J Surg Med. 2024;8(1):12-16.
https://doi.org/10.28982/josam.8020

Research article
https://jsurgmed.com/

JOSAM
Journal of Surgery and Medicine
e-ISSN: 2602-2079

Deep-learning-based diagnosis and grading of vesicoureteral reflux: A novel approach for improved clinical decision-making

Osman Ergün [1], Tekin Ahmet Serel [1], Sefa Alperen Öztürk [1], Hüseyin Bulut Serel [2], Sedat Soyupek [1], Burak Hoşcan [3]

[1] Süleyman Demirel University, Medical School, Department of Urology, Isparta, Turkey
[2] Getir Company, Software Engineering, Ankara, Turkey
[3] Bilim University, Vocational School of Health Services, Antalya, Turkey

ORCID of the author(s)

OE: https://orcid.org/0000-0001-7611-0933
TAS: https://orcid.org/0000-0002-2135-353X
SAÖ: https://orcid.org/0000-0003-4586-9298
HBS: https://orcid.org/0000-0002-8955-7942
SS: https://orcid.org/0000-0002-7345-3452
BH: https://orcid.org/0000-0002-7887-9575

Corresponding Author
Osman Ergün
Süleyman Demirel University Faculty of Medicine Hospital, Urology Academic Floor Secretariat, Eastern Campus, Çünür, Isparta, Turkey
E-mail: osmanergun77@gmail.com

Ethics Committee Approval
Ethical approval was not required, as the data was sourced from an open-access archive. Informed consent was not required due to design of the study. This study does not contain identifying information of the patients. However, it is crucial to emphasize that all data used in this study has been anonymized and treated with the utmost confidentiality.

Conflict of Interest
No conflict of interest was declared by the authors.

Financial Disclosure
The authors declared that this study has received no financial support.

Published
2024 January 17

Abstract

Background/Aim: Vesicoureteral reflux (VUR) is a condition that causes urine to flow in reverse, from the bladder back into the ureters and occasionally into the kidneys. It becomes a vital cause of urinary tract infections. Conventionally, VUR's severity is evaluated through imaging via voiding cystourethrography (VCUG). However, there is an unresolved debate regarding the precise timing and type of surgery required, making it crucial to classify VUR grades uniformly and accurately. This study's primary purpose is to leverage machine learning, particularly convolutional neural network (CNN), to effectively identify and classify VUR in VCUG images. The aspiration is to diminish classification discrepancies between different observers and to create an accessible tool for healthcare practitioners.

Methods: We utilized a dataset of 59 VCUG images with diagnosed VUR sourced from OpenI. These images were independently classified by two seasoned urologists according to the International Reflux Classification System. We utilized TensorFlow, Keras, and Jupyter Notebook for data preparation, segmentation, and model building. The CNN Inception V3 was employed for transfer learning, while data augmentation was used to improve the model's resilience.

Results: The deep-learning model attained exceptional accuracy rates of 95% and 100% in validation and training, respectively, after six cycles. It effectively categorized VUR grades corresponding to the global classification system. Matplotlib tracked loss and accuracy values, while Python-based statistical analysis assessed the model's performance using the F1-score.

Conclusion: The study's model effectively categorized images, including those of vesicoureteral reflux, which has significant implications for treatment decisions. The application of this artificial intelligence model may help reduce interobserver bias. Additionally, it could offer an objective method for surgical planning and treatment outcomes.

Keywords: vesicoureteral reflux, voiding cystourethrography, artificial intelligence, convolutional neural network, deep learning

How to cite: Ergün O, Serel TA, Öztürk SA, Serel HB, Soyupek S, Hoşcan B. Deep-learning-based diagnosis and grading of vesicoureteral reflux: A novel approach for improved clinical decision-making. J Surg Med. 2024;8(1):12-16.

Introduction

Vesicoureteral reflux (VUR) is a condition where urine flows backward from the bladder into the ureters and, occasionally, the kidneys. It is a significant contributory factor to urinary tract infections and primarily arises from anatomical or functional abnormalities [1]. The ureter does not close during voiding or under conditions of high intravesical pressure, such as neurogenic bladder, urethral stenosis, and posterior urethral valve [2]. Untreated, VUR may result in complications like reflux nephropathy, kidney damage leading to renal failure, end-stage renal disease, or growth retardation in children [3].

Voiding cystourethrography (VCUG) is a common radiological test employed to detect vesicoureteral reflux (VUR). VCUG enables comprehensive insights into the presence, absence, and severity of VUR, as defined by international standards [4-7]. It is often ordered by specialists such as urologists, pediatric urologists, pediatricians, pediatric nephrologists, and pediatric surgeons. For patients with persistent high-grade reflux (grades 4/5), surgical correction should be considered. However, consensus is lacking on the timing and selection of surgical methods. Reimplantation is generally preferable for higher reflux grades, whereas endoscopic injections can produce satisfactory results for lower grades. Thus, standardizing and accurately grading VUR is crucial. It strongly impacts treatment choices and promotes clear communication among health professionals.

Machine learning, a facet of artificial intelligence, involves teaching a computer to develop programs based on given data and anticipated results. This approach is increasingly prevalent in the biomedical field. While humans can easily understand the context and importance of an image, translating this ability to machine comprehension is a complex task. The convolutional neural network (CNN), a prominent machine learning technique, has made significant strides in the field of medical imaging. CNN architectures are widely employed in image detection and classification [8]. In simple terms, an application, often shortened to ‘app’, is a type of software or collection of programs designed to assist users in completing specific tasks. The use of CNNs holds notable implications for healthcare professionals in clinical practice.

This research focused on using deep-learning methods to identify and categorize vesicoureteral reflux (VUR) in images from Voiding Cystourethrogram (VCUG) studies. The intention was to decrease discrepancies in classification between observers and to create a corresponding application.

Materials and methods

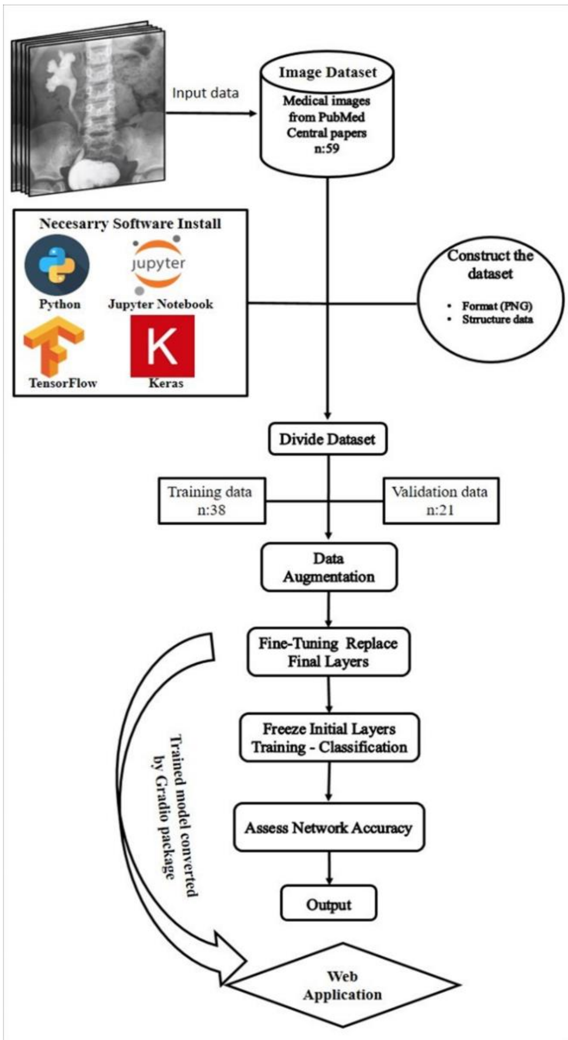
Data collection

This study used a dataset of 59 images, all previously diagnosed with VUR using VCUG. The images were sourced from OpenI, a PubMed Central online medical image archive. Because the data was obtained from an open-access archive, there was no need for ethical approval. All analyses were carried out on this readily available data.

Image classification

Two seasoned urologists (OE, SAÖ), with a decade of experience each, classified all images based on the International Reflux Classification System, independently and unaware of the other’s determinations. In instances of conflict, a senior urologist (SS) with 20 years of experience was called in. Images causing unresolved disagreements were omitted from the study. The study’s flow is illustrated in Figure 1.

Figure 1: The flowchart of the study.



Data preparation

For data analysis, we used TensorFlow, Keras, and Jupyter Notebook. Jupyter Notebook, an open-source web application, enabled us to create and share documents containing text and live code [9,10].

Data splitting

The images were divided at random into two groups: a training group of 38 cases and a validation group of 21 cases. This method aimed to represent the dataset accurately and evenly [11,12].

Model construction

A Jupyter Notebook environment was set up using prerequisites from the Keras library. We used a convolutional neural network model called Inception V3 for this task [13]. The Cross-Entropy loss function was used for classification, and we applied the Adam optimizer with standard settings [14].

Transfer learning

We designed a two-step transfer learning strategy, which included freezing and training the final layers. We then added extra layers with random initialization to the pre-trained Inception V3 model, originally trained on ImageNet. Finally, we fine-tuned the model with a weight of 0.0001.

Data augmentation

The Keras Image Data Generator was used for image enhancement to improve the model's strength. This generator was set up with the training directory, file requirements, picture dimensions, and batch size.

Model deployment

To develop an easy-to-use application, we used the Gradio package in Python. This facilitated the transformation of the trained model into a web application. You can access the web application at https://huggingface.co/spaces/Ragio/VUR_grade_prediction.

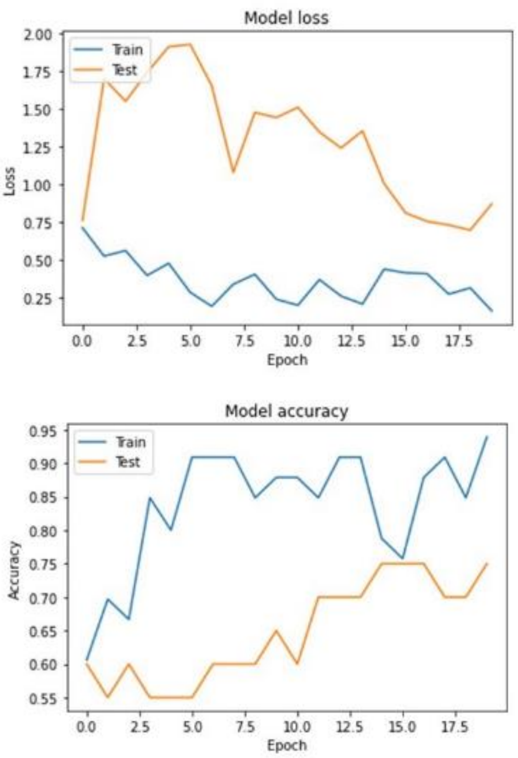
Statistical analysis

All statistical analyses were conducted using Python 3.6. The model's performance was evaluated using the F1-score, which ranges from 0 to 1, with 1 demonstrating flawless positive predictive value and sensitivity.

Results

Once the code is run, the model begins its training process. After six epochs, the model yields a training accuracy of 100% and a validation accuracy of 95%. The validation accuracy is lower due to the smaller number of images (n=21) in this group. If both validation and training values decrease, it demonstrates effective learning by the model. Matplotlib is a Python package utilized for plotting and generating figures in multiple formats. It captures the loss and accuracy values in arrays (Figure 2).

Figure 2: Loss and accuracy arrays of the created model.

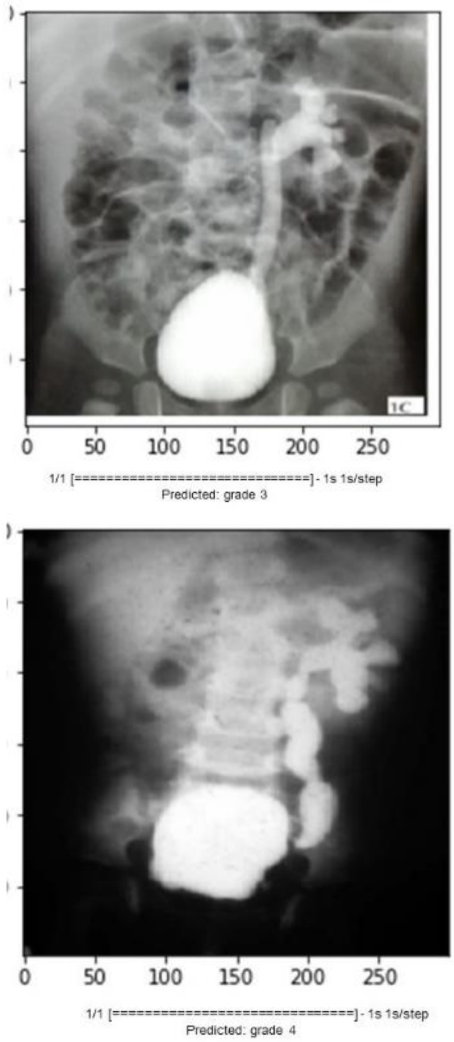


The hold-out test is employed to gauge the efficiency of our trained model as well as to evaluate the training and validation of the images. Keras, a Python-based deep-learning library, is utilized in training models within neural networks and executing the data generator on a batch of test images. It operates on a directory of issues using a for-loop or evaluates one point at a time.

Our model successfully estimated all vesicoureteral reflux (VUR) grades according to the International Reflux Classification System. Specifically, it categorized: Grade I as the presence of contrast medium solely in the non-dilated ureter; Grade 2 as the presence of contrast in both the ureter and the renal pelvis, with no significant dilatation; Grade 3 as mild dilation in the ureter and the pyelocalyceal system; Grade 4 as tortuosity and moderate dilation in the ureter, along with blunted renal fornices; and Grade 5 as tortuosity and severe dilation in the ureter, dilated pyelocalyces, loss of fornices, and papillary impression.

Figure 3 visually represents the estimated grade 4 and grade 5 VUR. The potential to distinguish between these two grades is indicated with numbers in parentheses, which range from 0 to 1.

Figure 3: Estimated values of grade III and IV vesicoureteral reflux.



Discussion

Vesicoureteral reflux (VUR) is a perilous disease that poses significant risks to a child's kidney health, alongside various other health complications [15]. Early diagnosis of this disease is critical to prevent recurring urinary tract infections and further kidney damage in children [16]. Urologists universally advise that all children with recurrent febrile urinary tract infections undertake Voiding Cystourethrograms (VCUGs). The VCUG's findings allow them to gauge the severity of VUR, enabling them to prescribe accurate treatments. This can range from frequent follow-ups, prophylactic antibiotics, or even necessitate endoscopic or surgical procedures.

Studies have shown that the combination of human analysis and deep-learning systems is more effective than using either independently [17]. The employment of CNNs in reviewing radiological images can decrease the workload of healthcare professionals and lower clinical practice costs. Such technology can also reduce variability in observation and promote uniform interpretation. As a result, integrating deep learning into decision-making processes can significantly benefit specialists. Developers are creating deep-learning algorithms to analyze medical images, such as ultrasound and MRI scans, to help identify and categorize urological diseases [18-24]. Additionally, tools powered by deep learning are under examination for their potential to assist in surgical planning and decision-making in the field of urology.

Technologies play a crucial role in diagnosing and managing urological diseases. However, their accuracy and therapeutic effectiveness are often reliant on the operator's skills and the urologist's expertise. Pediatric urologists demonstrate an increased interest in standardizing classification methods and reporting in the pediatric urology field. Prenatal hydronephrosis, seen in about 5% of pregnancies, commonly leads to urology clinic referrals. Recognizing this, Lorenzo et al. [25] developed a deep-learning model to predict the need for surgical intervention in children diagnosed with hydronephrosis during prenatal screening. This early study offers substantial potential for improving clinical decision-making by identifying which patients are more likely to require surgery. Another study reported that deep learning could accurately detect VUR and hydronephrosis [26]. This underscores the strength of deep learning in quickly developing an accurate differentiation algorithm for recognizing hydronephrosis and VUR using minimal code and training cases.

Serrano-Durba et al. [27] developed a deep-learning model to predict the results of endoscopic treatment for VUR. Compared to traditional statistical methods, the deep-learning model proved superior in all evaluated variables, including sensitivity, specificity, and positive and negative predictive values. A related study involved 96 children with VUR and aimed to generate a deep-learning model for predicting the outcomes of various VUR treatments. The authors observed that deep learning outperformed traditional statistical methods [28]. The model accurately predicted VUR resolution and suggested that deep learning could potentially enhance traditional methods for more precise clinical outcome predictions.

There is no substantial evidence showing significant benefits of correcting persistent, low-grade reflux (grades I-III) when there are no symptoms and kidney function remains normal. Only those enduring persistent high-grade reflux (grades IV/V) should contemplate surgical correction. Reimplantation tends to yield better results than endoscopic corrections for higher reflux grades. Hence, patients with persistent high-grade reflux should be offered reimplantation, while endoscopic correction might be more fitting for lower reflux grades.

Current data insinuates that about one-third of Voiding Cystourethrogram (VCUG), results display inconsistent grading among clinicians, especially with moderate (grade 3-4) Vesicoureteral Reflux (VUR) [29]. Khondker et al. [30] developed a deep-learning model to gauge the reliability of reflux grading by assessing VCUGs for four features: ureteral tortuosity, proximal, distal, and maximum ureteral dilatation. This feature set was used to train the model to predict VUR grades.

The team reported that the developed model determined VUR grades with human-like accuracy, and there was a strong correlation between VUR grade and the four features mentioned above. Unlike the International Reflux Classification System, their model ignored the appearances of renal calyces in grade classification, posing a notable limitation. Additionally, the use of data obtained by measurements defined by established mathematical relationships might pose a disadvantage over the International Reflux Classification System.

Our study underscores the accurate recognition of VUR grades through the application of a deep-learning model, indicating the model's ability to correctly classify and categorize images, particularly those with grade 3 and 4 VUR. This accuracy is consequential as it can influence treatment decisions. The use of deep learning also curbs interobserver bias, helps cut costs, and reduces patients' radiation exposure. In addition, it facilitates objective surgical planning and the attainment of treatment goals.

We have developed a deep-learning model and an accompanying web application. This app enables healthcare professionals to expedite image interpretation, thereby speeding up diagnosis times. Moreover, it lessens healthcare workers' workload and enhances patient care quality. This application also proves useful for preliminary diagnosis in non-specialist environments.

We have converted our deep-learning model into an interactive web application using Gradio that is accessible to users worldwide. This user-friendly interface lets us integrate our model smoothly into a web application, providing real-time predictions based on data users input. Gradio's customization options and user-friendly design can create a smooth user experience, enabling easy interaction with our model. The web application is unrestricted, providing open access to all users.

Limitations

This study has several limitations. First, the number of images available for analysis was limited. However, we used a novel method that includes stock images from OpenI to construct the training set for the artificial intelligence. This innovative approach greatly improves the data's generalizability. Second, it is important to note that reflux can occur during both bladder filling and voiding. Reflux during voiding has a higher chance of resolution than during bladder filling. Regrettably, due to the nature of our data source, we are unable to determine whether the observed reflux in the images occurred during the bladder filling or voiding phase.

Conclusions

Vesicoureteral Reflux (VUR) is a serious disease that can cause significant mortality and morbidity, making early and accurate diagnosis crucial, particularly in pediatric patients. In our study, we developed a deep-learning model and application aimed at diagnosing and grading VUR in voiding cystourethrography (VCUG) images. Our model proves highly accurate in both diagnosis and staging of this disease, supported by a user-friendly web application that we also developed.

References

1. Cetin N, Gencler A, Kavaz Tufan A. Risk factors for development of urinary tract infection in children with nephrolithiasis. J Paediatr Child Health. 2020;56(1):76-80.
2. Weitz M, Schmidt M. To screen or not to screen for vesicoureteral reflux in children with ureteropelvic junction obstruction: a systematic review. Eur J Pediatr. 2017;176(1):1-9.
3. Wadie GM, Moriarty KP. The impact of vesicoureteral reflux treatment on the incidence of urinary tract infection. Pediatr Nephrol. 2012;27(4):529-38.
4. Siomou E, Giapros V, Serbis A, Makrydimas G, Papadopoulou F. Voiding urosonography and voiding cystourethrography in primary vesicoureteral reflux

associated with mild prenatal hydronephrosis: a comparative study. Pediatr Radiol. 2020;50(9):1271-6.
5. Baydilli N, Selvi I, Pinarbasi AS, Akinsal EC, Demirturk HC, Tosun H, et al. Additional VCUG-related parameters for predicting the success of endoscopic injection in children with primary vesicoureteral reflux. J Pediatr Urol. 2021;17(1):68.e1-68.e8.
6. Schaeffer AJ, Greenfield SP, Ivanova A, Cui G, Zerin JM, Chow JS, et al. Reliability of grading of vesicoureteral reflux and other findings on voiding cystourethrography. J Pediatr Urol. 2017;13(2):192-8.
7. Çelebi S, Özaydın S, Baştaş CB, Kuzdan Ö, Erdoğan C, Yazıcı M, et al. Reliability of the Grading System for Voiding Cystourethrograms in the Management of Vesicoureteral Reflux: An Interrater Comparison. Adv Urol. 2016;2016:1684190.
8. Ilhan HO, Sigirci IO, Serbes G, Aydin N. A fully automated hybrid human sperm detection and classification system based on mobile-net and the performance comparison with conventional methods. Med Biol Eng Comput. 2020;58(5):1047-68.
9. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. TensorFlow: large-scale machine learning on heterogeneous distributed systems V2 (Version 2). 2016, March 16. http://arxiv.org/abs/1603.04467.
10. Kluyver T, Ragan-Kelley B, Pérezc F, Granger B, Bussonnier M, Frederic J, et al. Jupyter Notebooks – a publishing format for reproducible computational workflows. In: Loizides F, Schmidt B, eds. Positioning and Power in Academic Publishing: Players, Agents and Agendas. Amsterdam, IOS Press; 2016:87-90. doi: 10.3233/978-1-61499-649-1-87.
11. Johnson J. Cnn-benchmarks. Accessed 13 March 2023. https://github.com/jcjohnson/cnn-benchmarks/;2023.
12. Wang S, Summers RM. Machine learning and radiology. Med Image Anal. 2012;16(5):933-51.
13. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE; 2016:2818-26.
14. Kingma DP, Ba J. Adam: a method for stochastic optimization. In: Bengio Y, LeCun Y, eds. Proceedings of the 3rd International Conference on Learning Representations. San Diego, CA: USA; 2015:1-15. https://arxiv.org/abs/1412.6980
15. Meena J, Hari P. Vesicoureteral reflux and recurrent urinary tract infections. Asian J. Pediatric Nephrol. 2019;2:61-70.
16. De Palma D. Radionuclide Tools in Clinical Management of Febrile UTI in Children. Semin Nucl Med. 2020;50(1):50-5.
17. Fazal MI, Patel ME, Tye J, Gupta Y. The past, present and future role of artificial intelligence in imaging. Eur J Radiol. 2018;105:246-50.
18. Batuello JT, Gamito EJ, Crawford ED, Han M, Partin AW, McLeod DG, et al. Artificial neural network model for the assessment of lymph node spread in patients with clinically localized prostate cancer. Urology. 2001;57(3):481-5.
19. Poulakis V, Witzsch U, de Vries R, Emmerlich V, Meves M, Altmannsberger HM, et al. Preoperative neural network using combined magnetic resonance imaging variables, prostate-specific antigen, and Gleason score for predicting prostate cancer biochemical recurrence after radical prostatectomy. Urology. 2004;64(6):1165-70.
20. Wu S, Chen X, Pan J, Dong W, Diao X, Zhang R, et al. An artificial intelligence system for the detection of bladder cancer via cystoscopy: a multicenter diagnostic study. J Natl Cancer Inst. 2022;114(2):220–7.
21. Fenstermaker M, Tomlins SA, Singh K, Wiens J, Morgan TM. Development and Validation of a Deep-learning Model to Assist with Renal Cell Carcinoma Histopathologic Interpretation. Urology. 2020;144:152-7.
22. 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.
23. Katz JE, Abdelrahman L, Nackeeran S, Ezeh U, Visser U, Deane LA. The Development of an Artificial Intelligence Model Based Solely on Computer Tomography Successfully Predicts Which Patients Will Pass Obstructing Ureteral Calculi. Urology. 2023;174:58-63.
24. Çelik Ö, Aslan AF, Osmanoğlu UÖ, Çetin N, Tokar B. Estimation of renal scarring in children with lower urinary tract dysfunction by utilizing resampling technique and machine learning algorithms. J Surg Med. 2020;4(7):573-7.
25. Lorenzo AJ, Rickard M, Braga LH, Guo Y, Oliveria JP. Predictive Analytics and Modeling Employing Machine Learning Technology: The Next Step in Data Sharing, Analysis, and Individualized Counseling Explored with a Large, Prospective Prenatal Hydronephrosis Database. Urology. 2019;123:204-9.
26. Serel A, Ozturk SA, Soyupek S, Serel HB. Deep Learning in Urological Images Using Convolutional Neural Networks: An Artificial Intelligence Study. Turk J Urol. 2022 Jul;48(4):299-302.
27. Serrano-Durba A, Serrano AJ, Magdalena JR, Martín JD, Soria E, Domínguez C, et al. The use of neural networks for predicting the result of endoscopic treatment for vesico-ureteric reflux. BJU Int. 2004;94:120–2.
28. Seckiner I, Seckiner SU, Erturhan S, Erbagci A, Solakhan M, Yagci F. The use of artificial neural networks in decision support in vesicoureteral reflux treatment. Urol Int. 2008;80(3):283-6. doi: 10.1159/000127342.
29. Khondker A, Kwong JCC, Yadav P, Chan JYH, Singh A, Skreta M, et al. Multi-institutional Validation of Improved Vesicoureteral Reflux Assessment With Simple and Machine Learning Approaches. J Urol. 2022 Dec;208(6):1314-22.
30. Khondker A, Kwong JCC, Rickard M, Skreta M, Keefe DT, Lorenzo AJ, et al. A machine learning-based approach for quantitative grading of vesicoureteral reflux from voiding cystourethrograms: Methods and proof of concept. J Pediatr Urol. 2022 Feb;18(1):78.e1-78.e7.