

REVIEW ARTICLES PREGLEDNI ČLANCI

ARTIFICIAL INTELLIGENCE IN URODYNAMIC STUDIES

UPOTREBA VEŠTAČKE INTELIGENCIJE U URODINAMIČKIM ISPITIVANJIMA

Filip DOŽIĆ^{1,2}, Žarko DIMITRIĆ¹, Đorđe FILIPOVIĆ^{1,2}, Dimitrije JEREMIĆ^{1,2}, Jovan RADOJEVIĆ^{1,2} and Stevan STOJANOVIĆ^{1,2}

ORCID NUMBER

Filip Dožić – 0009-0001-9660-6909

Žarko Dimitrić – 0009-0001-8053-3389

Đorđe Filipović – 0009-0008-2532-0699

Dimitrije Jeremić – 0009-0002-8732-3170

Jovan Radojević – 0009-0009-3826-0813

Stevan Stojanović – 0000-0002-9067-7933

University Clinical Center of Vojvodina, Urology Clinic, Novi Sad¹

University of Novi Sad, Faculty of Medicine Novi Sad²

Review article

Pregledni članak

UDK 616.6-07:004.8

<https://doi.org/10.2298/MPNS2606129D>

Abstract

Introduction. Urodynamic studies are a fundamental diagnostic tool in functional urology, crucial for assessing urinary tract function. Over the past decade, artificial intelligence has been progressively applied across several scientific disciplines, with healthcare representing a significant area of growth. This review aimed to examine the published literature on the use of machine learning in urodynamics over the preceding five years. **Material and Methods.** A comprehensive literature review was conducted using *Google Scholar* and *PubMed* to identify studies published between 2020 and 2025 on AI applications aimed at reducing the invasiveness of urodynamic testing, enhancing diagnostics, and predicting treatment outcomes. **Results.** Recent studies demonstrate that artificial intelligence and machine learning significantly enhance diagnostic precision and predictive capability in urodynamics. Algorithms including convolutional neural networks, support vector machines, logistic regression, and decision trees accurately classify detrusor overactivity, bladder outlet obstruction, and other urodynamic abnormalities. **Discussion.** Deep learning models excel in analysing complex high-dimensional data, achieving reporting accuracies of up to 100%, whilst conventional machine learning methods provide greater interpretability. AI systems using uroflowmetry data reached diagnostic accuracies of up to 84%, surpassing expert performance. **Conclusion.** Overall, artificial intelligence and machine learning improve the reliability and efficiency of urodynamic assessments, although larger clinical studies are required for validation.

Key words: Urodynamics; Artificial Intelligence; Decision Support Systems, Clinical; Machine Learning

Sažetak

Uvod. Urodinamičke studije predstavljaju važan dijagnostički alat u funkcionalnoj urologiji i od ključnog su značaja za procenu funkcije urinarnog trakta. Tokom poslednje decenije, primetna je sve učestalija upotreba veštačke inteligencije u različitim naučnim disciplinama, uključujući i zdravstvenu zaštitu. Cilj ovog rada bio je da se sprovede pregled literature o primeni mašinskog učenja u urodinamičkim ispitivanjima u poslednjih pet godina. **Materijal i metode.** Sproveden je pregled literature o primeni veštačke inteligencije u urodinamičkim ispitivanjima dostupni u bazi *Google Scholar* i *PubMed* u periodu od 2020. do 2025. godine. **Rezultati.** Sprovedene studije pokazuju da veštačka inteligencija i mašinsko učenje značajno unapređuju dijagnostičku preciznost i sposobnost predviđanja u urodinamičkim ispitivanjima. Algoritmi poput konvolucionih neuronskih mreža, mašina potpornih vektora, logističke regresije i stabala odlučivanja mogu tačno da klasifikuju hiperaktivnost detrusora, subvezikalnu opstrukciju i druge poremećaje. **Diskusija.** Modeli dubokog mašinskog učenja ističu se u analizi složenih, visokodimenzionalnih podataka, postižući tačnost i do 100%, dok konvencionalne metode mašinskog učenja nude veću interpretabilnost. AI sistemi zasnovani na uroflowmetrijskim podacima dostigli su dijagnostičku tačnost do 84%, nadmašujući učinak stručnjaka. **Zaključak.** Veštačka inteligencija i mašinsko učenje poboljšavaju pouzdanost i efikasnost urodinamičkih procena, iako su potrebne veće kliničke studije radi validacije. **Ključne reči:** urodinamika; veštačka inteligencija; klinički sistemi za podršku odlučivanju; mašinsko učenje

Introduction

Urodynamic studies (UDS) are a fundamental diagnostic tool in functional urology, essential for assessing urinary tract function and facilitating the diagnosis of complex urological disorders. The International Continence Society standard urodynamic

evaluation encompasses procedures including free uroflowmetry, postvoid residual assessment, cystometry, and pressure-flow studies. Interpretation of UDS presents distinct challenges owing to the complexity of the data, requiring considerable skill and experience to achieve reliable results. The inherent subjectivity of UDS interpretation, together with heteroge-

✉ Corresponding author: Filip Dožić, E-mail: filipdozic@uns.ac.rs

Abbreviations

UDS	– urodynamic studies
AI	– artificial intelligence
ML	– machine learning
BPNN	– backpropagation neural network
CNN	– convolutional neural network
LR	– logistic regression
DL	– deep learning
SVM	– support vector machine
LUTS	– Lower Urinary Tract Symptoms
BCI	– bladder contractility index
BOOI	– bladder outlet obstruction index
LLM	– Large Language Models

neity in clinical practice, highlights the need for tools that improve standardisation and precision [1].

Over the past decade, artificial intelligence (AI) has been progressively applied across several scientific disciplines, with healthcare being a significant area of growth. Emerging technologies focus on machine learning (ML) algorithms, which are becoming increasingly important in the prediction, prevention, diagnosis, and treatment decision-making for a range of diseases. The escalating integration of electronic, digital, and automated technologies within healthcare also broadens the scope of potential ML applications. A substantial body of research exists on the application of machine learning across medical disciplines, including urology, cardiovascular medicine, and geriatrics, reflecting widespread and growing interest within the medical community [2]. Despite considerable emphasis on modern and non-invasive approaches to reduce patient discomfort during UDS, interpretation remains hampered by a substantial degree of inter-observer variability, which significantly undermines diagnostic consistency. ML and AI are being evaluated as means of improving the diagnostic precision of UDS and addressing these inconsistencies. Such developments may also enhance the performance of emerging non-invasive technologies, which typically yield less reliable data [3].

The potential of ML for urodynamic data analysis has been widely discussed. Current evidence suggests

that ML is well-suited for analysing urodynamic data; however, current results to date have not yet demonstrated clinical utility [4].

Material and Methods

To perform this narrative review, we conducted a comprehensive literature search using *PubMed* and *Google Scholar* to identify relevant studies published between January 2020 and October 2025 (**Table 1**). The search combined keywords relating to AI and urodynamics. AI-related search terms comprised: artificial intelligence, machine learning, deep learning, large language model, and ChatGPT. Urodynamics-related terms included: uroflowmetry, urodynamics, urology, lower urinary tract symptoms, overactive bladder, detrusor overactivity, detrusor underactivity, and bladder outlet obstruction.

Inclusion criteria were as follows: studies that utilised AI models to assess, predict, or diagnose lower urinary tract symptoms; studies reporting quantitative evaluations of model performance; and review articles, meta-analyses, or editorials. Exclusion criteria comprised: articles written in languages other than English, publications without available full texts, and articles not related to UDS. Following application of predefined criteria and removal of duplications, a total of 17 studies were included in the final analysis.

Results

Studies published over the past five years highlight the capacity of machine learning and artificial intelligence to improve the diagnostic precision and predictive capability of urodynamics.

Artificial intelligence algorithms may be broadly categorised into two types: conventional ML (excluding deep learning) and deep learning (DL). Conventional ML algorithms are highly interpretable, which makes them particularly useful for diagnostic and prognostic tasks in urinary dysfunctions where evi-

Table 1. The search strategy summary

Items	Specification
Date of search	January 2020 to October 2025
Databases and other sources searched	<i>PubMed</i> and <i>Google Scholar</i>
Search terms used	Combination of terms: Artificial intelligence OR machine learning OR deep learning OR large language model OR ChatGPT AND uroflowmetry OR urodynamics OR urology OR lower urinary tract symptoms OR overactive bladder OR detrusor overactivity OR detrusor underactivity OR bladder outlet obstruction
Exclusion criteria	Articles written in non-English languages; publications without available full texts; articles that were not related to UDS
Number of studies included	17

Table 2. Differences between Conventional ML and Deep learning models

	Interpretability	Sample size	Clinical utility
Conventional ML	highly interpretable	small datasets	useful for simple diagnostic and prognostic tasks
Deep learning model	limited interpretability	extensive datasets	analysing high-dimensional data

dence-based conclusions are essential. Deep learning is a subset of ML that employs neural networks. In contrast to conventional ML, DL algorithms are adept at extracting complex image features and processing large datasets, making them well-suited to identifying complex curve patterns within UDS. The principal differences between conventional ML and deep learning are summarised in **Table 2**.

However, DL models require large datasets for effective training. Conventional ML models, by contrast, have a simpler internal architecture with greater interpretability, and are frequently supported by visualisation tools that facilitate detailed evaluation. A limitation of conventional ML, however, is its reduced capacity to identify key features within large, high-dimensional, and unstructured clinical data, such as urodynamic tracing curves. Deep learning algorithms, conversely, are capable of analysing high-dimensional data, autonomously extracting features via deep neural networks, and identifying relationships between imaging characteristics and research objectives. Deep learning models are, however, more complex and frequently operate as a “black box”, limiting the real-time interpretability and transparency of their decision-making processes [1].

The studies analysed applied both conventional and DL models to urodynamic data, including: back-propagation neural network (BPNN) [5]; convolutional neural network (CNN) [6]; discrete wavelet transformation; exponential moving average [7]; and logistic regression (LR) [8].

Several studies examined different conventional machine learning algorithms for the development of diagnostic models assessing bladder compliance, including support vector machine (SVM), Random Forest, XGBoost, perceptron, and Naive Bayes [9].

Some studies on ML algorithms have demonstrated precision in identifying detrusor overactivity, detrusor underactivity, and other urodynamic abnormalities from tracings. A further body of literature indicates that ML algorithms integrating urodynamic parameters can effectively predict disease severity or outcomes in functional urologic conditions, such as overactive bladder and neurogenic lower urinary tract dysfunction [10].

In the study by Weaver et al., a DL model combined volume-pressure recordings with fluoroscopic images to enable automatic, precise classification of

bladder dysfunction severity in a population with spina bifida. Four models were evaluated for predicting the severity of bladder dysfunction: a clinical model using prospectively collected clinical data from videourodynamic studies; a deep learning CNN trained on raw volume-pressure recordings; a deep learning imaging model trained on fluoroscopic images; and an ensemble model averaging the risk probabilities from the volume-pressure and fluoroscopic models [11].

Hobbs et al. developed an ML approach for identifying detrusor overactivity in urodynamic tests. Files were divided into training and test sets, and each file was windowed to generate data subsets. Multiple classification approaches were evaluated, including k-nearest neighbors and SVM with linear, polynomial, and radial basis function kernels, achieving an area under the curve of 91.9%, with a sensitivity of 84.2% and specificity of 86.4% [12].

Zhou et al. developed deep learning models for diagnosing detrusor overactivity. A single-channel CNN and a 3-channel CNN were constructed. The single-channel CNN was trained using detrusor pressure (p Det) alone as the input signal, whilst the 3-channel CNN was trained simultaneously on vesical, abdominal, and detrusor pressure signals. Both CNNs comprised two convolutional CNN layer groups and a traditional BPNN; each layer group contained several convolutional kernels, convolutional layers, and pooling layers. The traditional BPNN served as the classifier, with Softmax activation function in the output layer enabling classification of different signal types, achieving an accuracy of 78.12% for patients without detrusor overactivity and 100% for those with it [3,13].

In the study by Ding et al. ML algorithms were applied to a preliminary analysis of urodynamic studies using 13 urodynamic parameters as inputs. Three ML approaches were evaluated: Decision Tree, LR and SVM. The LR and the SVM models demonstrated the best performance; the top-performing model achieved an average area under the curve of 0.90 (0.90 ± 0.08) [14].

Matsukawa et al. developed a DL model using a neural network for diagnosing lower urinary tract function in men with lower urinary tract symptoms (LUTS), utilising uroflowmetry data alone. A neural network was trained on uroflowmetry waveforms and subsequently used to diagnose LUTS in new patients

based on their waveforms. The architecture comprised alternating convolutional and maximum pooling layers, followed by a global average pooling layer to summarise the waveform's features, and two fully connected layers to convert these features into bladder contractility index (BCI) and bladder outlet obstruction index (BOOI) values. The AI system's estimated BCI and BOOI values showed significant positive correlations with pressure flow study values (BCI: $r = 0.60$, $P < 0.001$; BOOI: $r = 0.46$, $P < 0.001$). The mean diagnostic accuracy of the AI system, based on uroflowmetry data, was 84%, significantly higher than that of urologists (56%) [15].

Jin et al. constructed a long short-term memory deep-learning algorithm for uroflowmetry data. Gammie et al. (2023) identified some key advantages of AI in UDS, including effective performance in noisy surroundings, pattern recognition in unstructured set-

tings, extraction of clinically relevant variables, and potential improvement in objectivity among clinicians. UDS interpretation is highly susceptible to variability depending on the skills of the interpreter. Jin et al. examined sound-based prediction of urinary flow, bladder outlet obstruction, and LUTS, demonstrating that certain algorithms can enhance the interpretation of variations in urine sounds to an accuracy comparable to that of clinicians [16]. Results from all included studies are summarised in **Table 3**.

Large Language Models (LLMs) are a form of AI that utilise a deep neural network architecture and apply concepts from natural language processing. These models employ unsupervised learning to analyse large volumes of text data, learning language patterns, grammatical rules, and contextual nuances. Deep neural networks, consisting of numerous layers of interconnected nodes, analyse sequential data -

Table 3. Included studies overview

Authors	Year	Algorithm type	Import data	Aim	Accuracy
Zhou et al. [5]	2024	BPNN	11 urinary flow indicators, including P.Qbegin, P.Qmax, Qmax	Diagnose BOO	AUC: 0.949 ± 0.060 ; ACC: 94.4%; SEN: 100%; SPE: 89.3%
Pong et al. [6]	2022	CNN	RMS values, MFCC features (root mean square; mel-frequency cepstral coefficients)	Develop a vibration-based diagnostic tool for voiding dysfunction	ACC: 98%; precision: 98.26%; recall: 98.19%; F1-score: 98.19%
Zareen et al. [7]	2022	DWT + EMA filtering	Pves signal	Estimate detrusor pressure	Correlation: 0.895
Cetinel et al. [8]	2022	Multivariate LR	Symptoms, uroflow curve patterns, detrusor pressure	Distinguish BOO and DU	SEN: 82.5%, SPE: 60% for BOO
Ge Z et al. [9]	2022	(SVM), Random Forest, XGBoost, perceptron, logistic regression, and Naive Bayes	73 features (time-domain, frequency-domain, wavelet features)	bladder compliance	AUC: 0.873; ACC: 84%; SEN: 70.41%; SPE: 84.53%
Weaver et al. [11]	2023	CNNs for imaging + pressure-volume data	Volume-pressure measurements, fluoroscopic imaging	Classify bladder dysfunction severity	ACC: 70%, weighted kappa: 0.54
Hobbs et al. [12]	2022	SVM with RBF kernel	Time-based and frequency-domain features from Pves, Pabd, and Pdet	Detect DO	AUC: 0.919 ± 0.013 , SEN: 84.2%, SPE: 86.4%
Zhou et al. [13]	2023	Single-channel CNN, 3-channel CNN	Pves, Pabd, Pdet signals with wavelet-denoising preprocessing	Diagnose DO	ACC: 99.99% (train), validation: 99.72% (3-channel CNN)
Ding et al. [14]	2024	DT, LR, SVM	15 urodynamic parameters and demographics	Automate diagnostic system for LUTD	AUC: DT (0.63–0.98), LR (0.73–0.99), SVM (0.64–1.00)
Matsukawa et al. [15]	2021	NN (1D convolutional)	UFM waveform features	Classify LUTD (DU, BOO, DU + BOO)	ACC: 84%; SEN: 79.7%; SPE: 88.7%
Jin et al. [16]	2021	LSTM	Loudness and roughness from uroflow sound signals	Predict and classify bladder emptying patterns using sound signals	ACC: 95%; mean flowrate error: 6%; SEN: high

such as sentences - by predicting the next word on the basis of preceding words, thereby developing a comprehensive understanding of linguistic structure. Reliance on such models without appropriate oversight risks patients bypassing professional consultation, potentially leading to incorrect diagnosis or inappropriate treatment [17,18]. Based on current literature, no original research articles have evaluated LLMs in the diagnosis of urodynamic findings. A few studies examining the role of LLMs in urodynamics quality control have excluded urodynamics trace interpretation, owing to the inability of the ChatGPT version used to analyse images directly [4]. Large language models show potential for quality control and data interpretation, but their diagnostic application in urodynamics remains unexplored.

Discussion

These studies demonstrate that AI and ML models can enhance diagnostic precision and reliability. As these technologies advance, it is likely that AI and ML will be integrated into urodynamics software to deliver automated, real-time data analysis and interpretation. Such integration could streamline clinical processes, reduce diagnostic discrepancies, and improve the accuracy and applicability of urodynamic investigations in routine clinical practice. Nevertheless, further clinical trials are required to improve accuracy and validate these models prior to widespread adoption [3].

Artificial intelligence in urodynamics is advancing rapidly, yet a number of obstacles remain, and visual analysis of the urodynamic trace by a clinician remains the cornerstone of urodynamic diagnosis. **Figure 1** illustrates the potential incorporation of machine learning algorithms into urodynamic studies. Further research into the generalisability of models is needed, as the majority of studies in this field are retrospective, single-centre analyses conducted without external validation.

Furthermore, large volumes of labelled data are required for model training in both conventional ML and DL approaches. Existing studies lack sufficient sample sizes to exclude the possibility of model overfitting and limited generalisability, which severely restricts their clinical applicability. For example, a study that achieved nearly 100% accuracy in detecting detrusor overactivity (DO) enrolled only 92 patients, of whom 44 comprised the test - an inadequately small dataset for a deep learning method. Existing studies have also focused predominantly on a limited number of urodynamic trace parameters, omitting other urodynamic indicators that would be essential for clinical application.

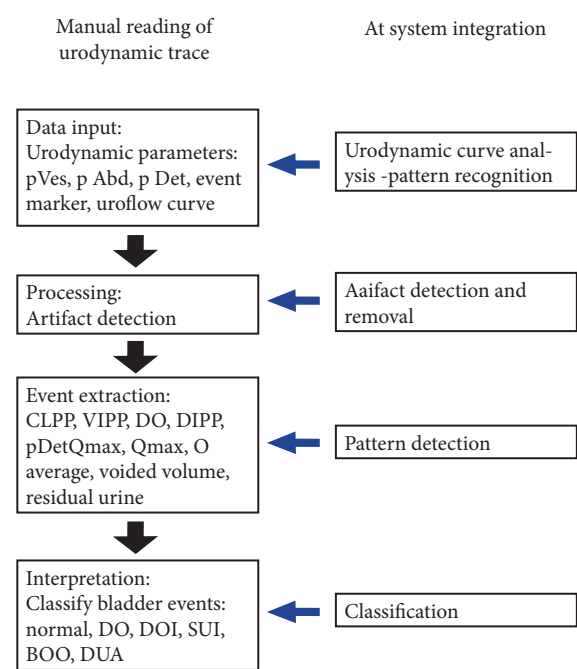


Figure 1. Potential introduction of machine learning algorithms to urodynamic studies

A further limitation is that the majority of studies apply highly specific inclusion and exclusion criteria and recruit patient cohorts that are markedly homogeneous. This highlights the critical importance of external validation studies to determine whether algorithmic models can consistently generalise to different patient populations.

Oversimplification of predictive models represents another drawback. Many models are designed only for binary classification tasks, such as determining whether detrusor overactivity is present or absent. In clinical practice, however, LUTS is frequently multifaceted: a patient may present with detrusor overactivity, detrusor underactivity and bladder outlet obstruction concurrently. Reducing a complex clinical picture to a binary outcome does not adequately reflect the reality of LUTS management.

A major obstacle to AI development in urodynamics is the absence of a consistent, high-quality, multicentre database. Existing urodynamic databases are predominantly single-centre, likely reflecting the lack of agreed diagnostic criteria for UDS and the inter-rater reliability of interpretations. This absence of uniformity further hampers urodynamic AI research. There is an immediate need to establish standardised diagnostic criteria and nomenclature for UDS to advance AI development in this field. Both clinical practice and the integration of AI into UDS would benefit from the adoption of consensus-driven standards. Developing a reliable multicentre urodynamic database is another important step; such a resource, en-

compassing diverse and representative datasets, would facilitate model training and validation. The ability to perform well across a heterogeneous patient population is one of the most important prerequisites for clinical adoption of AI models.

Finally, the risks associated with the misuse of AI in urodynamics must be carefully considered. Whilst AI holds considerable promise, inadequately validated models may lead to diagnostic errors, inappropriate treatment recommendations, or excessive reliance on automated systems. Rigorous validation and verification procedures are necessary to ensure the reliability and safety of AI tools before deployment in clinical settings. Accessing accurate medical information is essential; even a low rate of misinformation can have

serious consequences for patients. Given these limitations, close collaboration between AI developers and healthcare professionals is crucial to ensure the safe and effective use of AI in clinical practice.

Conclusion

Machine learning has demonstrated promising predictive and diagnostic capabilities in urodynamic studies, indicating considerable potential for near-future clinical application. However, challenges including a lack of multicentre studies, limited availability of large datasets, and the complexity of clinical conditions continue to restrict its widespread adoption.

References

1. Huang HH, Cheng PY, Tsai CY. Exploring artificial intelligence in functional urology: a comprehensive review. *Urol Sci*. 2025;36(1):2-10.
2. Wang Q, Wang X, Jiang X, Lin C. Machine learning in female urinary incontinence: a scoping review. *Digit Health*. 2024;10:20552076241281450.
3. Chew LE, Hannick JH, Woo LL, Weaver JK, Damaser MS. The future of urodynamics: innovations, challenges, and possibilities. *Neurourol Urodyn*. 2026;45(2):293-8.
4. Zeng X, Mo H, Shen H, Jin T. Evaluating the utility of ChatGPT in addressing conceptual and non-conceptual questions related to urodynamic quality control and trace analysis. *Sci Rep*. 2025;15(1):20091.
5. Zhou Q, Li G, Cui K, Mao W, Lin D, Yang Z, et al. Using machine learning to construct the diagnosis model of female bladder outlet obstruction based on urodynamic study data. *Investig Clin Urol*. 2024;65(6):559-66.
6. Pong YH, Tsai VF, Hsu YH, Lee CH, Wang KC, Tsai YT. Application of a deep learning neural network for voiding dysfunction diagnosis using a vibration sensor. *Applied Sciences*. 2022;12(14):7216.
7. Zareen F, Ouyang Z, Majerus SJ, Bruns TM, Damaser MS, Karam R. Detrusor pressure estimation from single-channel urodynamics. *Annu Int Conf IEEE Eng Med Biol Soc*. 2022;2022:3718-22.
8. Çetinel B, Önal B, Selçuk B, Can G, Aferin U, Yıldırım Ö. Urodynamic curve patterns may predict female bladder outlet obstruction and detrusor underactivity. *Urology*. 2022;165:150-6.
9. Ge Z, Tang L, Peng Y, Zhang M, Tang J, Yang X, et al. Design of a rapid diagnostic model for bladder compliance based on real-time intravesical pressure monitoring system. *Comput Biol Med*. 2022;141:105173.
10. Knorr JM, Werneburg GT. Machine learning and artificial intelligence to improve interpretation of urodynamics. *Curr Bladder Dysfunct Rep*. 2024;19(1):44-53.
11. Weaver JK, Martin-Olenski M, Logan J, Broms R, Antony M, Van Batavia J, et al. Deep learning of videourodynamics to classify bladder dysfunction severity in patients with spina bifida. *J Urol*. 2023;209(5):994-1003.
12. Hobbs KT, Choe N, Aksenov LI, Reyes L, Aquino W, Routh JC, et al. Machine learning for urodynamic detection of detrusor overactivity. *Urology*. 2022;159:247-54.
13. Zhou Q, Chen Z, Wu B, Lin D, Hu Y, Zhang X, et al. A pilot study: detrusor overactivity diagnosis method based on deep learning. *Urology*. 2023;179:188-95.
14. Ding Z, Zhang W, Wang H, Ke H, Su D, Wang Q, et al. An automatic diagnostic system for the urodynamic study applying in lower urinary tract dysfunction. *Int Urol Nephrol*. 2023;56(2):441-9.
15. Matsukawa Y, Kameya Y, Takahashi T, Shimazu A, Ishida S, Yamada M, et al. Development of an artificial intelligence diagnostic system for lower urinary tract dysfunction in men. *Int J Urol*. 2021;28(11):1143-8.
16. Jin J, Chung Y, Kim W, Heo Y, Jeon J, Hoh J, et al. Classification of bladder emptying patterns by LSTM neural network trained using acoustic signatures. *Sensors (Basel)*. 2021;21(16):5328.
17. Eckrich J, Ellinger J, Cox A, Stein J, Ritter M, Blaikie A, et al. Urology consultants versus large language models: potentials and hazards for medical advice in urology. *BJUI Compass*. 2024;5(5):438-44. Erratum in: *BJUI Compass*. 2024;5(12):1324-9.
18. Fogt F. The illusion of diagnostic agreement: institutional bias as a hidden force in modern medicine. *Med Pregl*. 2025;78(5-8):127-31.

Rad je primljen 30. XII 2025.

Recenziran 1. IV 2026.

Prihvaćen za štampu 13. V 2026.

BIBLID.0025-8105:(2026):LXXIX:3-6:129-134.