

SIUP | SOCIETA' ITALIANA DI UROLOGIA PEDIATRICA
41° CONGRESSO NAZIONALE
SORRENTO 6-8 NOVEMBRE 2025

26. HIERARCHICAL CLASSIFICATION OF TORSION AND NON-TORSION IN THE DIAGNOSIS OF ACUTE SCROTUM IN PEDIATRIC SETTINGS THROUGH MACHINE LEARNING APPROACHES.

Fabiola Cassaro ⁽¹⁾ - **Giovanni Cicceri** ⁽²⁾ - **Santi D'Antoni** ⁽¹⁾ - **Pietro Impellizzeri** ⁽¹⁾ - **Salvatore Vitabile** ⁽²⁾
- **Salvatore Arena** ⁽¹⁾ - **Carmelo Romeo** ⁽¹⁾

⁽¹⁾ UOC Chirurgia Pediatrica, AOU "G.Martino", Università degli Studi di Messina, Dipartimento di Patologia Umana dell'Adulto e dell'Età Evolutiva "Gaetano Barresi", Messina, Italia - ⁽²⁾ Dipartimento di Biomedicina, Neuroscienze e Diagnostica Avanzata, Università di Palermo, Palermo, Italia

INTRODUCTION AND AIM OF THE STUDY

Acute scrotum is a pediatric emergency that requires prompt diagnosis to avoid irreversible damage. The aim of this study was to develop and evaluate a hierarchical machine learning (ML)-based diagnostic workflow to accurately classify torsion versus non-torsion cases in patients treated for acute scrotum using clinical data and ultrasound (US) parameters.

MATERIALS AND METHODS

A retrospective study was conducted on 111 patients (0–19 years; median 12) diagnosed with acute scrotum. Clinical signs assessed included: pain onset <8 hrs, swelling, erythema, absent cremasteric reflex, increased consistency, palpation pain, and retraction. US signs included: increased volume, heterogeneity, absent vascular signals, and the “whirlpool sign.” Clinical and US signs were correlated with final diagnoses. A two-level ML classification approach was used: the first level trained models only on clinical features, to train several ML models (70/30 stratified split, GridSearchCV with 5-fold CV); the second level applied models to US features, limited to patients classified as non-torsion in level one.

RESULTS

In the first-level, the best ML classifier was SVM with RBF kernel, C=1, gamma='scale' hyperparameters. This model achieved an accuracy of 91.18%, precision of 90.48%, recall of 95.00%, F1-score of 92.68%, and AUROC of 91.07% on the test set. In the second-level, the same test set was maintained, and separate ML-based classifiers were trained exclusively using US-derived features. The best model (SVM), optimized with C=1 and gamma='auto', achieved an accuracy of 85.29%, precision of 82.61%, recall of 95.00%, F1-score of 88.37%, and AUROC of 87.68%.

INTERPRETATION OF RESULTS

One false negative occurred at the first level, but the second-level US-based model correctly identified the case as torsion, resulting in 100% recall overall for testicular torsion cases by compensating for the clinical model's limitation. To improve the interpretability of the best predictive ML model, XAI methods were applied using SHAP (SHapley Additive exPlanations) values. SHAP results showed age, increased testicular consistency, and absent cremasteric reflex as the most influential features, while erythema and tenderness had minimal impact. This level of explainability reinforces clinical confidence in the use of AI-based systems, particularly in emergency settings where decision-making must be rapid and justified.

CONCLUSIONS

A hierarchical ML approach combining clinical and US data ensured robust diagnostic support in pediatric acute scrotum. By cascading predictions, this strategy prevents false negatives in torsion diagnosis, offering a clinically valuable safeguard in emergency settings. The findings on XAI not only confirm the central role of certain clinical predictors but also provide a transparent rationale behind the ML model's decisions.