Automated prognosis of prenatal hydronephrosis

Lauren Erdman[1,2,3,4]*†, Marta Skreta[1,2,4]†, Mandy Rickard[5], Carson McLean[1,2,3], Aziz Mezlini[1,2], Anne-Sophie Blais[5], Michael Brudno[1,2,3,4], Anna Goldenberg[1,2,3,6□], Armando J Lorenzo[5□]

Affiliations:

[1] University of Toronto, Department of Computer Science, Toronto, Canada

[2] Hospital for Sick Children, Program in Genetics and Genome Biology, Toronto, Canada

[3] Vector Institute, Toronto, Canada

[4] Center for Computational Medicine, SickKids Hospital, Toronto, Canada

[5] Hospital for Sick Children, Department of Surgery, Division of Urology, Toronto, Canada

[6] Canadian Institute for Advanced Research (CIFAR) Child and Brain Development

† Co-first author

□ Co-senior author

Background: Prenatal hydronephrosis (PHN) is a term used to describe dilatation of the urinary tract and occurs in approximately 1-5% children while in utero[1]. PHN comprises 15-30% of patients seen in paediatric urology clinics. The majority of children will experience spontaneous resolution, but this can take up to 36 months or longer to occur[2,3]. Approximately 25-30% of children diagnosed with PHN will require a surgical intervention to manage an underlying pathology; however, it is challenging to distinguish between surgical and non-surgical cases early on. As a result, all children with PHN are investigated with serial ultrasound imaging and in some cases, invasive testing with ionizing radiation and radioisotope exposure. Still, only a fraction of these children receive surgery. To reduce the uncertainty surrounding PHN prognosis, Lorenzo et al were able to predict surgical prognosis with an AUC=0.9 from differential renal function, diuretic t1/2 time, and other clinical characteristics. However, this model relied on invasive testing, feature engineering, and expert annotation, precluding its integration into clinical practice[4]. These analyses indicate that PHN prognosis is predictable, but the degree to which it can be predicted directly from ultrasound images alone and without curated features or invasive testing remains untested.

Objective: Our goal is to automate the surgical prognosis of patients with PHN based on ultrasound kidney images. We are the first group to model this outcome without hand-curated features.

Methods: Ultrasound DICOM images were cropped to remove text annotations and ultrasound beam borders and their contrast was normalized using histogram equalization. The images were resized to 256x256 pixels. Individuals in our data set were divided into a training/validation across 5 folds and each fold was trained independently (n patients=276, n kidneys=1710 ; note that each patient has multiple ultrasound images acquired across repeated visits). A held-out test set (n patients=69, n kidneys=211) was generated with the most recent patients in the test set in order to evaluate our algorithm's performance on hypothetical future patients[5]. Our predictive model was a CNN with the task of predicting surgical prognosis from one or two views (sagittal/transverse) from kidney ultrasound images[6,7]. To consider information from one view, we tested a CNN with 7 convolutional layers and 2 linear layers. We then investigated the predictive gain from using information from 2 views using a Siamese CNN architecture. We also measured the impact of transfer-learning, comparing three auxiliary tasks to initialize our networks and then fine-tune the networks for surgical prognosis: MNIST classification, Optical Coherence Tomography (OCT) classification, and an ultrasound patching approach[8,9]. All models were optimized using stochastic gradient descent; we used a learning rate of 0.001 and a momentum of 0.9, as well as early-stopping to prevent overfitting.

Results: Our model predicts whether a given kidney will receive surgery with an AUC of 0.923 (95% CI, 0.887-0.955) and an AUPRC of 0.753 (95% CI, 0.692-0.821) on a held-out test set when considering both sagittal and transverse views. We use Gradient-weighted Class Activation Maps (Grad-CAMs) to show our classifier's focus on the degree of kidney and ureter dilatation, as well as the parenchyma (Figure 1)[10]. The AUC drops marginally when training on only one view (sagittal AUC/AUPRC: 0.915/0.766, transverse AUC/AUPRC: 0.889/0.763), but this difference is not significant. This finding is useful in practice: if a clinician doesn't have access to a given view, the model can still make a prediction with high accuracy. We find that transferring weights from pretraining tasks (PT) did not improve performance, and in some cases, it was marginally worse. This demonstrates the importance of selecting an appropriate pretraining task and that pretraining is not always useful in a medical context, a finding that has also been made by other groups[11].

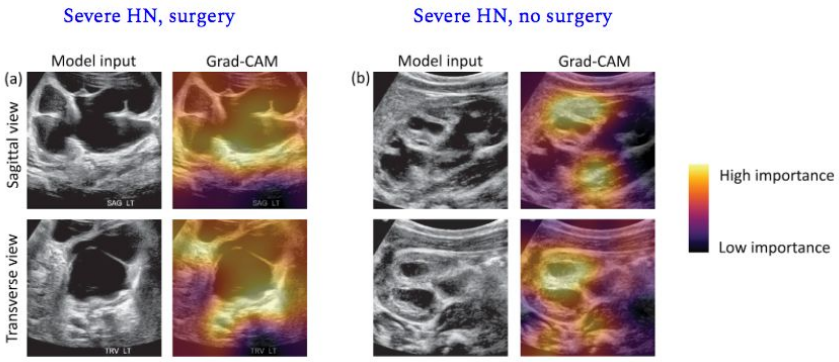

Figure 1. (a) Grad-CAM showing a kidney with severe PHN, correctly classified as requiring surgery. (b) Grad-CAM showing a kidney with severe PHN, correctly classified as not requiring surgery.

Implications for improving quality of care: In this work we developed a classifier which accurately discriminates between surgical and non-surgical PHN cases based on 2 ultrasound images of the kidney without any feature-engineering. Predicting surgical prognosis in patients with PHN will allow clinicians to reduce patient exposures to invasive testing and the number of follow-up visits, particularly in cases where the PHN will resolve on its own (Figure 2). Because this decision has many clinical components, including clinical features is likely to enhance the current predictive power of our classifier. Future work will include these clinical features, increase the size of our test group to include more positive cases (currently 8.5% in test vs 14.5% in training), and use a non-symmetric loss function to upweight positive cases in our classifier.

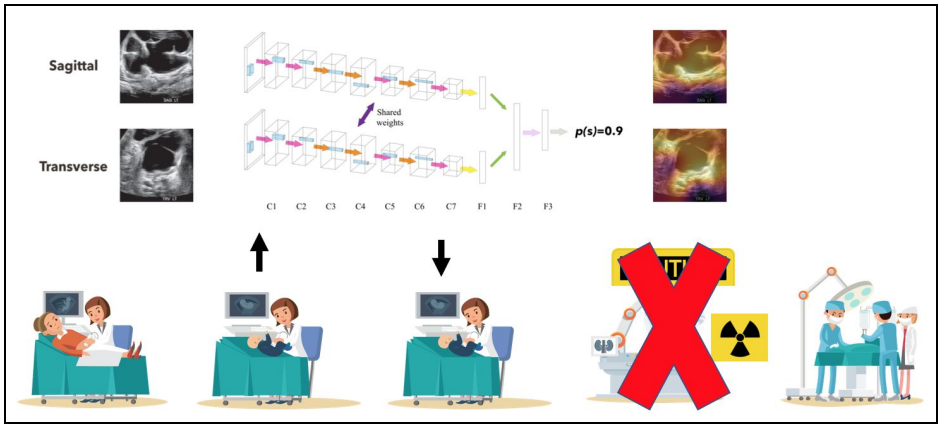

Figure 2. We propose that our model can be used by clinicians to streamline their workflow and make more confident prognoses earlier. During a patient ultrasound visit, the clinician can augment their decision-making process by incorporating our model's prediction as the probability that the screened kidney would require surgery. In this example, the model predicts a high probability of surgery, which may motivate the doctor to forego invasive tests and proceed to surgery earlier.

References

1. Fernbach, S. K., Maizels, M. & Conway, J. J. Ultrasound grading of hydronephrosis: introduction to the system used by the Society for Fetal Urology. *Pediatr. Radiol.* **23**, 478–480 (1993).
2. Sidhu, G., Beyene, J. & Rosenblum, N. D. Outcome of isolated antenatal hydronephrosis: a systematic review and meta-analysis. *Pediatr. Nephrol.* **21**, 218–224 (2006).
3. Braga, L. H., D'Cruz, J., Rickard, M., Jegatheeswaran, K. & Lorenzo, A. J. The Fate of Primary Nonrefluxing Megaureter: A Prospective Outcome Analysis of the Rate of Urinary Tract Infections, Surgical Indications and Time to Resolution. *J. Urol.* **195**, 1300–1305 (2016).
4. Lorenzo, A. J., Rickard, M., Braga, L. H., Guo, Y. & Oliveria, J.-P. 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* **123**, 204–209 (2019).
5. Jung, K. & Shah, N. H. Implications of non-stationarity on predictive modeling using EHRs. *J.*

Biomed. Inform. **58**, 168–174 (2015).

6. Kermany, D. Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification. *Mendeley Data* (2018).
7. LeCun, Y. & Cortes, C. MNIST handwritten digit database. (2010). Available at: http://yann.lecun.com/exdb/mnist/.
8. Noroozi, M. & Favaro, P. Unsupervised Learning of Visual Representations by Solving Jigsaw Puzzles. *arXiv [cs.CV]* (2016).
9. Kermany, D. S. *et al.* Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. *Cell* **172**, 1122–1131.e9 (2018).
10. Selvaraju, R. R. *et al.* Grad-cam: Visual explanations from deep networks via gradient-based localization. in *Proceedings of the IEEE International Conference on Computer Vision* 618–626 (2017).
11. Raghu, M., Zhang, C., Kleinberg, J. & Bengio, S. Transfusion: Understanding transfer learning with applications to medical imaging. *arXiv preprint arXiv:1902. 07208* (2019).