Published in INFOR, (forthcoming)

DEVELOPMENT OF A DECISION ALGORITHM TO SUPPORT EMERGENCY TRIAGE OF SCROTAL PAIN AND ITS IMPLEMENTATION IN THE MET SYSTEM

WOJTEK MICHALOWSKI

School of Management
University of Ottawa
Ottawa, Ontario, Canada
e-mail: wojtek@management.uottawa.ca

SZYMON WILK*

Institute of Computing Science
Poznan University of Technology
Poznan, Poland
e-mail: szymon.wilk@cs.put.poznan.pl

KEN FARION, JOHN PIKE, STEVEN RUBIN

Children's Hospital of Eastern Ontario
Ottawa, Ontario, Canada
e-mail: farion, pike, rubin@cheo.on.ca

ROMAN SŁOWIŃSKI

Institute of Computing Science
Poznan University of Technology
Poznan, Poland
e-mail: roman.slowinski@cs.put.poznan.pl

ABSTRACT

The acute scrotum is a clinical condition in boys and adolescent males that is normally first assessed in the Emergency Department of a hospital. We used data from the patients' charts and applied knowledge discovery technique based on rough set theory to develop a clinical decision algorithm for triaging patients with this condition. As demonstrated by a limited retrospective evaluation, the algorithm supports early triage decisions on the basis of readily available information, resulting in good triage accuracy. In order to make the algorithm usable in clinical practice, to integrate it with the workflow, and to make it available at the point of care, we

* This research was conducted while Dr. Wilk was Postdoctoral Fellow in the School of Management, University of Ottawa.

implemented it as an application in the mobile clinical decision support environment called MET (Mobile Emergency Triage). MET uses ontologies to represent domains of various acute presentations and triage support functionalities, and renders specific applications on demand from these ontologies.

Keywords: mobile health, clinical decision support system, decision algorithm, knowledge discovery, ontological model, rough set theory, acute scrotum

1. Introduction

The acute scrotum (scrotal swelling and/or pain) in boys and adolescent males is an infrequent but challenging problem that is typically first assessed by Emergency Department (ED) physicians. The usual causes of acute scrotum are torsion of a testicular appendage, testicular torsion, or infection of the testis/epididymis [1]. True testicular torsion is a surgical emergency, with a limited window of opportunity of 4-12 hours from the onset of pain to make the diagnosis and untwist the spermatic cord in order to salvage the testicle [2]. Making an accurate diagnosis of testicular torsion can be difficult clinically, and given the limited window of opportunity, surgical consultation occurs frequently.

Historically, management of acute scrotum involved emergent surgical exploration, especially if there was any diagnostic uncertainty. With the advent of diagnostic tools such as the nuclear medicine testicular flow scan and color Doppler Ultrasound, a more discriminating approach has become the norm. Experience has proven, however, that these diagnostic modalities are imperfect and cannot replace clinical acumen.

The ED triage function of the acute scrotum carried out by the ED physician is an important part of the management process. Following a limited evaluation of clinical signs, symptoms and tests under significant time pressure, the ED physician must decide whether the patient's problem falls into one of three broad triage classes:

- *Discharge*, which indicates the patient has a benign and resolving condition, and can be discharged home;
- *Clinic*, which indicates the patient has a non-urgent condition requiring further evaluation in the urology clinic in the coming days to weeks;

- *Consult*, which indicates the patient has a suspected testicular torsion or other surgical emergency, and should be immediately seen by the urologist.

The issue faced by the ED physician is to quickly and accurately identify those patients with potential testicular torsion, so the urologist can intervene before permanent loss of the testicle, while not over-burdening the surgeon with acute referrals for non-emergent problems, which are more common. This decision situation is an area in medicine where a triage support tool could be clinically useful. Such a tool could prompt the ED physician to systematically collect pertinent and readily available information and then provide a patient-specific recommendation based on this information.

In response to this perceived need, we have developed a scrotal pain decision algorithm[1] that captures and applies the clinical knowledge used in the triage decision-making process. The algorithm was created by analyzing historical ED data for discriminating patient information that would predict the most appropriate triage category (discharge, clinic, consult) and associated patient management plan. Following the structure of other decision algorithms implemented in clinical practice [3], we chose to use decision rules as a format of representing the clinical knowledge encapsulated in the scrotal pain decision algorithm.

To maximize the clinical usefulness of the decision algorithm, it has been designed as a support tool maintaining the following conditions recently identified as critical success factors [4]:

- The support should be computer-based;
- The support functionality must be aligned with the clinical workflow;
- Support "on demand" must be available at the point of care;
- Support should be provided in an advisory mode.

In meeting these conditions, we have implemented the scrotal pain triage decision algorithm as an application in MET (Mobile Emergency Triage) [5, 6]. MET is a fourth-generation mobile health [7] decision support environment that can be used for developing and executing applications supporting triage of various acute ED presentations [6]. MET uses ontologies to represent

[1] The term "decision algorithm" is used throughout this paper to describe a decision model representing the (clinical) knowledge necessary to solve a decision problem, coupled with a solver (classifier) that uses this knowledge in order to derive at a solution (recommendation).

domains of these presentations and the support functionality they require. The ontological models are then used to render on-demand a complete application suited for the specific presentation and specific access device (i.e., a handheld, tablet or a notebook computer). MET's design has been successfully validated during a clinical trial of the abdominal pain triage support application (MET-AP) [8, 9].

Thus, the goal of the research described in this paper is two-fold. First, we set out to develop a decision algorithm that would support the ED triage of patients with acute scrotum by recommending the appropriate management path for specific patients. The second goal was to implement this scrotal pain triage decision algorithm in the MET environment (as the MET-SP application) using an ontological model describing the problem domain and triage support functionality. The resultant MET-SP would allow physicians to evaluate a patient's condition, facilitate the collection of relevant data at a point of care, and provide a triage recommendation regarding possible management paths.

The paper is organized as follows. We start with a description of the retrospective chart data, knowledge discovery methodology based on rough set theory that was utilized with this data and the process followed to identify the most appropriate decision algorithm. This is followed by a description of the algorithm, the results of its preliminary evaluation, and the creation of MET-SP application. Finally, we conclude with a discussion.

2 METHODS

2.1 Retrospective Chart Study

The data used to build the scrotal pain decision algorithm were collected through a retrospective review of patients presenting to the ED of Children's Hospital of Eastern Ontario (CHEO) with acute scrotal pain and/or swelling between 1993 and 2002. This single centre was chosen and felt to appropriately represent a potential sample of pediatric patients with acute scrotal problems, as it is the only pediatric hospital providing urological consultation to a regional population of 400,000 children (2001 data). As epidemiological studies have not shown differences in the incidence of pediatric disease of the genitalia between Canadian, U.S., and European populations, it is reasonable to conclude that our sample is representative of other populations.

The initial selection of patient charts was based on both the presenting complaint as well as the discharge diagnosis in an attempt to capture as many patients with acute scrotum as possible. Each patient visit was then evaluated against a set of inclusion and exclusion criteria to identify patients with spontaneous acute scrotum where an etiology or diagnosis was not already known. In this way, the inclusion/exclusion criteria define the potential group of patients for which the decision algorithm could be clinically useful at the ED triage stage of management. Charts were included if the patient was between 1 and 17 years of age (older patients are referred to adult hospitals and not seen in CHEO) and the presenting complaint was scrotal pain or swelling (duration up to 10 days), as expressed by patient or parent, or found as part of the patient's presentation. Exclusion criteria included scrotal trauma and scrotal pain or swelling caused by an acute disease process for which the patient was already undergoing treatment (i.e., already diagnosed with epididymitis, but pain or swelling worsening). Data for thirteen commonly evaluated and typically documented clinical attributes associated with acute scrotum were abstracted from selected charts by a single abstractor. The abstractor was not blinded to the patient's final outcome, but did not assign the final patient category.

The abstracted attributes included historical symptoms, physical exam signs, and basic laboratory tests, as outlined in Table 1.

Table 1. Clinical attributes

	Code	Description
1	AGE	Age of the patient
2	CORD	Was the spermatic cord palpable?
3	LIE	Positional lie of the testicle
4	PAINONSET	Onset of pain – sudden or gradual?
5	PAINSITE	Site of pain
6	PAINTYPE	Type of pain – constant or intermittent?
7	REFLEX	Presence of the cremasteric reflex
8	SWELLING	Swelling of testes
9	TEMPR	Temperature
10	TENDER	Tenderness on palpation of the testes
11	VOMIT	History of vomiting
12	WBC	White blood cell count from blood sample
13	WBCHPF	White blood cells per high power field on urine microscopy

One pediatric urologist reviewed each selected patient and assigned the patient to one of the three triage categories using the final diagnosis established during ED visit and any subsequent follow-up documented in the chart. Such process allowed us to identify those cases where the initial ED triage was incorrect (e.g., discharged patients with persisting problem had another visit to the ED with the similar complaint) and established appropriate triage decision. This triage category, determined to be the correct course of action that should have occurred for the patient given the final diagnosis, became the gold standard used during development and evaluation of the scrotal pain decision algorithm.

2.2 Knowledge Discovery

The scrotal pain triage decision algorithm was built using a discovery-driven approach [10] in order to avoid the Feigenbaum bottleneck associated with manual management and update of a large body of extracted knowledge [11]. In considering the type of clinical data available and our past experience with developing MET-AP [12], we decided to use rough set theory [13, 14] for extracting knowledge from historical patient data recorded in the medical record. Rough set theory is a convenient discovery-driven method for the analysis of data where data objects are described by means of condition and decision attributes. Condition attributes describe the objects in terms of available information and a decision attribute partitions these objects into groups (decision classes). In our clinical data set, patients were considered as objects, clinical attributes as the condition attributes, and the gold standard triage decision as the decision attribute allowing us to partition the patients into three triage categories.

The usefulness of rough set theory for discovering knowledge from data comes from its capability to effectively handle inconsistently described objects (i.e., where patients have the same values for clinical attributes yet belong to different triage categories), and objects with incomplete information (i.e., written medical records are rarely standardized in format and content, thus resulting in a large number of missing values when abstracted) [15]. Rough set theory deals with incomplete information by assuming that a missing value for an attribute is equivalent to any possible value for that attribute. This simplification does not distinguish between the situation where values were collected but not recorded and the situation where the values were never collected (i.e., question never asked, examination or test not performed) because the physician felt the information was unnecessary, but such a distinction can rarely be inferred from patient charts. Nevertheless, it is a suitable methodology for developing decision algorithms from the type of extracted data that is available in a clinical domain. Rough set theory offers several measures to assess the informational value of individual condition attributes and their subsets [16]. One such measure is the Shapley value [17] that indicates how well a condition attribute explains relationships in data – the attribute with the highest value is considered to be the most informative. Rough set theory also facilitates representing patterns and relationships between condition and decision attributes in form of *decision rules*, yielding a *decision model*. A decision rule may be characterized by several properties – the most important ones are *coverage* and *consistency*.

The coverage of a rule is defined as the percentage ratio of the number of objects verifying the condition and the decision parts of a rule to the number of objects in the decision class indicated by a rule. In our scenario, the coverage of a rule refers to the proportion of patients in a given triage class that have values of clinical attributes specified in the condition part of the rule. Consider for instance the following rule with lets say, 12.8% coverage and 80% consistency:

"*if* pain is intermittent, and testis is tender in upper pole, *then* triage recommendation is clinic"

The value of the coverage equal to 12.8% means that 12.8% of patients who were sent to clinic had intermittent pain and tender testis in upper pole.

The consistency of a rule is defined as the percentage ratio of the number of objects verifying the condition and the decision parts of a rule to the number of objects verifying only the condition part. If the consistency of a rule is below 100%, it means that some objects in the data set that satisfy the condition part of the rule belong to a decision class other than the one indicated in the decision part (thus the decision suggested by a rule cannot be considered as certain). The value of the consistency equal to 80% for the above rule means that 80% of patients with intermittent pain and tender testis in upper pole should have been sent to clinic. Another way to interpret consistency is to state that triage recommendation provided by this rule would be 80% certain.

Rough set theory provides various techniques to generate decision models composed of rules. Depending on their characteristics, a resulting rule-based decision model can be composed of one the following sets:

- *Minimal set of rules*: includes the minimal number of rules required to cover all objects in the data set; this is the concise and complete model,

- *Exhaustive set of rules*: includes all possible rules that can be induced from the data; this is complete but not concise model,
- *Satisfactory set of rules*: includes all possible rules that satisfy pre-defined requirements usually associated with the rule's properties; depending on these requirements the model may be concise but not complete.

There are several strategies (also called classifiers or solvers) of applying a rule-based decision model to classify new objects (to predict the triage decision for a new patient). Usually these strategies use some measure of similarity or distance between a rule and a classified object [14, 18], and are called similarity-based (or distance-based) classifiers. When applying rules to classify an object, the classifier calculates the similarity between each rule in the decision model and the object. The measure considers the number of conditions in a rule that have the same values as values of the condition attributes describing an object (we say that they are verified by these values). If all conditions are verified (the object's values *exactly match* the rule), then the similarity is maximal (equal to 100%), otherwise, if some conditions are not verified (the object's values only *partially match* the rule), the similarity diminishes in relation to the number of conditions that are not verified, until it reaches a value of 0%.

In a usual classification strategy, the classifier identifies first those rules that are exactly matched by an object in question. If there are none, then it searches for the most similar partially matched rules (or rules with similarity above a certain threshold value). If more than one rule has been identified (either exactly or partially matched), the classifier aggregates their suggestions by calculating a *strength factor* for each decision class that is a sum of coverage of all identified rules that point to this class. The class with the greatest strength factor is labelled the recommended one; however, the classifier can also provide results and strength factors for the remaining classes. If no rule is exactly or partially matched (i.e., similarity equals 0%), the classifier is unable to generate a recommendation; such deadlock can be resolved by recommending the dominant decision class or some default decision class. This two-phase strategy employed by a similarity-based classifier allows applying decision rules on objects that are inconsistent with the decision model (there are no rules exactly matching these objects) or that are described by incomplete information.

The basic measure of performance of a decision algorithm (a model and a classifier) is the *accuracy of classification*, defined as a percentage of the ratio of the correctly classified objects to all classified objects. However, for the evaluation of a clinical decision algorithm, this measure is not sufficient and it is often augmented with *sensitivity* and *specificity* [19]. The experimental design using these two additional measures requires the binary classification of objects into *positive* and *negative* decision classes, where the positive class includes all objects of special interest, and the negative class includes the remaining ones. For the triage of scrotal pain, patients requiring consult are considered to belong to the positive decision class, and those who can be discharged home or sent to clinic belong to the negative decision class. Accordingly, sensitivity represents the probability that a “positive patient” is correctly classified, while specificity represents the probability that a “negative patient” is correctly classified. Both probabilities can be used to define a measure representing *the gain of the probability of the correct prediction* [20] (for simplicity we will use the term “gain” hereafter), defined as [*sensitivity* + *specificity* – 1]. This measure has been successfully used in clinical studies [21, 22] where it proved to be more informative than the simple accuracy of classification. A decision algorithm generating the highest gain is considered to have the best performance.

2.3 Evaluation of the Decision Algorithm

In order to evaluate quality of the decision algorithm developed with a help of rough set theory it was necessary to conduct an experiment that compared its performance to some well-established standard. In the study described in this paper we used retrospective chart data and compared results generated by the algorithm with results provided by color Doppler Ultrasound, which is perceived to be a very reliable diagnostic test to differentiate patients when there is uncertainty [23]. Data transcribed from the ED charts were used as input to the decision algorithm, and the suggested decision class was considered as output to be compared to the outcome of the diagnostic test (color Doppler Ultrasound) and the gold standard (verified triage decision).

In clinical practice color Doppler Ultrasound is used as a differential diagnostic tool and its findings can be condensed into a binary outcome:

- *Positive*, in which testicular torsion is highly suspected by absent or significantly diminished testicular blood flow and an emergent urological consult should be requested,

- *Negative*, in which a benign or non-emergent condition is evident by normal or increased testicular blood flow.

In order to compare the performance of color Doppler Ultrasound and the performance of the rule-based decision algorithm against the gold standard, we assumed a binary recommendation by the decision algorithm and by the gold standard – *consult* was treated as positive, and *discharge* and *clinic* were combined into negative class. We used all the performance measures described earlier (accuracy, specificity, sensitivity, and gain) to evaluate operation of the decision algorithm.

3. RESULTS

3.1. Data Used

A final sample of 171 ED patient visits for acute scrotal pain and/or swelling were identified over the ten-year period, following application of the search, inclusion and exclusion criteria. Basic characteristics of this population are given in Table 2. It can be seen that most of the patients treated in the ED came with non-emergent scrotal pain and could be sent to urology clinic.

Table 2. Characteristics of the data set

Characteristic	Value
Mean age	10.5 years
% of patients in the discharge triage category	11.7%
% of patients in the clinic triage category	59.6%
% of patients in the consult triage category	28.7%

As expected from abstracting historical records, the majority of charts had incomplete data. Table 3 outlines the proportion of charts in which data for each attribute was incomplete. Of note, five attributes: cord palpable (CORD), positional lie of the testicle (LIE), onset of pain (PAINONSET), cremasteric reflex (REFLEX), and white blood cells in high power field (WBCHPF), had missing data in more than 20% of cases, and the white blood cell count from blood sample (WBC) was missing in over 90% of cases.

Table 3. Proportion of cases with missing values for each clinical attribute

	Attribute code	Missing values [%]
1	AGE	0.0
2	CORD	60.2
3	LIE	24.6
4	PAINONSET	23.4
5	PAINSITE	0.6
6	PAINTYPE	4.1
7	REFLEX	33.9
8	SWELLING	9.4
9	TEMPR	11.1
10	TENDER	0.6
11	VOMIT	15.2
12	WBC	90.6
13	WBCHPF	33.9

Most of the attributes had nominal values while urology specialist discretized numeric values for age of the patient (AGE), temperature (TEMPR), white blood cells from blood sample (WBC), and white blood cells in high power field on urine microscopy (WBCHPF).

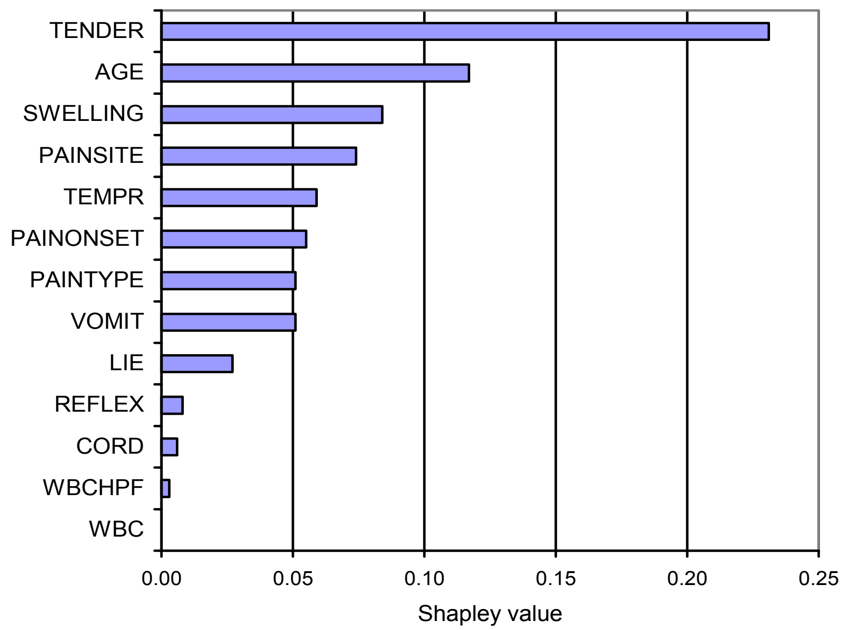
3.2. Development of the Scrotal Pain Decision Algorithm

Development of the scrotal pain decision algorithm was conducted in three phases:

1. Evaluation of the information values of the attributes;
2. Development of different decision models and cross-validation tests;
3. Selection of the best decision model and development of the decision algorithm.

In the first phase, we calculated the Shapley value for each of the clinical attributes. Results are given in Figure 1. This analysis revealed that tenderness of the testes (TENDER) was the most informative attribute; it also shows that the laboratory tests (WBCHPF and WBC) were of almost no use.

Figure 1. Information value of the clinical attributes



During the second phase, we developed several decision models using possible subsets of attributes starting from the one containing the top three attributes – TENDER, AGE and SWELLING, and subsequently adding less informative ones, until we ended with the subset containing all attributes except WBC. For each subset, we generated decision rules and checked their performance (in terms of their accuracy and gain) using 10-fold cross-validation test [24]. During this test, the analyzed data set was 10 times automatically divided into a learning subset and a testing subset. The learning subset was used for generating rules, while the testing subset was applied to check the performance of the rules and of the similarity-based classifier. The results were averaged over the 10 repetitions.

For a more thorough comparison, we decided to check two different approaches to construction of rule-based decision models in order to see which one would be more appropriate for the analyzed data. Namely, we created a decision model applying the LEM2 algorithm [18] that generates a minimal set of rules, and compared it to a decision model created with the EXPLORE [25] algorithm that generates exhaustive or satisfactory sets of rules. The latter algorithm searches the space of possible rules and retains those that satisfy pre-defined requirements stated as thresholds for minimal rule coverage or consistency (if no requirement is defined, the exhaustive set of rules is generated). In our case, the consistency was set to 100%, and the minimal coverage of rules was set according to the procedure described in [26]. The decision models were evaluated with the testing data using a similarity-based classifier [14]. When evaluating partially matched rules, the classifier was set to accept rules with the similarity score equal to 50% or more, and in a situation of deadlock no recommendation was generated (namely, such a situation was considered as misclassification). We decided against setting any default triage class during the tests in order to get clear indication of model's performance (asssigning the default class might increase the accuracy of classification for this class).

The results of cross-validation tests for different models are outlined in Table 4. Rows in this table refer to specific subsets of attributes, starting with a model using only the three attributes with the highest Shapley values, and followed by the models created by adding additional attributes in step-wise fashion until all but WBC were included. The most appropriate decision model in the cross-validation tests was generated by EXPLORE for the subset of 11 condition attributes (all attributes except WBC and WBCHPF). Gain measure for this model was equal to 0.60, and the overall classification accuracy to 72.5%. A slightly higher accuracy (73.0%) was achieved for the EXPLORE generated model using 12 attributes (all except WBC), however it has a lower gain resulting from lower specificity.

Table 4. Results of cross-validation tests for different decision models

	LEM2		EXPLORE	
Model	Accuracy	Gain	Accuracy	Gain
Model with 3 attributes	62.6%	0.30	13.4%	-0.70
Model with 4 attributes	61.4%	0.34	42.0%	-0.11
Model with 5 attributes	63.1%	0.36	47.3%	-0.01
Model with 6 attributes	61.9%	0.40	64.9%	0.26
Model with 7 attributes	66.6%	0.50	67.2%	0.42
Model with 8 attributes	69.6%	0.49	73.1%	0.52
Model with 9 attributes	69.5%	0.54	71.3%	0.51
Model with 10 attributes	67.8%	0.49	70.2%	0.56
Model with 11 attributes	66.0%	0.46	72.5%	0.60
Model with 12 attributes	66.6%	0.49	73.0%	0.59

In the third phase of the experiment we constructed the scrotal pain decision algorithm. It included a decision model identified in the second phase and a similarity based classifier. The model consisted of 123 rules, with a few examples presented in Table 5 for illustration. When constructing the algorithm, we decided to use the clinic triage class as the default one to be selected in case of the deadlock. Rationale for such a choice is that the algorithm should always arrive at a recommendation, and it should be a "conservative" one if there is no ability to come to a definite conclusion.

Rough set theory generates consistent rules that do not require explicit verification by the experts [27]. Fortunately, our model preserves the links between rules and the input data, so it is always possible to identify the data objects (patients) that led to a particular rule and give clinicians the opportunity to review and confirm the clinical validity of the rule.

Table 5. Sample decision rules from the scrotal pain decision model

if age is 11 years or older, and pain is located in both testes, and testes are not tender, *then* triage recommendation is discharge (coverage = 5.0%)
if age is below 11 years, and pain is located in both testes, and temperature is lower than 37°C, *then* triage recommendation is discharge (coverage = 10.0%)
if pain is intermittent, and testis is tender in upper pole, *then* triage recommendation is clinic (coverage = 12.8%)
if onset of pain is gradual, and testis is tender in upper pole, and temperature is between 37 and 39°C, *then* triage recommendation is clinic (coverage = 14.7%)
if vomiting is present, and left testis is swollen, and lie is elevated, *then* triage recommendation is consult (coverage = 14.3%)
if pain is constant, and vomiting is present, and left testis is swollen, and cord is abnormal, *then* triage recommendation is consult (coverage = 22.5%)

3.3. Evaluation of a Scrotal Pain Decision Algorithm

The scrotal pain decision algorithm was evaluated on new data (not used for its development and testing) transcribed retrospectively from 30 charts of patients admitted to the ED of CHEO between 1997 and 2002. This testing data has general characteristics (mean age of a patient, distribution of patients among triage classes) that are similar to the characteristics of a learning set used for algorithm development (see Table 6).

Table 6. Characteristics of the new testing data set

Characteristic	Value
Mean age	10.1 years
% of patients in the discharge triage category	13.3%
% of patients in the clinic triage category	53.3%
% of patients in the consult triage category	33.4%

The transcription process followed the same regimen as described for the retrospective chart study, but was limited only to those patients that had color Doppler Ultrasound results recorded in the chart. In order to compare the performance of color Doppler Ultrasound to the performance of the decision algorithm against the gold standard, we imposed a binary classification on the decision algorithm recommendations and gold standard triage classes – *consult* was treated as positive, and *discharge* and *clinic* were combined into negative class.

Binary classification enabled us to calculate the measures of sensitivity and specificity for the scrotal pain decision algorithm and color Doppler Ultrasound (see Table 7). The decision algorithm achieved accuracy of 76.7% and turned out to be a very good predictor for patients requiring consult (sensitivity equal to 0.90). Its specificity was lower which indicates that the algorithm acted conservatively and assigned the consult triage recommendation to some patients who did not require it. No deadlock occurred when applying the algorithm to the testing data set.

Table 7. Performance of the scrotal pain decision algorithm and color Doppler Ultrasound

	Decision algorithm	Color Doppler Ultrasound
Sensitivity	0.90	0.50
Specificity	0.70	0.95
Gain	0.60	0.45
Accuracy	76.7%	80.0%

Compared to the decision algorithm, the classification accuracy of color Doppler Ultrasound was slightly higher (80% versus 76.7%), but its gain measure was lower (0.45 versus 0.60). Performance of color Doppler Ultrasound on a testing data set was surprisingly poor (sensitivity 0.50), compared to results reported elsewhere (sensitivity 0.99, specificity 0.89) [28]. However, even taking into account these reported results, the decision algorithm is still a viable alternative to the color Doppler Ultrasound in that it achieved comparable ability to identify the critical group of patients (i.e., those with *positive* outcome) with no need for expensive and time-consuming diagnostic testing. The algorithm was less accurate in identifying *negative* patients, but considering its advisory character, such inaccuracy seems to be acceptable.

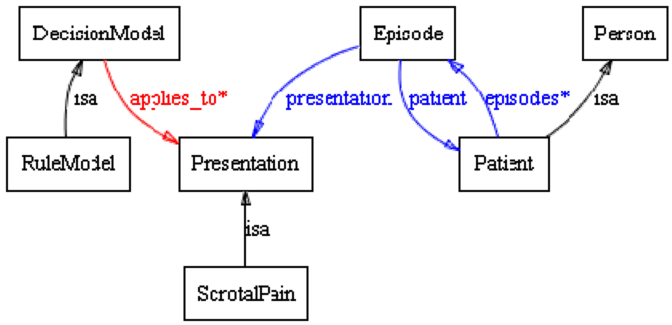
3.4. Development of the MET-SP Application

The design of the MET clinical decision support environment was inspired by research on decision support systems with separated domain ontologies and solvers [29], where the domain ontology enumerates and organizes classes of concepts in a problem area (i.e., the concepts of a patient, presentation, or decision model), and the solver (or classifier) is a general strategy that can be applied to any domain ontology in order to solve the described problem. Specifically, the domain ontology is applied as a schema to build a decision model that is later processed by the solver to arrive at the suggested recommendation. This separation enhances reusability, as different decision models can use one solver. This is important for the MET environment, as all decision algorithms integrated so far share the same solver (the similarity-based classifier). Moreover, the separation also increases the versatility and flexibility of MET, as defining a new application usually requires extending the ontology to cover its domain.

MET includes a domain ontology common for all emergent acute presentations, which is then specialized for each application implementing a specific decision algorithm. Specialization is easily accomplished utilizing the definition of the specific presentation class, which contains the clinical attributes appropriate for triaging patients with the particular presentation. Implementing a decision algorithm requires the creation of an instance of the class representing the particular decision model in the domain ontology. If a new decision algorithm uses a solver that is not available in MET environment, it has to be added.

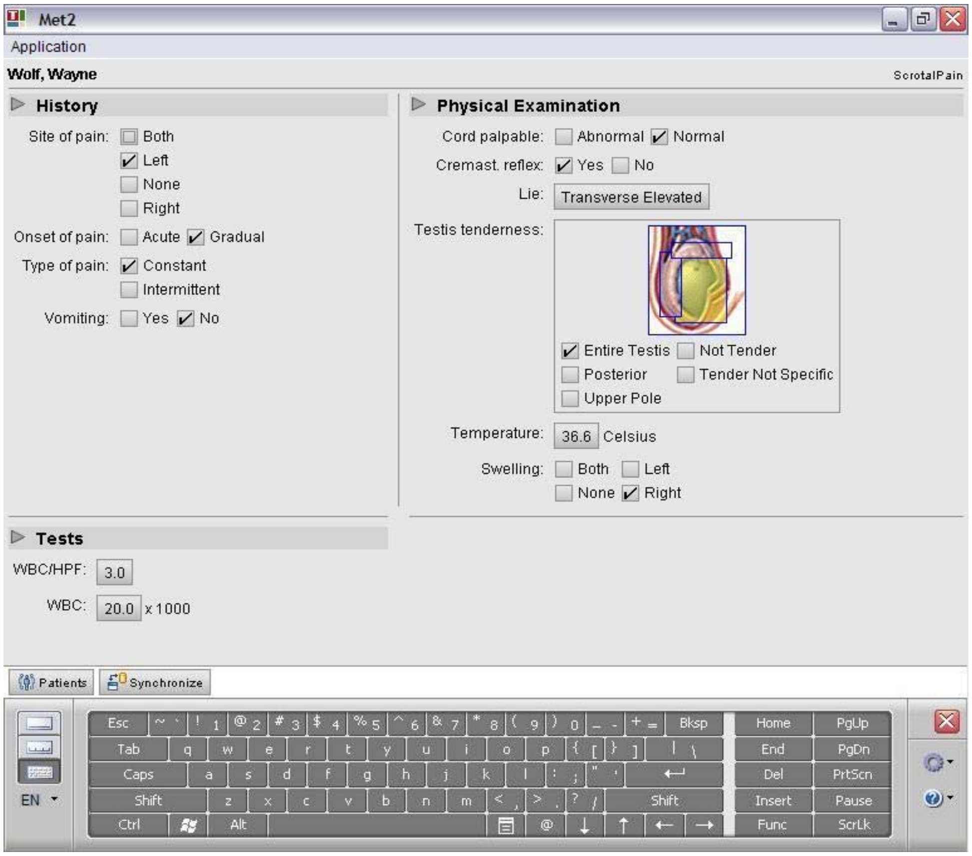
Figure 2 presents the general ontology for the MET-SP application, where the *ScrotalPain* class includes all the attributes needed for triage of acute scrotum. For MET-SP we created the instance of *RuleModel* (derived from *DecisionModel*) with all 123 rules discovered during the analysis of retrospective data. The similarity-based classifier required for MET-SP application had been previously implemented and was reused without need for customization.

Figure 2. Domain ontology specialized for the MET-SP application



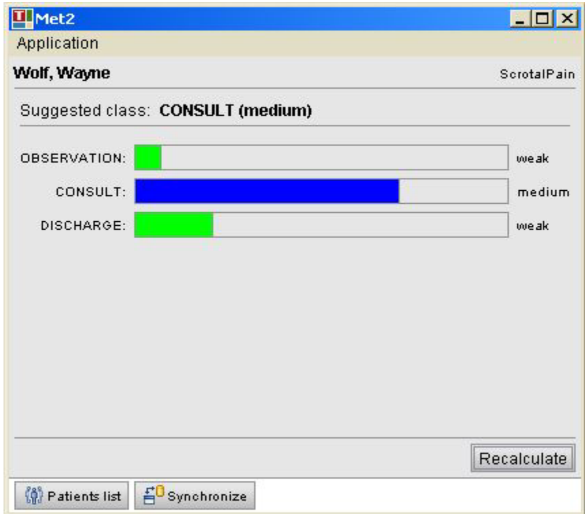
The MET environment was developed for the well-defined domain of emergency triage and all supported applications have to fit the same workflow. This allowed us to prepare generic interface templates for all supported computing platforms and applications. These templates were carefully designed [30] following high level methodologies suggested for clinical systems (user-centered and task-centered design [31]). An example of a template for the MET-SP application is given in Figure 3.

Figure 3. User interface of the MET-SP for a tablet computer



The advisory character of MET-SP empowers clinicians to make independent triage decisions [32, 33]. In order to reinforce this concept of “advisor” rather than “decider”, we present the user with results for all possible outcomes along with strength factors. (see Figure 4). This offers a more complete assessment of the clinical situation and lets the physician evaluate the recommendations against his/her intuition and current hypothesis.

Figure 4. User interface for reporting triage suggestions



To date, the MET-SP system has not been deployed or prospectively evaluated in clinical practice. However, the MET-AP system for triaging pediatric abdominal pain was successfully tested during a clinical trial in the ED at CHEO. The trial ran for 7 months and the application was used 24 hours a day, 7 days a week by more than 150 different members of the ED medical staff, including physicians and residents [8, 9]. This trial demonstrated that the MET clinical decision support environment in general, and the MET-AP in particular, were aligned with the emergency workflow and accepted by clinicians. These results suggest that MET-SP, developed according to the same design principles as MET-AP, should be equally acceptable in clinical practice.

5. DISCUSSION

The research described in this paper shows how we have created a decision algorithm to support early ED triage of an acute pediatric condition using a versatile knowledge discovery methodology (rough set theory). It also discusses how ontological modeling helps to implement the algorithm as an application in the MET environment.

Our retrospective evaluation of the algorithm, in comparison with color Doppler Ultrasound, demonstrated that it is a viable aid to triaging patients in the ED with acute scrotal pain/swelling. Despite the limitations associated with the retrospective data from a single centre used to develop and validate the decision algorithm, our initial results support the general approach we have adopted in constructing decision algorithms and implementing them in the MET clinical decision support environment. We are also convinced that an additional advantage associated with this type of computer-based support is a result of enforcing structured data collection. Other researchers have shown that such functionality alone positively impacts overall diagnostic accuracy [34].

The success of our research relies on two factors:

1. The use of an hybrid approach to the decision algorithm development, which combines a flexible data mining methodology (rough set theory) with evaluation of the information value of the clinical attributes, and a comprehensive way of creating an underlying decision model;
2. The use of an ontological model to represent the clinical problem domain and required triage support functionality, thus enabling us to construct a high level abstract representation that can be later rendered for a specific application and a computing platform.

This is a novel approach to the development of clinical decision support system, as it creates seemingly individual applications (abdominal pain, scrotal pain, etc) that function in a broader triage support context and render complete models (decision model and an interface) "on demand".

The MET-SP is intended for use in the ED by clinicians with varying level of clinical acumen – junior medical residents to senior staff physicians. However, our earlier experience suggests that more experienced users can provide more accurate values of clinical attributes and thus obtain more reliable triage recommendations. From this perspective, experienced clinicians may benefit more from using the triage support application than the less experienced ones, while the latter group may gain more from the structured data collection. We plan to verify these hypotheses in future research as we prospectively evaluate the decision algorithm and the system implementation in clinical care.

Acknowledgements

Research reported in this paper was supported by NSERC-CHRP grant. The authors would like to acknowledge contributions of M. Kersten, N. Milman, R. Payne, B. Plouffe and D. Weiss who were members of MET Research Team during work on MET-SP. Comments of the editors and the reviewers helped to improve quality of the paper.

REFERENCES

1. Lewis AG, Bukowski TP, Jarvis PD, Wacksman J, Sheldon CA. Evaluation of acute scrotum in the emergency department. Journal of Pediatric Surgery 1995;30(2):277-81; discussion 281-2.

2. Knight PJ, Vassy LE. The diagnosis and treatment of the acute scrotum in children and adolescents. Annals of Surgery 1984;200(5):664-73.

3. Rae F. The Ottawa ankle rule. Journal of Emergency Medicine 2001;18(2):147.

4. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. BMJ 2005;330(7494):765.

5. Michalowski W, Slowinski R, Wilk S. MET System: A new approach to m-health in emergency triage. Journal on Information Technology in Healthcare 2004;2(4):237-249.

6. Michalowski W, Slowinski R, Wilk S, Farion K, Pike J, Rubin S. Design and development of a mobile system for supporting emergency triage. Methods of Information in Medicine 2005;44(1):14-24.

7. Daou Systems Inc. Going mobile: from ehealth to mhealth. A Daou Systems White Paper. In; 2001.

8. Michalowski W, Slowinski R, Wilk S, Farion K. Mobile emergency triage: Lessons from a clinical trial. In: Kendall EJ, editor. Decision Sciences Institute 2004. Proceedings. Boston: DSI; 2004. p. 6601-6606.

9. Farion K, Michalowski W, Slowinski R, Wilk S, Rubin S. Rough set methodology in clinical practice: controlled hospital trial of the MET system. In: Tsumoto S, Słowiński R, Komorowski J, Grzymała-Busse JW, editors. Rough Sets and Current Trends in Computing. 4th International Conference, RSCTC 2004; 2004 June 2004; Uppsala, Sweden: Springer-Verlag; 2004. p. 805-814.

10. Hanson CW, 3rd, Marshall BE. Artificial intelligence applications in the intensive care unit. Critical Care Medicine 2001;29(2):427-35.

11. Steel L. Second generation expert systems. Future Generation Computer Systems 1985;1(4):213-221.

12. Wilk S, Slowinski R, Michalowski W, Greco S. Supporting triage of children with abdominal pain in the Emergency Room. European Journal of Operational Research 2005;160(3):696-709.

13. Greco S, Matarazzo B, Slowinski R. Rough set theory for multicriteria decision analysis. European Journal of Operational Research 2001;129:1-47.

14. Stefanowski J. Classification support based on the rough sets. Foundations of Computing and Decision Sciences 1993;18(3-4):371-380.

15. Greco S, Matarazzo B, Slowinski R. Dealing with missing data in rough set analysis of multi-attribute and multi-criteria decision problems. In: Zanakis SH, Doukidis G, Zopounidis C, editors. Decision Making: Recent Developments and Worldwide Applications: Kluwer Academic Publishers; 2000. p. 295-316.

16. Greco S, Matarazzo B, Slowinski R. Fuzzy measures as a technique for rough set analysis. In: Proceedings of the 6th European Congress on Intelligent Techniques and Soft Computing (EUFIT'98). Aachen; 1998. p. 99-103.

17. Shapley LS. A value for n-person games. In: Kuhn HW, Tucker AW, editors. Contributions to the Theory of Games II: Princeton University Press; 1953. p. 307-317.

18. Grzymala-Busse JW. LERS - A system for learning from examples based on rough sets. In: Słowiński R, editor. Intelligent Decision Support - Handbook of Applications and Advances of the Rough Set Theory. Dordrecht/Boston: Kluwer Academic Publishers; 1992. p. 3-18.

19. Owens DK, Sox HC. Medical decision making: Probabilistic medical reasoning. In: Shortliffe EH, Perreault LE, Wiederhold G, Fagan LM, editors. Medical Informatics. Computer Applications in Health Care and Biomedicine. 2 ed: Springer-Verlag; 2001. p. 76-131.

20. Bairagi R, Suchindran CM. An estimator of the cutoff point maximizing the sum of sensitivity and specificity. Indian Journal of Statistics 1989;51:263-269.

21. Grzymala-Busse JW, Goodwin LK, Grzymala-Busse WJ, Zhang X. An approach to imbalanced data sets based on changing rule strength. In: Learning from Imbalanced Data Sets. AAAI Workshop at the 17th Conference on AI, AAAI-2000. Austin; 2000. p. 69-74.

22. Grzymala-Busse JW, Hippe ZS. Postprocessing of rule sets induced from a melanoma data Set. In: 26th International Computer Software and Applications Conference on Prolonging Software Life: Development and Redevelopment; 2002; 2002. p. 1146-1151.

23. Weber DM, Rosslein R, Fliegel C. Color Doppler sonography in the diagnosis of acute scrotum in boys. European Journal of Pediatric Surgery 2000;10(4):235-41.

24. Mitchell T. Machine learning: McGraw Hill; 1997.

25. Stefanowski J, Vanderpooten D. Induction of decision rules in classification and discovery-oriented perspectives. International Journal of Intelligent Systems 2001;16(1):13-28.

26. Stefanowski J. On rough set based approaches to induction of decision rules. In: Skowron A, Polkowski L, editors. Rough Sets in Data Mining and Knowledge Discovery. Heidelberg: Physica-Verlag; 1998. p. 500-529.

27. Siminski R, Wakulicz-Deja A. Application of decision units in knowledge engineering. In: Tsumoto S, Słowiński R, Komorowski J, Grzymała-Busse JW, editors. Rough Sets and Current Trends in Computing. 4th International Conference, RSCTC 2004; 2004 June 2004; Uppsala, Sweden: Springer-Verlag; 2004. p. 721-726.

28. Baker LA, Sigman D, Mathers RI, Benson J, Docimo SG. An analysis of clinical outcomes using color Doppler testicular ultrasound for testicular torsion. Pediatrics 2000;105:604-607.

29. Musen MA. Scalable software architectures for decision support. Methods of Information in Medicine 1999;38(4-5):229-238.

30. Michalowski W, Kersten M, Wilk S, Slowinski R. Designing man-machine interactions for mobile clinical systems: MET triage support using Palm handhelds. European Journal of Operation Research 2005 (in press).

31. Rector AL, Horan B, Fitter M, Kay S, Newton PD, Nowlan WA, et al. User centered development of a general practice medical workstation: The PEN&PAD experience. In: Bauersfeld P, Bennett J, Lynch G, editors. Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI 92); 1992; Monterey, USA: New York: ACM; 1992. p. 447-453.

32. Musen MA, Shahar Y, Shortliffe EH. Clinical decision support systems. In: Shortliffe EH, Perreault LE, Wiederhold G, Fagan LM, editors. Medical Informatics. Computer Applications in Health Care and Biomedicine. 2 ed: Springer-Verlag; 2001. p. 573-609.

33. Ramnarayan P, Britto J. Paediatric clinical decision support systems. Archives of Disease in Childhood 2002;87(5):361-2.

34. Korner H, Sondenaa K, Soreide JA, Andersen E, Nysted A, Lende TH. Structured data collection improves the diagnosis of acute appendicitis. British Journal of Surgery 1998;85(3):341-4.