ORIGINAL PAPER

The use and applicability of machine learning algorithms in predicting the surgical outcome for patients with benign prostatic enlargement. Which model to use?

Panagiotis Mourmouris [1], Lazaros Tzelves , Georgios Feretzakis [2, 3], Dimitris Kalles [2],
Ioannis Manolitsis [1], Marinos Berdempes [1], Ioannis Varkarakis [1], Andreas Skolarikos [1]

[1] 2nd Department of Urology, National and Kapodistrian University of Athens, Sismanogleio General Hospital, Athens, Greece;
[2] School of Science and Technology, Hellenic Open University, 26335 Patras, Greece;
[3] Department of Quality Control, Research and Continuing Education, Sismanogleio General Hospital, 15126 Marousi, Greece.

Summary

Objectives: Artificial intelligence (AI) is increasingly used in medicine, but data on benign prostatic enlargement (BPE) management are lacking. This study aims to test the performance of several machine learning algorithms, in predicting clinical outcomes during BPE surgical management.

Methods: Clinical data were extracted from a prospectively collected database for 153 men with BPE, treated with transurethral resection (monopolar or bipolar) or vaporization of the prostate. Due to small sample size, we applied a method for increasing our dataset, Synthetic Minority Oversampling Technique (SMOTE). The new dataset created with SMOTE has been expanded by 453 synthetic instances, in addition to the original 153. The WEKA Data Mining Software was used for constructing predictive models, while several appropriate statistical measures, like Correlation coefficient (R), Mean Absolute Error (MAE), Root Mean-Squared Error (RMSE), were calculated with several supervised regression algorithms - techniques (Linear Regression, Multilayer Perceptron, SMOreg, k-Nearest Neighbors, Bagging, M5Rules, M5P - Pruned Model Tree, and Random forest).

Results: The baseline characteristics of patients were extracted, with age, prostate volume, method of operation, baseline Q_{max} and baseline IPSS being used as independent variables. Using the Random Forest algorithm resulted in values of R, MAE, RMSE that indicate the ability of these models to better predict % Q_{max} increase. The Random Forest model also demonstrated the best results in R, MAE, RMSE for predicting % IPSS reduction.

Conclusions: Machine Learning techniques can be used for making predictions regarding clinical outcomes of surgical BPRE management. Wider-scale validation studies are necessary to strengthen our results in choosing the best model.

KEY WORDS: *Artificial intelligence; Benign prostatic enlargement; Machine learning; Transurethral resection; Transurethral vaporization.*

Submitted 8 August 2021; Accepted 22 September 2021

INTRODUCTION

The holy grail of surgery, in every surgical field, is the ability to make accurate predictions of magnitude and direction of possible outcomes, after performing each treatment-procedure for every individual patient.

The best method to accomplish the aforementioned goal, especially in urology, was the development of nomograms, which are based on conventional statistical methods (1). Such statistical methods are performed on a specific dataset with the main purpose to identify potential relationships (2). These techniques are usually applied on local datasets, but to be valid, a set of assumptions should be met, which commonly are underestimated in medical literature (3). With the increase in volume and availability of data, a novel tool has emerged and has the potential to surpass all others, setting new standards in the management of patients. This novel tool is machine learning, a major *artificial intelligence* (AI) field, which develops models based on large volumes of data in order to detect relationships or make predictions (3). The strict assumptions, which determine statistics applicability, do not pose a limit for AI and *machine learning* (ML) techniques, offering the advantage of greater flexibility and access to more healthcare-related data, which commonly do not comply with these rules (3).

In the past two decades, AI has been increasingly applied in everyday urological clinical practice and has shown promising results (1, 4). Most of available data of AI applications in urology, deal mainly with oncologic patients and associated health issues. In *benign prostatic enlargement* (BPE), AI has been recently used for predicting the severity of obstruction using diagnostic tests (5, 6). *Torshizi et al.* (6) attempted to infer symptom score and also provide a treatment suggestion for BPE, using a fuzzy-ontology system, which relies on a logic of imprecise information or variables used to make inferences (6, 7). A reported accuracy of 90%, when compared to expert opinion for making this decision, implies that AI can be helpful in benign urological conditions.

Back in 2001, *Megherbi et al.* (8), evaluated four AI algorithms regarding their predictive ability of surgical treatment success for BPE, using either transurethral resection or *visual laser ablation of the prostate* (VLAP) (8). The small number of patients, along with the vague definition of outcome, limit the applicability of these findings.

Several techniques exist for the surgical management of BPE, including transurethral vaporization using normal saline and bipolar energy (TUVis), transurethral resection using normal saline and bipolar energy (TURis) and transurethral resection using monopolar energy (TURP), with results showing similar efficacy in most trials at a short-term follow-up of 12 months, using conventional statistical analysis (9).

The aim of this study is to test and compare several machine learning algorithms, regarding their predictive ability for assessing treatment outcomes for BPE (IPSS score and Q_{max} changes), using baseline patient characteristics and one of the treatment methods (TUVis, TURis, TURP).

METHODS

Patients

Patients suffering from BPE, who were admitted at our tertiary care Urology Department between September 2017 and March 2019, were operated with one of three available methods (transurethral vaporization-TUVis, transurethral resection using bipolar energy-TURis, transurethral resection using monopolar energy-TURP), according to patient choice, physician surgical competence and equipment availability. Data were extracted retrospectively, using a prospectively collected database, and the study protocol was duly approved by the institutional review board of the hospital (19836/07.10.2020). All patients signed informed consent before being treated for their condition and were treated according to the principles of the Helsinki Declaration (10).

Patients were included in the study if they had prostate volume > 30 ml, indication for surgical management (urinary retention, failure of medical management, recurrent hematuria or urinary tract infections), absence of diagnosed prostatic adenocarcinoma and/or pathologic digital rectal examination, IPSS> 7, and Q_{max} < 15 ml/sec.

Data collected

Baseline demographic data (age, medical history, use of antiplatelets, indication for surgery, ASA score) and BPE-specific data (IPSS/Q_{max}/post-voiding residual (PVR) pre- and postoperatively, prostate volume, PSA, procedural time, haemoglobin, and sodium changes and complication rates) were collected. Functional outcomes were assessed based on follow-up visits at 12 months after surgery.

Operative technique

Surgery was performed under spinal anesthesia in all cases, using a 26 Fr continuous flow resectoscope (Olympus TURis 2.0, Iglesias type) for bipolar resection and vaporization and a 28 Fr non-rotating continuous flow resectoscope (Karl Storz) for monopolar resection. Glycine 1.5% solution was used as irrigation flow for monopolar TUR-P and N/S 0.9% for bipolar TUR-P and vaporization. During vaporization of the prostate, an electrode with a mushroom-like shape was used, and energy settings were set at 270-290 watt for vaporization and 120-140 watt for coagulation. For transurethral resection, the method of Mauermayer or Nesbit was followed (11), while for vaporization, the hovering technique was used during which the electrode comes in direct contact with the prostatic tissue.

Data analysis

Basic descriptive statistics (mean, standard deviation, range) for the numerical variables (age, prostate volume, baseline Q_{max}, baseline IPSS, % Q_{max} Increase, % IPSS reduction) have been used. Several independent variables and outcome measures have been tested, but we present only those predictors resulting in significant outcomes.

The WEKA Data Mining Software was used for this study. This comprises an open-source machine learning toolkit containing a wide range of learning algorithms (12). Since no credible validation can be made to assess the performance of the final model (13), if the total dataset is used to train a model and then reused for testing, we set aside some data which must not be used during training. The dataset set aside makes up the test set, which allows us to compare actual values of the test data to the values predicted from the WEKA-based models.

The most widely used method to take advantage of the dataset is cross-validation, where we can use all of the data in test sets, but not simultaneously. Therefore, our data were divided into a number of equal-sized subsets, called folds. If we have k folds, then this is called k-fold cross-validation. Each fold is used once for testing on the model built using the remaining k-1 folds. Cross-validation is widely regarded as a reliable way to assess the quality of results from machine learning techniques; in our analysis, we have used 10-fold cross-validation (14). While k-fold cross-validation is a standard method for making good use of available data, there are still various statistical measures which can be computed, and which reveal different interpretations/aspects.

In order to find the best regression model for numeric prediction, we consider the performance measures of Correlation coefficient (R), Mean Absolute Error (MAE), Root Mean-Squared Error (RMSE), as reported by WEKA software (13) as described in Appendix A (**Supplementary Materials**). The supervised regression algorithms - techniques that are used in this research are: Linear Regression, Multilayer Perceptron, SMOreg, k-Nearest Neighbors, Bagging, M5Rules, M5P - Pruned Model Tree, and Random forests. Although the technical details of these techniques are beyond the scope of this article, a summary of them can be found in Appendix B (**Supplementary Materials**).

Due to the small size of the initial data set, we examined the performance of aforementioned algorithms by applying a method for increasing our sample size, the Synthetic Minority Oversampling Technique (SMOTE), which is a statistical method for uniformly increasing the number of cases in a data set to render it more balanced. However, in our case, we just used SMOTE to increase our dataset by generating extra artificial instances in a statistically sound way.

The new (artificial) instances that were generated by the SMOTE are not just duplicates of existing minority instances. Instead, this method takes feature space samples for each target class and its nearest neighbours. After that, new instances are produced that combine features of the target case with those from its neighbours (15).

	Total		Per system					
			TUVis		TURis		TURP	
Variable	Range	Mean/SD	Range	Mean/SD	Range	Mean/SD	Range	Mean/SD
Age (years)	47-91	70.39/8.67	47-91	69.87/9.41	47-89	70.87/8.73	51-88	70.43/7.68
Prostate volume(ml)	20-175	59.48/24.44	31-98	59.88/20.51	20-175	63.48/29.20	20-105	54.81/21.78
Baseline Q_{max} (ml/sec)	3.40-11.90	7.24/1.75	3.40-9.30	6.52/1.53	3.40-11.90	7.83/1.90	4.50-9.90	7.38/1.51
Baseline IPSS	16-29	21.81/2.97	16-28	22.85/3.05	17-29	21/2.60	17-28	21.57/2.90
Percentage Q_{max} increase(%)	60-394	160.39/62.98	89-388	181.27/60.0	60-394	149.60/72.4	69-296	149.67/47.89
Percentage IPSS reduction(%)	29.4-76.5	59.5/7.1	29.4-75	57.3/7.6	44.0-75.0	63.2/7.1	50-76.5	58.07/4.55

Table 1.
Baseline patient characteristics.

	Total		Per system					
			TUVis		TURis		TURP	
Variable	Range	Mean/SD	Range	Mean/SD	Range	Mean/SD	Range	Mean/SD
Age (years)	47-91	70.44/8.46	47-91	70.75/8.78	47-89	71.41/8.77	51-88	69.09/7.57
Prostate volume (ml)	20-175	58.94/22.94	31-98	61.24/20.66	20-175	62.56/25.83	20-105	52.74/20.64
Baseline Q_{max} (ml/sec)	3.40-11.90	7.16/1.50	3.40-9.30	6.67/1.24	3.4-11.9	7.68/1.69	4.50-9.90	7.13/1.37
Baseline IPSS	16-29	21.90/2.81	16-28	22.79/2.83	17-29	20.76/2.49	17-28	22.06/2.71
Percentage Q_{max} increase (%)	60.00-394.00	162.10/53.12	89.00-388.00	173.62/47.96	60.00-394.00	152.78/63.82	59.00-296.00	159.81/42.71
Percentage IPSS reduction (%)	29.4-76.5	59.28/6.28	29.4-75.0	57.46/6.5	44-75	63.07/6.15	50.0-76.5	57.30/4.02
* SMOTE: Synthetic Minority Oversampling Technique.								

Table 2.
*Augmented dataset statistics after applying SMOTE *.*

RESULTS

A total of 153 patients with BPE were included (52 in TUVis group, 52 in bipolar-TURis group and 49 in monopolar TURP group). Baseline patient characteristics and % Q_{max} Increase, % IPSS reduction, are shown in Table 1. Machine learning techniques were applied in all outcomes gathered from chart review (functional outcomes-IPSS/PVR/Q_{max} change after surgery, haemoglobin drop postoperatively, sodium drop postoperatively, procedural time) using method of operation, age, prostate volume, ASA score, indication for surgery, use of antiplatelets, baseline Q_{max} and baseline IPSS as predictors, but in this study, only metrics of significant findings are reported.

In order to better depict the increase in Q_{max} and reduction in IPSS, we use percentages rather than absolute differences.

After applying the SMOTE, the new dataset contains an extra 453 synthetic instances, in addition to the original 153 patients' data.

The new allocation of the 606 instances is: 205 in TUVis group, 205 in bipolar-TURis group and 196 in monopolar TURP group. Baseline patient characteristics and % Q_{max} increase, % IPSS reduction, are shown in Table 2.

According to Table 3, considering all three metrics (R, MAE, RMSE) for % Q_{max} increase, Random Forest algorithm outperforms other models, with values of correlation coefficient (R) 0.9697, MAE 7.78 and RMSE 13.26. The values of MAE and RMSE are percentage points since the target variable % Q_{max} increase denotes the corresponding percentage increase of Q_{max} after applying the corresponding system approach on a specific patient.

As shown in Table 4, considering all three metrics (R, MAE, RMSE) for % IPSS reduction, the Random Forest model again outperforms other models, with values of

Table 3.
Percentage Q_{max} increase prediction using various machine learning methods.

| Method	R	MAE	RMSE
Linear regression	0.9004	17.8	23.12
Multilayer perceptron	0.9088	16.7	22.2
SMO reg	0.895	17.7	23.85
lazy.IBk	0.935	9.24	18.80
meta.Bagging	0.9526	11.13	16.25
M5Rules	0.9274	14.95	19.88
Trees.M5P	0.9253	14.9	20.15
trees.RandomForest	0.9697 *	7.78 *	13.26 *

R: Correlation coefficient; MAE: Mean Absolute Error; RMSE: Root Mean-Squared Error.
*Best results are marked by * in each column.*

Table 4.
Percentage IPSS reduction prediction using various machine learning methodslearning methods.

Method	R	MAE	RMSE
Linear regression	0.4493	4.14	5.61
Multi layer perceptron	0.5751	3.95	5.36
SMOreg	0.4199	4.17	5.7
lazy.IBk	0.8793	1.53*	3.07
meta.Bagging	0.7906	2.73	3.91
M5Rules	0.678	3.35	4.62
trees.M5P	0.7231	3.17	4.36
trees.RandomForest	0.8989 *	1.63	2.80 *

R: Correlation coefficient; MAE: Mean Absolute Error; RMSE: Root Mean-Squared Error.
*Best results are marked by * in each column.*

Figure 1.
Development of a random forest from decision trees.
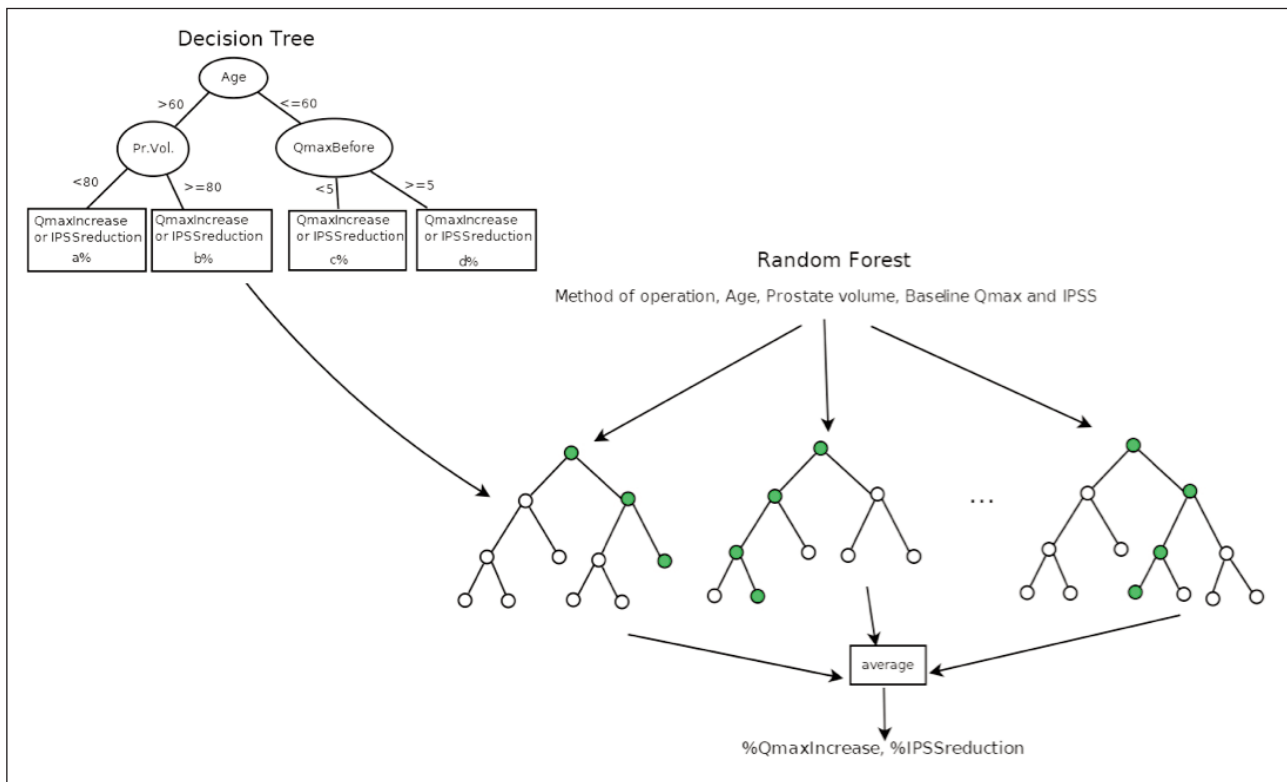
correlation coefficient 0.8989, MAE 1.63, and RMSE 2.80, with the only exception that k-Nearest Neighbors model has smaller but very close value of MAE. The values of MAE and RMSE are percentage points since the target variable % IPSS reduction denotes the corresponding percentage decrease of IPSS after applying the corresponding system approach.

A Decision Tree algorithm is easily understood and ideal for obtaining non-linear relationships between independent and dependent variables. Random forest is a collection of decision trees constructed in a specific random manner. Random Forest usually performs better than a single Decision Tree in terms of accuracy and reduced overfitting. The major advantages of Random Forests are that they can handle both linear and non-linear relationships as well, they are not significantly impacted by outliers and they effectively balance the bias-variance trade-off. Figure 1 shows an example of how a Random Forest is constructed from Decision Trees.

Correlation coefficient (R) is used to measure the strength of a linear relationship between two variables, in our case the predicted and the actual values of the target variables % Q_{max} increase and % IPSS reduction. The closer the value of the correlation coefficient is to 1, the better the regression model is.

Mean Absolute Error (MAE) is the average error between the absolute value of the predicted and actual value for each pair. *Root Mean-Squared Error* (RMSE) It shows how far predicted values fall from measured actual values using Euclidean distance.

Concerning the values of the MAE and RMSE, the closer their values to zero, the better the model's performance, since both metrics are proportional to the difference between the actual and predicted values.

Readers can find on the website (16) two WEKA data set sample files (.arff) for experimental purposes to create their own models based on their local facility data. Furthermore, we have uploaded the two experimental models for % Q_{max} increase and % IPSS reduction prediction with considered independent variables the method of operation, age, prostate volume, baseline Q_{max} and IPSS.

DISCUSSION

The ultimate goal of AI is to create systems which are able to perform intellectually challenging tasks, similar to those performed by humans. Today, the closest we get to such systems, is usually aided by non-linear mathematic and statistical models (17) and mostly drawn from the machine-learning sub-field of AI, though significant developments also occur in other sub-fields too, such as natural language processing and visual perception with deep learning (7).

A substantial number of such models attempt to assist medical practitioners, using a variety of sources for data and feedback, such as handwritten notes and books, medical imaging scanning and tissues grading. So, it is the impact on everyday clinical practice that will likely guide the training of these models and also decide the success of these AI technologies. In our study, we tested several machine learning algorithms, in order to find the

one with the least error in predicting IPSS reduction and Q_{max} increase, taking into consideration patient parameters that are widely available and easily assessed during daily urological practice in a usual clinical setting.

Physicians could use these algorithms preoperatively and in conjunction with clinical judgement and discussion with patients, decide whether to perform surgery or not. There are, so far, some, but sparse, data about the implementation of this technological advance in urology, with the majority of existing studies in urological literature, focusing on the effect of these systems in improving prediction accuracy in prostate cancer diagnosis and management.

There is still an unmet need for better prostate cancer detection in order to avoid unnecessary biopsies. A recent paper investigated different *prostate-specific antigen* (PSA) assays and developed a novel predictive tool based on *artificial neural networks* (ANN), concluding that AI technology can aid in minimizing variability of each PSA assay but only if a separate ANN system is utilized for every PSA assay and not one for all (18).

As for the mpMRI diagnostic optimization, alongside their fusion biopsy implications, there is an increasing body of literature that reports on system development to integrate pre-processing, segmentation, and registration in order to fully automate the procedure, with promising outcomes so far (19-21).

Besides cancer-related research, AI systems have also been utilized in other aspects of urological pathology. In urinary stone disease, there are reports that AI systems have been implemented in order to predict stone composition (22), surgical outcomes of *percutaneous nephrolithotomy* (PNL) (23), and *shock wave lithotripsy* (SWL) (24), with excellent accuracy. Similarly in patients with vesicoureteral reflux, as reported by *Seckiner et al.* (25), the ANN reported 98.5% sensitivity, 92.5% specificity, 97% positive predictive value, and 96% negative predictive value, which can definitely be considered very promising. Contradictive results were published for the role of AI systems in predicting surgical outcomes, mainly in robotic surgery (4, 26).

The necessity to personalize treatment in patients with cancer and the high heterogeneity of neoplastic diseases is a potential reason that led scientists to focus mainly on this field of medicine and less on benign conditions like BPE.

Notably, the implementation of AI techniques in BPE diagnosis and, especially, treatment is at its early stages, with currently scarce reports about the utilization of AI systems in BPE patients. *Torshizi et al.* presented a hybrid fuzzy- ontology intelligent system with multiple layers that consisted of two modules: the first was evaluating symptoms severity, whereas the second was evaluating the management options. Nevertheless, this system did not evaluate the outcomes of different surgical entities according to individual patient characteristics (6). Furthermore, the evaluation of bladder outlet obstruction symptoms has been the topic of another relatively recent study, where the detection rate of BPE in these patients using an ANN was 72%, and where the authors concluded that the pressure-flow study could not be omitted and replaced with the intelligent system.

The management of BPE depends on disease stage, symptom intensity, patient preference and health status. Common indications for surgical management include failure of medical treatment for moderate- severe *lower urinary tract symptoms* (LUTS), recurrent urinary retention or infections, hematuria, bladder stones, kidney damage. A common perception is that prostate volume correlates with symptom severity and with health-related quality of life, but this is not backed up by the relevant literature (27). A clinical dilemma occurs in patients who do not fulfill criteria and absolute indications for surgery, while both physicians and patients need to know an estimation of functional outcomes post-operatively.

Diagnostic tests are not highly specific for attributing LUTS to BPE, except for urodynamic testing, which is an invasive, costly and time-consuming examination. Choo et al developed a nomogram, which permits prediction of benign outlet obstruction-related surgery, with satisfactory metrics (28) based on clinical and urodynamic parameters. Since urodynamics is not available at every clinical setting, these nomograms may not be applicable for a substantial percentage of patients.

According to *Pielke* (1984), a model can be considered predictive if two conditions are satisfied: (a) the standard deviations of the predictions and observations are approximately the same, and (b) RMSE is less than the standard deviation of the observations (29). Our results indicate that the Root Mean Squared Error (13.26) for the model % Q_{max} increase Random Forest is much smaller than the value of the standard deviation (53.12) of the actual values of the dependent variable % Q_{max} increase. Furthermore, the standard deviation of the predicted values is 57.53 percentage points (p.p.), which is close to the corresponding standard deviation of the actual values (62.79 p.p.).

The results for the second model (% IPSS reduction Random Forest), indicate that the *Root Mean Squared Error* (2.80) for the best model is also much smaller than the value of the standard deviation (6.28) of the actual values of the dependent variable % IPSS reduction, and the standard deviation of the predicted values is 5.13 percentage points (p.p.) is very close to the corresponding standard deviation of the actual values (5.31 p.p.). Therefore, our proposed model meets the two conditions to be considered predictive, both regarding % Q_{max} increase and % IPSS reduction.

Personalized medicine is touted as the future in healthcare settings, especially after the development of large-scale databases with patient –omic characteristics (proteomics, genomics, metabolomics etc). Predictive analytics on data of such volume and complexity seems to be feasible using AI techniques with the ability to adapt and ''*learn*'' from data during the whole process, giving endless opportunities both for patient outcomes improvement and cost savings for healthcare systems (30).

A limitation of our study is that, due to the limited sample size, our models may not be immediately applicable to all urology departments. For that reason, it will be preferable that our methodology is implemented in the data of each local facility, or ideally, on a larger pool of data collected from multiple sites, so as to have a greater potential for learning and test whether the mean absolute

error can be reduced. Another drawback of this study is that laser methods for prostate resection were not studied due to the lack of appropriate equipment during the period of data collection.

Moreover, using more clinical-related data in the future, such asomic data, could pave the way for producing better predictive models. The retrospective collection of data is also a limitation, but since this was performed through a prospectively collected database, confounding is partially alleviated.

CONCLUSIONS

BPE is a very common clinical condition, with various treatment modalities available for patients. At the same time, AI models increasingly provide surgeons with accurate decision-making tools.

As *health information system* (HIS) use is expanded in a healthcare facility, it will be easier to utilize data collected for the HIS using artificial intelligence techniques to benefit patients.

This study presents a methodology for predicting clinical outcomes in BPE management, according to pre-operative characteristics and a variety of relatively standard and widely available AI techniques. Results are promising to regard IPSS and Q_{max} improvement, but more validation studies are needed before a wider scale application of these findings.

REFERENCES

1. Drouin SJ, Yates DR, Hupertan V, et al. A systematic review of the tools available for predicting survival and managing patients with urothelial carcinomas of the bladder and of the upper tract in a curative setting. World J Urol. 2013; 31:109-16.

2. Azzolina D, Baldi I, Barbati G, et al. Machine learning in clinical and epidemiological research: Isn't it time for biostatisticians to work on it? Epidemiol. Biostat. Public Heal. 2019; 16:e13245-1-3.

3. Rajula HSRV G, Manchia M, Antonucci N, Fanos V. Comparison of conventional statistical methods with machine learning in medicine: diagnosis, drug development, and treatment. Medicina 2020; 56:455.

4. Hung AJ, Chen J, Gill IS. Automated performance metrics and machine learning algorithms to measure surgeon performance and anticipate clinical outcomes in robotic surgery. JAMA Surg. 2018; 153:770-1.

5. Sonke GS, Heskes T, Verbeek AL, et al. Prediction of bladder outlet obstruction in men with lower urinary tract symptoms using artificial neural networks. J Urol. 2000; 163:300-5.

6. Torshizi AD, Zarandi MH, Torshizi GD, Eghbali K. A hybrid fuzzy-ontology based intelligent system to determine level of severity and treatment recommendation for Benign Prostatic Hyperplasia. Comput Methods Programs Biomed. 2014; 113:301-13.

7. Zadeh LA. Fuzzy sets. Information and Control. 1965; 8:338-53.

8. Megherbi D, Boulenouar A, Kaula N, et al. Comparison of artificial intelligence and machine learning algorithms as a predictor of surgical outcomes in benign prostatic hyperlasia cases (BPH). Proc. SPIE 4389, Component and Systems Diagnostics, Prognosis, and Health Management, (20 July 2001).

9. Gravas S CJ, Gacci M, Gratzke C, et al. Management of nonneurogenic male LUTS. EAU guidelines. ISBN 978-94-92671-07-3. 2020.

10. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. Jama. 2013; 310:2191-4.

11. Mauermayer W. Transurethral surgery: Springer-Verlag Berlin Heidelberg; 1983.

12. Hall M, Frank E, Holmes G, et al. The WEKA data mining software: An update. SIGKDD Explor Newsl. 2008; 11:10-8.

13. Ian H. Witten EF, Mark A. Hall, Christopher J. Pal. Data Mining, Fourth Edition: Practical Machine Learning Tools and Techniques (4th. ed.): Morgan Kaufmann Publishers Inc., San Francisco, CA, USA; 2016.

14. Bengio Y, Grandvalet Y. No unbiased estimator of the variance of K-fold cross-validation.J Mach Learn Res. 2004; 5:1089-1105.

15. Microsoft Azure, SMOTE [Internet]. Cited at 03 July 2021. Available from: https://docs.microsoft.com/en-us/azure/machine-learning/algorithm-module-reference/smote

16. The use and applicability of machine learning algorithms in predicting the surgical outcome for patients with benign prostatic enlargement. Which model to use? Available online: http://www.learningalgorithm.eu/datafiles_Urol.html (accessed on 4July 2021).

17. Chen J, Remulla D, Nguyen JH, et al. Current status of artificial intelligence applications in urology and their potential to influence clinical practice. BJU Int. 2019; doi: 10.1111/bju.14852.

18. Stephan C, Cammann H, Meyer HA, et al. An artificial neural network for five different assay systems of prostate-specific antigen in prostate cancer diagnostics. BJU International. 2008; 102:799-805.

19. Song Y, Zhang YD, Yan X, et al. Computer-aided diagnosis of prostate cancer using a deep convolutional neural network from multiparametric MRI. J Magn Reson Imaging. 2018; 48:1570-7.

20. Wildeboer RR, van Sloun RJG, Wijkstra H, Mischi M. Artificial intelligence in multiparametric prostate cancer imaging with focus on deep-learning methods. Computer methods and programs in biomedicine. 2020; 189:105316.

21. van Sloun RJG, Wildeboer RR, Mannaerts CK, et al. Deep learning for real-time, automatic, and scanner-adapted prostate (zone) segmentation of transrectal ultrasound, for example, magnetic resonance imaging-transrectal ultrasound fusion prostate biopsy. Eur Urol Focus. 2021; 7:78-85.

22. Kriegshauser JS, Paden RG, He M, et al. Rapid kV-switching single-source dual-energy CT ex vivo renal calculi characterization using a multiparametric approach: refining parameters on an expanded dataset. Abdom Radiol (NY). 2018; 43:1439-45.

23. Aminsharifi A, Irani D, Pooyesh S, et al. Artificial neural network system to predict the postoperative outcome of percutaneous nephrolithotomy. J Endourol. 2017; 31:461-7.

24. Seckiner I, Seckiner S, Sen H, et al. A neural network - based algorithm for predicting stone - free status after ESWL therapy. Int Braz J Urol. 2017; 43:1110-4.

25. Seckiner I, Seckiner SU, Erturhan S, et al. The use of artificial neural networks in decision support in vesicoureteral reflux treatment. Urol Int. 2008; 80:283-6.

26. Hung AJ, Chen J, Che Z, et al. Utilizing machine learning and

P. Mourmouris, L. Tzelves, G. Feretzakis, D. Kalles, I. Manolitsis, M. Berdempes, I. Varkarakis, A. Skolarikos

automated performance metrics to evaluate robot-assisted radical prostatectomy performance and predict outcomes. J Endourol. 2018; 32:438-44.

27. Lepor H. Evaluating men with benign prostatic hyperplasia. Reviews in urology. 2004; 6 Suppl 1(Suppl 1):S8-s15.

28. Choo MS, Yoo C, Cho SY, et al. Development of decision support formulas for the prediction of bladder outlet obstruction and prostatic surgery in patients with lower urinary tract symptom/benign pro-
static hyperplasia: part I, development of the formula and its internal validation. 2017; 21(Suppl 1):S55-65.

29. Pielke RA. Mesoscale Meteorological Modelling. Academic Press, Orlando, 612pp.

30. Stanfill MH, Marc DT. Health information management: implications of artificial intelligence on healthcare data and information management. Yearb Med Inform. 2019; 28:56-64.

Correspondence
Mourmouris Panagiotis, MD
thodoros13@yahoo.com
Manolitsis Ioannis, MD
giannismanolit@gmail.com
Berdempes Marinos, MD
marinosberdebes@hotmail.com
Varkarakis Ioannis, MD
medvark3@yahoo.com
Skolarikos Andreas, MD
andskol@yahoo.com
2nd Department of Urology, National and Kapodistrian University of Athens, Sismanogleio General Hospital, Athens (Greece)

Tzelves Lazaros, MD (Corresponding Author)
lazarostzelves@gmail.com
2nd Department of Urology, National and Kapodistrian University of Athens, Sismanogleio General Hospital, Athens
Sismanogleiou 1, 15126, Marousi (Greece)

Feretzakis Georgios, MD
georgios.feretzakis@ac.eap.gr
Kalles Dimitris, MD
Greece kalles@eap.gr
School of Science and Technology, Hellenic Open University, 26335 Patras