

---

# Artificial Intelligence for Pediatric Vesicoureteral Reflux Assessment on Voiding Cystourethrography: A Scoping Evidence Map of Diagnostic Performance and Clinician-Assisted Interpretation

---

[Yusuf Atakan Baltrak](#)\* and Hasan Deliaga

Posted Date: 1 June 2026

doi: 10.20944/preprints202606.0066.v1

Keywords: vesicoureteral reflux; voiding cystourethrography; artificial intelligence; deep learning; pediatric urology; scoping review



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC, OpenAlex.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Review

# Artificial Intelligence for Pediatric Vesicoureteral Reflux Assessment on Voiding Cystourethrography: A Scoping Evidence Map of Diagnostic Performance and Clinician-Assisted Interpretation

Yusuf Atakan Baltrak <sup>1,\*</sup> Hasan Deliaga <sup>2</sup>

<sup>1</sup> Department of Pediatric Urology, Adana City Training and Research Hospital, Adana, Türkiye

<sup>2</sup> Department of Pediatric Urology, Bursa Yuksek Ihtisas Education and Research Hospital, Bursa, Türkiye

\* Correspondence: dratakanbaltrak@gmail.com

## Highlights

### What are the main findings?

- AI models for pediatric VUR assessment on VCUG show promising diagnostic and reliability signals, especially in externally validated and AI-assisted interpretation studies.
- Current evidence remains heterogeneous and does not prove autonomous replacement of expert clinician or radiologist interpretation.

### What are the implications of the main findings?

- AI is best positioned as a clinician-led decision-support tool to improve grading consistency and high-grade VUR recognition.
- Prospective multicenter blinded reader studies using locked external datasets are needed before routine clinical implementation.

## Abstract

**Background/Objectives:** Vesicoureteral reflux (VUR) is an important pediatric urologic condition in which accurate grading on voiding cystourethrography (VCUG) may influence counseling, surveillance, antimicrobial prophylaxis, and surgical decision-making. Visual grading is partly subjective, particularly around the grade III-IV boundary. This review mapped the available evidence on AI-based pediatric VUR assessment on VCUG, with emphasis on diagnostic performance, clinician comparison, AI-assisted interpretation, and implementation gaps. **Methods:** A computer-assisted scoping evidence map was conducted using an exported screening dataset containing 500 records and an extraction dataset containing 13 included studies. Studies were classified as direct VCUG-based VUR grading, low/high-grade classification, AI-assisted clinician interpretation, or indirect clinical prediction. Because the exported files did not document independent duplicate screening, full-text verification, or database-specific deduplication, the synthesis was designed as an evidence map rather than a formal diagnostic meta-analysis. **Results:** Thirteen studies published between 2019 and 2025 were included. Reported AI performance was generally promising but heterogeneous. The strongest externally validated Deep-VCUG study reported external AUC values of 0.944 for unilateral reflux and 0.924 for bilateral reflux. A qVUR model achieved AUC 0.84 and improved grading reliability 3.6-fold. A multicenter VCUG-DAM study showed marked improvement in clinician AUC with AI support. Direct AI-versus-clinician evidence remained limited. **Conclusions:** AI may support pediatric VCUG interpretation by improving grading consistency and high-grade VUR recognition. However, current evidence does not establish autonomous superiority over radiologists. AI should currently be considered a decision-support tool rather than a replacement for expert clinician interpretation.

**Keywords:** vesicoureteral reflux; voiding cystourethrography; artificial intelligence; deep learning; pediatric urology; scoping review

---

## 1. Introduction

Vesicoureteral reflux (VUR) is characterized by retrograde urine flow from the bladder into the ureter and renal collecting system. Accurate diagnosis and grading are clinically relevant because reflux severity may influence surveillance, parental counseling, antimicrobial prophylaxis, endoscopic injection, ureteral reimplantation, and long-term renal risk assessment [1,2].

Voiding cystourethrography (VCUG) remains the conventional imaging method for diagnosing and grading VUR. The International Reflux Grading System classifies reflux from grade I to grade V according to reflux extension, ureteral dilatation, tortuosity, renal pelvic dilatation, and calyceal distortion [1]. Although widely used, this grading system depends on visual interpretation and may show clinically relevant variability between readers.

Reader variability is especially important around the grade III-IV boundary. For pediatric urologists, this distinction may affect counseling, follow-up intensity, prophylaxis discussions, and the threshold for intervention. Therefore, a tool that improves consistency at this threshold could be clinically useful even if it does not replace radiologist interpretation.

AI, machine learning, and deep learning have recently been applied to pediatric VCUG interpretation. These approaches include quantitative feature-based models, random forest classifiers, support vector machines, convolutional neural networks, ensemble learning systems, dual-stream architectures, and dynamic multi-image weighting systems. Reported performance is often promising, but the literature is heterogeneous in target task, imaging input, architecture, validation strategy, comparator design, and outcome reporting.

The key clinical question is therefore not simply whether an AI model can produce a high AUC in a selected dataset. A more clinically relevant question is whether AI can improve meaningful VCUG interpretation compared with clinicians, and whether the evidence is strong enough to support clinical implementation. This review aimed to map the available evidence, separate direct grading studies from indirect prediction studies, evaluate evidence quality, and clarify the current role of AI-assisted pediatric VUR interpretation.

## 2. Materials and Methods

### 2.1. Design and Reporting Framework

A scoping evidence synthesis design was selected because the literature is heterogeneous in population, imaging input, AI architecture, validation strategy, comparator design, target outcome, and performance reporting. The manuscript was organized according to PRISMA Extension for Scoping Reviews (PRISMA-ScR) principles [3]. Because the supplied dataset did not include formal database exports, independent duplicate screening records, or a full-text verification field, this article is best interpreted as a computer-assisted scoping evidence map and not as a completed conventional systematic review with meta-analysis.

### 2.2. Information Sources and Search Strategy

Two exported files generated with Elicit were used as the primary evidence source for the synthesis: a screening export containing 500 records and an extraction export containing detailed data for 13 included studies. The exact export date was not available in the supplied manuscript file. Elicit was used as an organizational support tool; the clinical interpretation, evidence categorization, and final conclusions were author verified. For transparency, the search strategy below is provided as a reproducible example for future formal database verification:

Example PubMed/MEDLINE search strategy: (“vesicoureteral reflux” OR “vesico-ureteral reflux” OR VUR) AND (“voiding cystourethrography” OR “voiding cystourethrogram” OR VCUG) AND (“artificial intelligence” OR “machine learning” OR “deep learning” OR “convolutional neural network” OR CNN OR “computer-aided diagnosis”) AND (child OR children OR pediatric OR paediatric OR infant).

For a future formal systematic or scoping review, database-specific exports from PubMed/MEDLINE, Embase, Scopus, Web of Science, and Google Scholar should be added, deduplicated, independently screened, and incorporated into an expanded PRISMA-ScR flow diagram.

### 2.3. Eligibility Criteria

- Included pediatric patients or pediatric imaging data.
- Evaluated AI, machine learning, deep learning, or computer-aided diagnostic approaches.
- Addressed VUR detection, VUR grading, low/high-grade classification, AI-assisted interpretation, or clinically related VUR prediction.
- Used VCUG or accepted VUR-related imaging or clinical data.
- Reported diagnostic performance, reliability, agreement, or clinician-comparison metrics.
- Used human pediatric datasets.

Studies were excluded if they were unrelated to pediatric VUR, lacked AI methodology, used non-human data only, were editorials, isolated case reports, or technical reports without clinically interpretable outcomes.

### 2.4. Study Selection and Data Charting

Study selection and data extraction were based on the supplied screening and extraction exports. The screening export classified 13 records as included and 487 as excluded. Extracted variables included title, authors, year, DOI, journal or conference venue, AI performance metrics, radiologist or clinician performance metrics, AI algorithm details, direct comparison results, study population, VUR grading system, and dataset or validation characteristics.

The supplied exports did not document independent dual-reviewer verification. This limitation is disclosed explicitly. Before conversion into a full systematic review, eligibility decisions and extracted data should be independently verified by at least two reviewers, with disagreements resolved by consensus or a third reviewer.

### 2.5. Evidence Classification and Confidence Mapping

To reduce methodological heterogeneity, studies were classified into four groups: direct VCUG grading studies, supportive low/high-grade classification studies, AI-assisted interpretation studies, and indirect clinical prediction studies. Direct grading and AI-assisted interpretation studies were prioritized in the narrative synthesis. Confidence was described qualitatively based on study type, sample size, validation strategy, comparator design, and reporting detail available in the extraction export.

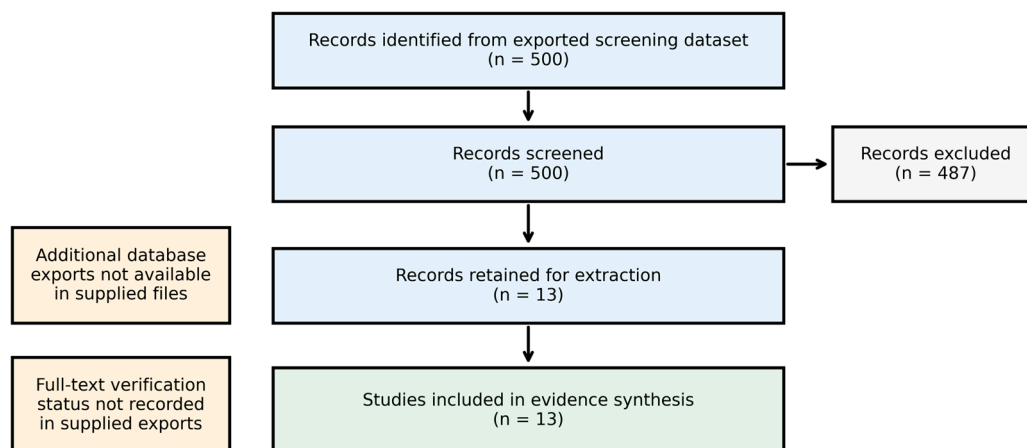
### 2.6. Synthesis Approach

Because of heterogeneity across AI architectures, unit of analysis, target outcomes, validation designs, and performance metrics, a pooled diagnostic meta-analysis was not performed. Findings were summarized narratively and tabulated by clinical evidence role.

### 3. Results

#### 3.1. Study Selection

The screening export contained 500 records. Thirteen studies were included and 487 records were excluded as not directly relevant or below the screening threshold. No additional database records were available in the supplied files.



PRISMA-ScR flow based on the supplied screening and extraction exports.

**Figure 1.** PRISMA-ScR study selection flow diagram. The supplied screening export identified 500 records; 13 studies were included and 487 records were excluded. The diagram should be expanded if formal database exports are added.

#### 3.2. Characteristics of Included Studies

The 13 included studies were published between 2019 and 2025. Dataset size ranged from small public or single-center image collections to multicenter datasets containing thousands of VCUG images. Some studies analyzed images, others renal units, and others patients. AI approaches included ensemble learning, hybrid convolutional neural networks, dual-stream CNN architectures, random forest models, support vector machines, multilayer perceptrons, gradient boosting, optimal classification trees, and dynamic multi-image weighting systems.

**Table 1. Included studies and evidence role based on the supplied extraction export.**

Study	Year	Venue	DOI	Evidence category	Confidence
Li et al. [4]	2024	eClinicalMedicine	10.1016/j.eclinm.2024.102466	Direct VCUG grading / AI-assisted interpretation	High/moderate
Khondker et al. [5]	2022	Journal of Urology	10.1097/JU.0000000000002987	Quantitative grading / reliability	High/moderate

Wu et al. [6]	2025	Research	10.34133/research.0771	AI-assisted interpretation	High/moderate
Khondker et al. [7]	2022	Journal of Pediatric Urology	10.1016/j.jpuro.2021.10.009	Quantitative grading	Moderate
Eroglu et al. [8]	2021	Computer Methods and Programs in Biomedicine	10.1016/j.cmpb.2021.106369	Direct VCUG grading	Moderate
Chen et al. [9]	2025	Journal of Imaging Informatics in Medicine	10.1007/s10278-025-01438-1	Direct VCUG grading	Moderate
Wang et al. [10]	2024	Journal of Pediatric Urology	10.1016/j.jpuro.2023.11.003	Indirect prediction	Low/moderate
Ergun et al. [11]	2024	Journal of Surgery and Medicine	10.28982/josam.8020	Direct VCUG grading	Low/moderate
Kabir et al. [12]	2024	Journal of Pediatric Urology	10.1016/j.jpuro.2023.10.030	Low/high-grade classification	Low/moderate
Alqaraleh et al. [13]	2025	Data and Metadata	10.56294/dm2025756	Classification study	Low/moderate
Estrada et al. [14]	2019	Journal of Urology	10.1097/JU.000000000000186	Indirect clinical prediction	Moderate
Kose et al. [15]	2020	BioMed Research International	10.1155/2020/1895076	Indirect clinical prediction	Low/moderate
Wahed et al. [16]	2024	NETAPPS Conference	10.1109/NETAPPS63333.2024.10823509	Conference / limited report	Low

### 3.3. AI Diagnostic Performance

AI algorithms generally demonstrated promising performance, but reported results varied according to study design, validation rigor, unit of analysis, and target task. The strongest externally validated evidence came from Li et al., whose Deep-VCUG model reported external AUC values of 0.944 for unilateral reflux and 0.924 for bilateral reflux [4]. Khondker et al. reported AUC 0.84 and demonstrated a 3.6-fold improvement in grading reliability compared with traditional assessment [5]. Wu et al. showed strong performance for side-specific reflux recognition and substantial improvement in clinician performance with AI support [6].

Smaller studies reported very high or near-perfect metrics, including training accuracy approaching 100% and AUC values near 1.0 [11,13,16]. These findings should be interpreted cautiously because of small sample size, limited external validation, and possible overfitting.

**Table 2. Clinically relevant AI performance findings.**

Study	Main finding	Clinical interpretation
Li 2024 [4]	External AUC 0.944 for unilateral VUR and 0.924 for bilateral VUR	Strong externally validated signal
Khondker 2022 [5]	AUC 0.84; reliability improved 3.6-fold	AI may improve grading consistency

Wu 2025 [6]	Clinician AUC markedly improved with AI assistance	Strong signal for clinician support
Eroglu 2021 [8]	Accuracy 96.9%	Promising, but validation details should be interpreted cautiously
Chen 2025 [9]	Patient-level accuracy 0.84; AUC 0.82	Clinically plausible side-specific performance
Small/public datasets [11,13,16]	Near-perfect or limited-report metrics	Exploratory only; overfitting and reporting-bias risk

### 3.4. Radiologist and Clinician Comparator Evidence

Radiologist-specific comparator evidence was limited. Most studies used radiologist labels as reference standards but did not independently quantify radiologist performance. Khondker et al. reported inter-rater reliability of 0.44 and median agreement of 0.71, supporting the concern that conventional VUR grading is variable [5]. Wu et al. reported baseline clinician AUC values before AI assistance and demonstrated substantial improvement after AI support [6].

**Table 3. AI-assisted clinician interpretation studies.**

Study	Baseline clinician performance	AI-assisted performance	Interpretation
Li 2024 [4]	Clinician comparison performed	Improved junior and senior clinician performance	AI may support readers across experience levels
Khondker 2022 [5]	Similar accuracy; lower reliability	Reliability improved 3.6-fold	AI may standardize grading
Wu 2025 [6]	Left VUR AUC 0.6288; right VUR AUC 0.7305	Left VUR AUC 0.9641; right VUR AUC 0.9506	Strongest signal for AI-assisted interpretation

### 3.5. Direct Versus Indirect Evidence

Direct VCUG grading studies are the most clinically relevant for pediatric VUR interpretation [4,8,9,11]. Supportive low/high-grade or quantitative VUR studies are relevant for severity stratification but may not fully reproduce standard grade I-V reporting [5,7,12,13]. AI-assisted interpretation studies are the most relevant for real-world workflow because they evaluate whether AI improves clinician performance [4,6]. Indirect prediction studies, including models predicting dilating VUR, recurrent urinary tract infection risk, or likely benefit from VCUG, should not be interpreted as proof that AI can grade reflux severity on VCUG [10,14,15].

### 3.6. Clinical Management Implications

No included study definitively demonstrated that AI changes management decisions such as antibiotic prophylaxis, endoscopic injection, ureteral reimplantation, follow-up intensity, or renal outcome surveillance. Current evidence primarily demonstrates improvement in image classification, grading consistency, and clinician diagnostic performance. Therefore, AI should currently be interpreted as a decision-support tool rather than a management-changing intervention.

## 4. Evidence Quality Map

**Table 4. Evidence quality and implementation concerns.**

Domain	Main concern	Clinical impact
Search reproducibility	Screening and extraction exports were used; conventional database exports were not available	Limits claims of comprehensive systematic search
Full-text assessment	Full-text verification status was not available in the supplied export information	Requires manual verification before a formal systematic review
Study heterogeneity	Different target tasks, datasets, units of analysis, and outcomes	Prevents reliable meta-analysis
Comparator evidence	Limited direct AI-versus-radiologist or AI-versus-clinician studies	Restricts superiority claims
Validation quality	External validation was limited to selected studies	Real-world performance remains uncertain
Small datasets	Near-perfect metrics were reported in some small cohorts	Overfitting risk
Clinical management evidence	No outcome-based management studies	Limits implementation claims
Reporting quality	Incomplete sensitivity, calibration, confidence interval, and decision-curve reporting	Reduces comparability and implementation readiness

## 5. Discussion

### 5.1. Principal Findings

This scoping evidence synthesis shows that AI algorithms have promising performance for pediatric VUR assessment on VCUG, but current evidence does not support autonomous replacement of radiologist interpretation. The most clinically relevant signal is not AI superiority over radiologists, but the possibility that AI may improve grading consistency and reduce interobserver variability.

This distinction is clinically important around the grade III-IV boundary, where grading differences can alter counseling, surveillance intensity, and discussion of intervention. AI may be most useful when it provides standardized support for image interpretation and reduces uncertainty in high-grade VUR recognition.

### 5.2. Interpretation in the Context of Current Evidence

The strongest evidence comes from multicenter or externally validated studies [4,6]. In contrast, near-perfect results from small datasets should be interpreted cautiously because of overfitting risk, limited case diversity, and insufficient external validation [11,13,16].

An important contribution of this review is the separation of direct VCUG grading studies from indirect clinical prediction models. Models predicting dilating VUR, recurrent urinary tract infection, or the probability of benefiting from VCUG may be clinically useful, but they should not be interpreted as proof that AI can accurately grade reflux severity on VCUG [10,14,15].

### 5.3. Clinical Implementation

The most realistic near-term role of AI in pediatric VUR imaging is decision support. AI may help standardize grading, support less experienced readers, improve communication between radiology and pediatric urology teams, and reduce uncertainty in high-grade VUR recognition. Final interpretation and management decisions should remain clinician-led.

For implementation, future systems should be tested in reader studies that compare AI-alone, clinician-alone, and AI-assisted clinician interpretation on the same locked external dataset. Studies should report not only AUC and accuracy, but also sensitivity for high-grade VUR, calibration, decision-curve analysis, confidence intervals, reporting time, error patterns, and performance across age, sex, unilateral/bilateral disease, and imaging quality subgroups.

### 5.4. Reporting Standards for Future AI studies

Future research should follow established AI and diagnostic accuracy reporting frameworks, including CLAIM, TRIPOD+AI, and STARD-AI, to improve reproducibility, transparency, and clinical applicability [17–20]. Predefined model locking, external validation, data governance, and clear human-AI interaction protocols are especially important before clinical deployment.

### 5.5. Limitations

This review has several limitations. First, although the supplied screening and extraction exports provide a reproducible basis for the current synthesis, formal database exports from PubMed/MEDLINE, Embase, Scopus, Web of Science, and Google Scholar were not available. Second, full-text verification status was not available in the supplied export information and should be completed manually before a formal systematic review. Third, the included studies were heterogeneous in AI architecture, target task, validation strategy, and comparator design. Fourth, only a small number of studies directly compared AI with clinician performance. Fifth, no study definitively showed that AI changes management decisions or improves patient outcomes.

### 5.6. Future Directions

Future studies should use prospective multicenter reader-based designs. The ideal study should include AI-alone, radiologist-alone, pediatric urologist-alone, and AI-assisted reader arms evaluated on the same locked external dataset. Future studies should also evaluate grade III-IV discordance, high-grade VUR sensitivity, calibration, decision-curve analysis, reporting time, PACS integration, management impact, and patient outcomes.

## 6. Conclusions

AI algorithms show promising diagnostic performance for pediatric VUR assessment on VCUG, particularly in externally validated and multicenter studies. However, current evidence does not prove autonomous superiority over radiologists. AI should currently be considered a decision-support tool rather than a replacement for expert clinician interpretation. Its most clinically relevant role may be improving grading consistency, supporting high-grade VUR recognition, and reducing uncertainty around the grade III-IV boundary. Prospective multicenter blinded reader studies are required before AI-assisted VCUG grading can be recommended for routine clinical implementation.

**Supplementary Materials:** The following supporting information can be downloaded at the website of this paper posted on Preprints.org: Supplementary File S1: PRISMA-ScR checklist; Supplementary File S2: screening export; Supplementary File S3: extraction export. If the CSV exports are not uploaded as supplementary files, the Data Availability Statement should be revised accordingly.

**Author Contributions:** Conceptualization, Y.A.B.; methodology, Y.A.B.; evidence synthesis, Y.A.B.; writing-original draft preparation, Y.A.B.; writing-review and editing, Y.A.B. The author has read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable. This article is an evidence synthesis of previously published studies and does not include new patient-level data.

**Informed Consent Statement:** Not applicable. This article does not include new patient-level data or identifiable patient information.

**Data Availability Statement:** The data summarized in this review were derived from the supplied screening and extraction exports and previously published studies. The exports should be submitted as supplementary files where journal policy and licensing permit; otherwise, they are available from the corresponding author upon reasonable request.

**Acknowledgments:** The author thanks all researchers whose work contributed to the evidence base summarized in this review.

**Use of AI Tools:** Elicit was used to support literature screening and extraction organization. Generative AI tools were used for language editing, formatting, and manuscript organization. The author verified the scientific interpretation, conclusions, and final submitted content.

**Conflicts of Interest:** The author declares no conflicts of interest.

## References

1. Lebowitz, R.L.; Olbing, H.; Parkkulainen, K.V.; Smellie, J.M.; Tamminen-Mobius, T.E. International system of radiographic grading of vesicoureteric reflux. *Pediatr. Radiol.* **1985**, *15*, 105–109. <https://doi.org/10.1007/BF02388714>.
2. Peters, C.A.; Skoog, S.J.; Arant, B.S.; Copp, H.L.; Elder, J.S.; Hudson, R.G.; Khoury, A.E.; Lorenzo, A.J.; Pohl, H.G.; Shapiro, E.; Snodgrass, W.T.; Diaz, M. Summary of the AUA guideline on management of primary vesicoureteral reflux in children. *J. Urol.* **2010**, *184*, 1134–1144. <https://doi.org/10.1016/j.juro.2010.05.065>.
3. Tricco, A.C.; Lillie, E.; Zarin, W.; O'Brien, K.K.; Colquhoun, H.; Levac, D.; Moher, D.; Peters, M.D.J.; Horsley, T.; Weeks, L.; et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. *Ann. Intern. Med.* **2018**, *169*, 467–473. <https://doi.org/10.7326/M18-0850>.
4. Li, Z.; Tan, Z.; Wang, Z.; Tang, W.; Ren, X.; Fu, J.; Wang, G.; Chu, H.; Chen, J.; Duan, Y.; et al. Development and multi-institutional validation of a deep learning model for grading of vesicoureteral reflux on voiding cystourethrogram: a retrospective multicenter study. *EclinicalMedicine* **2024**, *69*, 102466. <https://doi.org/10.1016/j.eclinm.2024.102466>.
5. Khondker, A.; Kwong, J.C.C.; Yadav, P.; Chan, J.Y.H.; Singh, A.; Skreta, M.; et al. Multi-institutional validation of improved vesicoureteral reflux assessment with simple and machine learning approaches. *J. Urol.* **2022**, *208*, 1314–1322. <https://doi.org/10.1097/JU.0000000000002987>.
6. Wu, M.; Li, Z.; Liu, Y.; et al. Dynamic multi-image weighting for automated detection and diagnosis of abnormal urinary tract on voiding cystourethrography with a deep learning system: a retrospective, large-scale, multicenter study. *Research* **2025**, *8*, 0771. <https://doi.org/10.34133/research.0771>.
7. Khondker, A.; Kwong, J.C.C.; Rickard, M.; Skreta, M.; Keefe, D.T.; Lorenzo, A.J.; Erdman, L. A machine learning-based approach for quantitative grading of vesicoureteral reflux from voiding cystourethrograms: methods and proof of concept. *J. Pediatr. Urol.* **2022**, *18*, e1–e78–78.e7. <https://doi.org/10.1016/j.jpuro.2021.10.009>.
8. Eroglu, Y.; Yildirim, K.; Cinar, A.; Yildirim, M. Diagnosis and grading of vesicoureteral reflux on voiding cystourethrography images in children using a deep hybrid model. *Comput. Methods Programs Biomed.* **2021**, *210*, 106369. <https://doi.org/10.1016/j.cmpb.2021.106369>.

9. Chen, G.; Su, L.; Wang, S.; Liu, X.; Wu, W.; Zhang, F.; Zhao, Y.; Zhu, L.; Zhang, H.; Wang, X.; Yu, G. Automated grading of vesicoureteral reflux (VUR) using a dual-stream CNN model with deep supervision. *J. Imaging Inform. Med.* **2025**, *38*, 3517–3525. <https://doi.org/10.1007/s10278-025-01438-1>.
10. Wang, H.-H.S.; Li, M.; Cahill, D.; Panagides, J.; Logvinenko, T.; Chow, J.; Nelson, C. A machine learning algorithm predicting risk of dilating VUR among infants with hydronephrosis using UTD classification. *J. Pediatr. Urol.* **2024**, *20*, 271–278. <https://doi.org/10.1016/j.jpuro.2023.11.003>.
11. Ergun, O.; Serel, T.A.; Ozturk, S.A.; Serel, H.B.; Soyupek, S.; Hoscan, B. Deep-learning-based diagnosis and grading of vesicoureteral reflux: a novel approach for improved clinical decision-making. *J. Surg. Med.* **2024**, *8*, 12–16. <https://doi.org/10.28982/josam.8020>.
12. Kabir, S.; Pippi Salle, J.L.; Chowdhury, M.E.H.; Abbas, T.O. Quantification of vesicoureteral reflux using machine learning. *J. Pediatr. Urol.* **2024**, *20*, 257–264. <https://doi.org/10.1016/j.jpuro.2023.10.030>.
13. Alqaraleh, M.; Alzboon, M.S.; Al-Batah, M.S.; Al Aesa, L.Y.; Abu-Arqoub, M.H.; Marie, R.R.; Alsmadi, F.H. Machine learning-based quantification of vesicoureteral reflux with enhancing accuracy and efficiency. *Data Metadata* **2025**, *4*, 756. <https://doi.org/10.56294/dm2025756>.
14. Estrada, C.R.; Nelson, C.P.; Wang, H.H.; Bertsimas, D.; Dunn, J.; Li, M.; Zhuo, D.; et al. Targeted workup after initial febrile urinary tract infection: using a novel machine learning model to identify children most likely to benefit from voiding cystourethrogram. *J. Urol.* **2019**, *202*, 144–152. <https://doi.org/10.1097/JU.000000000000186>.
15. Kose, T.; Ozgur, S.; Cosgun, E.; Keskinoglu, A.; Keskinoglu, P. Effect of missing data imputation on deep learning prediction performance for vesicoureteral reflux and recurrent urinary tract infection clinical study. *BioMed Res. Int.* **2020**, *2020*, 1895076. <https://doi.org/10.1155/2020/1895076>.
16. Wahed, M.A.; Alzboon, M.; Alqaraleh, M.; et al. Enhancing diagnostic precision in pediatric urology: machine learning models for automated grading of vesicoureteral reflux. In Proceedings of the 2024 7th International Conference on Internet Applications, Protocols, and Services (NETAPPS), 2024. <https://doi.org/10.1109/NETAPPS63333.2024.10823509>.
17. Mongan, J.; Moy, L.; Kahn, C.E., Jr. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): a guide for authors and reviewers. *Radiol. Artif. Intell.* **2020**, *2*, e200029. <https://doi.org/10.1148/ryai.2020200029>.
18. Tejani, A.S.; Klontzas, M.E.; Gatti, A.A.; et al. Checklist for Artificial Intelligence in Medical Imaging: CLAIM 2024 update. *Radiol. Artif. Intell.* **2024**. <https://doi.org/10.1148/ryai.240300>.
19. Collins, G.S.; Moons, K.G.M.; Dhiman, P.; Riley, R.D.; Beam, A.L.; Van Calster, B.; et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ* **2024**, *385*, e078378. <https://doi.org/10.1136/bmj-2023-078378>.
20. Sounderajah, V.; Guni, A.; Liu, X.; et al. The STARD-AI reporting guideline for diagnostic accuracy studies using artificial intelligence. *Nat. Med.* **2025**. <https://doi.org/10.1038/s41591-025-03953-8>.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.