



Feasibility of a deep learning-based diagnostic platform to evaluate lower urinary tract disorders in men using simple uroflowmetry

Seokhwan Bang^{1,*†}, Sokhib Tukhtaev^{2,*}, Kwang Jin Ko¹, Deok Hyun Han¹, Minki Baek¹,
Hwang Gyun Jeon¹, Baek Hwan Cho², Kyu-Sung Lee¹

¹Department of Urology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, ²Medical AI Research Center, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea

Purpose: To diagnose lower urinary tract symptoms (LUTS) in a noninvasive manner, we created a prediction model for bladder outlet obstruction (BOO) and detrusor underactivity (DUA) using simple uroflowmetry. In this study, we used deep learning to analyze simple uroflowmetry.

Materials and Methods: We performed a retrospective review of 4,835 male patients aged ≥ 40 years who underwent a urodynamic study at a single center. We excluded patients with a disease or a history of surgery that could affect LUTS. A total of 1,792 patients were included in the study. We extracted a simple uroflowmetry graph automatically using the ABBYY Flexicapture® image capture program (ABBYY, Moscow, Russia). We applied a convolutional neural network (CNN), a deep learning method to predict DUA and BOO. A 5-fold cross-validation average value of the area under the receiver operating characteristic (AUROC) curve was chosen as an evaluation metric. When it comes to binary classification, this metric provides a richer measure of classification performance. Additionally, we provided the corresponding average precision-recall (PR) curves.

Results: Among the 1,792 patients, 482 (26.90%) had BOO, and 893 (49.83%) had DUA. The average AUROC scores of DUA and BOO, which were measured using 5-fold cross-validation, were 73.30% (mean average precision [mAP]=0.70) and 72.23% (mAP=0.45), respectively.

Conclusions: Our study suggests that it is possible to differentiate DUA from non-DUA and BOO from non-BOO using a simple uroflowmetry graph with a fine-tuned VGG16, which is a well-known CNN model.

Keywords: Artificial intelligence; Bladder outlet obstruction; Detrusor underactivity; Lower urinary tract symptoms

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received: 9 November, 2021 • **Revised:** 23 January, 2022 • **Accepted:** 24 February, 2022 • **Published online:** 25 March, 2022

Corresponding Author: Kyu-Sung Lee <https://orcid.org/0000-0003-0891-2488>

Department of Urology, Samsung Medical Center, Sungkyunkwan University School of Medicine, 81 Irwon-ro, Gangnam-gu, Seoul 06351, Korea
TEL: +82-2-3410-3554, FAX: +82-2-3410-3027, E-mail: ksleedr@skku.edu

Baek Hwan Cho <https://orcid.org/0000-0001-7722-5660>

Medical AI Research Center, Samsung Medical Center, Sungkyunkwan University School of Medicine, 81 Irwon-ro, Gangnam-gu, Seoul 06351, Korea
TEL: +82-2-3410-0885, FAX: +82-2-3410-0878, E-mail: baekhwan.cho@samsung.com

*These authors contributed equally to this study and should be considered co-first authors.

†Current affiliation: Department of Urology, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Korea.

INTRODUCTION

Lower urinary tract symptoms (LUTS) is a common disease with multifactorial causes. The most common cause of LUTS in men is benign prostate hyperplasia (BPH). Up to 50% of men over 50 years of age and 80% of men over 80 years of age experience LUTS caused by BPH [1]. Detrusor underactivity (DUA) is another very common cause of LUTS. One review found that between 9% and 28% of patients with LUTS under 50 years of age had DUA, while 48% of those over 70 years of age had DUA [2]. LUTS is a concept that includes voiding dysfunction and storage dysfunction, each feature represented by DUA and bladder outlet obstruction (BOO), respectively [3]. It is critical to distinguish between these two diseases because their treatments and clinical responses differ.

Urodynamic studies (UDSs) are the gold standard for the diagnosis and evaluation of LUTS. However, the use of UDS is limited by its invasiveness. Porru et al. [4] found that 4% to 45% of patients experience UDS complications, mostly urinary tract infection and hematuria. In addition, several patients report feeling shame and discomfort during the test and post-test anxiety [5].

Simple uroflowmetry, one component of UDS, is a simple, noninvasive diagnostic screening procedure used to calculate the flow rate of urine over time. Uroflowmetry produces a uroflowmetry graph that contains information regarding the voiding volume and maximum urine flow rate (Q_{max}) [6]. Several previous trials have attempted to categorize simple uroflowmetry graphs into several groups; however, there has been insufficient evidence and objective standards, including lack of pressure data, to achieve this end. There is a lack of evidence that uroflowmetry can distinguish obstructed voiding and DUA. However, as we have mentioned, this distinction is crucial in determining the appropriate treatment for LUTS.

Medical image analysis, which uses deep learning algorithms, has recently become more popular for developing technologies such as image recognition [7,8]. Many studies have used deep learning algorithms to classify and diagnose several diseases based on images [9]. For instance, convolutional neural networks (CNNs) have recently focused on optimizing technology for analyzing, patterning, and predicting trends. In 2012, the CNN proposed by Krizhevsky et al. [10] emphasized its high performance in image recognition at classification task. Since then, researchers in the medical domain have been exploiting deep learning algorithms for various tasks to fully or partially automate the disease diagnosis.

This study sought to develop a fully automated device to distinguish DUA and BOO using patterns of simple uroflowmetry with a deep learning method.

MATERIALS AND METHODS

1. Ethics statement

This study was performed at a single center and was conducted according to the tenets of the Declaration of Helsinki. The Institutional Review Board of Samsung Medical Center approved this study (approval number: 2019-12-062). Informed consent was waived by the Institutional Review Board of Samsung Medical Center (Seoul, Korea) because of the study's retrospective design.

2. Patients

We retrospectively reviewed the clinical data of 4,835 men who underwent a pressure-flow study at Samsung Medical Center between December 2006 and December 2017. We analyzed all patients who were ≥ 40 years of age and who underwent a pressure-flow study and focused on the pattern of uroflowmetry regardless of storage function. Those with diseases that can affect lower urinary tract function, bladder cancer, and prostate cancer were excluded. Patients who underwent previous prostate, bladder, and/or urethral surgeries and those with indwelling catheters (or needing regular catheterization) were also excluded. Patients with a history of cerebrovascular accident, neurologic disorders, and spinal or pelvic bone trauma that could affect LUTS were excluded. Patients who had voided volumes less than 150 mL during simple uroflowmetry were also excluded. Finally, we excluded 77 patients whose study graphs were insufficient for analysis. Therefore, 1,792 patients were ultimately included (Fig. 1).

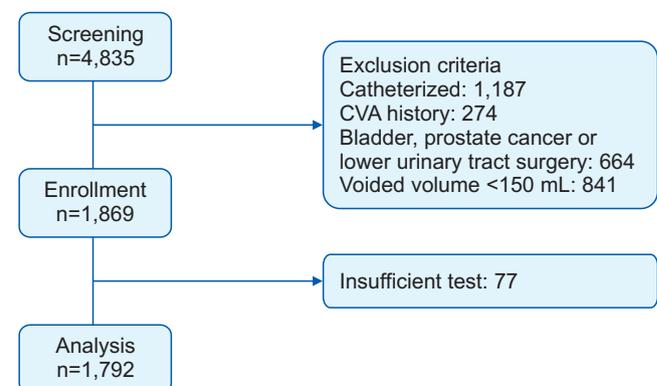


Fig. 1. Study design. CVA, cerebrovascular accident.

3. Urodynamic examination

The UDS were performed by experts according to the International Continence Society Good Urodynamics Practices protocol using an Aquarius TT UDS system (Laborie Medical Technologies, Toronto, ON, Canada) and a DORADO-KT (Laborie Medical Technologies) [11]. The UDS are recorded in four versions (7 Rel Z, 8 Rel A, 11 Rel 6, 12 Rel 0), each of which has a different output format.

DUA was defined as a bladder contractility index ($BCI = PdetQ_{max} + 5Q_{max}$) <100 [12]. BOO was defined as a BOO index ($BOOI = PdetQ_{max} - 2Q_{max}$) >40 [12].

4. Data pre-processing

The patients’ personal information and identification numbers were deleted according to the regulations. The uroflowmetry graph was extracted separately. The original graph was composed of data, and numerical information (and data that were not necessary for deep learning procedure). We separated the graph data using ABBYY Flexicapture® (ABBYY, Moscow, Russia), a program that permits the automated extraction of necessary parts from an image, except text. Using the ABBYY program, we extracted a uroflowmetry graph from the simple uroflowmetry test sheet (Fig. 2).

Deep learning models typically require a fixed image specification for training. Szegedy et al. [13] gained more accuracy with a 299×299 pixels input size, keeping the computational effort constant. Zoph et al. [14] used both 299×299

and 331×331 pixels for training ImageNet models. Similarly, we resized the resolution of all images to 299×299 pixels. Owing to the limited number of uroflowmetry graphs datasets, we performed a data augmentation technique for better classification performance of the trained models. The aim of data augmentation is to expand the size of a training dataset by generating modified images in the dataset. The nature of uroflowmetry graphs is greatly different from natural images such as dogs, cars, and pedestrian images. Thus, it is impractical to apply popular data augmentation techniques such as flipping and rotation because the spatial correlation of the uroflowmetry graph should be maintained. Therefore, we applied the cropping approach only as data augmentation, where we cropped the left and right top/bottom areas along with the central area that maintained approximately 90% of the original graph.

5. Deep neural network model implementation

We adopted ResNet-18 [15], Inception-V3 [16], and VGG16 [17] for the classification of the uroflowmetry images. After initializing with ImageNet-pretrained models, we extensively tuned hyper-parameters such as the learning rate, batch size, and activation functions in the training process. We trained DUA classification models and BOO classification models separately with the corresponding datasets.

To evaluate our models, 5-fold cross-validation was performed. Pre-processed images were randomly divided into

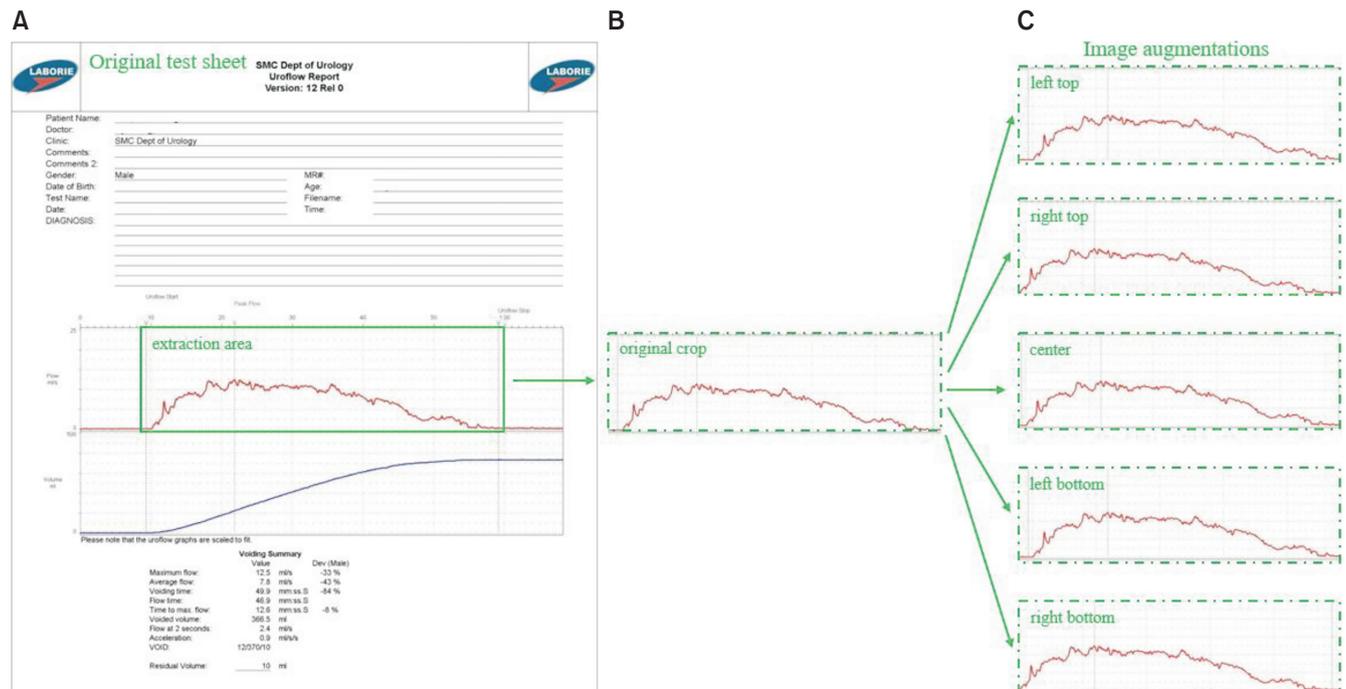


Fig. 2. An outline of uroflowmetry graph extraction and data augmentation pipeline The ABBYY program provides the extraction area from the original test sheet (A), then image augmentations (C) are made using the original crop (B).

five non-overlapping subsets: four subsets were used for training and one was left for validation. This process was repeated for all five subsets so that each subset was evaluated as a test set once. The results were averaged and recorded. The average value of area under the receiver operating characteristic (AUROC) curve derived from 5-fold cross-validation and accordant mean average precision (mAP) values for both DUA and BOO datasets were chosen as evaluation metrics [18].

Keras, a high-level Python API, was used as our deep learning platform, enabling fast experimentation. The networks were implemented in the Ubuntu 16.04 LTS environment, equipped with a 1080Ti GeForce GPU series.

ResNet-18 has been heavily involved in the deep learning community for the last half decade, allowing researchers to train deeper networks with the help of simply adding identity mappings to every few stacked layers. We chose ResNet-18 because it is light and suitable for our dataset at hand. Similarly, Inception-V3 is a CNN model that gained popularity in the deep learning community for its approach toward keeping the compute cost constant. Moreover, Inception-V3 is known to improve the training ability of a network through variations in properties. We employed Inception-V3 to determine whether it could capture low-level features of our uroflowmetry graphs. The last network we experimented with was VGG16, developed by the Visual Geometry Group of the University of Oxford. It presented a thoroughly evaluated network of increased depth, sticking to 3×3 convolutional filters. The model is relatively more straightforward than the ResNet and Inception counterparts and has achieved promising results in various tasks.

Therefore, we adopted VGG16 for the DUA and BOO datasets as well. Since VGG16 outperformed the former networks, we present detailed explanations of hyperparameter tunings of the VGG network alone. The model was optimized for DUA classification using a stochastic gradient descent with a learning rate of 0.003. Likewise, the hyperparameters of BOO classification were tuned as same as for DUA except for a learning rate of 0.01. The input size of 299×299 pixels showed better results compared to smaller analogs for both datasets.

6. Statistical analysis

Data analysis was performed using the Statistical Package for the Social Sciences (SPSS® Statistics version 25.0; SPSS Inc., IBM Corp., Chicago, IL, USA), and a Student's t-test was used to compare patient characteristics. Statistical significance was set at a p-value of <0.05.

RESULTS

As shown in Table 1, among the 1,792 patients, 482 (26.90%) had BOO, and 893 (49.83%) had DUA. There were significant differences between BOO and non-BOO patients in UDS parameters except time to voiding time. In DUA and non-DUA patients, there were significant differences in all the pressure-flow study parameters, except age and voiding volume.

As a result of deep learning evaluations, the mean 5-fold cross-validation AUROC metrics for DUA classification trained with ResNet-18 and Inception-V3 networks were 0.699 and 0.648, respectively. As mentioned, the best score of

Table 1. Baseline patient characteristics

Characteristic	BOO		p-value	DUA		p-value
	No (n=1,310)	Yes (n=482)		No (n=899)	Yes (n=893)	
Age, y	66.41	64.01	<0.001	64.39	64.93	0.229
BOOI	18.06	61.08	<0.001	33.01	26.22	<0.001
BCI	98.86	114.68	<0.001	127.66	78.38	<0.001
Voiding efficacy	86.35	77.78	<0.001	86.16	81.92	<0.001
Qmax, mL/s	13.95	9.99	0.001	14.67	11.09	0.001
Average flow, mL/s	6.38	4.58	<0.001	6.86	4.95	<0.001
Voiding time, s	66.13	72.83	0.022	54.38	81.58	<0.001
Flow time, s	50.00	57.28	<0.001	44.86	58.66	0.001
Time to peakflow, s	20.92	24.50	0.001	16.65	27.16	<0.001
Voided volume, mL	272.91	233.89	<0.001	262.04	262.79	0.881
Residual volume, mL	48.67	77.30	<0.001	45.82	66.99	<0.001

Values are presented as mean value only.

BOO, bladder outlet obstruction; DUA, detrusor underactivity; BOOI, bladder outlet obstruction index; BCI, bladder contractility index; Qmax, maximum urine flow rate.

Student t-test.

0.733 was obtained with a fine-tuned VGG16 network. The BOO classification trained with ResNet-18 and Inception-V3 networks were 0.661 and 0.560, respectively. The VGG16 network trained with the BOO dataset also achieved a higher

discrimination rate of 0.722 than ResNet-18 and Inception-V3. Figs. 3 and 4 show the ROC curves and PR curves of the VGG16 network for the DUA and BOO datasets, respectively. We also calculated the sensitivity and specificity

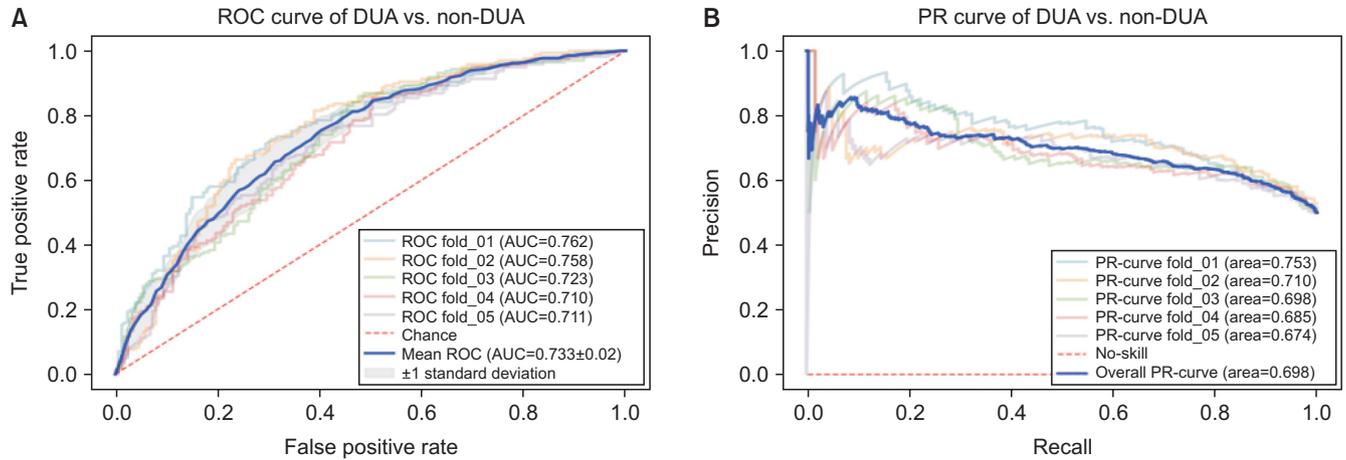


Fig. 3. The mean ROC curve (A) and the mean PR curve (B) of VGG16 network for DUA vs. non-DUA classification. ROC, receiver operating characteristic; DUA, detrusor underactivity; AUC, area under the curve; PR, precision-recall.

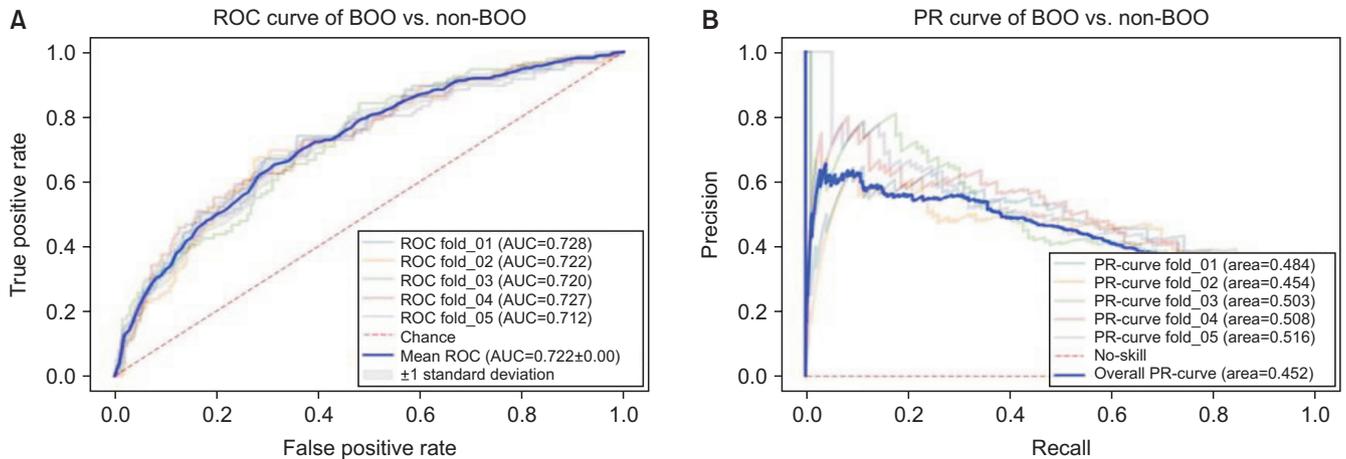


Fig. 4. The mean ROC curve (A) and the mean PR curve (B) of VGG16 network for BOO vs. non-BOO classification. ROC, receiver operating characteristic; BOO, bladder outlet obstruction; AUC, area under the curve; PR, precision-recall.

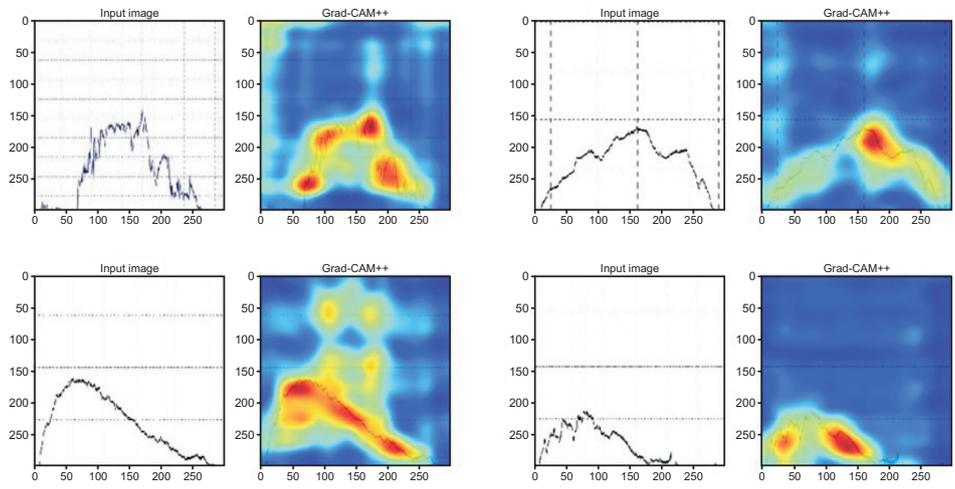


Fig. 5. Model explainability with Grad-CAM++. The first row presents samples from the VGG16 model trained with the DUA dataset while the second row depicts samples from the VGG16 model trained with the BOO dataset. BOO, bladder outlet obstruction.

values of the DUA and BOO models. The sensitivity and the specificity of VGG16 network for DUA dataset accounted for 65.9% and 68.9% at the maximum Youden's index [19]. The sensitivity and the specificity of VGG16 network for BOO dataset accounted for 65.1% and 68.9% at the maximum Youden's index. Furthermore, because a fine-tuned VGG16 was the winner among the three experimental models, we only depicted the visualizations of a GRAD-CAM++ [20]. Visual explanation techniques such as GRAD-CAM++ are used to produce rough localization mappings by highlighting important regions in the image. GRAD-CAM++ provides feature maps with respect to a specific class score to generate visual explanations. Fig. 5 illustrates some samples from uroflowmetry images and their respective mappings next to them. Evidently, GRAD-CAM++ activated the signal graphs compared to background regions. This implies that models learned to identify clinically proper regions in the images.

DISCUSSION

Since the introduction of simple uroflowmetry in 1948 [6], several attempts have been made to establish a pattern of analysis for this technique. Van de Beek et al. [21] attempted to classify and predict uroflowmetry. In this study, the group attempted to formalize uroflowmetry and identify diagnostic patterns among specialists. However, the predictive rate was only 36%. Gacci et al. [22] published a common flow pattern in 2007. They formulated uroflowmetric parameters and searched for the items of diagnostic suspicion of uroflowmetry curves. However, their agreement was not satisfactory, as it had a kappa value of 0.05. Moreover, the analysis was based on the lack of reproducibility and the characteristics of simple uroflowmetry, which vary greatly depending on the environment.

There have also been other attempts to predict or diagnose BOO. Bladder wall thickness (BWT) was predicted to be increased by BOO as one of the indicators that can be measured by ultrasound [23]. Manieri et al. [24] first discussed this possibility. Using 5 mm as the reference point and a significant difference ($r > 0.6$), this group found that 63% of the normal group had values < 5 mm, while 88% of patients with BOO had values > 5 mm. In contrast, Hakenberg et al. [25] found that the BWT increased slightly with age, but not significantly.

The penile cuff test was also applied to measure BOO. This test measures the detrusor contractility by detecting the iso-volumetric bladder pressure [26]. An inflatable cuff is placed around the penis shaft and expands automatically until the urine flow is interrupted. The next cuff then de-

flates rapidly to restart the flow. This cycle can be repeated until the urination ends. The pressure required to interrupt urinary flow during the cycle is considered to represent bladder pressure (Pcuff.int) [27]. However, this method has several limitations, including its high cost and the need for patients to be seated when they take the test. The seated nature of the test may introduce bias, as most men void while standing [28]. We attempted to mitigate these limitations using deep learning.

The prediction of BPH through AI has also been suggested by other researchers. Torshizi et al. [29] predicted severity of BPH based on fuzzy-ontology, and the accuracy was about 90%. However, the results of this study presented the severity based on the results obtained through questionnaire and clinical examination, and our study showed a big difference in that it looked at the possibility of diagnosis only by graph analysis. In addition, a non-invasive prediction of LUTS using ANN (artificial neural network) was also presented [30], but its accuracy did not satisfied expectations. In this study, we tried to overcome such limitations using CNN, and the prediction of DUA is the first attempt.

In this study, we proposed the use of a deep learning tool as a diagnostic alternative to invasive UDSs. To our knowledge, this is a novel approach. We believe that it can be used as the basis for the development of a tool to compensate for the defect of the UDS. This study sought to determine if one could use graph patterns to predict disease. We compared patients with and without DUA and those with and without BOO. We did not account for patients who may have both DUA and BOO. Given the large number of other patients with LUTS, the study attempted to identify these complex diseases. We used CNN to confirm the accuracy of predictions for patients with BOO and DUA using only a simple uroflowmetry graph. The raw signal data of the urodynamic test results graph was not provided from the urodynamic test device. Hence, an image capture software program, AB-BYY Flexicapture[®], was used to extract 1,792 data samples and there was no error case. Thus, we believe that the image capture process was robust. This research is meaningful in that it used a deep learning method to approach areas that have not been investigated using prototype trials. We consider that this is a meaningful work that will serve as a cornerstone for further research. We experimented with known algorithms offered for classification tasks such as ResNet-18, Resnet-50, Inception-V3, Efficientnet-B0, however, final predictions were not as good as VGG16's (data not shown). Besides, with our dataset, VGG19 attained the same result as its VGG16 variants, therefore we decided to select the lighter one. As this is a feasibility study of deep learn-

ing models on urodynamic test data, a further study with a larger dataset will be needed. Also, we will consider experimenting with recent models in our future study.

This study has several limitations. First, the prediction rate of this study is only slightly over 70%, which indicates that a higher prediction rate is required. Additionally, the mean AUC scores of fine-tuned VGG16 can be ameliorated by increasing the number of training images. Second, the capacity to set the basis for model predictions is confined due to the absence of external data. Although visual interpretations of GRAD-CAM++ in Fig. 5 provided some evidence that the model discriminated between the signal graph and gridlines in the background, the full interpretability needs to be addressed in future work. And third, this study is excluded patients who had both BOO and DUA. We included patients who had only BOO or only DUA. In further studies, it is needed to be include this complexed situation to developed useful device to diagnose BOO and DUA.

CONCLUSIONS

Our study suggests possibility of automated and non-invasive device to differentiate DUA from non-DUA and BOO from non-BOO using a simple uroflowmetry graph with a fine-tuned VGG16, which is a well-known CNN model.

CONFLICTS OF INTEREST

The authors have nothing to disclose.

FUNDING

This work was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean government (Ministry of Science and ICT) (No. 2017R1E1A1A01077487, 2020R1F1A1070952).

AUTHORS' CONTRIBUTIONS

Research conception and design: Seokhwan Bang and Kyu-Sung Lee. Data acquisition: Sokhib Tukhtaev and Baek Hwan Cho. Statistical analysis: Seokhwan Bang, Sokhib Tukhtaev, and Baek Hwan Cho. Data analysis and interpretation: Seokhwan Bang and Deok Hyun Han. Drafting of the manuscript: Seokhwan Bang and Sokhib Tukhtaev. Critical revision of the manuscript: Deok Hyun Han and Kyu-Sung Lee. Obtaining funding: Kyu-Sung Lee and Baek Hwan Cho. Administrative, technical, or material support: Minki Baek and Hwang Gyun Jeon. Supervision: Kwang Jin

Ko and Deok Hyun Han. Approval of the final manuscript: Kyu-Sung Lee.

REFERENCES

- Egan KB. The epidemiology of benign prostatic hyperplasia associated with lower urinary tract symptoms: prevalence and incident rates. *Urol Clin North Am* 2016;43:289-97.
- Osman NI, Esperto F, Chapple CR. Detrusor underactivity and the underactive bladder: a systematic review of preclinical and clinical studies. *Eur Urol* 2018;74:633-43.
- Han DH, Jeong YS, Choo MS, Lee KS. The efficacy of transurethral resection of the prostate in the patients with weak bladder contractility index. *Urology* 2008;71:657-61.
- Porru D, Madeddu G, Campus G, Montisci I, Scarpa RM, Usai E. Evaluation of morbidity of multi-channel pressure-flow studies. *Neurourol Urodyn* 1999;18:647-52.
- Yeung JY, Eschenbacher MA, Pauls RN. Pain and embarrassment associated with urodynamic testing in women. *Int Urogynecol J* 2014;25:645-50.
- Chancellor MB, Rivas DA, Mulholland SG, Drake WM Jr. The invention of the modern uroflowmeter by Willard M. Drake, Jr at Jefferson Medical College. *Urology* 1998;51:671-4.
- Shen D, Wu G, Suk HI. Deep learning in medical image analysis. *Annu Rev Biomed Eng* 2017;19:221-48.
- Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. *Med Image Anal* 2017;42:60-88.
- Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 2016;316:2402-10.
- Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst* 2012;25:1097.
- Schäfer W, Abrams P, Liao L, Mattiasson A, Pesce F, Spangberg A, et al. Good urodynamic practices: uroflowmetry, filling cystometry, and pressure-flow studies. *Neurourol Urodyn* 2002;21:261-74.
- Abrams P. Bladder outlet obstruction index, bladder contractility index and bladder voiding efficiency: three simple indices to define bladder voiding function. *BJU Int* 1999;84:14-5.
- Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. *ArXiv*. 1512.00567 [Preprint]. 2015 [cited 2021 Jun 23]. Available from: <https://arxiv.org/abs/1512.00567>.
- Zoph B, Vasudevan V, Shlens J, Le QV. Learning transferable architectures for scalable image recognition. *ArXiv*. 1707.07012 [Preprint]. 2018 [cited 2021 Jun 23]. Available from: <https://arxiv.org/abs/1707.07012>.

- arxiv.org/abs/1707.07012.
15. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *ArXiv*. 1512.03385 [Preprint]. 2015 [cited 2021 Aug 5]. Available from: <https://arxiv.org/abs/1512.03385>.
 16. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. *ArXiv*. 1409.4842 [Preprint]. 2014 [cited 2021 Aug 5]. Available from: <https://arxiv.org/abs/1409.4842>.
 17. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *ArXiv*. 1409.1556 [Preprint]. 2015 [cited 2021 Aug 5]. Available from: <https://arxiv.org/abs/1409.1556>.
 18. Hajian-Tilaki K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian J Intern Med* 2013;4:627-35.
 19. Youden WJ. Index for rating diagnostic tests. *Cancer* 1950;3:32-5.
 20. Chattopadhyay A, Sarkar A, Howlader P, Balasubramanian VN. Grad-CAM++: improved visual explanations for deep convolutional networks. *ArXiv*. 1710.11063 [Preprint]. 2018 [cited 2018 Nov 9]. Available from: <https://arxiv.org/abs/1710.11063>.
 21. Van de Beek C, Stoevelaar HJ, McDonnell J, Nijs HG, Casparie AF, Janknegt RA. Interpretation of uroflowmetry curves by urologists. *J Urol* 1997;157:164-8.
 22. Gacci M, Del Popolo G, Artibani W, Tubaro A, Palli D, Vittori G, et al. Visual assessment of uroflowmetry curves: description and interpretation by urodynamists. *World J Urol* 2007;25:333-7.
 23. Lee HN, Lee YS, Han DH, Lee KS. Change of ultrasound estimated bladder weight and bladder wall thickness after treatment of bladder outlet obstruction with dutasteride. *Low Urin Tract Symptoms* 2017;9:67-74.
 24. Manieri C, Carter SS, Romano G, Trucchi A, Valenti M, Tubaro A. The diagnosis of bladder outlet obstruction in men by ultrasound measurement of bladder wall thickness. *J Urol* 1998;159:761-5.
 25. Hakenberg OW, Linne C, Manseck A, Wirth MP. Bladder wall thickness in normal adults and men with mild lower urinary tract symptoms and benign prostatic enlargement. *Neurourol Urodyn* 2000;19:585-93.
 26. Van Mastrigt R, Pel JJ. Towards a noninvasive urodynamic diagnosis of infravesical obstruction. *BJU Int* 1999;84:195-203.
 27. Griffiths CJ, Rix D, MacDonald AM, Drinnan MJ, Pickard RS, Ramsden PD. Noninvasive measurement of bladder pressure by controlled inflation of a penile cuff. *J Urol* 2002;167:1344-7.
 28. Mangera A, Chapple C. Modern evaluation of lower urinary tract symptoms in 2014. *Curr Opin Urol* 2014;24:15-20.
 29. Torshizi AD, Zarandi MH, Torshizi GD, Eghbali K. A hybrid fuzzy-ontology based intelligent system to determine level of severity and treatment recommendation for Benign Prostatic Hyperplasia. *Comput Methods Programs Biomed* 2014;113:301-13.
 30. Sonke GS, Heskes T, Verbeek AL, de la Rosette JJ, Kiemeny LA. Prediction of bladder outlet obstruction in men with lower urinary tract symptoms using artificial neural networks. *J Urol* 2000;163:300-5.