World Journal of Urology (2026) 44:16
https://doi.org/10.1007/s00345-025-06097-z

RESEARCH

Artificial inteligence reading of cystometric traces provides good correlation with human diagnosis

Jose Emilio Batista-Miranda[1] · Jose Maria Quinteiro Gonzalez[2] · Juan Francisco Monzón-Falconi[1] · Melanie Tatiana Lopez de Mesa[1] · Anaïs Bassas-Parga[1] · Luis Miguel Hernandez Acosta[2] · Daniel Quinteiro Donaghy[2]

Received: 23 February 2025 / Accepted: 12 November 2025

Abstract

Aim Urodynamic studies are essential for diagnosing lower urinary tract dysfunction but are expert-dependent and time-consuming. Artificial intelligence (AI), notably machine learning (ML) and deep learning (DL) may help automate and standardize interpretation, reducing inter-observer variability and improving efficiency.

Objective To evaluate the correlation between artificial intelligence (AI) based classification and human expert diagnosis of detrusor overactivity (DO) in cystometry (CMG), with explicit handling of artifacts and quantification of parameters.

Study design Retrospective, single-center, observational diagnostic-accuracy (cross-sectional) study with a consecutive cohort of adults who underwent cystometry in 2023, in which AI outputs were compared with a reference standard (three-urologist consensus). We evaluated 517 cystometry (CMG) tracings: 200 used to train AI models and 317 reserved for testing. Two approaches were assessed: (i) image-based CNN-VGG16 deep learning, which achieved 75% accuracy for detecting detrusor overactivity (DO) but did not yield quantitative metrics and (ii) wavelet-based ML (Daubechies transforms), which improved accuracy to 84.2%, with 82.6% specificity and 86.3% sensitivity, while providing detailed contraction descriptors. An Isolation Forest anomaly-detection stage identified and managed artifacts (e.g., coughs, open lines, catheter movement). Integrating signal processing (time–frequency denoising and rule-based thresholds) with AI classification supported robust CMG event recognition, enabling clearer identification of DO, estimation of bladder compliance from DO-free segments, and mitigation of artifacts. Both branches produced classifications in less than 20 s per study.

Conclusion Combining algorithmic outputs with expert supervision could deliver practical, faster, and more reproducible urodynamic reporting, while preserving clinical accountability and transparency and generalizability.

Keywords Urodynamic · Machine learning · Overactive detrusor

Introduction

Lower urinary tract symptoms (LUTS) encompass a range of clinical manifestations, classified by the International Continence Society (ICS) into filling phase symptoms such as nocturia, increased frequency, urgency, and incontinence emptying phase symptoms including weak stream, intermittency, and straining to void and post-micturition symptoms, like terminal dribbling and the sensation of incomplete emptying. These symptoms can significantly impact urinary function and are closely linked to diminished quality of life and adverse psychological outcomes [1–3].

Lower urinary tract symptoms have traditionally been related to bladder outlet obstruction (BOO), most frequently when histological benign Prostatic Hyperplasia (BPH) progresses through benign prostatic enlargement (BPE) to BPO, (Benign Prostatic Obstruction) However, increasing numbers of studies have shown that LUTS are often unrelated to the prostate, Bladder dysfunction may also cause LUTS, including detrusor overactivity/OAB, detrusor underactivity (DU)/underactive bladder (UAB), as well as other

Jose Emilio Batista-Miranda, Juan Francisco Monzón-Falconi and Melanie Tatiana Lopez de Mesa are equally contributed.

✉ Juan Francisco Monzón-Falconi
jfrancisco.falconi@hotmail.com

Jose Emilio Batista-Miranda
jebatista@uroclinica.net

1 Teknon Medical Center, Barcelona, Spain

2 University of Las Palmas, Gran Canaria, Spain

structural or functional abnormalities of the urinary tract and its surrounding tissues. Lower urinary tract disorders such as overactive bladder (OAB), detrusor underactivity (DU), and urinary incontinence (stress, urgency, mixed) are highly prevalent and impose a substantial clinical and economic burden [3, 4]. The EpiLUTS study analyzed approximately 30,000 adults over the age of 40 in the United States, the United Kingdom, and Sweden, with an average age of 56.6 years. The results indicated that between 72% and 76% of participants experienced at least one LUTS symptom with a frequency of "sometimes" or more, and that around 48%–53% reported symptoms "often." Most of these symptoms cause moderate to severe discomfort, especially when they occur more frequently [5]. Urodynamic studies are essential diagnostic tools for evaluating lower urinary tract dysfunction.

The evaluation typically begins with noninvasive uroflowmetry, followed by invasive procedures such as cystometry and pressure-flow study. Cystometry (CMG) is recognized as the reference standard for assessing the filling phase. Additional tests, including concurrent electromyography (EMG) of the pelvic floor muscles can provide further clinical insight. However, interpreting these studies is complex, highly dependent on the operator's expertise, and often time-consuming: an estimated between 10 and 20 min, depending on the reviewer's expertise and complexity of traces. The primary objective of urodynamic testing is to reproduce the patient's symptoms while collecting physiological data that help elucidate the underlying pathophysiology and inform treatment decisions, Clinicians typically review the full filling phase and voiding phase, correlating signal morphology with recorded events (urgency, cough tests, leakage) to determine the presence of detrusor overactivity (DO), estimate bladder compliance ($\Delta V/\Delta Pdet$ in contraction-free segments), and identify stress or straining phenomena, Although indispensable, this interpretive process is manual, expertise-dependent, and time-consuming, which motivates efforts to standardize and support it computationally. Despite standardization initiatives, UDS interpretation remains subject to operator dependence and inter-rater variability, Moreover, signal artifacts such as cough, movement, open lines, or poor balloon/rectal channel quality complicate automated and human interpretation alike, and are a frequent source of false positives/negatives if not explicitly modeled, These pain points (subjectivity, artifacts, and workload) justify exploring AI-assisted pipelines that enhance objectivity, reproducibility, and efficiency [6–9].

Urology is increasingly moving toward the integration of artificial intelligence (AI), particularly in imaging and pathology interpretation. AI offers the ability to process large datasets, identify complex patterns, and support diagnostic decision-making. Machine learning has reached expert-level accuracy in various diagnostic fields like breast cancer and retinal disease, suggesting promise for complex signals such as UDS. In urodynamics, research shows that time and frequency-domain analysis, data windowing, and multi-channel models (Pves, Pabd, Pdet) improve detection of DO, with better AUC and specificity. However, results are affected by artifacts and subjective labeling; even advanced models struggle with generalizability when artifact-heavy cases are excluded. Explicit artifact management and multi-signal integration remain essential for reliable AI-assisted UDS interpretation. Within this framework, the application of AI techniques including computer vision, machine learning (ML), and deep learning (DL) to automate cystometry (CMG) interpretation may help address some of the main limitations of urodynamics: namely, the operator-dependent variability in the interpretation of findings and the time required for analysis. Automating CMG interpretation could therefore provide new opportunities to standardize and enhance urodynamic assessment. However, AI applied to CMG must also contend with challenges such as signal noise, artifacts (e.g., coughs, catheter events), baseline drift, and the need to accurately quantify clinically relevant parameters, including detrusor overactivity and bladder compliance. While previous work has explored the use of AI in urology CMG-focused pipelines that explicitly incorporate artifact management and quantitative parameter extraction remain limited [9–16].

Against this background, our study integrates signal processing techniques with ML classification while explicitly modeling artifacts, aiming to reduce subjectivity, improve reproducibility, and provide quantitative outputs (e.g., contraction timing/duration and compliance-friendly segments), all benchmarked against expert reviewers. This approach addresses precisely the shortcomings of current practice and builds on the growing body of AI research in urology/urodynamics, thereby meeting the reviewers' request for clearer justification.

Objective

Objective. To evaluate the correlation between artificial intelligence (AI)–based classification and human expert diagnosis of detrusor overactivity (DO) in cystometry (CMG), with explicit handling of artifacts and quantification of relevant parameters.

Materials and methods

Definitions Vesical pressure (Pves), abdominal/rectal pressure (Pabd), and detrusor pressure (Pdet = Pves − Pabd) were acquired according to ICS standards. Detrusor overactivity (DO) was operationally defined as involuntary detrusor contractions during filling associated with patient-reported urgency; a Pdet threshold of 15 cmH_2O was used to mark candidate contractions; Bladder compliance was estimated as ΔV/ΔPdet within filling segments free of DO [2, 14]. In practice, accurate compliance estimation requires excluding involuntary detrusor contractions that artificially elevate Pdet at the end of filling.

Study design: Retrospective, single-center, observational diagnostic-accuracy (cross-sectional) study with a consecutive cohort of adults who underwent cystometry in 2023, in which AI outputs were compared with a reference standard (three-urologist consensus) using a mutually exclusive training/test split.

Dataset and setting. We reviewed CMG tracings from 517 consecutive adult studies performed in 2023 at a single center using the same equipment and standardized technique according with ICS. Pediatric patients and studies with simultaneous EMG were excluded. All traces were anonymized.

Reference labels. Three functional urologists independently reviewed each trace; disagreements were resolved by consensus, yielding 284 "stable detrusor" and 233 "detrusor overactivity" labels for the image-based dataset. Sex and age were recorded descriptively only and were not used for model training or inference.

Train/test split. This study evaluated 517 cystometry (CMG) tracings, of which 200 images were used to train the AI models and 317 were used for testing. Splits were mutually exclusive and preserved label prevalence (no leakage).

Two AI techniques were assessed: (i) CNN-VGG16 deep learning, which achieved 75% accuracy in detecting detrusor overactivity but did not provide quantitative analysis (e.g., contraction time, volume, or duration); and (ii) wavelet-based ML (Daubechies transforms), which improved accuracy to 84.2%, with 82.6% specificity and 86.3% sensitivity, while also providing detailed contraction data. We implemented a two-branch pipeline:

Deep learning (DL) branch. A CNN based on VGG16 processed CMG images to classify DO vs. stable detrusor; Grad-CAM++ visualizations highlighted regions influencing the decision. This branch provided categorical outputs (presence/absence of DO) but no quantitative metrics.

Wavelet-based ML branch. Pressure signals were transformed from time to time-frequency space using Daubechies wavelets with soft-threshold denoising. Sections before infusion onset and after maximum cystometric capacity were removed; the first 75 mL of infusion were excluded to avoid empty-bladder noise. A rule-based stage applied an effective Pdet threshold of 15 cmH_2O on the reconstructed signal, allowing estimation of contraction timing, duration, and volume at contraction onset. This branch yielded quantitative measures and a categorical label.

Artifact detection and management. We explicitly modeled artifacts via an Isolation Forest anomaly detector operating on Pves/Pabd concordance and local temporal windows. Simultaneous spikes in Pves and Pabd above a dynamic threshold were flagged as cough events; isolated Pabd anomalies without corresponding Pdet changes were flagged as artifacts (e.g., open line, rectal contraction, balloon leakage). Contrary to a common misconception, a decrease in Pabd is unlikely to result from probe descent, as this maneuver typically leads to an increase in Pabd. Therefore, sustained reductions in Pabd were interpreted as indicative of probable balloon leakage or other quality concerns, and these data segments were subsequently downweighed in the analysis. Minor residual noise that persisted following the masking process was deliberately retained during both training and testing phases to evaluate model robustness.

Evaluation metrics and endpoints

Primary endpoint. Diagnostic agreement (accuracy) against the consensus of three urologists.

Secondary endpoints. Specificity, sensitivity, and time to result per trace.

Exploratory outputs. Detection of stress urinary incontinence (SUI), intentional straining, and full bladder compliance profiling were not prespecified primary endpoints; any algorithmic flags related to these phenomena are reported as exploratory and were not clinically validated in this study.

Ethics. All patients provided informed consent for research use of de-identified data; protocol approved by the regional ethics committee (Study 2024-06-URO-CMT).

Results

We analyzed 517 adult CMG studies acquired under ICS standards. Demographics (31.5% men; 68.5% women) are provided for context and were not used as model inputs. Expert review by three functional urologists produced the reference labels: 233 tracings with detrusor overactivity (DO) and 284 with stable detrusor. Model performance is reported against this consensus on a held-out test set (n=317), after training on 200 studies (mutually exclusive split with preserved class prevalence). (Fig. 1)

The image-based CNN-VGG16 deep learning branch achieved a 75% accuracy rate with a sensibility of 87.9% (±5.4) and specificity is 66.1% (±7.1) in classifying test tracings as detrusor overactivity (DO) or stable, as measured against expert consensus. (Table 1). This model provides only a binary output, indicating the presence or absence

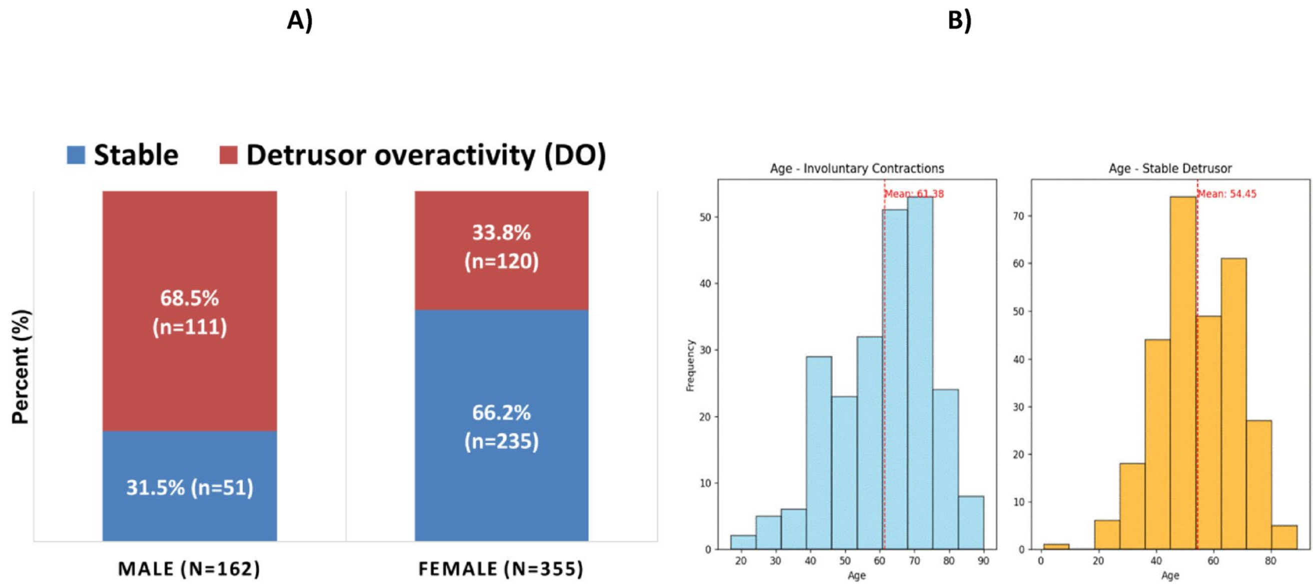

Fig. 1 , (A) Proportion of DO vs. stable detrusor by sex (100% stacked bars; labels show % and counts; (B) Age distribution by diagnostic label (boxplots with mean and SD/95% CI) Note: Age and sex were not used as model inputs

Table 1 Performance summary of the two AI branches against three-urologist consensus (test set $n=317$)

Branch	Inputs	Output type	Accuracy	Sensitivity	Specificity	Latency / tracing	One-line takeaway
CNN-VGG16 (image-based DL)	CMG images (filling phase), standardized; no explicit artifact masks	Binary label (DO vs. stable); Grad-CAM++ heatmaps; no quantitative metrics	75%	87.9% (±5.4)	66.1% (±7.1)	<20 s	Fast, explainable classification, but no contraction-level quantification and more susceptible to image artifacts.
Wavelet-based ML (Daubechies)	Raw Pves/Pabd/Pdet signals; preprocessed + Isolation Forest artifact handling; reconstructed Pdet	Binary label + quantitative metrics (onset vs. volume, duration, peak Pdet, AUC; compliance windows)	84.2%	86.3% (±4.4)	82.6% (±4.4)	<20 s	Higher accuracy and clinically interpretable metrics; artifact modeling reduces cough/open-line false positives.

Notes. Ground truth: three-urologist consensus (233 DO / 284 stable).

of DO, without delivering detailed contraction-level metrics such as onset relative to infused volume, duration, or peak detrusor pressure (Pdet). Notably, Grad-CAM++ heatmaps consistently emphasized regions that were critical to the model's decision-making process, including segments associated with leak markers or sudden increases in Pdet, lending face validity to its predictions despite the lack of quantitative data (see Fig. 2A–B). However, the approach is limited by the absence of explicit artifact modeling and the inability to provide quantitative outputs.

Wavelet-based ML (signal-based) performance. The time–frequency branch (Daubechies transforms with soft-threshold denoising) achieved 84.2% accuracy, with specificity 82.6% (±4.4) and sensitivity 86.3% (±4.4) on the held-out test set. After reconstruction and rule-based thresholding at Pdet ≥ 15 cmH_2O, it produced quantitative contraction descriptors (onset relative to infused volume, duration, peak Pdet, and area over threshold) and automatically identified DO-free filling segments for estimating bladder compliance ($\Delta V/\Delta Pdet$) (Fig. 3A–C). In practice, denoising plus rule-based detection reduced false positives from cough/open-line events and yielded more clinically interpretable outputs than the image-only branch (Table 1).

Latency and reporting. End-to-end time to yield the diagnosis in less than <20 s per trace in both branches. An auxiliary reporting module generated a draft summary for clinician verification and editing, on the other hand human time takes around 10 min to verify quality control of the study and interpret the result of the cystometry and write a diagnosis.

Before wavelet-based inference, we applied an unsupervised Isolation Forest on short sliding windows to detect non-physiologic segments using Pves–Pabd concordance (Fig. 3B). The algorithm isolates outliers via random partitioning; windows with short average path length were flagged as anomalies. Synchronous Pves + Pabd spikes above a dynamic threshold were labeled cough, whereas isolated Pabd excursions without proportional Pdet change

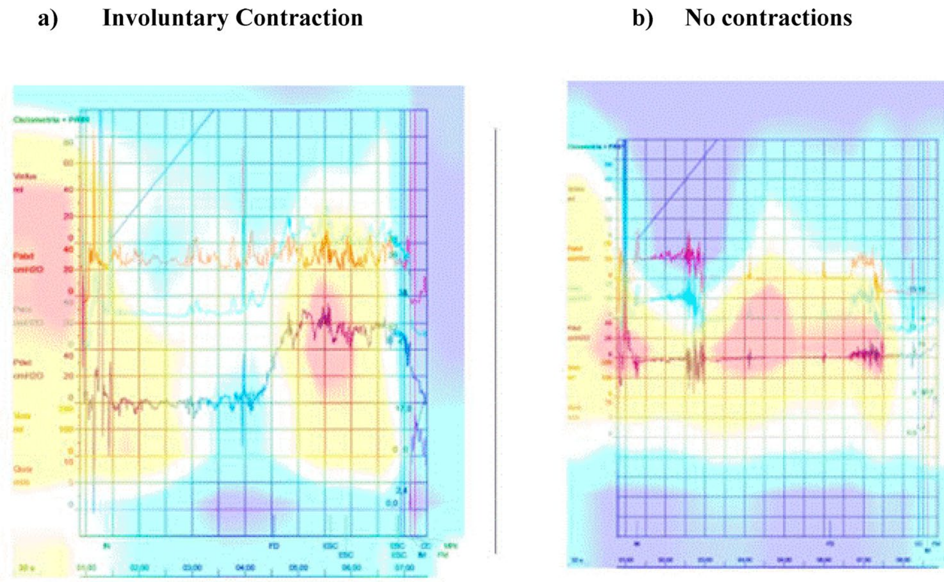

Fig. 2 CNN-VGG16 treated image showing yellow- red areas where the system detected the changes, accounting for contractions

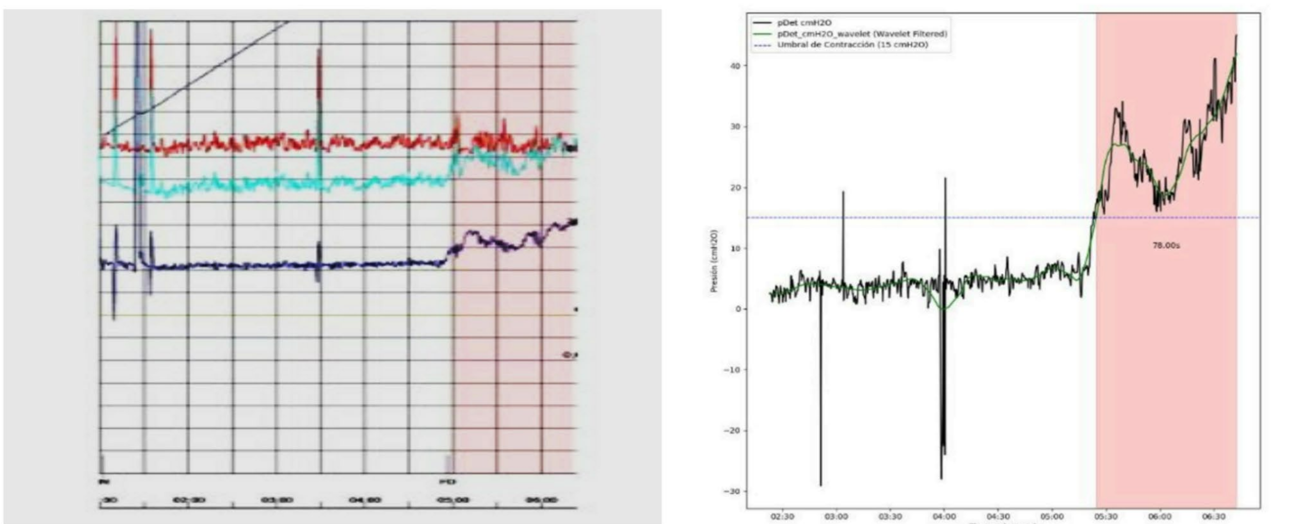

Fig. 3 Original CMG trace (left) and Daubechies Wavelet analysis (right) showing a smoothed Pdet tracing (involuntary contraction marked in pink)

were attributed to open-line events, rectal contractions, or balloon/rectal-channel quality issues. Sustained decreases in Pabd were interpreted as balloon leakage/quality problems—not "probe descent," which typically increases Pabd—and these segments were down-weighted in subsequent analyses (Fig. 4). This preprocessing reduced false positives and improved robustness, particularly in tracings with multiple artifact flags. As prespecified, brief residual noise persisting after masking was retained in both training and testing to assess model resilience.

Misclassifications in both branches clustered in tracings with multiple artifact flags or with borderline Pdet excursions near the 15 cmH_2O threshold. These conditions increased disagreement with the consensus label and highlight the value of explicit artifact modeling and potential hybrid strategies that combine denoised signal features with image-level context.

The algorithms in this study were designed to classify DO vs. stable detrusor, even exploratory flags consistent with stress urinary incontinence (SUI) during cough testing and intentional straining were emitted when channel patterns and annotations permitted; because these were not prespecified endpoints and lacked standardized leak markers, no diagnostic metrics are reported for these phenomena, Compliance estimates derive from DO free segments and are reported as quantitative outputs rather than as a prespecified diagnostic endpoint. In addition to contraction-level descriptors, the wavelet-based branch enabled more robust

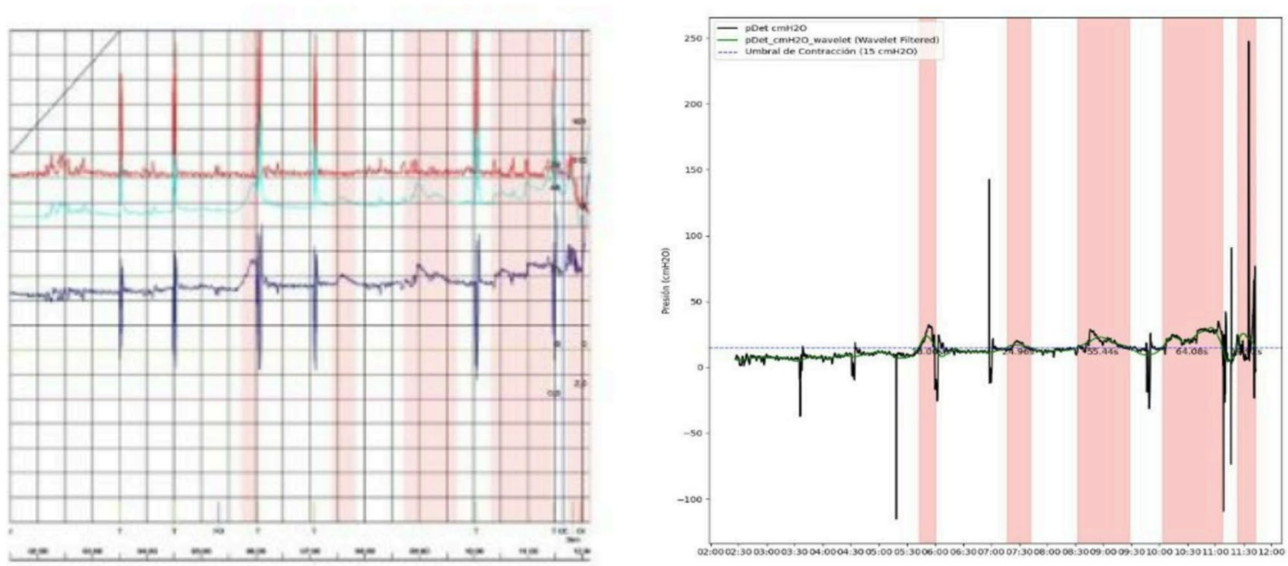

Fig. 4 Original CMG trace (left) and Daubechies Wavelet analysis (right) showing a smoothed a Pdet tracing (involuntary contraction periods marked in pink) differentiated of a cough signal

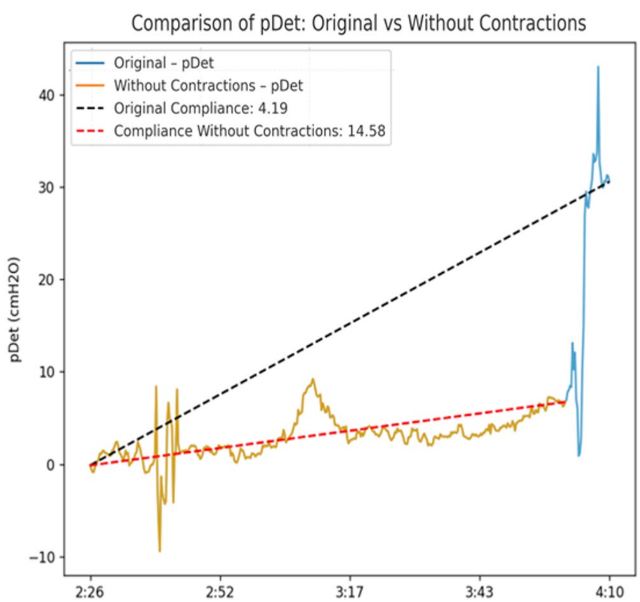

Fig. 5 Example of bladder compliance calculation with and without correction for involuntary detrusor contractions. The blue line shows the original detrusor pressure (pDet), while the orange line represents the corrected signal with contractions removed. The dashed black line indicates the compliance slope calculated directly from the original trace, which is artificially low due to a terminal contraction. The dashed red line shows the corrected compliance, excluding artifacts, providing a more accurate estimate of bladder compliance

estimation of bladder compliance. Figure 5 illustrates a representative case where a terminal involuntary contraction artificially lowered the compliance slope to 4.2 ml/cmH_2O; after correction, the recalculated compliance rose to 14.6 ml/cmH_2O, within normal limits. This example highlights the importance of explicit artifact handling, as uncorrected curves would lead to false classification of low compliance. (Fig. 5)

Discussion

Artificial Intelligence (AI), a concept which dates to the 1950s, can be defined as a set of mathematical algorithms and computer programs that learn to perform tasks requiring types of intelligence usually found in human beings. There are different AI learning techniques used to diagnose overactive detrusor, studies have been published using Artificial Neural Networks (ANN), Support Vectors (SVM), and Convolutional Neural Networks (CNN). Currently, machine learning and deep learning algorithms are primarily used to enhance the interpretation of urodynamics in examinations [1, 3, 11].

We evaluated AI–human agreement using a three-urologist consensus (DO vs. stable) as the reference standard on a held-out test set (train/test 200/317). Performance was computed per tracing: the CNN (VGG16) image branch achieved 75% accuracy (binary label with Grad-CAM++ explanations, no quantitative metrics), while the wavelet-based signal branch reached 84.2% accuracy with specificity 82.6% (±4.4) and sensitivity 86.3% (±4.4), and provided contraction-level descriptors and DO-free segments for compliance estimation. Operating points were not tuned by subgroup during testing, and we intentionally retained short residual noise after masking to probe robustness. Post hoc, we reviewed errors by sex: disagreements clustered where artifact burden was high or Pdet hovered near 15 cmH_2O in both sexes; cough-with-leak sequences (more frequent in female tracings) biased the image branch toward false positives, while sustained abdominal strain in some male tracings produced threshold-borderline excursions. These findings motivate sex-stratified operating points or artifact-aware calibration in future work (without using sex as an explicit input feature).

There is a relationship between age and the presence of involuntary detrusor contractions, which is more common in older individuals. These results suggest that detrusor overactivity may be influenced by factors related to aging and gender, consistent with previous studies which found that physiological changes related to aging, such as neuromuscular degeneration and alterations in detrusor muscle contractility, could contribute to a higher prevalence of involuntary contractions in older individuals. Age is a significant risk factor for lower urinary tract disorders [6–8]. This correlation guides future research and personalized therapeutic approaches.

UDS interpretation is intrinsically challenging owing to pitfalls from the patient, the operator, and the test itself. Examples include: (i) capacity and compliance definitions that lack fixed endpoints and require excluding detrusor contractions, which complicates standardization and comparability; (ii) ambiguity around detrusor leak-point pressure (D-LPP) the traditional 40 cmH_2O cutoff is not universally validated, and neurogenic DO leak-point pressure >75 cmH_2O has been associated with hydronephrosis but with level-3 evidence; and (iii) the need for repeat fillings because several UDS parameters show limited test–retest agreement. These issues affect clinical reads and any algorithm trained on such data [18].

Each AI technique has its own advantages for CMG review. CNN deep learning has demonstrated satisfactory accuracy at 75% and the ability to detect significant changes in the tracings. However, one of the main limitations of this approach is its inability to provide detailed quantitative analysis, such as contraction time, volume, and duration. The Daubechies wavelet method, when added, achieved higher accuracy (84%) in graph classification and analysis of all quantitative data, thereby increasing interpretability. This method offers a valuable screening tool and initial classification that can subsequently be reviewed by an expert physician to validate the final diagnosis. Other previously used SVM-based models found lower sensitivity and specificity. For example, the study conducted by Hobbs et al. [9] found initial performance based on time or frequency features of entire wavelets, achieving a time-domain agreement of 62.4% ± 5.2%, a frequency-domain agreement using FFT of 74.0% ± 6.3%, and later reaching sensitivity of 68% and specificity of 84% using windowing to improve the analysis.

So, we achieve better diagnostic results without windowing; Allowing an analysis of the complete wavelets including artifacts, which means our methodology emphasizes the importance of subtle differences, providing an advantage over deep learning classification approaches or classical methods such as SVM (Support Vector Machine).

These findings indicate that, although the CNN-based model demonstrated improved specificity compared with SVM approaches, further methodological refinement remains necessary to achieve sensitivity levels sufficient for clinical applicability. Furthermore, prior investigations have reported that employing CNN-VGG16 for the detection of detrusor dysfunction in urodynamic studies may be enhanced through the integration of hybrid approaches—such as the combination of wavelet transforms with deep learning—thereby improving both specificity and sensitivity in the identification of urodynamic events [9, 12–14].

Zhou et al. [11], who develop a pilot study on 2023, using deep learning (CNN) and (WSTD) Wavelet soft threshold denoising, found their model achieve an specificity of 90,63% but only 50% of sensitivity, this shows that with the model they use based on CNN the specificity significantly improves in comparison with SVM, but to achieve adequate sensitivity for a correct diagnosis of the test, it is necessary to perfect the method to be used so that the AI can correctly identify the artifacts and make a diagnosis in terms of sensitivity [12].

Also, this finding correlates with those obtained in our study, where artifact identification was addressed by implementing the Isolation Forest anomaly detection method. This method allowed for the identification of non-physiological events such as changes in abdominal and vesical pressures, marking signal anomalies when they coincided within a specific threshold in a time window. Specifically, isolation Forest on short windows using Pves–Pabd concordance flagged nonphysiological segments: synchronous Pves + Pabd spikes were labeled as cough, whereas isolated Pabd excursions without proportional Pdet change suggested open-line/rectal contractions/balloon issues. Sustained Pabd drops were treated as probable balloon leakage (not "probe descent," which typically increases Pabd) and downweighed. This reduced false positives and improved robustness but does not replace the need for standardized acquisition and repeat testing when discrepant, as recommended in the UDS pitfalls literature [1][17, 19, 20].

It is important to mention that artifact treatment is crucial to improving the accuracy of AI models, as both systems (CNN and wavelets) showed that the main sources of classification error were due to the presence of multiple artifacts or borderline values. Among the most common artifacts

[1] This article is based on work that received the Best Abstract Award at the International Continence Society (ICS) Congress 2024.

identified were probe movements, catheter expulsion, syringe valve openings used to correct tube movements, or pressure spikes caused by infusion or vibrations produced by the passage of the infusion line, Our study supports the need to integrate advanced artifact detection and correction strategies within AI models to ensure greater accuracy in the interpretation of urodynamic studies.

Given accuracy < 90%, both branches should be viewed as decision support rather than stand-alone diagnostics. The signal branch's quantitative outputs (onset, duration, peak Pdet, area over threshold; compliance-friendly segments) can facilitate standardized reporting and longitudinal follow-up, while the image branch offers fast triage and explainability. motion/position artifacts and EMG/electrode issues also alter traces, reinforcing the need for rigorous quality control [18–20].

On the other hand, both branches produced results in <20 s per tracing, suggesting a path to shorter review cycles versus manual reads (typically several minutes). The signal branch's quantitative outputs (onset, duration, peak Pdet, area over threshold; compliance-friendly segments) can support standardized reporting and longitudinal comparisons, while the image branch offers fast triage and visual explainability.

Limitation

Single-center design, absence of external validation. Future work should pursue multicenter external validation, calibrated operating points tailored to clinical use-cases, true signal + image fusion to test for incremental gains in accuracy and robustness, and prospective studies quantifying both time savings and diagnostic impact across patient subgroup.

Conclusion

The integration of curve analysis and machine learning contributes to the classification of urodynamic events in CMG, enabling more accurate detection of involuntary detrusor contractions and low bladder compliance. Daubechies Wavelet has a higher accuracy (84%) in classifying graphs and analyze all the quantitative data, thus increasing interpretability. This method surpasses traditional approaches, addressing challenges posed by common artifacts in Urodynamics. The application of advanced computer vision techniques and specific algorithms has proven fundamental in enhancing the objectivity and accuracy of evaluations. These elements support semi-automated, faster (<20 s) traces review while preserving transparency (Grad-CAM++ for image decisions) and focusing quantification on physiologically meaningful segments, These advancements provide a detailed analysis of quantitative information in clinical practice, quality control of urodynamics studies and facilitating semi-automated chart reviews and enabling more reliable diagnoses and personalized treatments for lower urinary tract disorders. The combination of AI techniques with expert supervision could offer a practical system for generating high-quality urodynamic reports, reducing interpretation time and making it applicable in all healthcare settings.

Author contributions All authors reviewed the manuscript.

Data availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Abrams P, Cardozo L, Fall M et al (2002) The standardization of terminology of lower urinary tract function: report from the standardization sub-committee of the international continence society. Neurourol Urodyn 21:167
2. D'Ancona C, Haylen B, Oelke M, Abranches-Monteiro L, Arnold E, Goldman H, Hamid R, Homma Y, Marcelissen T, Rademakers K, Schizas A, Singla A, Soto I, Tse V, de Wachter S, Herschorn S, Standardisation Steering Committee ICS and the ICS Working Group on Terminology for Male Lower Urinary Tract (2019) & pelvic floor symptoms and Dysfunction. The international continence society (ICS) report on the terminology for adult male lower urinary tract and pelvic floor symptoms and dysfunction. Neurourol Urodyn 38(2):433–477
3. Song G, Wang M, Chen B, Long G, Li H, Li R, Liu Z, Wei C, Wang T, Wang S, Liu J, Zhang Y, Liu X (2021) Lower urinary tract symptoms and sexual dysfunction in male: A systematic review and Meta-Analysis. Front Med (Lausanne) 8:653510. https://doi.org/10.3389/fmed.2021.653510
4. EAU Guidelines. End. presented at the EAU Annual Congress Paris April 2024. ISBN 978-94-92671-23-3
5. Coyne KS, Sexton CC, Thompson CL, Milsom I, Irwin D, Kopp ZS, Chapple CR, Kaplan S, Tubaro A, Aiyer LP, Wein AJ (2009) The prevalence of lower urinary tract symptoms (LUTS) in the USA, the UK and Sweden: results from the Epidemiology of LUTS (EpiLUTS) study. BJU Int. 104(3):352 – 60. https://doi.org/10.1111/j.1464-410X.2009.08427. x
6. Smith M, Patel S (2020) Machine learning for predicting outcomes in urological surgery. J Surg Rob 29(2):88–95
7. Matsukawa Y, Kameya Y, Takahashi T, Shimazu A, Ishida S, Yamada M, Sassa N, Yamamoto T (2023) Characteristics of uroflowmetry patterns in men with detrusor underactivity revealed by artificial intelligence. Int J Urol 30(10):907–912. https://doi.org/10.1111/iju.15233
8. Bentellis I, Guérin S, Khene ZE, Khavari R, Peyronnet B (2021) Artificial intelligence in functional urology: how it May shape the future. Curr Opin Urol 31(4):385–390

9. Hobbs KT, Choe N, Aksenov LI et al (2022) Machine learning forurodynamic detection of detrusor overactivity. Urology, 159:47–254
10. Wang H-HS, Cahill D, Panagides J et al (2021) Pattern recognition algorithm to identify detrusor overactivity on urodynamics. Neurourol Urodyn 40:428–434
11. Cullingsworth ZE, Kelly BB, Deebel NA et al (2018) Automated quantification of low amplitude rhythmic contractions (LARC) during real-world urodynamics identifies a potential detrusor overactivity subgroup. Wagg AS, editor. PLoS One. 13: e0201594
12. Liu X, Zhong P, Gao Y, Liao L (2024) Applications of machine learning in urodynamics: A narrative review. Neurourol Urodyn. 43(7):1617–1625. https://doi.org/10.1002/nau.25490. Epub 2024 Jun 4. PMID: 38837301
13. Gammie A, Arlandis S, Couri BM, Drinnan M, Carolina Ochoa D, Rantell A, de Rijk M, van Steenbergen T, Damaser M (2024) Neurourol Urodyn 43(6):1337–1343 Epub 2023 Nov 3. PMID: 37921238; PMCID: PMC11610238. Can we use machine learning to improve the interpretation and application of urodynamic data? ICI-RS 2023
14. Romanzi LJ, Groutz A, Heritz DM, Blaivas JG (2001) Involuntary detrusor contractions: correlation of urodynamic data to clinical categories. Neurourol Urodyn. 20(3):249 – 57. https://doi.org/10.1002/nau.1002. PMID: 11385691
15. Zhou Q, Chen Z, Wu B, Lin D, Hu Y, Zhang X, Liu J (2023) A pilot study: detrusor overactivity diagnosis method based on deep learning. Urology 179:188–195. https://doi.org/10.1016/j.urology.2023.04.030Epub 2023 Jun 13. PMID: 37315592
16. Choi RY, Coyner AS, Kalpathy-Cramer J, Chiang MF, Campbell JP (2020) Introduction to machine learning, neural networks, and deep learning. Transl Vis Sci Technol 9(2):14
17. Bang S, Tukhtaev S, Ko KJ et al (2022) Feasibility of a deep learning-based diagnostic platform to evaluate lower urinary tractdisorders in men using simple uroflowmetry. Investig ClinUrol 63(3):301–308
18. Knorr JM, Werneburg GT (2024) Machine learning and artificial intelligence to improve interpretation of urodynamics. CurrBladder Dysfunct Rep 19:44–53
19. Finazzi Agrò E, Bianchi D, Iacovelli V (2020) Pitfalls in urodynamics. Eur Urol Focus 6(5):820–822. https://doi.org/10.1016/j.euf.2020.01.005Epub 2020 Jan 22. PMID: 31982363
20. Rosier PF, Schaefer W, Lose G et al (2017) Urodynamics, uroflowmetry, cystometry, and pressure-flow study. Neurourol Urodyn 36:1243–1260. https://doi.org/10.1002/nau.23124. Internatio nal Continence Society Good Urodynamic Practices and Terms 2016