Ultrasound based computer-aided-diagnosis of kidneys for pediatric hydronephrosis

Juan J. Cerrolaza, Craig A. Peters, Aaron D. Martin, Emmarie Myers, Nabile Safdar, et al.

Ultrasound Based Computer-Aided Diagnosis of Kidneys for Pediatric Hydronephrosis
Juan J. Cerrolaza*a, Craig A. Peters,a,b Aaron D. Martina,b, Emmarie Myersa, Nabile Safdaracd, Marius George Lingurarad
aSheikh Zayed Institute for Pediatric Surgical Innovation, Children’s National Medical Center, Washington, DC, USA; bDept. of Pediatric Urology, Children’s National Medical Center, Washington, DC, USA; cDept. of Diagnosis Imaging and Radiology, Children’s National Medical Center, Washington, DC, USA; dDept. of Radiology and Pediatrics, School of Medicine and Health Sciences, George Washington University, Washington, DC, USA.

ABSTRACT

Ultrasound is the mainstay of imaging for pediatric hydronephrosis, though its potential as diagnostic tool is limited by its subjective assessment, and lack of correlation with renal function. Therefore, all cases showing signs of hydronephrosis undergo further invasive studies, like diuretic renogram, in order to assess the actual renal function. Under the hypothesis that renal morphology is correlated with renal function, a new ultrasound based computer-aided diagnosis (CAD) tool for pediatric hydronephrosis is presented. From 2D ultrasound, a novel set of morphological features of the renal collecting systems and the parenchyma, is automatically extracted using image analysis techniques. From the original set of features, including size, geometric and curvature descriptors, a subset of ten features are selected as predictive variables, combining a feature selection technique and area under the curve filtering. Using the washout half time (T1/2) as indicative of renal obstruction, two groups are defined. Those cases whose T1/2 is above 30 minutes are considered to be severe, while the rest would be in the safety zone, where diuretic renography could be avoided. Two different classification techniques are evaluated (logistic regression, and support vector machines). Adjusting the probability decision thresholds to operate at the point of maximum sensitivity, i.e., preventing any severe case be misclassified, specificities of 53%, and 75% are achieved, for the logistic regression and the support vector machine classifier, respectively. The proposed CAD system allows to establish a link between non-invasive non-ionizing imaging techniques and renal function, limiting the need for invasive and ionizing diuretic renography.

Keywords: Hydronephrosis, kidney, ultrasound imaging, computer-aided diagnosis, machine learning.

INTRODUCTION

Hydronephrosis is one of the most common finding in pediatric urology, affecting 2-2.5% of children [1]. The accumulation of urine in the renal pelvis as a result of obstruction to outflow causes the dilation of the renal collecting system and the distortion of the renal parenchyma [2]. Based on these morphological evidences, ultrasound (US) imaging is the mainstay for early diagnosis of hydronephrosis, yet is limited by its subjective assessment, and lack of correlation with functional imaging modalities, like diuretic renogram (MAG-3 scan [3]). Based on simple visual inspection, the commonly used Society for Fetal Urology (SFU) grading system for hydronephrosis [4] is subjective, and the categorization obtained does not necessarily reflects the actual renal function of the kidney, as shown in Fig. 1. As a result, the SFU system is not universally accepted as useful [5] and functional renal imaging has emerged as the gold standard for assessing hydronephrosis. A renal MAG-3 scan provides valuable and objective information on renal function, though its invasive nature, involving bladder catheterization and ionizing radiation exposure, makes it especially undesirable for children.

The main goal of this work is to create a new US based computer-aided diagnostic framework for pediatric hydronephrosis. The underlying hypothesis of this study is that hydronephrosis can be objectively quantified using modern image analysis of US images, and that there is a correlation between the degree of hydronephrosis and the severity of the obstruction as measured by diuretic renography.

*JCerrola@cnmc.com; phone 1 202 476 1272; innovationinstitute.childrensnational.org
Figure 1. Lack of correlation between SFU grading system and renal function. According to the SFU grading system, the upper image would be labeled as a moderate hydronephrotic case, while the lower would correspond to a more severe case. However, functional information reveals that the former has a greater obstruction.

Figure 2. Cumulative frequency distribution of the population. The ages considered in the data set ranges from newborns (< 5 months) to 14 years of ages. The washout half times goes from < 6 min (minor obstruction) to more than 120 min.

From the manual segmentation of renal parenchyma and collecting system in 2DUS images, an optimal set of morphological descriptors of the kidney are automatically extracted and used as inputs features of a machine learning algorithm. This classifier is able to predict the degree of hydronephrosis of the renal unit, identifying with maximum sensitivity the severe cases that require immediate medical attention, while maximizing the number of non-critical cases where diuretic renography can be safely avoided.

**METHODOLOGY**

The data set considered in this study consists of 50 hydronephrotic cases belonging to 45 children (31 male and 14 female; studies of both kidneys are available from some patients) of different ages, from newborns to 14 years of age. The cumulative histogram of the age distribution is depicted in Fig. 2. For each case, concurrent renal 2D ultrasound imaging and diuretic renography are available, allowing to study the correlation between renal function and morphology. The degree of obstruction of the kidney is described by means of the washout half time ($T_{1/2}$): the time required to clear half of the accumulated tracer from the renal pelvis after reaching its peak and being stimulated by administration of flurosemide [2] Fig. 2. also illustrates the distribution of $T_{1/2}$ in the data set. Using this parameter as a reference, thresholds of functional significance are defined to determine those patients where diuretic renography could be safely avoided, and which would benefit from invasive functional imaging. In particular, a threshold of 30 minutes is considered in this study.
From the sequence of 2DUS scans, a 2DUS slice containing a whole longitudinal section of the kidney and its collecting system is selected and manually segmented in consensus by two experienced clinicians, a radiologist and an urologist, minimizing the subjectivity of the delineation process in US images. Though the development of accurate and automatic segmentation tools in US is especially challenging, the promising results reported by some recent works [2, 7, 8] suggest the possibility of automating this preliminary stage in the near future. Fig. 3 (b) shows the consensus masks obtained for the case depicted in Fig. 3 (a).

![Figure 3](https://www.spiedigitallibrary.org/conference-proceedings-of-spie)

**Figure 3.** 2DUS slice selection and segmentation. a) Selected slice containing the whole longitudinal section of the kidney. b) Consensus manual delineation of renal parenchyma and collecting system.

![Figure 4](https://www.spiedigitallibrary.org/conference-proceedings-of-spie)

**Figure 4.** Morphological descriptors. Sample of some of the morphological descriptors considered as potential predictive features of hydronephrosis severity: eccentricity of the circumscribed ellipse \((1-b^2/a^2)^{1/2}\), maximum and minimum parenchyma thickness \((P_{max}, P_{min})\), curvature statistics of the collecting system \((C_{cs})\) and kidney surface \((C_{k})\) (i.e., mean, maximum, minimum, kurtosis, etc), point of maximum curvature of the kidney medial axis \(KMA\), \((P_{KMA max})\).

The binary masks of the kidney and its collecting system, are processed by means of image analysis techniques to automatically extract a set of morphological parameters of the kidney. These quantitative shape descriptors can be intuitively grouped in the following three categories. **Size descriptors:** Parameters describing the size of the collecting system and renal parenchyma, such as relative area, perimeter, or parenchymal thickness. **Geometric shape descriptors:** Generic geometric parameters as the circularity ratio, eccentricity of the circumscribed ellipse, or convexity. The Hu set
of seven invariant moments [9] of both, the parenchyma and the collecting system, are also considered here. These features, independent of position, size and orientation, have been extensively applied to image registration and reconstruction, and are of great interest for image pattern recognition. We have also considered other shape descriptors as the kidney medial axis [10], and other derived features such as the asymmetry index (i.e. ratio between the segments that joint the two extremes of the kidney medial axis and the point of maximum curvature of this axis). **Curvature descriptors**: Including the traditional extrinsic curvature of planar curves, and the more recent local area integral invariants [11].

The potential utility of such morphological descriptors for assessing hydronephrosis has been stressed by several authors, attempting to analyze US imaging of hydronephrosis in more objective ways. The “hydronephrosis index” defined by Saphiro et al. [12] as the ratio between the area of the renal parenchyma and the total area of the kidney, has been shown to correlate with the SFU. Instead of a single feature, more than 130 quantitative shape descriptors are considered in this study (including the aforementioned “hydronephrosis index”), some of them are graphically illustrated in Fig. 4. These quantitative shape descriptors are used as potential predictive variables to evaluate the diagnostic performance of two different supervised classification models: logistic regression analysis, and support vector machine with radial basis function kernel. Both approaches have been widely used in the field of computer-aided-diagnosis with satisfactory results. Logistic regression (LOG) is a binary classifier that performs probability estimation using a logistic formula. On the other hand, support vector machines (SVM) operate by finding an hyperplane leaving the largest possible fraction of points of the same class on the same side, while maximizing the distance of either class from the hyperplane.

Prior to evaluating the classification capabilities of both methods, the high dimensional space of predictive features is reduced by finding the subset of relevant morphological descriptors that facilitate the classification process. By means of the feature selection technique proposed by Cai et al. [13], optimal subsets of 9 variables are obtained, using the area under the curve (AUC) criteria to select the final predictive set (Table 1). A detailed description of the resulting set is provided in the next section. The parameters optimization is performed via grid search, and the leave-one-out strategy is employed for training and testing purposes. Receiver operating characteristic (ROC) curve analysis is used to identify the operating points with the highest accuracy, sensitivity and specificity of each classifier. The ultimate goal is to identify those probability thresholds that maximize the sensitivity of detecting severe cases of hydronephrosis; that is, no case with a washout time above the defined threshold is misclassified. Thus, diuretic renography could be safely avoided in those cases classified as non-severe, reducing the number of unnecessary ionizing imaging studies.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Descriptor</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>Size</td>
<td>Minimum parenchymal thickness. $P_{min}/a$</td>
</tr>
<tr>
<td>P2</td>
<td>Size</td>
<td>Kurtosis of the parenchymal thickness. $Kurt(P/a)$</td>
</tr>
<tr>
<td>P3</td>
<td>Geometric</td>
<td>Asymmetry index of the kidney medial axis. $(P1_{KMA} - P_{KMA_{max}})/(P2_{KMA} - P_{KMA_{max}})$</td>
</tr>
<tr>
<td>P4</td>
<td>Geometric</td>
<td>1st Hu invariant moment of the renal parenchyma.</td>
</tr>
<tr>
<td>P5</td>
<td>Geometric</td>
<td>3rdHu invariant moment of the renal parenchyma.</td>
</tr>
<tr>
<td>P6</td>
<td>Curvature</td>
<td>Minimum curvature of the kidney contour. $min(C_K)$</td>
</tr>
<tr>
<td>P7</td>
<td>Curvature</td>
<td>Entropy of the curvature of the kidney contour. $Ent(C_K)$</td>
</tr>
<tr>
<td>P8</td>
<td>Curvature</td>
<td>Minimum curvature of the collecting system. $min(C_CS)$</td>
</tr>
<tr>
<td>P9</td>
<td>Curvature</td>
<td>Entropy of the parenchymal local area integral. $Ent(LAI(P))$</td>
</tr>
</tbody>
</table>

*The parenchymal thickness has been normalized by the major axis of the ellipse circumscribing the kidney (see Fig. 4).*

**RESULTS**

Following the procedure described above, a set of 10 features is selected, including the following morphological descriptors: first and third invariant moment of the parenchyma; inner curvature of the kidney; minimum and entropy of the curvature of the kidney; minimum curvature of the collecting system; entropy of the local area integral information of the collecting system; minimum, and kurtosis of parenchyma thickness; kurtosis of the distance between the kidney and the collecting system. Fig. 5 shows the ROC curves for the two classifiers under study. The AUC for SVM is 0.89, and 0.93 for LOG. The performance of each one of the classifiers is evaluated in terms of its specificity, sensitivity, and
accuracy, whose numerical values are listed in Table 2. The accuracy, sensitivity and specificity for SVM and LOG are 0.86, 0.57, and 0.97, and 0.82, 0.71, and 0.86, respectively. The ROC shown in Fig. 5 allows to identify the optimal operating points that provide 100% sensitivity while maximizing the specificity of the system (i.e., minimizing the false positive rate). The performance of the binary classifiers achieved after modifying the probability decision thresholds to operate at these points of maximum sensitivity is also shown in Table 1, where the best case in terms of specificity (percentage of patients in the safe zone) is marked with (*). In particular, it is the LOG classifier that provides the best result with a specificity of 0.75, while the specificity for the SVM is 0.53.

![Receiver Operating Characteristics Graphs](image)

**Figure 5.** Receiver operating characteristic graph. ROC curves obtained for the two machine learning methods under study, logistic regression and support vector machines. The graph also shows the area under the curve obtained for each case and the point of maximum sensitivity.

**Table 2.** Area under the curve, sensitivity, specificity, and accuracy obtained for the two classifiers studied, SVM, and LOG. After modifying the decision threshold to operate at the point of maximum sensitivity, the best result in terms of specificity is marked with (*).

<table>
<thead>
<tr>
<th></th>
<th>Original Values</th>
<th>Max. Sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AUC</td>
<td>Sensitivity</td>
</tr>
<tr>
<td>SVM</td>
<td>0.89</td>
<td>0.57</td>
</tr>
<tr>
<td>LOG</td>
<td>0.93</td>
<td>0.71</td>
</tr>
</tbody>
</table>

**CONCLUSIONS**

A new US based CAD tool for pediatric hydronephrosis is presented in this study. Alternatively to the subjective SFU grading system based on simple visual inspection, we propose a novel objective and reproducible scoring system for hydronephrosis able to relate measurable morphological findings from US imaging with renal function information provided by MAG-3 scans. Using the washout half time as clinical indicative of renal obstruction, the population is divided into two groups of patients: those with severe obstruction and those with satisfactory drainage where further invasive studies can be safely avoided. Beyond the typical parameters based on the area and size of the kidney, more than 130 anatomical features are automatically extracted from the 2DUS images via image analysis techniques. Many of these features are novel and tailored to capture the relation between renal parenchyma and collecting system. Based on the washout half time, a selection of these features are used as predictive variables to classify hydronephrotic patients as severe, where a further invasive imaging would be warranted, or non-critical, where diuretic renography could be safely avoided. Two different machine learning algorithms are tested: logistic regression, and support vector machines. The results obtained from a dataset of 50 cases with different degree of renal obstruction are very promising. Operating at the
point of maximum sensitivity to guarantee that no case with severe hydronephrosis is misclassified, up to 75% of non-severe cases are successfully identified. The results obtained with 2DUS images demonstrate the feasibility of developing reliable and objective diagnostic tools to support the routine clinical evaluation of kidneys from non-invasive imaging modalities. The new CAD system proposed here does not intend to replace MAG-3, but to present an accurate, quantitative and reproducible diagnostic tool for hydronephrosis based on simple 2DUS images that improves the current standards of management of hydronephrotic patients and limits the need for diuretic renography. In future work we expect to improve system performance by extending the study to 3D US images.

REFERENCES