Development of an Automatic Interpretation Algorithm for Uroflowmetry Results: Application of Artificial Intelligence

Min Soo Choo[1], Ho Young Ryu[2], Sangchul Lee[3]

[1]Department of Urology, Seoul Metropolitan Government - Seoul National University Boramae Medical Center, Seoul, Korea
[2]Department of Urology, Ewha Womens University Medical Center, Seoul, Korea
[3]Department of Urology, Seoul National University Bundang Hospital, Seongnam, Korea

Purpose: To develop an automatic interpretation system for uroflowmetry (UFM) results using machine learning (ML), a form of artificial intelligence (AI).

Methods: A prospectively collected 1,574 UFM results (1,031 males, 543 females) with voided volume > 150 mL was labelled as normal, borderline, or abnormal by 3 urologists. If the 3 experts disagreed, the majority decision was accepted. Abnormality was defined as a condition in which a urologist judges from the UFM results that further evaluation is required and that the patient should visit a urology clinic. To develop the optimal automatic interpretation system, we applied 4 ML algorithms and 2 deep learning (DL) algorithms. ML models were trained with all UFM parameters. DL models were trained to digitally analyze 2-dimensional images of UFM curves.

Results: The automatic interpretation algorithm achieved a maximum accuracy of 88.9% in males and 90.8% in females when using 6 parameters: voided volume, maximum flow rate, time to maximal flow rate, average flow rate, flow time, and voiding time. In females, the DL models showed a dramatic improvement in accuracy over the other models, reaching 95.4% accuracy in the convolutional neural network model. The performance of the DL models in clinical discrimination was outstanding in both genders, with an area under the curve of up to 0.957 in males and 0.974 in females.

Conclusions: We developed an automatic interpretation algorithm for UFM results by training AI models using 6 key parameters and the shape of the curve; this algorithm agreed closely with the decisions of urology specialists.

Keywords: Machine learning; Deep learning; Uroflowmetry; Automatic interpretation; Medical decision making

Corresponding author: Sangchul Lee ⓘD https://orcid.org/0000-0003-0844-6843
Department of Urology, Seoul National University Bundang Hospital, 82 Gumi-ro 173beon-gil, Bundang-gu, Seongnam 13620, Korea
Email: uromedi@naver.com

INTRODUCTION

Uroflowmetry (UFM) is a simple test that measures the urine stream in volume per unit time [1]. The main advantage of UFM is that it is non-invasive and relatively inexpensive [2,3]. Therefore, it is considered an indispensable, first-line screening method for most patients with suspected lower urinary tract dysfunction [2]. Most commercially available office uroflowmeters are based on weight transducers, which measure the voided volume (VV) and calculate the flow rate by detecting the difference over time [4]. These flowmeters provide both a graphical presentation of the uroflow and a range of electronically read parameters [4].

To obtain representative results in UFM, adequate privacy should be provided, and patients should be asked to void when they feel a "normal" desire to do so [2]. However, UFM is usually carried out on an outpatient basis, in specified procedure areas without basic privacy, and often involves having the person urinate into the uroflowmeter at a predetermined time [5]. This process is unnatural and requires "on-demand" voiding often with either low or very high bladder filling, which compromises the results [5]. It has therefore been recommended that UFM should be repeated, which is time-consuming and costly for both patients and health care providers [6].

Sound-based UFM represents a new approach to recording urinary flow patterns and measuring urinary flow parameters in a non-invasive manner by analysing the sound generated by a stream of urine striking the water surface in the toilet bowl. We developed a novel mobile acoustic UFM that works as a microphone built into a smartphone. In a previous study, researchers developed a device called sonouroflow with a similar concept [7]. At the technical level, sonouroflow processed each voiding session as a whole and analyzed only the sound pressure level. However, our acoustic UFM method analyzes a session in detail by dividing it into hundreds of sections. In addition, our method estimates variables and analyzes data by applying various signal processing methods in the time domain and frequency domain as well as the sound pressure level. Clinical trials confirmed that our device was non-inferior in performance to a conventional UFM. We released the developed acoustic UFM program as an application through the App Store and Google Play.

Patients can use our acoustic UFM application for free with only a smartphone. Using this program, they can perform UFM tests in any location (at home, at work, on vacation, or anywhere else) at any time just as comfortably as they would void in daily life. However, even if users obtain accurate and representative measurements, it is difficult to judge whether the results are normal or whether they reflect abnormal conditions that require urological management [8]. To identify abnormal results and recommend that those users visit a urologic clinic, we aimed to develop an automatic interpretation system for UFM results by applying machine learning (ML) and deep learning (DL), 2 subsets of artificial intelligence (AI).

MATERIALS AND METHODS

Patients
With the approval of the Seoul National University Bundang Hospital Institutional Review Board (IRB No.: B-1912-583-001), 3,000 patients over 20 years of age who were scheduled to undergo UFM in the outpatient urology clinic based on clinical judgement were prospectively included in this study from January to December 2019. Before being included as subjects, patients agreed to join this study of their own volition and completed an written informed consent form. This study was conducted in accordance with the ethical principles stated in the Declaration of Helsinki. All patients with incomplete UFM results were excluded. Patients with a UFM volume of less than 150 mL were also excluded.

Using a web-based reading tool, 3 urologists read the UFM measurements independently. The participants included a senior urologist with more than 10 years of clinical experience, a urologist with more than 5 years of clinical experience, and a junior urologist with less than 5 years of experience. The 3 independent researchers classified each result as normal, borderline, or abnormal by visually inspecting the pattern of the flow curve and evaluating the relevant quantitative parameters of UFM as defined by the International Continence Society (ICS): voiding time (VT), flow time (FT), time to maximum flow rate (TQmax), maximum flow rate (Qmax), average flow rate (Qavg), and VV.

Abnormality was defined as a condition in which a urologist judged from the UFM results that further evaluation was required and that the patient should visit a urologic clinic.

ML algorithms
To develop the optimal automatic interpretation system, we applied 4 ML algorithms (logistic regression, decision tree, support vector machine, and random forest algorithms) and 2 DL

algorithms (a convolutional neural network [CNN] and a recurrent neural network [RNN]).

ML models were trained with all parameters of UFM results. DL models were trained to digitally analyze 2-dimensional (2D) images of the UFM curve. DL modelling was performed by converting the 2D image of the UFM curve into the time-series value of the instantaneous flow rate at time "t."
Through supervised ML algorithms, the UFM results of a randomly selected 80% of cases per gender were selected as a training set, and algorithms were developed to classify them into 3 groups: normal, borderline, and abnormal. The developed algorithms were validated externally with a test set consisting of the remaining 20% of results to evaluate the consistency and discrimination of the model.

Statistics

The evaluation variables are represented by descriptive statistics. The interobserver consistency of the investigators' readings was assessed in terms of the interclass correlation coefficients by calculating Cronbach alpha. Additionally, in order to determine the extent of agreement between the investigators (intraobserver agreement), Cohen kappa values were calculated [9]. A scatter plot matrix was used to visualize relationships between pairs of variables in a grid format. Each scatter plot shows the correlation between 2 variables. In addition, the kernel density estimation curve for each variable was drawn, and different colours were displayed for each group to provide additional information.

The consistency between the clinical decision and the interpretation by the ML algorithm was calculated as the accuracy, defined as the percentage of correct interpretations out of the total number of results. The area under the receiver operating characteristic curve was used to assess the discrimination performance of the model as a summary performance measure [10].

RESULTS

A total of 3,741 UFM cases were screened, and 1,269 tests were excluded according to exclusion criteria. After excluding 894 cases with a VV of less than 150 mL, we ultimately analyze 1,574 cases (1,031 in males and 543 in females).

The mean ages of the male and female patients were 66.5 ± 10.5 years and 63.6 ± 12.1 years, respectively. The UFM results of male cases were labelled normal in 521 cases (50.5%) and ab-

normal in 232 cases (22.4%), with unanimous decisions in 51.4% of cases. For female cases, 420 (77.3%) were normal and 60 (11.0%) were abnormal, with a 70.5% unanimity rate.

The internal consistency of the UFM readings was high (Cronbach alpha 0.88 [0.87–0.89] in males and 0.85 [0.83–0.86] in females). Moderate interobserver agreement was reached with regard to the normalcy of the UFM curve, with a kappa value of 0.43–0.55 in male cases and 0.39-0.56 in female cases.

Regarding the correlation between continuous variables as observed from the scatter plot matrix, VV and Qavg were positively correlated, and a negative correlation of Qmax with VT or FT was observed (Fig. 1). In addition, in the scatter plots with different colours between groups, it was possible to observe a clear distribution difference between the normal and abnormal groups.

For ML, 824 male and 338 female cases (80% of the data) were used as the training set, and 207 male and 82 female cases (the remaining 20% of the data) were used as the test set. When ML with logistic regression was performed with only one feature, 57.0%–83.8% accuracy was achieved. The index with the highest accuracy as a single variable was the Qmax value. When 2 features were used, 71.1%–85.2% accuracy was achieved. The variables that showed the best accuracy in a 2-feature model were Qmax and VV. When the number of features was increased one by one from 4 (VV, Qmax, TQmax, and Qavg) to 7 (VV, Qmax, TQmax, Qavg, VT, FT, and DT), the accuracy plateaued, increasing only from 86.5% to 88.9%. The interpretation algorithm showed the best accuracy when using a combination of 6 parameters: VV, Qmax, TQmax, Qavg, VT, and FT (Fig. 2).

In male cases, the interpretation accuracy of the ML models was 87.4%–88.9%, with the random forest model showing the highest accuracy. In the DL models, which were trained using the shape of the UFM curve, the accuracy slightly increased, reaching 0.918 for the CNN model and 0.908 for the RNN model. In female cases, the interpretation accuracy of the ML models was 87.2%–90.8%, with the random forest and logistic regression models tied for the highest accuracy. Interestingly, in the female cases, the DL models showed a dramatic improvement in accuracy over the other models, with the CNN model achieving 95.4% and the RNN model achieving 94.5% accuracy. The performance in clinical discrimination was outstanding in both genders, with a maximum area under the curve of 0.957 in males and 0.974 in females (Fig. 3).
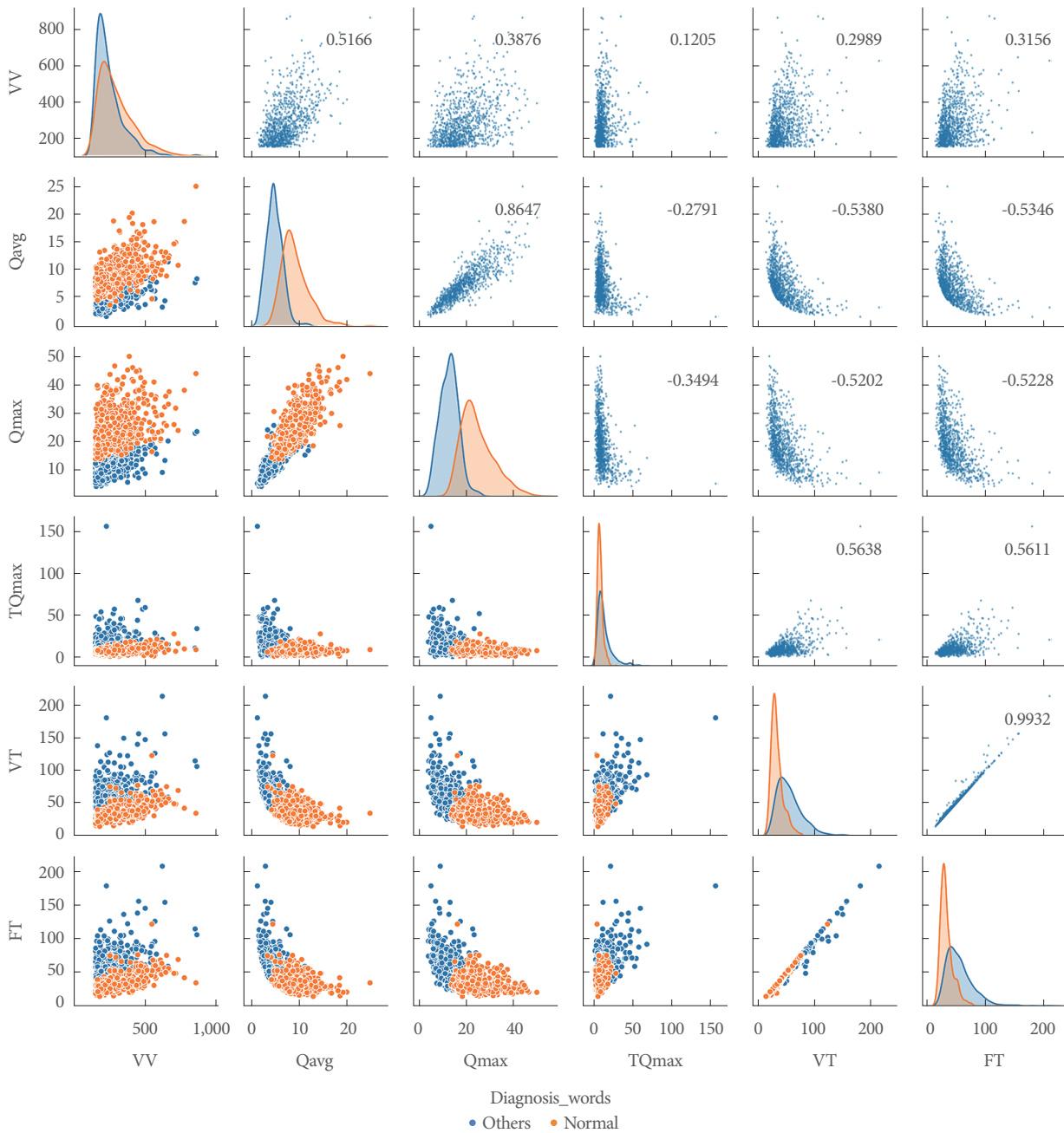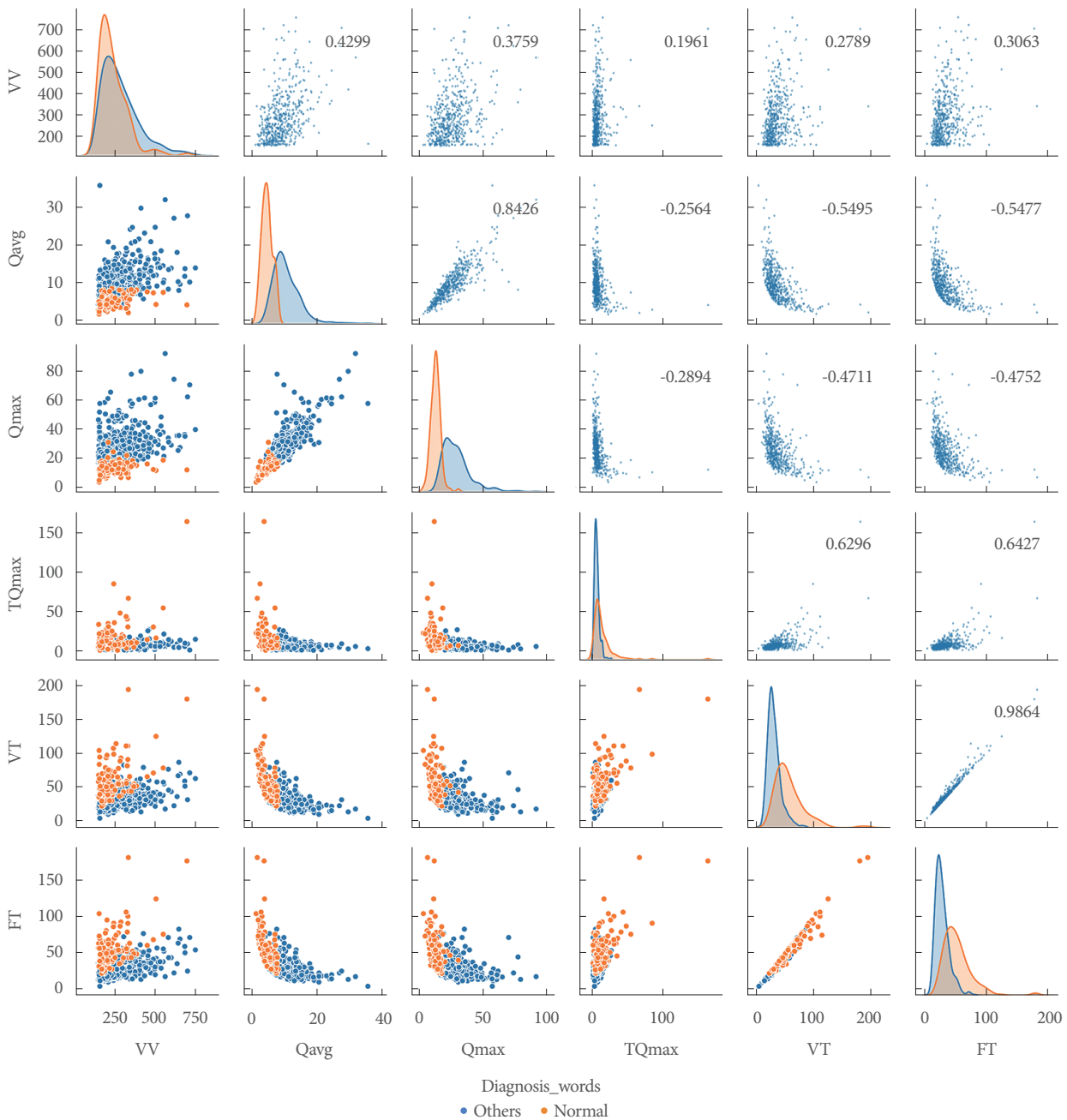
Fig. 1. Scatter plot matrix of continuous variables of uroflowmetry with the kernel density estimation curve. (A) Male patients. (B) Female patients. VV, voided volume; Qavg, average flow rate; Qmax, maximum flow rate; TQmax, time to maximum flow rate; VT, voiding time; FT, flow time.

(*continued*)

DISCUSSION

We developed an AI algorithm that automatically interprets the results of UFM using 6 key parameters and the shape of the curve; we confirmed that the results generated by this algo-rithm were in very close agreement with the decisions of urology specialists. There was no significant difference in consistency according to the ML method in male cases, but in female cases, the accuracy increased dramatically when DL models were added to recognize the shape of the UFM curve. To the best of

Fig. 1. Scatter plot matrix of continuous variables of uroflowmetry with the kernel density estimation curve. (A) Male patients. (B) Female patients. VV, voided volume; Qavg, average flow rate; Qmax, maximum flow rate; TQmax, time to maximum flow rate; VT, voiding time; FT, flow time. (continued)

our knowledge, this is the first study to develop an automatic UFM reading algorithm using AI models.

For female cases, there was a dramatic increase in accuracy when DL models were added to recognize the 2D images of the UFM curve. However, when male cases were analyzed using DL models, there was no significant change in accuracy. One

plausible explanation is that the 6 numerical parameters extracted from the UFM curve are sufficient to reflect the patient's condition in male cases; on the other hand, in female patients, the parameters of UFM may be insufficient, and the shape of the curve must be referenced when reading UFM results. Few studies on the diagnostic application of UFM are available in
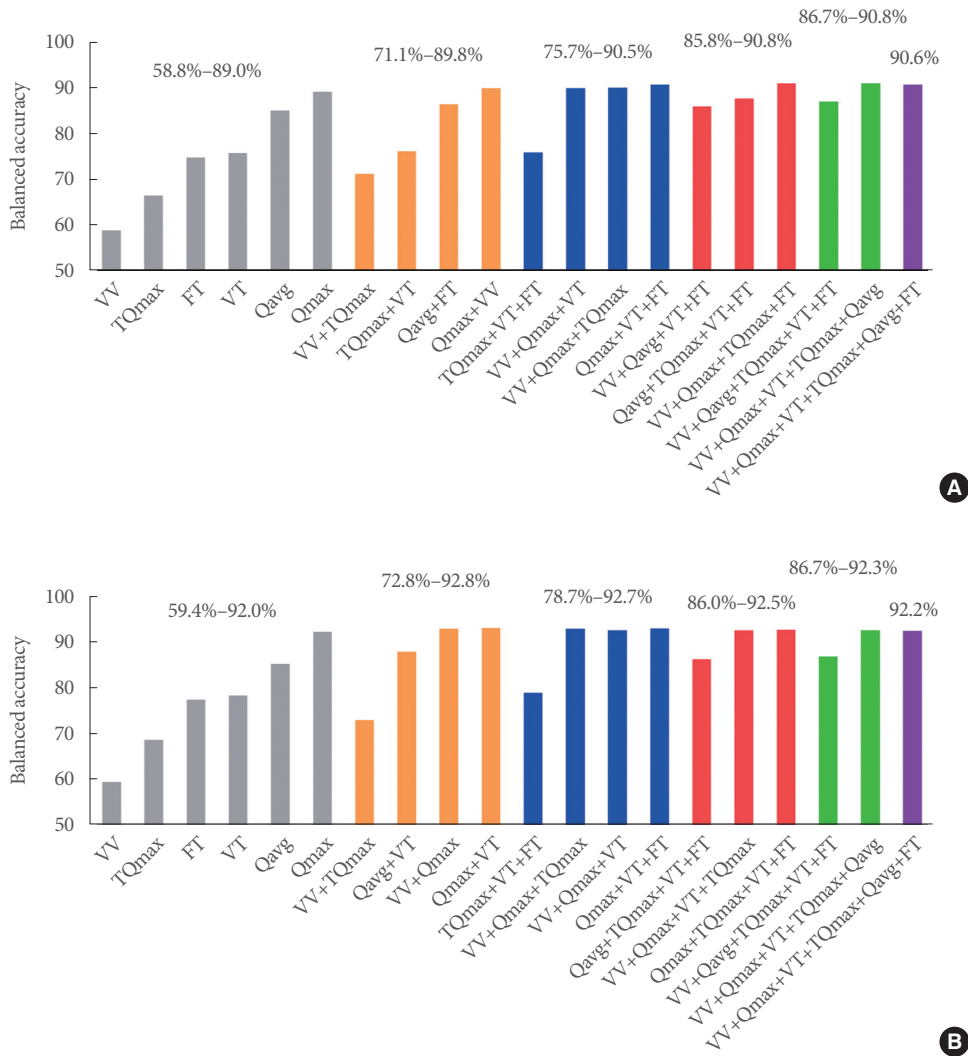
Fig. 2. Change in accuracy of an automatic interpretation system trained with logistic regression methods for uroflowmetry results as the number of available parameters increases. (A) Male patients. (B) Female patients. VV, voided volume; Qavg, average flow rate; Qmax, maximum flow rate; TQmax, time to maximum flow rate; VT, voiding time; FT, flow time.

women, and there is no clarity regarding reference values, their variations, and which factors influence these values [11]. Particular caution is recommended when interpreting the UFM results of female patients [12]. Further research may be needed to determine whether there are better numerical parameters and/or any unknown features that can be used for the interpretation of female UFM results.

Several studies have examined the limitations of clinic-based UFM [13,14]. The difficulty of providing a space with adequate privacy to relax and the demand for the patient to void without the normal desire to void are unrepresentative of daily voiding patterns, and it is also not feasible to repeat measurements in the clinic due to time constraints [15]. A potential solution to

this phenomenon of "bashful bladder" under forced conditions is for the patient to measure his or her own urinary flow at home [15]. Several home UFM techniques have been introduced, such as timing methods, funnel devices, and electronic devices [16-18]. However, these techniques also do not provide a complete alternative due to the economic barrier imposed by the high cost of electronic devices, as well as the possibility of inaccuracy when patients calculate the values manually. For these reasons, the best option to date is our sound-based UFM system that works with a smartphone.

Our mobile app-based acoustic UFM is an easy-to-use, noninvasive method to estimate a patient's urodynamics simply by recording sounds with a smartphone during voiding. A novel
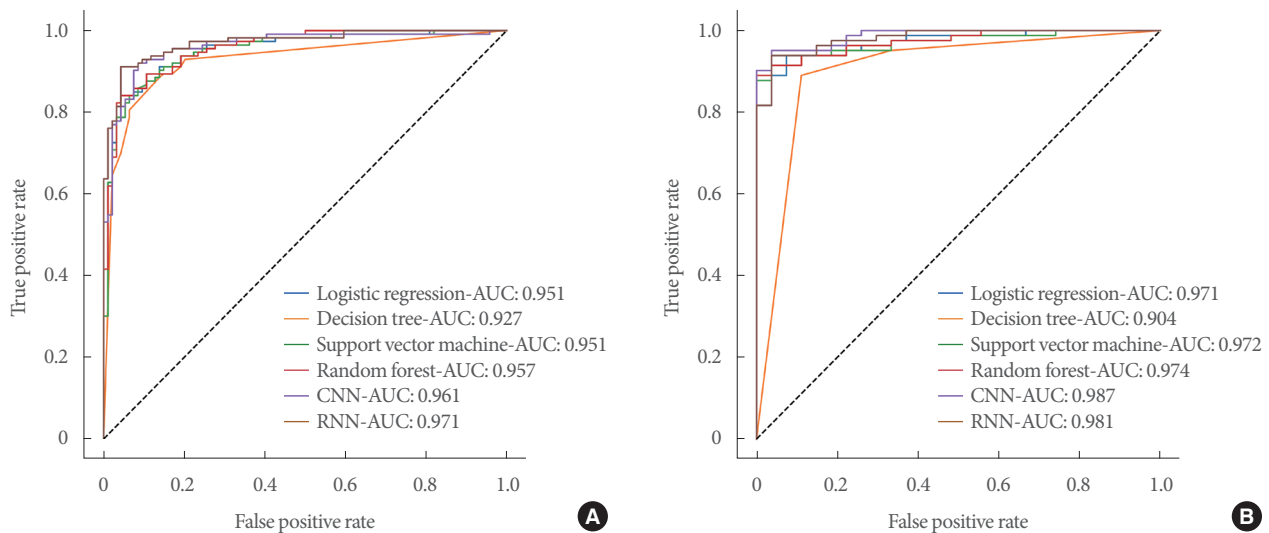
Fig. 3. Analysis of area under the receiver operating characteristic curve (AUC) to assess the performance of automatic interpretation systems according to machine learning methods. (A) Male patients. (B) Female patients. CNN, convolutional neural network; RNN, recurrent neural network.

acoustic AI engine is applied to suppress sound artifacts, offset environmental characteristics, and improve prediction accuracy. This acoustic UFM system was built from 35,000 sessions of voiding data from 4,700 people in various real acoustic environments. It has already been validated clinically and is listed by the U.S. Food and Drug Administration as a uroflowmeter and medical device data system. Our mobile acoustic UFM system can be used to check and monitor the rate and volume of urinary flow in daily, natural settings. It can also track longitudinal trends and includes an automatic voiding diary for daily usage. This smartphone application might improve the shortcomings of current voiding diaries, such as incomplete records with missing values and low compliance [19]. In this study, we developed an automatic reading algorithm using ML to carry out mobile acoustic UFM.

The common disadvantage of all home UFM techniques, including the sound-based UFM we developed, is the lack of measurement of postvoid residual (PVR). Although the ICS recommends reporting PVR in UFM results, the key parameters of UFM are Qmax, VV, and flow pattern [2,20]. However, the most relevant parameter related to bladder outlet obstruction is Qmax [21]. Rather than having low clinical significance, PVR may be considered an independent parameter not included in UFM, hence the term "UFM with PVR." If PVR is included, it can greatly increase the potential clinical impact, but UFM itself is sufficient as a screening test. An abnormal result from home UFM would presumably lead to a subsequent clinic

visit, where stand-alone measurement of PVR could be obtained after normal voiding [15].

In this study, the interobserver reliability of UFM readings by the 3 researchers was relatively low. Other studies have also reported on the variability of interpretation [22]. In particular, the degree of agreement in the diagnosis of bladder outlet obstruction is very poor, at K=0.20 [23]. Urologists' degree of experience may be the most important factor in these differences [24]. Since there are no absolute values defining normal limits, the interpretation of UFM results must be subjective and empirical [25,26]. For this reason, it is believed that the only way to automatically read UFM results is to use the opinions of clinical experts as a reference and attempt to replicate them. Although the degree of consensus among clinical experts' opinions is not high, the agreement between AI readings and majority expert readings was excellent, at 95%. Urologists label UFM results mainly by reviewing the shape of UFM curves, but the numerical values of the UFM parameters actually represent the meaning of the shape quite well.

In conclusions, we developed an AI system that applies 4 ML and 2 DL algorithms and automatically interprets the results of UFM using 6 key parameters and the shape of the curve, in this study. We confirmed that the agreement between the automated readings and the judgement of urology specialists was very high. In females, the accuracy of the readings increased dramatically when DL models were added to recognize the shape of the UFM curve. Further research may be needed to deter-

mine whether there are currently unrecognized parameters of the shape of the UFM curve that would improve the interpretation of UFM results.

AUTHOR CONTRIBUTION STATEMENT

· Conceptualization: *SL*
· Data curation: *MSC, HYR, SL*
· Formal analysis: *MSC, HYR, SL*
· Funding acquisition: *SL*
· Methodology: *MSC, HYR*
· Project administration: *SL*
· Visualization: *MSC, HYR, SL*
· Writing-original draft: *MSC*
· Writing-review & editing: *MSC, SL*

ORCID

Min Soo Choo   0000-0002-7852-5822
Sangchul Lee   0000-0003-0844-6843

REFERENCES

1. Gammie A, Drake MJ. The fundamentals of uroflowmetry practice, based on International Continence Society good urodynamic practices recommendations. Neurourol Urodyn 2018;37(S6):S44-9.

2. 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.

3. Lee KS, Koo KC. Clinical factors associated with the feeling of incomplete bladder emptying in women with little postvoided residue. Int Neurourol J 2020;24:172-9.

4. Jørgensen JB, Jensen KM. Uroflowmetry. Urol Clin North Am 1996;23:237-42.

5. Krhut J, Gärtner M, Sýkora R, Hurtík P, Burda M, Luňáček L, et al. Comparison between uroflowmetry and sonouroflowmetry in recording of urinary flow in healthy men. Int J Urol 2015;22:761-5.

6. Gärtner M, Krhut J, Hurtik P, Burda M, Zvarova K, Zvara P. Evaluation of voiding parameters in healthy women using sound analysis. Low Urin Tract Symptoms 2018;10:12-6.

7. Zvarova K, Ursiny M, Giebink T, Liang K, Blaivas JG, Zvara P. Recording urinary flow and lower urinary tract symptoms using sonouroflowmetry. Can J Urol 2011;18:5689-94.

8. Lee YJ, Kim MM, Song SH, Lee S. A novel mobile acoustic uroflowmetry: comparison with contemporary uroflowmetry. Int Neurourol J 2021;25:150-6.

9. Costantini E, Mearini E, Pajoncini C, Biscotto S, Bini V, Porena M. Uroflowmetry in female voiding disturbances. Neurourol Urodyn 2003;22:569-73.

10. Liu YB, Yang SS, Hsieh CH, Lin CD, Chang SJ. Inter-observer, intra-observer and intra-individual reliability of uroflowmetry tests in aged men: a generalizability theory approach. Low Urin Tract Symptoms 2014;6:76-80.

11. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-74.

12. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. J Thorac Oncol 2010;5:1315-6.

13. Sorel MR, Reitsma HJB, Rosier PFWM, Bosch RJLHR, de Kort LMO. Uroflowmetry in healthy women: a systematic review. Neurourol Urodyn 2017;36:953-9.

14. Sand PK, Ostergard DR. Uroflowmetry in the female. In: Sand PK, Ostergard DR, editors. Urodynamics and the evaluation of female incontinence: a practical guide. London: Springer; 1995. p. 6-10.

15. Boci R, Fall M, Waldén M, Knutson T, Dahlstrand C. Home uroflowmetry: improved accuracy in outflow assessment. Neurourol Urodyn 1999;18:25-32.

16. Porru D, Scarpa RM, Prezioso D, Bertaccini A, Rizzi CA. Home and office uroflowmetry for evaluation of LUTS from benign prostatic enlargement. Prostate Cancer Prostatic Dis 2005;8:45-9.

17. Bray A, Griffiths C, Drinnan M, Pickard R. Methods and value of home uroflowmetry in the assessment of men with lower urinary tract symptoms: a literature review. Neurourol Urodyn 2012;31:7-12.

18. Hansen MV, Zdanowski A. The use of a simple home flow test as a quality indicator for male patients treated for lower urinary tract symptoms suggestive of bladder outlet obstruction. Eur Urol 1997;32:34-8.

19. Jiang YH, Chen SF, Kuo HC. Frontiers in the clinical applications of botulinum toxin A as treatment for neurogenic lower urinary tract dysfunction. Int Neurourol J 2020;24:301-12.

20. Pel JJ, van Mastrigt R. Development of a low-cost flow meter to grade the maximum flow rate. Neurourol Urodyn 2002;21:48-54.

21. Sonke GS, Robertson C, Verbeek AL, Witjes WP, de la Rosette JJ, Kiemeney LA. A method for estimating within-patient variability in maximal urinary flow rate adjusted for voided volume. Urology 2002;59:368-72.

22. Gratzke C, Bachmann A, Descazeaud A, Drake MJ, Madersbacher S, Mamoulakis C, et al. EAU guidelines on the assessment of non-neurogenic male lower urinary tract symptoms including benign prostatic obstruction. Eur Urol 2015;67:1099-109.

23. Oelke M, Höfner K, Jonas U, de la Rosette JJ, Ubbink DT, Wijkstra H. Diagnostic accuracy of noninvasive tests to evaluate bladder outlet obstruction in men: detrusor wall thickness, uroflowmetry, postvoid residual urine, and prostate volume. Eur Urol 2007;52: 827-34.

24. Jørgensen JB, Mortensen T, Hummelmose T, Sjørslev J. Mechanical versus visual evaluation of urinary flow curves and patterns. Urol Int 1993;51:15-8.

25. 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.

26. Grino PB, Bruskewitz R, Blaivas JG, Siroky MB, Andersen JT, Cook T, et al. Maximum urinary flow rate by uroflowmetry: automatic or visual interpretation. J Urol 1993;149:339-41.