Exploring the Predictive Role of Inflammatory Markers in Neuropathic Bladder-Related Kidney Damage with Machine Learning

Su Özgür¹, Sevgin Taner², Gülnur Gülnaz Bozcuk³, Günay Ekberli⁴

¹Ege University, Translational Pulmonary Research Center-EgeSAM, İzmir, Turkey
²Adana City Training and Research Hospital, Clinic of Pediatric Nephrology, Adana, Turkey
³Adana City Training and Research Hospital, Clinic of Pediatrics, Adana, Turkey
⁴Adana City Training and Research Hospital, Clinic of Pediatric Urology, Adana, Turkey

ABSTRACT

Aim: The main objective of this study was to predict upper urinary tract damage utilizing novel approaches, such as machine learning models, by incorporating simple predictors alongside established radiological and clinical factors.

Materials and Methods: In this retrospective study, a total of 191 patients who underwent blood tests, urine analysis, imaging, and urodynamic studies (UDS) in order to assess their nephrological and urological status were included. Basic statistical analyses were conducted using IBM SPSS Version 25. A significance level of p<0.05 was employed to establish statistical significance. The machine learning analyses were performed on Ddsv4-series Azure Virtual Machines, equipped with 32 vCPUs with a memory capacity of 128 GiB.

Results: In the model where clinical and imaging data were jointly assessed, the k-nearest neighbor (KNN) model demonstrated the highest performance, achieving values of 0.813 area under the curve and 0.854 accuracy. For the KNN Model, the best predictors for kidney function loss were as follows: neutrophil/lymphocyte (1.0577), abnormal bladder in ultrasound (1.054), vesicoureteral reflux (0.901), ferritin (0.898), neutrophil/albumin (0.678), platelet/lymphocyte (0.619), increased detrusor leakage pressure (0.435), age (0.3505), decreased bladder capacity in urodynamics (0.3009), and white blood cell (0.266).

Conclusion: Based on our findings, initial patient evaluation through basic blood and urine tests, ultrasonography, UDS, and voiding cystourethrography is crucial for identifying risk factors and preventing renal damage. Complete blood count-derived inflammatory biomarkers offer cost-effective and accessible alternatives to other radiological tools in primary care settings. These machine learning models may hold clinical relevance in pre-clinical or resource-limited hospitals, by guiding clinicians in implementing preventative measures.

Keywords: Neuropathic bladder dysfunction, kidney damage, inflammatory markers, machine learning, k-nearest neighbor, random forest
in patients with NBD (3). High intravesical pressure transmitted to the upper urinary tract causes decreased glomerular filtration and impaired urine flow from the renal collecting system to the bladder. As a common result of his pathology, progressive damage and kidney failure develop (4). An initial evaluation of patients with ultrasonography (US), voiding cystourethrography (VCUG), 99m Technetium Dimercaptosuccinic acid (DMSA), and blood and urine test can guide clinicians in terms of early diagnosis, treatment and the possible avoidance of negative consequences in the future. Identifying risk factors and indicators of upper urinary tract damage and progression to chronic kidney disease is an important cornerstone in the monitoring of these patients. Knowing the prognostic indicators in terms of upper system damage can guide medical and surgical treatment in order to prevent kidney damage and can also protect those patients who do not require medical or surgical treatment from undergoing unnecessary and troublesome advanced imaging.

Machine learning is increasingly being utilized in healthcare services. It demonstrates better accuracy in diagnosing diseases such as VUR and urinary tract infection (UTI), which are often challenging to differentiate in the clinic, by employing decision support algorithms. It provides significant support to clinicians in early diagnosis. While some biomarkers are utilized in clinical practice for the early diagnosis of NBD in the literature, there are no machine learning studies employing innovative approaches. This was the first study on this topic using machine learning techniques.

It is crucial to extract meaningful information from complex patterns in clinical data. Our objectives in this patient group were to make an early diagnosis of UUTD and, more importantly, to identify those patients at risk of UUTD. In this way, it may be possible to prevent kidney damage by using more aggressive diagnostic and treatment methods in this patient group.

**Materials and Methods**

The electronic medical records of patients diagnosed with neuropathic bladder in the Pediatric Urology and Nephrology units at Adana City Training and Research Hospital were retrospectively reviewed. Ethical approval was obtained from the Adana City Training and Research Hospital Clinical Research Ethics Committee (approval no.: 1367, date: 08.04.2021).

A total of 191 patients who underwent blood tests, urine analysis, imaging, and urodynamic studies (UDS) for evaluating nephrological and urological status were included in this study. Those patients under the age of one year were excluded in terms of a time period for renal scar development. Demographic characteristics, medical history, and laboratory and imaging results were documented. Differentiated kidney functions and scarring were evaluated via Tc-99m DMSA scan. The presence of more than a 10% decrease in differentiated kidney functions in scintigraphic evaluation was considered as loss of kidney function (5).

Accordingly, those patients with loss of kidney function were categorized as Group 1, and those patients without loss of kidney function were categorized as Group 2.

**K-Nearest Neighbor Algorithm**

K-nearest neighbors (KNN) is a standard machine learning method which has been extended to large-scale data mining efforts. Test samples are classified to the class most frequently occurring among the KNN in a multidimensional parameter space. Despite its simplicity, this method has a sound theoretical basis in non-parametric density estimation and can often outperform much more sophisticated methods. The method requires only the choice of k, the number of neighbors to be considered when making the classification. Small values of k will select the closest training points, which are best able to estimate the correct classification at the test point. Usually, k is chosen as the value which minimizes the classification error on some independent validation data or through cross-validation procedures (Figure 1) (6).

**Random Forest**

Random forest (RF) is an ensemble learning algorithm widely used for both classification and regression tasks in machine learning. It operates by constructing a multitude of decision trees during training and outputs the mode of the classes (classification) or the mean prediction (regression) of the individual trees (Figure 2) (7).

**Variable Importance**

Variable importance in machine learning models refers to the measure of the impact which individual input features (variables) have on the model’s predictive performance or the outcome of interest. It quantifies the degree to which each variable contributes to the model’s ability to make accurate predictions. Variable importance helps in understanding which features are the most influential in making decisions, allowing practitioners to focus on the most relevant factors and so potentially improves model interpretability and generalization.
Various techniques can be employed to gauge variable importance, including permutation importance, feature importance scores derived from algorithms like RF, or the analysis of coefficients in linear regression. These methodologies assign scores or rankings to each feature based on the degree to which the model’s performance is compromised when a particular feature is altered or omitted (8).

**Performance Metrics**

Performance metrics play a crucial role in evaluating the effectiveness of classification models, particularly those used in machine learning, statistics, and data analysis, with a focus on binary classification scenarios (Table I).

**Accuracy:** As a fundamental metric, accuracy calculates the ratio of correct predictions to the total number of predictions. Widely employed in medical machine learning applications, it should be interpreted cautiously, especially when handling imbalanced datasets.

**Area under the receiver operating characteristic curve (AUC-ROC):** Assessing a model’s ability to distinguish between positive and negative classes, the AUC-ROC metric utilizes a visual representation through the ROC curve, where a higher value signifies enhanced class discrimination.

**Recall:** Also known as sensitivity or the true positive rate, recall quantifies the proportion of true positives relative to all actual positive instances. It is crucial in scenarios where the identification of all positive instances is of the utmost importance.

**Precision:** Precision gauges the ratio of true positive predictions among all positive predictions, emphasizing the minimization of false positive errors. This metric is particularly valuable when the cost associated with false positives is significant.

**F1 Score:** Representing the harmonic mean of precision and recall, the F1 Score provides a balanced single metric which accounts for both false positives and false negatives. It proves especially beneficial in addressing class imbalance.

**Matthews Correlation Coefficient (MCC):** Serving as a metric to evaluate the quality of binary classifications, MCC considers true positives, true negatives, false positives,
and false negatives. It offers a balanced measure, ensuring reliability even in the face of imbalanced datasets (9).

**Statistical Analysis**

The sociodemographic and disease-related characteristics of the patients are presented using percentages (%), numbers (n), mean, and standard deviation (SD), as well as median, minimum, and maximum values. To compare the sociodemographic and disease-related characteristics between scar-positive and scar-negative cases, chi-square, Student’s t-tests, and Mann-Whitney U tests were employed. IBM SPSS Version 25 was utilized for basic statistical analyses. A significance level of p<0.05 was considered when determining statistical significance.

The machine learning analyses were conducted using Dds4-series Azure Virtual Machines, featuring a vCPU count of 32 and a memory capacity of 128 GiB. The results and parameters of the best model obtained from the analyses conducted in Azure Automated ML (KNN and RF) are presented.

In the analyses of this study, both classical algorithms (RF, KNN, logistic regression) and innovative algorithms (XGBoost, LightGBM, Gradient Boosting) were employed. The findings section focuses on those models which achieved high performance in line with the research purpose.

**Results**

A total of 191 patients who underwent blood tests, urine analysis, imaging, and UDS to assess nephrological and urological status were evaluated. The cases included in the study consisted of 62.3% (119) girls and 37.7% (72) boys (p>0.05).

The median age of children in Group 1 was 95 (3-207) months, whereas in Group 2 it was 63.5 (1-225) months, with a statistically significant difference (p=0.005). In Group 1, the median leak point pressure was 40 (7-100) cm H₂O, compared to 30 (5-120) cm H₂O in Group 2 (p>0.05). The estimated glomerular filtration rate (eGFR) level in Group 1 [94.4 (8.02-215.48) mL/dk/1.73 m²] was significantly lower than that in Group 2 [193.28 (22.72-476.54) mL/dk/1.73 m²] (p<0.001). The median hemoglobin level in Group 1 [10.95 (8-15.8) g/dL] was notably lower (p=0.002). In Group 1, the median urea level was 30 (5-275) mg/dL, while in Group 2, it was 22 (5-323) mg/dL (p<0.001). The creatinine level in Group 1 [0.47 (0.11-5.2) mg/dL] was higher than that in Group 2 [0.23 (0.04-3)] (p<0.001). The neutrophile/lymphocyte ratio (p>0.05), platelet/lymphocyte ratio (p>0.05), neutrophile/albumin ratio (p>0.05), and systemic immune index (p>0.05) were similar between the groups. However, the frequencies of MSUG (p<0.001) and VUR (p<0.001) were higher in Group 1 (Tables II and III).

Two different approaches were employed in the prediction models. In Model 1, imaging findings were considered, while in Model 2, only demographic and clinical variables were utilized in cases where imaging information was unavailable. The variables that most significantly contributed to the differential diagnosis are presented in both Model 1 and Model 2 (Table IV, Figures 3 and 4).
<table>
<thead>
<tr>
<th>Variables</th>
<th>Group 1: Loss of kidney function (+) n=51 (26.7%)</th>
<th>Group 2: Loss of kidney function (-) n=140 (73.3%)</th>
<th>Test statistics; p value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean ± SD</td>
<td>Median (Min.-Max.)</td>
<td>Mean ± SD</td>
</tr>
<tr>
<td>Leak point pressure (cm H₂O)</td>
<td>42.43±27.75</td>
<td>40 (7-100)</td>
<td>39.21±27.81</td>
</tr>
<tr>
<td>eGFR (mL/dk/1.73 m²)</td>
<td>97.72±66.76</td>
<td>94.4 (8.02-215.48)</td>
<td>204.66±85.27</td>
</tr>
<tr>
<td>Hemoglobin (g/dL)</td>
<td>11.2±1.55</td>
<td>10.95 (8-15.8)</td>
<td>11.9±1.66</td>
</tr>
<tr>
<td>MCV (fL)</td>
<td>77.17±9.15</td>
<td>76.65 (29.5-90.7)</td>
<td>77.03±7.56</td>
</tr>
<tr>
<td>WBC (x10³/mm³)</td>
<td>9.1±3.6</td>
<td>8.6 (4.5-20.0)</td>
<td>8.6±2.7</td>
</tr>
<tr>
<td>Neutrophil (x10³/mm³)</td>
<td>4.6±2.5</td>
<td>4.2 (1.8-15.2)</td>
<td>3.9±1.7</td>
</tr>
<tr>
<td>Lymphocyte (x10³/mm³)</td>
<td>3.3±15.2</td>
<td>3.2 (0.8-7.2)</td>
<td>3.6±1.8</td>
</tr>
<tr>
<td>Platelet (x10³/mm³)</td>
<td>325±102</td>
<td>338 (164-633)</td>
<td>410±621</td>
</tr>
<tr>
<td>Ferritin (mg/L)</td>
<td>61.63±72.08</td>
<td>35.85 (4-389)</td>
<td>51.66±88.6</td>
</tr>
<tr>
<td>Total protein (mg/dL)</td>
<td>40.23±33.36</td>
<td>47 (4.5-84)</td>
<td>31.86±31.23</td>
</tr>
<tr>
<td>Albumin (mg/dL)</td>
<td>20.19±19.03</td>
<td>4.4 (2.4-50.8)</td>
<td>17.3±18.22</td>
</tr>
<tr>
<td>Urea (mg/dL)</td>
<td>47.92±49.51</td>
<td>30 (5-275)</td>
<td>25.16±27.96</td>
</tr>
<tr>
<td>Creatinine (mg/dL)</td>
<td>0.86±1.01</td>
<td>0.47 (0.11-5.2)</td>
<td>0.29±0.29</td>
</tr>
<tr>
<td>Uric acid (mg/dL)</td>
<td>5.02±2.15</td>
<td>4.61 (2.04-10.08)</td>
<td>3.55±1.18</td>
</tr>
<tr>
<td>Cystatin C (ng/mL)</td>
<td>1.5±1.12</td>
<td>1 (0.41-6)</td>
<td>0.91±0.53</td>
</tr>
<tr>
<td>Parathormone level (pg/mL)</td>
<td>177.5±336.54</td>
<td>79.1 (12.8-2190)</td>
<td>44.29±34.03</td>
</tr>
<tr>
<td>Neutrophil lymphocyte ratio</td>
<td>1.67±1.32</td>
<td>1.28 (0.42-6.88)</td>
<td>1.39±0.86</td>
</tr>
<tr>
<td>Platelet lymphocyte ratio</td>
<td>0.12±0.07</td>
<td>0.1 (0.03-0.45)</td>
<td>0.16±0.45</td>
</tr>
<tr>
<td>Neutrophil albumin ratio</td>
<td>531.52±511.13</td>
<td>253.46 (41.3-1750)</td>
<td>541.66±489.49</td>
</tr>
<tr>
<td>Systemic immune index</td>
<td>556.18±509.07</td>
<td>416.2 (143.58-2447.5)</td>
<td>686.67±2067.07</td>
</tr>
</tbody>
</table>

U: Mann-Whitney U test statistics, p<0.05 Significance level
Min.-Max.: Minimum-maximum, SD: Standard deviation, MCV: Mean corpuscular volume, WBC: White blood cells, eGFR: Estimated glomerular filtration rate
The variable importance obtained from KNN for Model 1 and RF for Model 2 are presented in Table IV. Accordingly, the ten variables which best predicted loss of kidney function for Model 2 are as follows: neutrophile/lymphocyte (1.0577), abnormal bladder in US (1.054), VUR (0.901), ferritin (0.898), neutrophile/albumin (0.678), platelet/lymphocyte (0.619), increased detrusor leakage pressure (0.435), age (0.3505), decreased bladder capacity in urodynamics (0.3009), and WBC (0.266). We used AUC (0.854), Accuracy (0.813), Precision (0.885), and Recall (0.625) as model performance criteria for the machine learning algorithm comparisons for Model 1 (KNN). Similarly, for Model 2 (RF), the variables determined to predict loss of kidney function were Ferritin (0.490), Age (0.172), neutrophile/albumin (0.093), neutrophile/lymphocyte (0.088), platelet/lymphocyte (0.082), sex (0.045), and WBC (0.018) (Figures 3 and 4). In order to assess the performance of machine learning algorithms, we employed criteria including AUC (0.833),
accuracy (0.625), precision (0.823), and recall (0.667) for Model 3 (Table IV).

**Discussion**

Studies have been conducted in order to assess the risk factors regarding UUTD (10-12) in patients with NBD. McGuire et al. (11) were the first to consider bladder changes as a risk factor for renal dilatation and VUR in children. Sixty-eight percent of patients with end-filling detrusor pressure more than 40 H$_2$O had VUR and 81% had dilatation. Timberlake et al. (12) evaluated ultrasonographic findings and urodynamic parameters for the detection of risk factors in NBD. Higher detrusor pressure was reported to be related with renal scars. Additionally, trabeculation on US and the presence of VUR were also reported to be associated with renal scar (12).

Ekberli and Taner (13) conducted a retrospective review of patients diagnosed with neuropathic bladder. Logistic regression analysis was employed in order to identify significant predictors of scar formation in DMSA. Their results revealed a strong correlation between age, bladder changes observed in ultrasound and voiding cystogram, as well as high leak point pressure obtained during UDS, and upper urinary tract damage (13).

Li et al. (14) developed a predictive model for upper urinary tract damage in children with neurogenic bladder (NB). Their study revealed that recurrent UTI, bladder compliance, detrusor leak point pressure, overactive bladder, and clean intermittent catheterization are significant determinants of UUTD, and the efficacy of the model was validated. Univariate and multivariate logistic regression analyses were performed on the training cohort in order to identify predictors and create a nomogram. The nomogram exhibited strong discrimination, as indicated by the AUC-ROC in the training cohort [0.806, 95% confidence interval (CI): 0.737-0.874] and the validation cohort (0.831, 95% CI: 0.753-0.909). We achieved ROC values within the range of 83.3% to 85.4%, indicating a remarkably high level of predictive performance in our study (14).

In our study, leak point pressure was detected to be higher in the patient group with renal function loss. Also, abnormal ultrasonographic changes of the urinary tract,
the presence of VUR, recurrent UTI and bladder changes in VCUG can be considered as contributing factors for renal damage.

In addition to radiological evaluations, the accurate measurement of kidney function is essential in avoiding glomerular and tubular compromise. eGFR is defined as the GFR together with the serum creatinine value using the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) creatinine-based formula and the updated Schwartz “bedside” formula for children (15,16). The GFR <90 mL/m² for 3 months with or without signs of renal damage (Shwartz) is defined as chronic renal disease. While evaluating our results, in line with the literature knowledge, eGFR values were found to be significantly lower in the group with loss of function than in the group without.

Another important parameter as renal lesions and chronic renal disease marker is protein excretion in urine. In children, protein excretion of <100 mg/m²/day or <4 mg/m²/hour in a 24 hr urine collection is considered normal. It is important to investigate proteinuria (up to 5 mg/kg/day in NB) as a marker of renal lesion (up to 5 mg/kg/day in NB) (17). In our study, there was a significant difference in proteinuria between the groups, consistent with the literature.

There is an increasing trend towards evaluating the potential value of hematological pro-inflammatory markers in the diagnosis and prognosis of various chronic diseases given the link between inflammation and changes in peripheral blood cells (18,19). Neutrophil-to-lymphocyte ratio (NLR), platelet-to-lymphocyte ratio (PLR), and systemic immune inflammation index are complete blood count (CBC)-derived inflammatory biomarkers which have been widely used in adult study populations (20,21).

Ohtaka et al. (22) noted considerable variations in the NLR at 1 and 3 months following renal transplantation. Patients with malignancies post-renal transplant exhibited a persistent increase in their NLR. Their research proposes that keeping track of the NLR in kidney transplant recipients may aid in the timely identification of malignancies. In our research, we identified neutrophil/lymphocyte, neutrophil/albumin, platelet/lymphocyte, and platelet/albumin ratios as significant predictors in anticipating kidney function loss (22).

Hobbs et al. (23) developed a machine learning algorithm which identified detrusor overactivity in UDS within the spina bifida population. Their time-based model exhibited the highest AUC at 91.9%, along with sensitivity values of 84.2%. Given the significant variability in individual interpretations of UDS data, there is a pressing need to standardize UDS interpretation. The successful development and implementation of machine learning algorithms in this population has the potential to be scaled to encompass all patients, both children and adults, experiencing lower urinary tract symptoms undergoing UDS. We introduced significant clinical models aimed at the early prediction of renal function loss, demonstrating high accuracy (81.3%) and AUC (85.4%) values in this study (23).

There are a limited number of studies recommending the use of the above-mentioned markers in the pediatric population (24,25). Nicoară et al. (24) in their study aimed to assess the relationship between CBC-derived inflammatory biomarkers and the presence of MetS in obese children. As a conclusion, they reported these markers which are inexpensive and universally available in primary care settings, to be an attractive alternative or addition to the frequently assessed inflammatory biomarkers (24). Another study in the pediatric obese patient population also revealed leukocyte, lymphocyte, erythrocyte, and platelet levels as being significantly higher in overweight/obese children but there was no significant difference between the groups regarding NLR and PLR (25).

In univariate analyses, statistical significance not identified between groups (loss of kidney function positive and negative cases) can gain significance through methods adept at extracting meaningful information from complex data structures, such as machine learning. Algorithms such as RF and KNN, which demonstrated the best performance in this research, can exhibit high efficacy in small datasets and scenarios where data relationships are less complex. Also, models generated using these approaches are often more straightforward to explain compared to boosting algorithms (26-28).

In this study, we achieved high-performance results by presenting models which both included and excluded imaging, when addressing a crucial clinical scenario such as the prevention of kidney loss. We believe that the increasing adoption of innovative approaches in clinical settings can enhance the precision of treatment plans for patients.

Study Limitations

This study was retrospective. Prospective studies are required to enhance the sensitivity of the models and enable real-time predictions.
The primary objective of this research was to identify crucial biomarkers for differential diagnosis in the NB and to promptly reveal predictive variables without directly impacting the patient. The results (XGBoost, LightGBM, Gradient Boosting, etc.) obtained from advanced machine learning methods employed in model development were not included.

**Conclusion**

According to our results, it can be claimed that an initial evaluation of patients with basic blood and urine tests, US, UDS and VCUG is essential for the detection of risk factors and the prevention of renal damage. CBC-derived inflammatory biomarkers are inexpensive and more accessible compared to other radiological tools in primary care settings. These findings may have clinical relevance in pre-clinical settings or hospitals with limited resources and can guide clinicians when taking preventative measures.

**Ethics**

**Ethics Committee Approval:** Ethical approval was obtained from the Adana City Training and Research Hospital Clinical Research Ethics Committee (approval no.: 1367, date: 08.04.2021).

**Informed Consent:** Informed consent was obtained from all individual participants included in the study.

**Authorship Contributions**


**Conflict of Interest:** No potential conflict of interest was reported by the authors.

**Financial Disclosure:** The author(s) received no financial support for the research, authorship, and/or publication of this article.

**References**