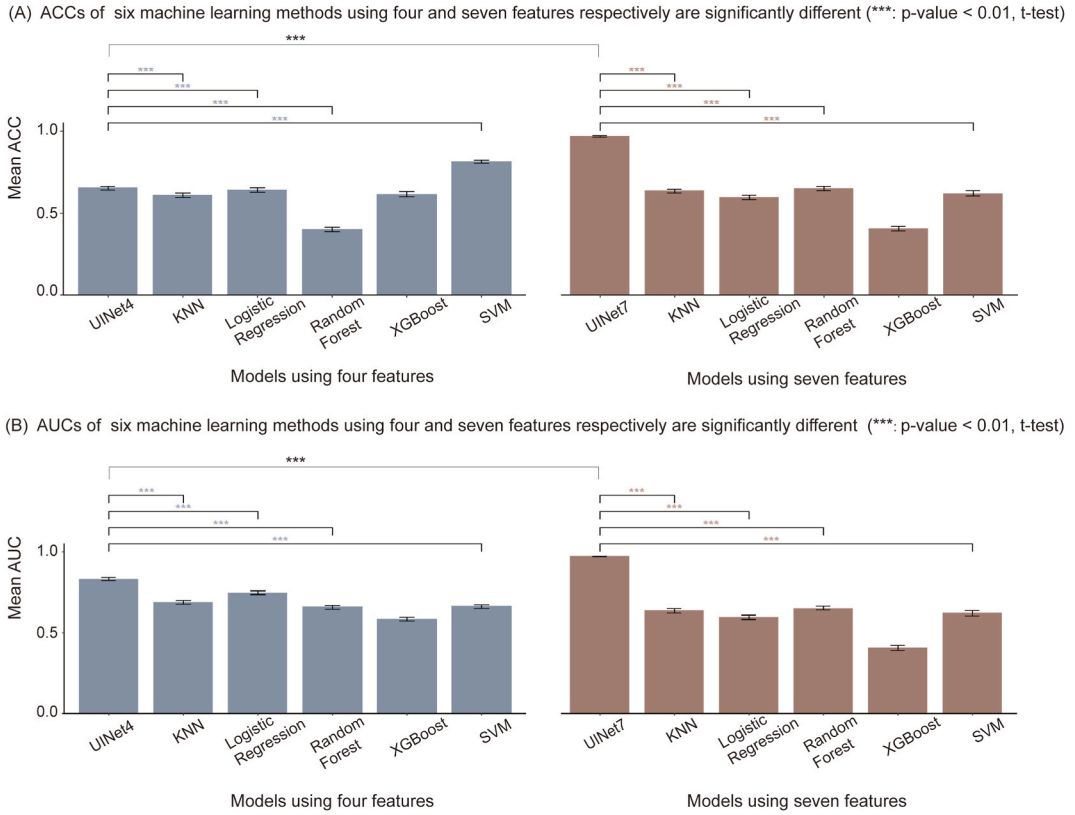


Fig. 4. There is a significant difference in the ACC (A) and AUC (B) metrics of the six machine learning algorithms when using four features and seven features by *t*-test. Whether using four features or using seven features, the models obtained by the neural networks (UINet4 and UINet7) are significantly higher than the other five machine learning methods. Compared with four features, the ACC and AUC of UINet7 are significantly higher.

clear that the neural network model proposed in this study outperforms the five representative machine learning models for predicting UI. Seven significant features were identified by the captum model, of which four were consistent with those identified by the LASSO regression analysis and univariate analysis. The accuracy of our model was 0.86, when these four features were used for prediction, and the accuracy reached 0.97, when all seven features were used. The model, benefiting from Captum, revealed that postoperative UI after RARP was significantly dependent on seven preoperative MRI measurements. As a result, these are the potential factors affecting UI. It also reflects the advantages and intelligence of deep neural networks, providing a reference for future clinical applications [32].

4.1. Data Considerations: characteristics of the sample

Our sample originated from a large medical center in China, which enriched the data on the Chinese population that was lacking in similar existing studies. This lays the foundation for future studies of population heterogeneity and diversity. People often associate intelligent medicine with big data [33]. To ensure that the training samples were used to their full potential when modeling neural networks, we set random number seeds for the sample size. At the same time, we adopted a three-fold cross-validation method for machine learning to enhance data availability. Several studies have investigated this topic with a sample size equivalent to the seven crucial factors revealed in our study [34–36]. Several of the MRI measurement values (including MUL) obtained in these studies are identical to those found in our study. However, these studies focused on very few (mostly fewer than 10) MRI measurements. More importantly, the sample sizes of these studies [34–36] ranged from 64 to 123 patients and were remarkably similar to those of our study. Therefore, our study has more advantages than the ones mentioned since it has significantly more anatomical preoperative MRI components [37]. Although this study did not involve big data analysis, the difficulty of data analysis was significantly increased by the fact that each sample had a large number of parameters and that dozens of critical anatomical factors needed to be examined. Fortunately, the neural network method has been well adapted to sample data and scientifically analyzed, demonstrating the advantages of artificial intelligence for this type of small sample size research. This provides a reference for the application of artificial intelligence to relatively small samples.

4.2. Key parameters: the interpretability of models

The vast array of parameters used in this analysis was derived from MRI-based anatomical data, emphasizing the pivotal role of the involved anatomical structures. Several meticulously selected variables (e.g., age, MUL, PAL, UW, DILAM, APMUT, and MUV) were nested in the UINet7 model. These parameters spanned three-dimensional prostate meridians, attributes connected to the pelvic basin musculature, and those pertinent to the urethra. Fig. 2 shows an overview of the cumulative effects of these parameters on the incidence of postoperative UI. Among the anatomical parameters derived from MRI, the prostate cancer meridian functions as an index of prostate size. We inferred an association between postoperative UI and complete prostate resection during RARP.

The muscles considered in our study encompassed three distinct structures: the obturator internus muscle, levator ani muscle, and anal sphincter. The thicknesses of these muscles were measured across three planes using MRI. A considerable body of evidence has confirmed the correlation between pelvic floor musculature and postoperative UI [38]. In line with the established research practices, our study incorporated the aforementioned trio of muscles. The levator ani muscle assumed precedence, and the DILAM within our UINet7 model extended the findings of previous research. Postoperative UI is a consequence of muscle weakening in this area [38]. Furthermore, research has definitively indicated a significant relationship between the levator ani muscle and postoperative UI after RARP [39]. The use of MRI to measure the width of the levator ani muscle aids in predicting postoperative UI.

Additionally, the pelvic fascia diaphragm and obturator muscle fascia are closely associated with incontinence [40]. Regarding the anal sphincter, scholarly discourse suggests that electromyographic activity serves as an indicator of UI stress [41]. Consistent with our study approach, the research conducted by Takuya Sadahira et al. used preoperative MRI scans to evaluate the anal sphincter. They did not report any successful outcomes associated with this parameter, which is consistent with ours [39].

With reference to the urethra, our primary focus within the MRI scans was on the dimensions, specifically the length, width, and thickness of the urethral wall. Key anatomical parameters related to the urethra were discerned via the captum, echoing the findings of previous studies. Numerous studies have identified MUL as a major factor influencing postoperative UI [11]. A thorough investigation and meta-analysis described the effects on UC after radical prostatectomy, and a longer preoperative MRI scan was associated with faster recovery after surgery [42]. Both UW and MUV were correlated with MUL. While the MUL is two-dimensional, the UW and MUV serve as three-dimensional equivalents that collectively describe the state of the urethra. Moreover, in the sagittal plane of the MRI, the PAL denotes the anatomical position of the vesicoureteral anastomosis at the RARP. According to Fukui et al., a shorter preoperative MRI-evidenced PAL is associated with improved postoperative UC [43].

4.3. Clinical operability of outcome indicators

Regarding the selection of outcome measures, the number of urine pads used to determine UC recovery was evaluated using the EPIC-26 scale. We employed a questionnaire-based approach to assess the condition of urinary incontinence in patients. This method is not only straightforward and easy to implement but also offers strong clinical feasibility and practicality. At present, the definition of outcome measures in many international studies [11,19,36,37,43–45] used 0 pads per day, and many other studies adopt the epic-26 scale [37,43,44]. These are consistent with the definition of outcome measures and even the evaluation scale. The outcomes of the aforementioned investigations are consistent with our findings, which is worth emphasizing. This finding highlights the effect of preoperative MUL on postoperative UC.

4.4. Applying captum for Deeper clinical Insights and limitations

In 2022, Jia et al. [46] used Captum to identify the important clinical parameters affecting BleNet, a neural network model for predicting late severe bleeding after transcatheter aortic valve replacement. However, Jia et al. [46] did not further investigate the important clinical features identified by Captum or analyze whether these clinical features could be helpful in clinical decision-making. In addition, they did not compare the similarities and differences between the features found by Captum and the univariate analyses. Their research did not explain how Captum interpreted deep learning models or how they could help with clinical decision-making. In this study, the same method was applied to clinical problems. We attempted to combine the captum interpretation deep learning model with practical clinical issues. Although the conclusion drawn was in line with the actual clinical situation, we suggest that further research is needed to verify this conclusion. This will also provide more references for the application of interpretable deep learning models to medical clinical problems and contribute to the intelligent transformation and application of medical clinical practice.

Our study has some limitations. First, the evaluation time of UI in this study was not completely unified; therefore, it can only reflect short-term postoperative UC. Secondly, the results of the algorithm in this study can only show the correlation; the specific effect value still requires optimization in the subsequent calculation model. In addition, although this study aimed to apply captum to clinical practice problems, the study's findings would be more strongly supported by big data if available. Therefore, there is still great room of improvement in the sample size used in our study. In addition, although the captim model has been published internationally, its interpretability requires further research to prove its rigor in dealing with scientific issues.

4.5. Implications for clinical practice and future research

In terms of academic research, this article fills the gap in in-depth research on UI features after RARP owing to limitations in clinical data and applicable models. By examining anatomical measurements with these seven characteristics, this result has a reference value for accurately optimizing surgical plans in clinical practice and may improve surgical techniques separately. However, based on our predictions, the results will help patients complete preoperative informed consent, psychological development planning, postoperative care, and rehabilitation in advance, allowing them to choose the best plan to recover from UC as soon as possible. Additionally, appropriate assistance can be provided as a reference value for formulating the best treatment strategy for patients by thoroughly evaluating the difficulty of postoperative UC and predicting postoperative UC. In summary, this study is an important attempt to integrate artificial intelligence techniques into clinical practice with potential transformative significance. It also provides a reference for how artificial intelligence can transform into a medical issue.

In the future, it will help accurately optimize surgical techniques, and it will be possible to improve surgical techniques solely by examining anatomical measurements using these seven features. Our accurate prediction results may assist patients in obtaining preoperative informed consent, planning for psychological support, and advanced postoperative nursing and rehabilitation, thereby allowing them to choose the best plan for regaining UC as quickly as possible.

Furthermore, after a thorough assessment of postoperative UC difficulties and prediction of postoperative UC, the optimal treatment strategy for patients can be appropriately determined.

5. Conclusions

In this study, we discovered that for UC after RARP, only seven MRI measurements were required for neural network modeling. This study represents an attempt to apply the capture neural network model to analyze medical problems, thereby providing a significant reference for the application of artificial intelligence in clinical practice. We anticipate further research to provide more robust scientific evidence for the clinical application of interpretable deep learning models, thereby guiding medical practice and promoting the advancement of medicine.

Author contribution statement

Jiakun Li: Conceived and designed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Xuemeng Fan: Conceived and designed the experiments; Performed the experiments; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Tong Tang, Erman Wu: Conceived and designed the experiments; Wrote the paper.

Dongyue Wang, Hui Zong, Xianghong Zhou, Yifan Li, Chichen Zhang, Yihang Zhang: Performed the experiments; Contributed reagents, materials, analysis tools or data.

Rongrong Wu: Performed the experiments; Analyzed and interpreted the data.

Cong Wu: Conceived and designed the experiments; Wrote the paper.

Lu Yang, Bairong Shen: Conceived and designed the experiments; Contributed reagents, materials, analysis tools or data; Wrote the paper.

1 - Conceived and designed the experiments.

2 - Performed the experiments.

3 - Analyzed and interpreted the data.

4 - Contributed reagents, materials, analysis tools or data.

5 - Wrote the paper.

Data availability statement

Data will be made available on request.

Ethics declarations

This study was reviewed and approved by the Research Ethics Committee of West China Hospital of Sichuan University (approval number: No. 2019-869). Informed consent was not required for this study because the research data and images were anonymised, and the institutional ethics committee determined it was not necessary based on the study's nature.

Consent for publication

Not applicable.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was supported by the Sichuan Department of Science and Technology (No. 2021ZYD0018) and the National Nature Science Foundation of China (No. 32070671 and No. 32270690).

The authors thank Dr. Changzhong Chen, Chi Chen, and Xin-Lin Chen (EmpowerStats X&Y Solutions, Inc., Boston, MA, USA) for providing the statistical methodology consultation. The authors gratefully thank Jing Zhao for contribution to language revision.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2023.e20337.

References

[1] K.D. Miller, L. Nogueira, T. Devasia, et al., Cancer treatment and survivorship statistics, CA Cancer J Clin 72 (5) (2022) 409–436, 2022.
[2] G.P. Haas, N. Delongchamps, O.W. Brawley, C.Y. Wang, G. de la Roza, The worldwide epidemiology of prostate cancer: perspectives from autopsy studies, Can. J. Urol. 15 (1) (2008) 3866–3871.
[3] F. Bray, J. Ferlay, I. Soerjomataram, R.L. Siegel, L.A. Torre, A. Jemal, Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA Cancer J Clin 68 (6) (2018) 394–424.
[4] N. Mottet, R. van den Bergh, E. Briers, et al., EAU-EANM-ESTRO-ESUR-SIOG guidelines on prostate cancer-2020 update. Part 1: screening, diagnosis, and local treatment with curative intent, Eur. Urol. 79 (2) (2021) 243–262.
[5] K.S. Virgo, R.B. Rumble, J. Talcott, Initial management of noncastrate advanced, recurrent, or metastatic prostate cancer: ASCO guideline update, J. Clin. Oncol. (2023) JCO2300155.
[6] E. Schaeffer, S. Srinivas, E.S. Antonarakis, et al., NCCN guidelines insights: prostate cancer, version 1.2021, J Natl Compr Canc Netw 19 (2) (2021) 134–143.
[7] J.A. Ligibel, K. Bohlke, A.M. May, et al., Exercise, diet, and weight management during cancer treatment: ASCO guideline, J. Clin. Oncol. 40 (22) (2022) 2491–2507.
[8] V.R. Patel, H.M. Abdul-Muhsin, O. Schatloff, et al., Critical review of 'pentafecta' outcomes after robot-assisted laparoscopic prostatectomy in high-volume centres, BJU Int. 108 (6 Pt 2) (2011) 1007–1017.
[9] N. Eisemann, S. Nolte, M. Schnoor, A. Katalinic, V. Rohde, A. Waldmann, The ProCaSP study: quality of life outcomes of prostate cancer patients after radiotherapy or radical prostatectomy in a cohort study, BMC Urol. 15 (2015) 28.
[10] C. Jeldres, J. Cullen, L.M. Hurwitz, et al., Prospective quality-of-life outcomes for low-risk prostate cancer: active surveillance versus radical prostatectomy, Cancer 121 (14) (2015) 2465–2473.
[11] V. Ficarra, G. Novara, R.C. Rosen, et al., Systematic review and meta-analysis of studies reporting urinary continence recovery after robot-assisted radical prostatectomy, Eur. Urol. 62 (3) (2012) 405–417.
[12] M.G. Sanda, R.L. Dunn, J. Michalski, et al., Quality of life and satisfaction with outcome among prostate-cancer survivors, N. Engl. J. Med. 358 (12) (2008) 1250–1261.
[13] M.A. Liss, K. Osann, N. Canvasser, et al., Continence definition after radical prostatectomy using urinary quality of life: evaluation of patient reported validated questionnaires, J. Urol. 183 (4) (2010) 1464–1468.
[14] S. Punnen, J.E. Cowan, J.M. Chan, P.R. Carroll, M.R. Cooperberg, Long-term health-related quality of life after primary treatment for localized prostate cancer: results from the CaPSURE registry, Eur. Urol. 68 (4) (2015) 600–608.
[15] J. Ma, W. Xu, R. Chen, et al., Robotic-assisted versus laparoscopic radical prostatectomy for prostate cancer: the first separate systematic review and meta-analysis of randomised controlled trials and non-randomised studies, Int. J. Surg. 109 (2023) 1350–1359.
[16] S.A. Greenberg, J.E. Cowan, P.E. Lonergan, et al., The effect of preoperative membranous urethral length on likelihood of postoperative urinary incontinence after robot-assisted radical prostatectomy, Prostate Cancer Prostatic Dis. 25 (2) (2022) 344–350.

[17] M.C. van Dijk-de Haan, T.N. Boellaard, R. Tissier, et al., Value of different magnetic resonance imaging-based measurements of anatomical structures on preoperative prostate imaging in predicting urinary continence after radical prostatectomy in men with prostate cancer: a systematic review and meta-analysis, Eur Urol Focus 8 (5) (2022) 1211–1225.
[18] C. von Bodman, K. Matsushita, C. Savage, et al., Recovery of urinary function after radical prostatectomy: predictors of urinary function on preoperative prostate magnetic resonance imaging, J. Urol. 187 (3) (2012) 945–950.
[19] K. Matsushita, M.T. Kent, A.J. Vickers, et al., Preoperative predictive model of recovery of urinary continence after radical prostatectomy, BJU Int. 116 (4) (2015) 577–583.
[20] K. Kitamura, T. China, M. Kanayama, et al., Significant association between urethral length measured by magnetic resonance imaging and urinary continence recovery after robot-assisted radical prostatectomy, Prostate Int 7 (2) (2019) 54–59.
[21] H. Lamberg, P.R. Shankar, K. Singh, et al., Preoperative prostate MRI predictors of urinary continence following radical prostatectomy, Radiology 303 (1) (2022) 99–109.
[22] H. Veerman, M.J. Hagens, C.M. Hoeks, et al., A standardized method to measure the membranous urethral length (MUL) on MRI of the prostate with high inter- and intra-observer agreement, Eur. Radiol. 33 (2022) 3295–3302.
[23] V. Iacovelli, M. Carilli, M. Sandri, et al., The role of preoperative prostatic shape in the recovery of urinary continence after robotic radical prostatectomy: a single cohort analysis, Prostate Cancer Prostatic Dis. 26 (2022) 374–378.
[24] J.M. Gorriz, C. Jimenez-Mesa, F. Segovia, J. Ramirez, J. Suckling, A connection between pattern classification by machine learning and statistical inference with the general linear model, IEEE J Biomed Health Inform 26 (11) (2022) 5332–5343.
[25] Data sharing in the age of deep learning, Nat. Biotechnol. 41 (4) (2023) 433.
[26] A.J. Hung, J. Chen, S. Ghodoussipour, 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. 124 (3) (2019) 487–495.
[27] M. Sumitomo, A. Teramoto, R. Toda, 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. 27 (10) (2020) 922–928.
[28] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.
[29] J. Devaraj, R. Madurai Elavarasan, R. Pugazhendhi, et al., Forecasting of COVID-19 cases using deep learning models: is it reliable and practically significant, Results Phys. 21 (2021), 103817.
[30] N. Kokhlikyan, V. Miglani, M. Martin, E. Wang, O. Reblitz-Richardson, Captum: A Unified and Generic Model Interpretability Library for PyTorch, 2020.
[31] J.T. Wei, R.L. Dunn, M.S. Litwin, H.M. Sandler, M.G. Sanda, Development and validation of the expanded prostate cancer index composite (EPIC) for comprehensive assessment of health-related quality of life in men with prostate cancer, Urology 56 (6) (2000) 899–905.
[32] D. Shen, G. Wu, H.I. Suk, Deep learning in medical image analysis, Annu. Rev. Biomed. Eng. 19 (2017) 221–248.
[33] F. Artico, A.L. Edge Iii, K. Langham, The future of artificial intelligence for the BioTech big data landscape, Curr. Opin. Biotechnol. 76 (2022), 102714.
[34] Y. Kadono, S. Ueno, S. Kadomoto, et al., Use of preoperative factors including urodynamic evaluations and nerve-sparing status for predicting urinary continence recovery after robot-assisted radical prostatectomy: nerve-sparing technique contributes to the reduction of postprostatectomy incontinence, Neurourol. Urodyn. 35 (8) (2016) 1034–1039.
[35] M. Honda, B. Kawamoto, S. Morizane, et al., A prognostic model for predicting urinary incontinence after robot-assisted radical prostatectomy, Int J Med Robot 13 (3) (2017).
[36] Y.H. Ko, L.M. Huynh, K. See, C. Lall, D. Skarecky, T.E. Ahlering, Impact of surgically maximized versus native membranous urethral length on 30-day and long-term pad-free continence after robot-assisted radical prostatectomy, Prostate Int 8 (2) (2020) 55–61.
[37] Y.J. Lee, J.W. Jung, S. Lee, et al., Contemporary trends in radical prostatectomy and predictors of recovery of urinary continence in men aged over 70 years: comparisons between cohorts aged over 70 and less than 70 years, Asian J. Androl. 22 (3) (2020) 280–286.
[38] K. Singh, W.M. Reid, L.A. Berger, Magnetic resonance imaging of normal levator ani anatomy and function, Obstet. Gynecol. 99 (3) (2002) 433–438.
[39] T. Sadahira, Y. Mitsui, M. Araki, et al., Pelvic magnetic resonance imaging parameters predict urinary incontinence after robot-assisted radical prostatectomy, Low. Urin. Tract. Symptoms 11 (3) (2019) 122–126.
[40] A. Tienza, M. Hevia, A. Benito, J.I. Pascual, J.J. Zudaire, J.E. Robles, MRI factors to predict urinary incontinence after retropubic/laparoscopic radical prostatectomy, Int. Urol. Nephrol. 47 (8) (2015) 1343–1349.
[41] F. Gerbaud, A. Charlanes, C. Chesnel, et al., Effect of radical prostatectomy on involuntary pelvic floor muscle contraction, Neurourol. Urodyn. 38 (4) (2019) 1093–1099.
[42] S.F. Mungovan, J.S. Sandhu, O. Akin, N.A. Smart, P.L. Graham, M.I. Patel, Preoperative membranous urethral length measurement and continence recovery following radical prostatectomy: a systematic review and meta-analysis, Eur. Urol. 71 (3) (2017) 368–378.
[43] S.C. Kim, C. Song, W. Kim, et al., Factors determining functional outcomes after radical prostatectomy: robot-assisted versus retropubic, Eur. Urol. 60 (3) (2011) 413–419.
[44] S.K. Choi, S. Park, H. Ahn, Randomized clinical trial of a bladder neck plication stitch during robot-assisted radical prostatectomy, Asian J. Androl. 17 (2) (2015) 304–308.
[45] W. Song, C.K. Kim, B.K. Park, et al., Impact of preoperative and postoperative membranous urethral length measured by 3 Tesla magnetic resonance imaging on urinary continence recovery after robotic-assisted radical prostatectomy, Can Urol Assoc J 11 (3–4) (2017) E93–E99.
[46] Y. Jia, G. Luosang, Y. Li, et al., Deep learning in prediction of late major bleeding after transcatheter aortic valve replacement, Clin. Epidemiol. 14 (2022) 9–20.