Pediatric Urology

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

Armando J. Lorenzo, Mandy Rickard, Luis H. Braga, Yanbo Guo, and John-Paul Oliveria

OBJECTIVE	To explore the potential value of utilizing a commercially available cloud-based machine learning platform to predict surgical intervention in infants with prenatal hydronephrosis (HN).
MATERIALS AND METHODS	A prospective prenatal HN database was uploaded into Microsoft Azure Machine Learning Studio. Probabilistic principal component analysis was employed for data imputation. Multiple clinical variables were included in two-class decision jungle and neural network for model training, using surgical intervention as the primary outcome. Models were scored and evaluated after a 70/30 split of the data.
RESULTS	A total of 557 entries were included. The optimized model (decision jungle) achieved an area under the curve of 0.9, accuracy of 0.87, and precision of 0.80, employing a threshold of 0.5 to predict surgery. Average time to train, score and evaluate the model was 5 seconds. The predictive model was deployed as a web service in 35 seconds, generating a unique API key for app and webpage development. Individualized prediction based on the included variables was deployed as a web-based and batch execution Excel file in less than one minute.
CONCLUSION	This cloud-based ML technology allows easy building, deployment, and sharing of predictive analytics solutions. Using prenatal HN as an example, we propose an opportunity to address contemporary challenges with data analysis, reporting a creative solution that moves beyond the current standard. UROLOGY 123: 204−209, 2019. © 2018 Elsevier Inc.

The utilization of "big data," artificial intelligence (AI), machine learning (ML), and predictive analytics has been exploited for several years by many lucrative business endeavors to individualize and maximize their reach to potential consumers, monetizing based on the rich profiling generated by these vast amounts of data. Germane to what has likely been experienced by most readers, this includes applying predictive algorithms to internet browsing and social media activity in order to target specific advertisements to individuals.[1] Cloud computing and analytical services such as IBM's Watson, Amazon Web Service (AWS), Google's TensorFlow and Microsoft's Azure Machine Learning Studio—among others—have the potential to impact a variety of sectors by providing easy access to sophisticated cloud based ML capabilities to a larger breadth of users.

There is promising potential in modern healthcare to harness the "big data" movement by taking advantage of the large quantities of information generated on a daily basis in various settings. Currently, this is driving a shift in clinical practice from mere episodic analysis of separate datasets towards algorithms relying on regularly updated information in order to better predict patient outcomes.[2] The scope and depth of applications is vast. ML algorithms have already been employed to predict infants at risk for cardiac arrest,[3] identify individuals mostly likely to participate in clinical trials[4] as well as to monitor social media to predict flu outbreaks.[5]

Prenatal hydronephrosis (HN) is reported in up to 5% of pregnancies and is a common referral to pediatric urology clinics.[6] The expected trajectory of infants affected with prenatal HN is not always immediately clear and may range from spontaneous resolution to surgical intervention. While clinicians are often armed with evidence and data when speaking to the families of these infants, it is not always possible to predict which infants will progress to surgery, which will develop urinary tract infection (UTI), or which will spontaneously improve at baseline or early visits.

There is clear benefit to harnessing the power of AI and ML algorithms to assist with risk stratification of infants with prenatal HN. With an accurate, continuously updated dataset based on prospectively captured clinical variables and outcomes, it may be possible to prognosticate in real time which infants may require surgery, which might develop UTIs, and which may improve with no intervention based on early or baseline clinical variables. To this effect, herein, we explore the value of using a commercially available predictive analytical ML platform from Microsoft (Azure) to predict surgical intervention in infants with prenatal HN. We hypothesized that the likelihood of surgical intervention in these infants could be accurately and efficiently predicted using ML algorithms.

METHODS

Setting, Population, and Study Inclusion/Exclusion Criteria

Following ethics board approval, we reviewed our prospectively collected prenatal HN database from 2008 to 2016 at a single tertiary pediatric referral center (n = 661). Patients with underlying uropathies such as ureteroceles, ectopic ureters, neurogenic bladder, and posterior urethral valves (PUVs) were excluded, for a total of 557 children with HN. Clinical information was collected longitudinally, at each clinic visit, with the frequency of follow-up determined by clinician preference. The de-identified dataset was then uploaded into Microsoft Azure Machine Learning Studio. Probabilistic principal component analysis[7] was employed for data imputation.

Independent Variables and Outcomes of Interest

Data were extracted on the following variables: age, gender, affected side, HN grade (Society for Fetal Urology [SFU][8]), differential renal function (DRF), diuretic $t^1/_2$ time, degree of ureteral dilatation, and renal pelvis anteroposterior diameter (APd) measurements, which were collected at baseline and at each follow-up visit. Surgical interventions and UTI development were also captured. For infants with bilateral prenatal HN, measurements from the most severe kidney were employed.

All clinical variables were included in predictive models using surgical intervention as the primary outcome. Indications for surgery included worsening of HN on repeated ultrasounds (US), deterioration of differential renal function >5% between two renal scans, and/or prolonged (>20min) drainage time on serial renograms, as well as the development of symptoms including febrile UTI or calculi.

Machine Learning Models

Following two ML models were trained: a two-class boosted decision tree model and a two-class neural network model to predict which infants were most likely to undergo a surgical intervention. The employed decision tree learning and artificial neural networks are fundamental and commonly used approaches to ML. For the present study, Microsoft's Azure Machine Learning Studio was selected due to its wide accessibility and ease of application. Detailed information is available online: (https://docs.microsoft.com/en-ca/azure/).

Statistical Analyses

Descriptive statistics, including frequencies, means (±standard deviation), and medians (min−max) were calculated as appropriate. Patients' characteristics and outcomes were compared between groups using parametric tests. Statistical analyses were conducted using SPSS v. 22.0 (SPSS 22.0, SPSS Inc., Chicago, Illinois). All tests were two-sided, and P-values of <.05 were considered statistically significant. The models were generated within the Microsoft Azure Machine Learning Studio. The dataset was initially divided with a randomized 70/30 split into a training and testing dataset. The two trained models were then scored using the test dataset and evaluated in real time. The results of the optimized model were reported as a receiver operator characteristics (ROC) curve, accuracy, and precision.

RESULTS

Of 557 patients, 105 (19%) had surgery, with 17 patients undergoing more than one procedure, and 19% developed a UTI. Surgical procedures performed are listed in Table 1. The mean age at presentation was 3.8 ± 4 months with a mean follow-up time of 25 ± 20 months. Males comprised 78% of the study population, 36% of which were circumcised. The majority of patients had UPJO-like HN (66%) and 55% had high-grade HN. Patient baseline characteristics including HN etiologies are presented in Table 2.

After categorizing patients into two groups, such as those managed nonsurgically and those who underwent surgery, we noted no significant difference in gender, with 80% of the nonsurgical group comprised of males vs 69% males in the surgical groups

Table 1. Surgical interventions

Surgical Procedure	n = 122 (%)
Pyeloplasty	83 (68)
Ureteral reimplantation	18 (15)
Distal cutaneous ureterostomy	4 (3)
Side to side refluxing ureterocystostomy	5 (4)
Nephrectomy	2 (2)
Endoscopic injection with Deflux®	4 (3)
Circumcision	6 (5)

17 patients had more than one procedure.

Table 2. Patient characteristcs

Characteristic	n = 557 (%)
Age (months)	3.8 ± 4
Gender	
- Male	435 (78)
Circumcised	156 (36)
Etiology	
- UPJO-like	367 (66)
- UPJO+VUR	11 (2)
- Megaureter	81 (15)
- VUR+PHN	91 (16)
- UPJO+UVJO	7 (1)
SFU grades	
- Low (I/II)	254 (46)
- High (III/IV)	303 (54)
APd (mm)	11 ± 7
DRF (%)	48 ± 12
T $^1/_2$ (min)	31 ± 54
UTI	107 (19)
Surgery	105 (19)
Maximum follow-up (months)	25 ± 20

(P = .09). We did note an expected difference in HN grades between the nonsurgical (47% high grade) and surgical groups (88% high grade) (P < .01) as well as APd (10± 5 mm vs 19± 11 mm; P < .01), DRF (49 ± 12% vs. 46 ± 12%; P = .02) and $t^1/_2$ time (19 ± 22 min vs 51 ± 79 min; P < .01) for nonsurgical and surgical patients, respectively.

The optimized model achieved an area under the curve of 0.9, accuracy of 0.87, and precision of 0.80, employing a threshold of 0.5 to predict need for surgery (Fig. 1). Average run time to train, score, and evaluate the model was 5 seconds. The predictive model was deployed as a web service in 35 seconds, generating a unique API key for app and webpage development. Individualized prediction based on the included variables was deployed as a web-based and batch execution Excel file in less than one minute. Updating the model with new data was achieved within these timeframes.

DISCUSSION

We have demonstrated that it is possible to accurately train a model for prediction of surgical intervention using a prenatal HN dataset and a readily available cloud-based predictive analytical platform. While predicting outcomes for patients based on datasets is not a new concept, traditionally constructing and validating these predictive models require time-consuming calculations and sophisticated analytical strategies to be carried out by properly trained individuals. Furthermore, by the time these calculations were complete, and results were available and ready for presentation, the static dataset could already be outdated. The advantage of cloud based ML platforms is the ability to generate these models in a reproducible fashion in a user-friendly environment. Furthermore, it allows for continuous input of new data, permitting real-time constant model retraining, thus strengthening predictions based on new clinical information. The model can then provide an output to influence clinical decision making within seconds. This is in stark contrast to traditional nomograms, which can be cumbersome and more difficult to update.[9]

Utilizing ML and AI technologies in healthcare is increasingly being explored by multiple specialties to predict various patient outcomes and is becoming a realistic
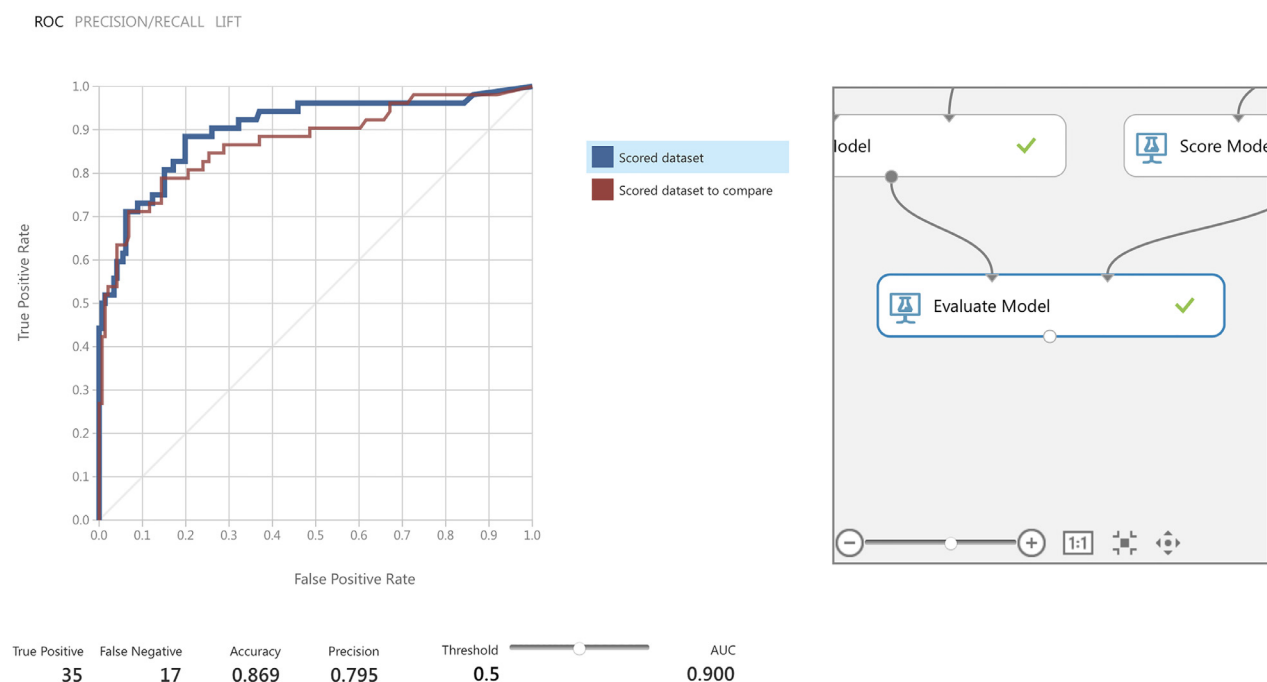
Figure 1. Azure output. (Color version available online.)

option as most institutions are transitioning to electronic medical records (EMRs). For example, on a hospital or department specific level, ML algorithms have been employed to accurately identify which individuals are most likely to agree to participate in clinical trials in an emergency department (ED) based on routinely captured demographic information[4]. Researchers in a neonatal intensive care unit were able to predict infants most at risk for cardiac arrest by AI interpretation of cardiorespiratory monitoring outputs up to 24 hours earlier than traditional monitoring would allow.[3] By taking advantage of the large amounts of readily available information generated during patient-physician encounters, there is potential to truly revolutionize and individualize healthcare.

In a study by Hao et al, ML algorithms were trained to accurately predict which ED patients were most at risk for re-visitation within 30 days of discharge based on a large dataset generated by EMR.[10] In another study by Cooper et al, a large national EMR generated dataset was used to preoperatively predict children at risk for surgical morbidity by comparing five statistical models.[11] While there are obvious advantages to vast datasets such as those generated by EMR and billing codes, one drawback is that there are gaps in the data, causing incomplete fields that may impact that validity of the results. The advantage of our dataset and the present study is that all data were obtained prospectively, in a single payer universal healthcare system using objective clinical variables. Therefore, while our dataset may be considered relatively "small" compared to enormous administrative ones, it is also more clean and homogenous.

ML and AI are also being utilized within our own pediatric urology subspecialty. Maximizing the data obtained from US imaging has been widely studied in pediatric urology as it may help stratify risk and help determine which children may benefit from invasive testing. Attempting to objectively assess US images has been previously explored and used to determine which children with PUV were at risk for developing renal failure.[12,13] Predictions of which patients might progress to surgery[14] as well as determination of surgical success by assessment of hydronephrotic kidneys have also been examined.[15−18] Logvinenko et al, attempted to determine if specific findings on renal US images could be used to predict subsequent abnormalities on VCUG using ML algorithms.[19] They reported that despite implementation of sophisticated prediction models, US was a poor screening test for VCUG abnormalities. In another study by Cerrolaza et al, renal US features in hydronephrotic kidneys were assessed to predict the need for diuretic renography.[20] It was reported that with the described semi-automated method analyzing US images, diuretic renograms could be decreased by 62%. While these findings highlight the potential value in exploring US image data, the process described in this study would be time consuming to carry out as segmentation of kidneys and delineation of areas of interest were done manually. Undoubtedly, future efforts at automatizing image analysis could add predictive value to models, harmonizing clinical information with in-depth evaluation of readily available radiological studies.

Although recent data would challenge the need for sophisticated analyses to evaluate outcomes (such as improvement after pyeloplasty), the value of using objective parameters for investigation and reporting is undeniable.[17]

Maximizing US image data poses an interesting area of study for the future. No matter how objective clinicians attempt to be when reviewing US images, subjective interpretation in unavoidable. Many people are familiar with the AI platform called Watson from IBM, its expanding application in many fields including health care, and its ability to "interpret" images and accurately diagnose tumors in oncology patients.[21] Hu et al developed an automatic method of image recognition of abdominal organs on CT scans that can effectively identify, localize, and segment four abdominal organs in under 2 minutes, which could aid clinical care and surgical planning.[22] Esteva et al described successfully utilizing convolutional deep neural network technology to classify and diagnose melanoma.[23] They reported that their model identified malignant melanoma as accurately as board certified dermatologists, which has the potential to significantly decrease morbidity and mortality in those patients, especially when considering expanding this technology to mobile devices. Taking advantage of existing technology and applying it to automatic interpretation of renal US imaging to assess hydronephrosis severity and parenchymal characteristics appears to be within reach. Doing so would remove the subjective interpretation and address most of the intrinsic limitations in the aforementioned studies. In addition, with standardized, automatic image assessment, subtle characteristics and nuances not previously recognized by clinicians could be combined with clinical variables which may strengthen outcome predictions and help to find associations between imaging modalities that have not been possible as of yet.

While the present study adds to the growing body of literature on ML, it is not without its limitations. We acknowledge that our current dataset is relatively small when compared to other larger administrative datasets usually employed for predictive analytics. However, as previously stated, despite its size, our dataset is clean, complete, and prospectively collected in a clinical setting with relevant variables and outcomes and does not rely on hospital or billing codes. Combining our dataset with the data from other similar institutions will allow for testing of our model and the possibility of more robust results and should be explored in the future. Indeed, another advantage of the employed technology is its ability to seamlessly enter data from different sources, repeating analyses in a matter of minutes. We also acknowledge that using surgical intervention as our outcome may be viewed as a limitation, as surgical indications may vary from one institution to the next and from surgeons to surgeons in the same hospital. However, our indications for surgical interventions as described above are arguably similar to those described at other centers, are based on objective clinical and diagnostic findings, and the included patients are rigorously monitored and undergo surgery for well-defined indications.

Despite the many possible shortcomings, we believe there is value in the present report. Herein, we have demonstrated a reproducible, accurate, fast method of appraising a prenatal HN dataset, which is applicable in everyday clinical use, with deployment of a key that can be introduced in webpages or applications, thus offering a predictive tool that could be employed and validated worldwide. The same techniques we have described could be used to predict other important outcomes (such as UTI risk and likelihood or timing of spontaneous resolution) in this population of infants, which is the matter of ongoing research efforts at our centers.

CONCLUSION

This powerful, cloud-based, ML technology allows easy building, deployment, and sharing of predictive analytics solutions. Using PHN as an example, we propose an opportunity to address current challenges with data analysis, with a creative solution that moves beyond the current standard, allowing for the creation and updating of predictive models based on large amounts of information.

References

1. Ghiassi M, Zimbra D, Lee S. Targeted twitter sentiment analysis for brands using supervised feature engineering and the dynamic architecture for artificial neural networks. *J Manag Inf Syst.* 2016;33:1034–1058.
2. Obermeyer Z EE. Predicting the future-big data, machine learning and clinical medicine. *N Engl J Med.* 2016;375:1216–1219.
3. Blount M, Ebling MR, Eklund JM, et al. Real-time analysis for intensive care: development and deployment of the artemis analytic system. *IEEE Eng Med Biol Mag.* 2010;29:110–118.
4. Ni Y, Beck AF, Taylor R, et al. Will they participate? Predicting patients' response to clinical trial invitations in a pediatric emergency department. *J Am Med Inform Assoc.* 2016;23:671–680.
5. Darcy A, Louie AK, Roberts LW. Machine learning and the profession of medicine; JAMA. 2016;315:551–552. https://jamanetwork.com/journals/jama/fullarticle/2488315
6. Arora S, Yadav P, Kumar M, et al. Predictors for the need of surgery in antenatally detected hydronephrosis due to UPJ obstruction − A prospective multivariate analysis. *J Pediatr Urol.* 2015;11.248.e1-5.
7. Tipping Michael BC. Probabilistic principal component analysis. *J R Stat Soc.* 1999;21:611–622.
8. Fernbach SK, Maizels M, Conway JJ. Ultrasound grading of hydronephrosis: Introduction to the system used by the society for fetal urology. *Pediatr Radiol.* 1993;23:478–480.
9. Vickers AJ, Kent M, Scardino PT. implementation of dynamically updated prediction models at the point of care at a major cancer center: Making nomograms more like Netflix. *Urology.* 2017;102:1–3.
10. Hao S, Jin B, Shin AY, et al. Risk prediction of emergency department revisit 30 days post discharge: A prospective study. *PLoS One.* 2014;9 e112944.
11. Cooper JN, Wei L, Fernandez SA, Minneci PC, Deans KJ. Pre-operative prediction of surgical morbidity in children: Comparison of five statistical models. *Comput Biol Med.* 2015;57:54–65.
12. Odeh R, Noone D, Bowlin PR, Braga LHP, Lorenzo AJ. Predicting risk of chronic kidney disease in infants and young children at diagnosis of posterior urethral valves: Initial ultrasound kidney characteristics and validation of parenchymal area as forecasters of renal reserve. *J Urol.* 2016;196:862–868.
13. Fischer K, Li C, Wang H, Song Y, Furth S, Tasian GE. Perinatal urology renal parenchymal area growth curves for children 0 to 10 months old. *J Urol.* 2016;195:1203–1208.
14. Rickard M, Lorenzo AJ, Braga LH. Renal parenchyma to hydronephrosis area ratio (PHAR) as a predictor of future surgical intervention for infants with high-grade prenatal hydronephrosis. *Urology.* 2017;101:85–89.
15. Kern AJM, Schlomer BJ, Timberlake MD, Peters CA, Hammer MR, Jacobs MA. Simple visual review of pre- to post-operative renal ultrasound images predicts pyeloplasty success equally as well as geometric measurements: A blinded comparison with a gold standard. *J Pediatr Urol.* 2017;14:401.e1–401.e7.
16. Rickard M, Braga LH, Oliveria J-P, Romao R, Demaria J, Lorenzo AJ. Percent improvement in renal pelvis antero-posterior diameter (PI-APD): Prospective validation and further exploration of cut-off values that predict success after pediatric pyeloplasty supporting safe monitoring with ultrasound alone. *J Pediatr Urol.* 2016;12:228.e1–228.e6.
17. Rickard M, Lorenzo AJ, Braga LH, Munoz C. Parenchyma-to-hydronephrosis area ratio is a promising outcome measure to quantify upper tract changes in infants with high-grade prenatal hydronephrosis. *Urology.* 2016;104:166–171.
18. Babu R, Sai V. Pelvis/cortex ratio: An early marker of success following pyeloplasty in children. *J Pediatr Urol.* 2010;6:473–476.
19. Logvinenko T, Chow JS, Nelson CP. Predictive value of specific ultrasound findings when used as a screening test for abnormalities on VCUG. *J Pediatr Urol.* 2015;11:176.e1–176.e7.
20. Cerrolaza JJ, Peters CA, Martin AD, Myers E, Safdar N, Linguraru MG. Quantitative ultrasound for measuring obstructive severity in children with hydronephrosis. *J Urol.* 2016;195:1093–1099.
21. Murdoch TB, Detsky AS. The inevitable application of big data to health care. *JAMA.* 2013;309:1351–1352.
22. Hu P, Wu F, Peng J, Bao Y, Chen F, Kong D. Automatic abdominal multi-organ segmentation using deep convolutional neural network and time-implicit level sets. *Int J Comput Assist Radiol Surg.* 2017;12:399–411.
23. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature.* 2017;542:115–118.

EDITORIAL COMMENT

The authors report a thought-provoking, preliminary exploration of the application of artificial intelligence (AI) and machine learning (ML) to predict the need for surgical intervention in children with prenatally-detected hydronephrosis.

The growing potential for AI applications in medicine is a reality. Nonetheless, a careful assessment of the AI literature in healthcare will reveal an almost exclusive focus on diagnostics; in other words, AI and ML are usually employed to advance interpretation, improve accuracy and outreach of diagnostic tests. The examples cited by the authors in the "discussion" section illustrate this point well.

With that in mind, the main question that arises is: what is the role of AI in treatment-related decisions, particularly those involving surgery? As the authors correctly mention, the indication for surgery in patients with prenatal hydronephrosis and possible ureteropelvic junction obstruction (UPJO) is not clear-cut and significant controversy still exists. In that way, single-center validation of ML as presented might pose a problem; as long as there is minimal consistency amongst practitioners working as part of the same group, indications for surgery "learned" by the model may prove to be faulty when applied to other institutions that follow a different set of criteria to operate. In this setting, a step where multicentric validation is performed is clearly needed.

The burden to generate good-quality evidence and establish best practices in the management of patients with prenatally-detected hydronephrosis is on our community of pediatric urologists; through careful study design and robust analysis procedures, we will hopefully have the ability to identify the variables that will populate future generalizable AI tools. This article is a great first step in that direction and the authors should be congratulated for it.

Rodrigo Romao, MD MSc, Departments of Surgery and Urology, IWK Health Centre, Dalhousie University

UROLOGY 123: 208−209, 2019. © 2018 Elsevier Inc.

AUTHOR REPLY

We sincerely appreciate the thoughtful comment accompanying our manuscript. We wholeheartedly agree that this is an exciting time for medicine, with sophisticated technology becoming more and more user friendly as well as increasingly accessible.

As highlighted, choosing an outcome such as surgical intervention may be perceived as surgeon or institution dependent. Nevertheless, we feel that our surgical indications, especially for pyeloplasty, are similar to many other pediatric institutions. We acknowledge that surgical indications may vary, as can be seen with related conditions such as uretero-vesical junction obstruction or vesicoureteral reflux. It is unlikely that advanced analytics will be able to standardize the multifactorial decision to proceed with surgery in the near future. It can, however, risk stratify patients while taking into account variability between providers or institutions. As models are perfected and databases grown, so will our ability to harness this technology to improve patient care.

We are grateful for the opportunity to publish this experience as one of the first in our sub-specialty. We hope that this information will spark interest in this technology and actively engage other in exploring further, introducing more sophisticated tools and fostering collaboration between experts in computer science and the growing artificial intelligence field with clinical experts in pediatric urology and radiology. While this technology can never replace informed clinical decision-making by experienced clinicians, we hope to be able to enhance patient, family and health care provider experience by fine tuning and training user-friendly models capable of consuming and quickly analyzing vast amounts of valuable clinical data to ultimately assist real-time decisions in every day practice.

Armando J. Lorenzo, Mandy Rickard, MN, NP, Division of Pediatric Urology, Department of Surgery, Hospital for Sick Children and University of Toronto, Toronto, Ontario, Canada

UROLOGY 123: 209, 2019. © 2018 Elsevier Inc.