Rough Set Analysis of Medical Datasets and A Case of Patients with Suspected Acute Appendicitis

Ulf S. Carlin, Jan Komorowski and **Aleksander Øhrn**[1]

Abstract. A significant area in the field of medical informatics is concerned with learning of medical models from low-level data. Ultimate goals of this activity are, among others, development of classifiers or predictors for unseen cases and analysis of the developed models so that new insight into the nature of the given problem can be obtained.

This article introduces a methodology based on rough sets [27] and Boolean reasoning [4] and illustrates its application on a dataset describing 257 patients with suspected acute appendicitis. Exactly the same dataset has previously been analyzed using logistic regression [9, 10], and the difference in performance between the two methods is found to be very small. However, the rough set approach offers in addition a set of decision rules that explicitly represent the discovered knowledge. These automatically synthesized rules perform better than a surgeon with a 2 to 6 year training.

The main attractions of rough sets for the medical informatics community should be their classificatory power and, most importantly, the possibility of mixing qualitative and quantitative parameters (both continuous and discrete) and combining explicit (user-defined) and data-generated models. There also exist now good toolkits running on Windows NT/95 that support the knowledge discovery *process* with rough sets. An example is the ROSETTA toolkit [18] which is also available in a public version [31].

1 Introduction

Acute appendicitis is one of the most common problems in clinical surgery in the western world [6, 7], and the diagnosis is sometimes difficult even for experienced surgeons. Two types of diagnostic errors have to be considered in the decision-making process: unnecessary operations are clearly desirable to avoid but a delayed diagnosis may lead to perforation of the appendix. Since perforation of the appendix leads to morbidity and occasionally death, a high rate of unnecessary surgical interventions is usually accepted. Analysis of collected data with the objective of improving various aspects of the diagnosis is therefore potentially valuable.

Rough set theory [23, 24] is a fairly new knowledge discovery technique that has been previously applied to the medical domain (see for instance, [19, 35, 36, 38, 39]). One advantage of the rough set approach is that a set of readable if-then rules is produced. Such rules have a potential to reveal new medical insight by pointing out strong patterns in the data material and may also collectively function as a classifier for unseen cases.

This paper gives first a short introduction to rough sets and a methodology for synthesis of models from low-level data. It then summarizes an analysis of a data set that describes patients with suspected acute appendicitis. The objective of the analysis has been to develop rules that could predict either the presence or absence of acute appendicitis on the basis of observed patient attributes. The same data set has previously been studied using logistic regression [9, 10]. A comparison between both methods of analysis is done and different aspects of their respective strengths and weaknesses are discussed.

The structure of the paper is as follows. In Sect. 2 basic notions of Pawlak's rough sets are recalled and illustrated with simple examples. The modelling process based on rough sets is briefly described in Sect. 3. The data material is presented in Sect. 4. Processing methodology is described in Sect. 5. Section 6 gives the results of processing and, finally, Sect. 7 offers an analysis and a discussion of the results. Only a basic acquaintance with propositional logic and some familiarity with mathematical notations for sets and functions are required to understand this presentation.

Related work Building models using rough sets provides powerful classifiers (and outcome predictors). Bazan [2] has shown in an independent study that rough set-based methods perform in most cases at least as well as and often better than all other major methodologies such as neural networks, decision trees, statistical methods, ILP, etc. His data sets included medical data that were obtained from University Medical Center of Oncology, Ljubljana, Slovenia. For example applications of rough sets in medicine see, for example, [8, 17, 19, 37, 38, 39].

2 An Overview of Rough Sets

A data set is represented as a table, where each row represents a case, a patient, or simply an object. Every column represents an attribute (a variable, an observation, a property, etc) that can be measured for each object. This table is called an *information system*. More formally, it is a pair $\mathcal{A} = (U, A)$,

[1] Knowledge Systems Group, Dept. of Computer and Information Science, Norwegian University of Science and Technology, N-7034 Trondheim, Norway, {usc, janko, aleks}@idi.ntnu.no Corresponding author: J. Komorowski.

U. S. Carlin, J. Komorowski, A. Øhrn
ECAI 98. *Workshop on Intelligent Data Analysis in Medicine and Pharmacology*

where U is a non-empty finite set of *objects* called the *universe* and A is a non-empty finite set of *attributes* such that $a : U \to V_a$ for every $a \in A$. The set V_a is called the *value set* of a.

Example 2.1 A very simple information system is shown in Tab. 1. There are seven cases or objects, and two condition attributes (*Age* and Lower Extremity Motor Score, *LEMS*). The reader will easily notice that cases x_3 and x_4 as well as x_5 and x_7 have exactly the same values of conditions. The cases are (pairwise) *indiscernible* using the available attributes. □

	Age	*LEMS*
x_1	$16-30$	50
x_2	$16-30$	0
x_3	$31-45$	$1-25$
x_4	$31-45$	$1-25$
x_5	$46-60$	$26-49$
x_6	$16-30$	$26-49$
x_7	$46-60$	$26-49$

Table 1. An example information system.

In most medical applications, there is an outcome or classification that is known. This *a posteriori* knowledge is expressed by one distinguished attribute called the decision attribute; the process is usually called supervised learning. Information systems of this kind are called *decision systems.* A decision system is any information system of the form $\mathcal{A} = (U, A \cup \{d\})$, where $d \notin A$ is the *decision attribute.* The elements of A are called *condition attributes* or simply *conditions.* The decision attribute may take several values, though binary outcomes are rather frequent.

Example 2.2 A small example decision table can be found in Tab. 2. The table has the same seven cases as in the previous example, but one decision attribute (*Walk*) with two possible outcomes has been added.

	Age	*LEMS*	*Walk*
x_1	$16-30$	50	*Yes*
x_2	$16-30$	0	*No*
x_3	$31-45$	$1-25$	*No*
x_4	$31-45$	$1-25$	*Yes*
x_5	$46-60$	$26-49$	*No*
x_6	$16-30$	$26-49$	*Yes*
x_7	$46-60$	$26-49$	*No*

Table 2. *Walk* – an example decision table.

The careful reader may again notice that cases x_3 and x_4 as well as x_5 and x_7 still have exactly the same values of conditions, but the first pair has a different outcome (different value of the decision attribute) while the second pair also has the same outcome. □

The definitions to be synthesized from decision tables will be of the rule form:

$$\texttt{if} Age = 16-30 \ \texttt{and} \ Lems = 50 \ \texttt{then} \ Walk = Yes$$

Among the properties of the constructed rule sets, minimality is one of the main issues. This is studied in the next section.

2.1 Indiscernibility

It is assumed that a decision system (i.e. a decision table) expresses all the knowledge about the model. This table may be unnecessarily large, in part because it is redundant in at least two ways. The same or indiscernible objects may be represented several times, or some of the attributes may be superfluous. We shall look into these issues now.

The notion of equivalence is recalled first. A binary relation $R \subseteq X \times Y$ which is reflexive (i.e. an object is in relation with itself xRx), symmetric (if xRy then yRx) and transitive (if xRy and yRz then xRz) is called an equivalence relation.

Let $\mathcal{A} = (U, A)$ be an information system, then with any $B \subseteq A$ there is associated an equivalence relation $IND_{\mathcal{A}}(B)$:

$$IND_{\mathcal{A}}(B) = \{(x, x') \in U^2 \mid \forall a \in B \ a(x) = a(x')\}$$

$IND_{\mathcal{A}}(B)$ is called the *B-indiscernibility relation.*

If $(x, x') \in IND_{\mathcal{A}}(B)$, then objects x and x' are indiscernible from each other by attributes from B. The equivalence classes of the B-indiscernibility relation are denoted $[x]_B$. The subscript $\mathcal{A}$ in the indiscernibility relation is usually omitted if it is clear which information system is meant.

The indiscernibility relation is an equivalence relation. It is straightforward to see that objects belonging to the same equivalence class of the indiscernibility relation are indiscernible. (Some extensions of standard rough sets do not require transitivity to hold. See, for example, [33]. Such relations is called tolerance or similarity relations, but these considerations are outside the scope of this paper.)

Example 2.3 Let us illustrate how a decision table such as Tab. 2 defines an indiscernibility relation. The subsets of the conditional attributes are $\{Age\}$, $\{LEMS\}$ and $\{Age, LEMS\}$. If we consider, for instance, $\{LEMS\}$ only, objects x_3 and x_4 belong to the same equivalence class and are indiscernible. (By the same token, x_5, x_6 and x_7 belong to another indiscernibility class.) The relation IND defines three equivalence classes listed below.

$$\begin{aligned} IND(\{Age\}) &= \{\{x_1, x_2, x_6\}, \{x_3, x_4\}, \\ &\quad \{x_5, x_7\}\} \\ IND(\{LEMS\}) &= \{\{x_1\}, \{x_2\}, \{x_3, x_4\}, \\ &\quad \{x_5, x_6, x_7\}\} \\ IND(\{Age, LEMS\}) &= \{\{x_1\}, \{x_2\}, \{x_3, x_4\}, \\ &\quad \{x_5, x_7\}, \{x_6\}\} \end{aligned}$$

□

2.2 Set Approximation

An equivalence relation induces a partitioning of the universe (the set of cases in our example). These partitions can be used to build new subsets, i.e. cases or concepts. Subsets that are most often of interest have the same value of the outcome attribute. It may happen, however, that a concept cannot be defined in a crisp manner. For instance, the set of patients with a positive outcome cannot be defined crisply using the attributes available in Tab. 2. The "problematic" patients are objects x_3 and x_4. Returning to the notion of (supervised) learning, it is not possible to induce a crisp (precise) definition of such patients from the table. It is here that the notion "rough set" emerges. Though we cannot define those patients crisply, it is possible to delineate the patients that certainly have a positive outcome, the patients that certainly do not have a positive outcome and, finally, the patients that belong to a boundary between the certain cases. If this boundary is non-empty, the set is rough. These notions are formally expressed as follows.

Let $\mathcal{A} = (U, A)$ be an information system and let $B \subseteq A$ and $X \subseteq U$. We can approximate X using only the information contained in B by constructing the *B-lower* and *B-upper approximations of X*, denoted $\underline{B}X$ and $\overline{B}X$ respectively:

$$\begin{aligned}\underline{B}X &= \{x \mid [x]_B \subseteq X\} \\ \overline{B}X &= \{x \mid [x]_B \cap X \neq \emptyset\}\end{aligned}$$

The objects in $\underline{B}X$ are certain members of X, while the objects in $\overline{B}X$ are possible members of X. The set $\overline{B}X - \underline{B}X$ is called the *B-boundary region of X*, and thus consists of those objects that we cannot decisively classify into X. The set $U - \overline{B}X$ is called the *B-outside region of X* and consists of those objects that certainly do not belong to X. A set is said to be *rough* (respectively *crisp*) if the boundary region is non-empty (respectively empty).[2]

Example 2.4 The most common case in supervised learning is to synthesize definitions of the outcome (or decision classes) in terms of the conditional attributes. The example Tab. 2 defines the following approximations of the $Walk$ outcome.

$$\begin{aligned}\text{Lower approximation} &= \{x_1, x_6\} \\ \text{Upper approximation} &= \{x_1, x_3, x_4, x_6\} \\ \text{Boundary region} &= \{x_3, x_4\} \\ \text{Outside region} &= \{x_2, x_5, x_7\}\end{aligned}$$

It follows that the outcome $Walk$ is a rough concept since the boundary region is not empty. This is shown graphically in Fig. 1. □

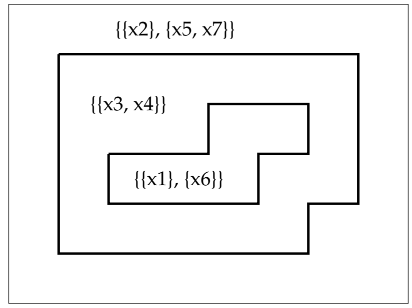

Figure 1. Approximating the set of walking patients in Tab. 2, using the two condition attributes Age and $LEMS$. Equivalence classes contained in the corresponding regions are shown.

2.3 Reducts

In the previous section we investigated one natural dimension of reducing data which is to identify equivalent objects, i.e. objects that are indiscernible using the available attributes. Savings are to be made since only one element of the equivalence class is needed to represent the entire class. The other dimension in reduction is to keep only these attributes that preserve the indiscernibility relation and, consequently, set approximation. The remaining attributes are redundant since their removal does not worsen the classification. There is usually several such subsets of attributes and those which are minimal are called reducts. Computing equivalence classes is straightforward. Finding minimal sets of reducts is NP-hard. In informal terms it means that for other than trivially small problems no computer will be ever able to find all solutions. It is, in fact, one of the bottlenecks of the rough set methodology. Fortunately, there exist good heuristics based on genetic algorithms that compute reducts in often acceptable time, unless the number of attributes is very high.

Example 2.5 Reducts are illustrated with the following decision system $\mathcal{A} = (U, \{Cand, Dipl, Exper, French, Refer\} \cup \{Decision\})$, where the names of the attributes *Candidate*, *Diploma*, *Experience*, and *References* are abbreviated for the lack of space. Let us consider only the conditional attributes. It appears that there is one minimal set of attributes $\{Exper, Refer\}$. The reader may check that the indiscernibility relation using the full set of attributes and the reduct is the same. (The actual construction of minimal sets of attributes will be soon revealed.) □

Cand	*Dipl*	*Exper*	*French*	*Refer*	*Decision*
x_1	*MBA*	*Medium*	*Yes*	*Excellent*	*Accept*
x_2	*MBA*	*Low*	*Yes*	*Neutral*	*Reject*
x_3	*MCE*	*Low*	*Yes*	*Good*	*Reject*
x_4	*MSc*	*High*	*Yes*	*Neutral*	*Accept*
x_5	*MSc*	*Medium*	*Yes*	*Neutral*	*Reject*
x_6	*MSc*	*High*	*Yes*	*Excellent*	*Accept*
x_7	*MBA*	*High*	*No*	*Good*	*Accept*
x_8	*MCE*	*Low*	*No*	*Excellent*	*Reject*

Table 3. *Hiring* – An example of an unreduced decision table.

Given an $\mathcal{A} = (U, A)$ the definitions of these notions are as follows. A *reduct* is a minimal set of attributes $B \subseteq A$ such that $IND(B) = IND(A)$. In other words, a reduct is a minimal set of attributes from A that preserves the partitioning of the universe (and hence the ability to perform classifications).

Reducts are strongly related [34] to prime implicants of Boolean formulae [4]. An *implicant* of a Boolean function f is any conjunction of literals (variables or their negations) such that if the values of these literals are true under an arbitrary valuation v of variables then the value of the function f under v is also true. A *prime implicant* is a minimal implicant.

Let $\mathcal{A}$ be an information system with n objects. The *discernibility matrix* of $\mathcal{A}$ is a symmetric $n \times n$ matrix with entries c_{ij} as given below. Each entry thus consists of the set of attributes upon which objects x_i and x_j differ.

[2] The letter B refers to the subset B of the attributes A. If another subset were chosen, e.g. $F \subseteq A$, the corresponding names of the relations would have been *F-boundary region, F-lower- and F-upper approximations.*

$$c_{ij} = \{a \in A \mid a(x_i) \neq a(x_j)\} \text{ for } i,j = 1,...,n$$

A *discernibility function* $f_{\mathcal{A}}$ for an information system $\mathcal{A}$ is a Boolean function of m Boolean variables $a_1^*, ..., a_m^*$ (corresponding to the attributes $a_1, ..., a_m$) defined as below, where $c_{ij}^* = \{a^* \mid a \in c_{ij}\}$. The set of all prime implicants of $f_{\mathcal{A}}$ determines the set of all reducts of $\mathcal{A}$.

$$f_{\mathcal{A}}(a_1^*, ..., a_m^*) = \bigwedge \left\{ \bigvee c_{ij}^* \mid 1 \leq j \leq i \leq n,\ c_{ij} \neq \emptyset \right\}$$

Example 2.6 The discernibility function for Tab. 3 is

$$\begin{aligned} f_{\mathcal{A}}(d,e,f,r) = & \\ & (e,r)(d,e,r)(d,e,r)(d,r)(d,e,)(e,f,r)(d,e,f) \\ & (d,r)(d,e,)(d,e)(d,e,r)(e,f,r)(d,f,r) \\ & (d,e,r)(d,e,r)(d,e,r)(d,e,f)(f,r) \\ & (e)(r)(d,f,r)(d,e,f,r) \\ & (e,r)(d,e,f,r)(d,e,f,r) \\ & (d,f,r)(d,e,f) \\ & (d,e,r) \end{aligned}$$

where "," stands for disjunction and each parenthesized tuple is a conjunct in the Boolean expression, and where the one–letter Boolean variables correspond to the attribute names in an obvious way. After simplification, the function is

$$f_{\mathcal{A}}(d,e,f,r) = er$$

(The notation er is a shorthand for $e \wedge r$).

Let us also notice that each row in the above discernibility function corresponds to one column in the discernibility matrix. This matrix is symmetrical with the empty diagonal. So, for instance, the last but one row says that the sixth object (more precisely, the sixth equivalence class) can be discerned from the seventh one by any of the attributes $Dipl, French$ or $Refer$ and by any of $Dipl, Exper$ or $French$. □

If we instead construct a Boolean function by restricting the conjunction to only run over column k in the discernibility matrix (instead of over all columns), we obtain the *k-relative discernibility function.* The set of all prime implicants of this function determines the set of all *k-relative reducts* of $\mathcal{A}$.

These reducts reveal the minimum amount of information needed to discern $x_k \in U$ (or more precisely $[x_k] \subseteq U$) from all other objects.

Using the notions introduced above, the problem of supervised learning is to find the value of the decision d that should be assigned to a new object which is described with the help of the conditional attributes. The set of attributes used to define the object should be, of course, minimal. For the example Tab. 3 it appears that $\{Exper, Refer\}$ and $\{Dipl, Exper\}$ are two minimal sets of attributes that uniquely define to which decision class an object belongs. The corresponding discernibility function is relative to the decision. The notions are now formalized.

Let $\mathcal{A} = (U, A \cup \{d\})$ be given. The cardinality of the image $d(U) = \{k | d(x) = k, x \in U\}$ is called the *rank of d* and is denoted by $r(d)$. Let us further assume that the set V_d of values of decision d is equal to $\{v_d^1, \ldots, v_d^{r(d)}\}$.

Example 2.7 Quite often the rank is two (e.g. $\{Yes, No\}$ or $\{Accept, Reject\}$). It can be an arbitrary number, however. For instance in the $Hiring$ example, we could have rank three if the decision had values in the set $\{Accept, Hold, Reject\}$. □

The decision d determines a partition $CLASS_{\mathcal{A}}(d) = \{X_{\mathcal{A}}^1, \ldots, X_{\mathcal{A}}^{r(d)}\}$ of the universe U, where

$$X_{\mathcal{A}}^k = \{x \in U \mid d(x) = v_d^k \text{ for } 1 \leq k \leq r(d)\}$$

$CLASS_{\mathcal{A}}(d)$ is called the *classification of objects in $\mathcal{A}$ determined by the decision d.* The set $X_{\mathcal{A}}^i$ is called the *i-th decision class of $\mathcal{A}$*. By $X_{\mathcal{A}}(u)$ we denote the decision class $\{x \in U \mid d(x) = d(u)\}$, for any $u \in U$.

Example 2.8 There are two decision classes in each of the running example decision systems, i.e. $\{Yes, No\}$ and $\{Accept, Reject\}$, respectively. The partitioning of the universe for the $Walk$ table is $U = X^{Yes} \cup X^{No}$ where

$$X^{Yes} = \{x_1, x_4, x_6\} \text{ and } X^{No} = \{x_2, x_3, x_5, x_7\}$$

and for the $Hiring$ table it is: $U = X^{Accept} \cup X^{Reject}$ where

$$X^{Accept} = \{x_1, x_4, x_6, x_7\} \text{ and } X^{Reject} = \{x_2, x_3, x_5, x_8\}$$

□

If $X_{\mathcal{A}}^1, \ldots, X_{\mathcal{A}}^{r(d)}$ are the decision classes of $\mathcal{A}$, then the set $\underline{B}X_1 \cup \ldots \cup \underline{B}X_{r(d)}$ is called the *B-positive region of $\mathcal{A}$* and is denoted by $POS_B(d)$.

Example 2.9 A quick check (left to the reader) reveals that $\underline{A}X^{Yes} \cup \underline{A}X^{No} \neq U$ while $\underline{A}X^{Accept} \cup \underline{A}X^{Reject} = U$. This is related to the fact that for the decision system in Tab. 2 a unique decision cannot be made for objects x_3 and x_4 while in the case of the other table all decisions are unique. □

This important property of decision systems is formalized as follows.

Let $\mathcal{A} = (U, A \cup \{d\})$ be a decision system. The *generalized decision in $\mathcal{A}$* is the function $\partial_A : U \longrightarrow \mathcal{P}(V_d)$ defined by

$$\partial_A(x) = \{i \mid \exists x' \in U\ x'\, IND(A)\, x \text{ and } d(x) = i\}$$

If $|\partial_A(x)| = 1$ for any $x \in U$ then a decision table $\mathcal{A}$ is called *consistent (deterministic)*, otherwise $\mathcal{A}$ is *inconsistent (non-deterministic)*.

It is easy to see that a decision table $\mathcal{A}$ is consistent if, and only if, $POS_A(d) = U$. Moreover, if $\partial_B = \partial_{B'}$, then $POS_B(d) = POS_{B'}(d)$ for any pair of non-empty sets $B, B' \subseteq A$.

Example 2.10 The *A-positive region of $\mathcal{A}$* in the $Walk$ decision system is a proper subset of U, while in the $Hiring$ decision system the corresponding set is equal to the universe U. The first system is non-deterministic, the second one deterministic. □

We have introduced above the notion of *k-relative* discernibility function. Since the decision attribute is so significant, it is useful to introduce a special definition for its case. Let $\mathcal{A} = (U, A \cup \{d\})$ be a consistent decision table and let $M(\mathcal{A}) =$

(c_{ij}) be its discernibility matrix. We construct a new matrix $M^d(\mathcal{A}) = (c^d_{ij})$ assuming $c^d_{ij} = \emptyset$ if $d(x_i) = d(x_j)$ and $c^d_{ij} = c_{ij} - \{d\}$, otherwise. Matrix $M^d(\mathcal{A})$ is called *the decision-relative discernibility matrix of* $\mathcal{A}$. Construction of *the decision-relative discernibility function* from this matrix follows the construction of the discernibility function from the discernibility matrix. It has been shown [34] that the set of *prime implicants* of $f^d_M(\mathcal{A})$ defines the set of all *decision-relative reducts* of $\mathcal{A}$.

Example 2.11 The *Hiring* decision table in Tab. 4 is now used to illustrate the construction of the corresponding decision-relative discernibility matrix and function. The rows are reordered for convenience putting the *Accept*-ed objects in the top rows. The corresponding discernibility matrix in Tab. 5

Cand	*Dipl*	*Exper*	*French*	*Refer*	*Decision*
x_1	*MBA*	*Medium*	*Yes*	*Excellent*	*Accept*
x_4	*MSc*	*High*	*Yes*	*Neutral*	*Accept*
x_6	*MSc*	*High*	*Yes*	*Excellent*	*Accept*
x_7	*MBA*	*High*	*No*	*Good*	*Accept*
x_2	*MBA*	*Low*	*Yes*	*Neutral*	*Reject*
x_3	*MCE*	*Low*	*Yes*	*Good*	*Reject*
x_5	*MSc*	*Medium*	*Yes*	*Neutral*	*Reject*
x_8	*MCE*	*Low*	*No*	*Excellent*	*Reject*

Table 4. *Hiring* – the reordered decision table.

is symmetrical and the diagonal is empty, and so are all the entries for which the decisions are equal. The resulting simplified decision-relative discernibility function is $f^d_M(\mathcal{A}) = (e \wedge d) \vee (e \wedge r)$. From the definition of the decision-relative matrix it follows that selecting one column of the indiscernibility matrix, e.g. corresponding to $[x_1]$, and simplifying it gives a minimal function that discerns $[x_1]$ from objects belonging to the corresponding decision class from objects belonging to the other decision classes. For example, the first column gives a Boolean function $(e,r)(d,e,r)(d,r)(d,e,f)$ which after simplification becomes $ed \vee rd \vee re \vee rf$. The reader can check that, for instance, "if $Refer = Excellent$ and $French = Yes$ then $Decision = Accept$" is indeed the case for x_1. It is rather illuminating to notice that if there is any other object for which $Refer = Excellent$ and $French = Yes$ hold then the decision will also be *Accept*. Indeed, this is the case for x_6. □

Example 2.12 There are two basic types of indiscernibility relations: object-relative and decision-relative. Reducts that are constructed for object-relative discernibility functions carry the minimum amount of information needed to discern that particular object (case, patient, etc). All other objects are "glued" into one class. Reducts constructed for decision-relative reducts convey the minimum information that is needed to make given decisions; objects that have the same decisions are "glued together" no matter whether they are or are not discernible. These two forms of indiscernbility can be combined to give four types of indiscernibility (and four types of reducts). □

2.4 Rough Membership

In the classical set theory, either an element belongs to a set or it does not. The corresponding membership function is

	$[x_1]$	$[x_4]$	$[x_6]$	$[x_7]$
$[x_1]$	$\emptyset$			
$[x_4]$	$\emptyset$	$\emptyset$		
$[x_6]$	$\emptyset$	$\emptyset$	$\emptyset$	
$[x_7]$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$
$[x_2]$	e,r	d,e	d,e,r	e,f,r
$[x_3]$	d,e,r	d,e,r	d,e,r	d,e,f
$[x_5]$	d,r	e	e,r	d,e,f,r
$[x_8]$	d,e,f	d,e,f,r	d,e,f	d,e,r

Table 5. *Hiring* – the decision-relative discernibility matrix. Collumns $[x_2], [x_3], [x_5]$ and $[x_8]$ are empty below the diagonal and are omitted here.

the characteristic function for the set, i.e. the function takes values 1 and 0, correspondingly. In the case of rough sets, the notion of membership is different. The *rough membership function* quantifies the degree of relative overlap between the set X and the equivalence class to which x belongs. It is defined as follows:

$$\mu^B_X(x) : U \longrightarrow [0,1] \text{ and } \mu^B_X(x) = \frac{|[x]_B \cap X|}{|[x]_B|}$$

The rough membership function can be interpreted as a frequency-based estimate of $\Pr(x \in X \mid x, B)$, the conditional probability that object x belongs to set X, given knowledge of the information signature of x with respect to (abbreviated wrt.) attributes B. This notion was introduced in [25].

The formulae for the lower and upper set approximations can be generalized to some arbitrary level of precision $\pi \in [\frac{1}{2}, 1]$ by means of the rough membership function [40], as shown below. Possible ties in the case of $\pi = 0.5$ can be resolved by assigning the objects in question to the interior of the set. Note that the lower and upper approximations as originally formulated are obtained as a special case with $\pi = 1.0$.

$$\underline{B}_\pi X = \{x \mid \mu^B_X(x) \geq \pi\}$$
$$\overline{B}_\pi X = \{x \mid \mu^B_X(x) > 1 - \pi\}$$

Rough sets can thus approximately describe sets of patients, events, outcomes, etc. that may be otherwise difficult to circumscribe.

2.5 Synthesis of Decision Rules

The reader has certainly realized that the reducts (of all the various types) serve the purpose of synthesizing *minimal* decision rules. Once the reducts have been computed, the rules are easily constructed by overlaying the reducts over the originating decision table and reading off the values.

Example 2.13 Given the reduct $\{Dipl, Exper\}$ in Tab. 4, the rule read off the first object is:

$$\texttt{if}\, Dipl = MBA\, \texttt{and}\, Exper = Medium\, \texttt{then}\, Decision = Accept$$

and similarly for the remaining objects. □

We shall make these notions precise. The rules are defined inductively in the usual manner.

Let $\mathcal{A} = (U, A \cup \{d\})$ be a decision system and let $V = \bigcup\{V_a \mid a \in A\} \cup V_d$. Atomic formulae over $B \subseteq A \cup \{d\}$ and V

are expressions of the form $a = v$; they are called *descriptors* over B and V, where $a \in B$ and $v \in V_a$. The set $\mathcal{F}(B, V)$ of formulae over B and V is the least set containing all atomic formulae over B and V and closed wrt. the propositional connectives $\wedge$ (conjunction), $\vee$ (disjunction) and $\neg$ (negation).

The semantics (meaning) of the formulae is also defined recursively. Let $\varphi \in \mathcal{F}(B, V)$. $|\varphi|_{\mathcal{A}}$ denotes the meaning of φ in the decision table $\mathcal{A}$ which is the set of all objects in U with the property φ. These objects are defined as follows:

1. if φ is of the form $a = v$ then $|\varphi|_{\mathcal{A}} = \{x \in U \mid a(x) = v\}$
2. $|\varphi \wedge \varphi'|_{\mathcal{A}} = |\varphi|_{\mathcal{A}} \cap |\varphi'|_{\mathcal{A}}$; $|\varphi \vee \varphi'|_{\mathcal{A}} = |\varphi|_{\mathcal{A}} \cup |\varphi'|_{\mathcal{A}}$; $|\neg\varphi|_{\mathcal{A}} = U - |\varphi|_{\mathcal{A}}$

The set $\mathcal{F}(B, V)$ is called the set of *conditional formulae of* $\mathcal{A}$ and is denoted $\mathcal{C}(B, V)$.

A *decision rule* for $\mathcal{A}$ is any expression of the form $\varphi \Rightarrow d = v$, where $\varphi \in \mathcal{C}(B, V)$, $v \in V_d$ and $|\varphi|_{\mathcal{A}} \neq \emptyset$. Formulae φ and $d = v$ are referred to as the *predecessor* and the *successor* of decision rule $\varphi \Rightarrow d = v$.

Decision rule $\varphi \Rightarrow d = v$ is *true* in $\mathcal{A}$ if, and only if, $|\varphi|_{\mathcal{A}} \subseteq |d = v|_{\mathcal{A}}$.

Example 2.14 Looking again at Tab. 4, some of the rules are, for example:

$$\begin{array}{lll} Dipl = MBA & \wedge\ Exper = Medium & \Rightarrow d = Accept \\ Exper = Low & \wedge\ Refer = Good & \Rightarrow d = Reject \\ Dipl = MSc & \wedge\ Exper = Medium & \Rightarrow d = Accept \end{array}$$

The first two rules are true in Tab. 4 while the third one is not true in that table. □

For a systematic overview of rule synthesis see [32].

3 The Modelling Process

We turn now to the discussion of the modelling and validation process. The process has three basic steps which are briefly discussed.

1. *Discretization:* Transforming non-categorical attributes in a decision table into categorical ones.
2. *Rule induction:* Synthesizing decision rules from a decision table.
3. *Rule application:* Applying the extracted decision rules to classify new cases.

The modelling procedure can be repeated in a systematic fashion, for instance, by employing a cross-validation scheme.

Discretization The rough set approach requires only indiscernibility. This means that there is no need to define an order or distance when attributes of different types are combined. On the other hand, non-categorical attributes must be discretized in a pre-processing step. The discretization step determines how coarsely we want to view the world. Taking, for instance, heart rate or body temperature, we have to establish cut-off points. In the simplest and most coarse case the result is two intervals: below and not less than some value. These intervals may be, however, refined. Since there sometimes exists good domain knowledge (and since the search for cut-off points is computationally very expensive), it is not unusual that domain experts prepare discretization manually. Discretization is a step that is not specific to the rough set approach, but that most rule or tree induction algorithms currently require for them to perform well. The interested reader is referred to [20] for a thourough exposition of discretization. The search for appropriate cut-off points essentially uses the approach of finding minimal Boolean expressions.

Rule Synthesis Several numerical factors can be associated with a synthesized rule. For example, the support of a decision rule is the number of objects that match the predecessor of the rule. Various frequency-related numerical quantities may be computed from such counts.

The main challenge in inducing rules from decision tables lies in determining which attributes that should be included in the conditional part of the rule. Although we can compute minimal decision rules, this approach results in rules that may contain noise or other peculiarities of the data set. Such detailed rules will overfit the data and will poorly classify unseen cases. More general, i.e. shorter rules should be rather synthesized. It implies that reduct approximations need to be found instead, i.e. attribute subsets that in a sense "almost" preserve the indiscernibility relation. One way of computing approximations is by computing reducts for random subsets of the universe of a given decision system and selecting the most stable reducts, i.e. reducts that occur in most of the subsystems. These reducts are usually inconsistent for the original table, but the rules synthesized from them are more tolerant to noise and other abnormalities; they perform better on unseen cases since they cover the most general patterns in the data. For a presentation of generating default rules see [15, 13, 14] and [12] who investigate synthesis of default rules or normalcy rules and some implementations of heuristics that search for such reducts.

One particularly successful method based on the resampling approach is called dynamic reducts. It is implemented in the ROSETTA system [18].

Rule Application When a set of rules have been induced from a decision table containing a set of training examples, they can be inspected to see if they reveal any novel relationships between attributes that are worth pursuing for further research. Furthermore, the rules can be applied to a set of unseen cases in order to estimate their classificatory power.

Several application schemes can be envisioned, but a simple one that has shown useful in practice is the following.

1. When a rough set classifier is presented with a new case, the rule set is scanned to find applicable rules, i.e. rules whose predecessors match the case.
2. If no rule is found (i.e. no rule "fires"), the most frequent outcome in the training data is chosen.
3. If more than one rule fires, these may in turn indicate more than one possible outcome. A voting process is then performed among the rules that fire in order to resolve conflicts and to rank the predicted outcomes. A rule casts as many votes in favour of its outcome as its associated support count. The votes from all the rules are then accumulated and divided by the total number of votes cast in order to arrive at a numerical measure of certainty for each outcome. This measure of certainty is not really a probability,

but may be interpreted as an approximation to such, if the model is well calibrated.

Summary Rough sets and information systems have been introduced and illustrated with simple examples. The interested reader is referred to Pawlak's monography [27]. Further readings on applications, theory and novel issues tackled by rough sets can be found in, for instance, [22], [30] and [26]. The latter contains an up-to-date bibliography of rough set research. Readers interested in hands-on experience with rough sets may down-load the public version of the ROSETTA system at [31].

4 Data Material

The data set studied in this paper consists of 257 patients with suspected acute appendicitis. For each patient the attributes listed in Tab. 6 and Tab. 7 were recorded. The binary and numerical attributes are summarized in Tab. 6 and Tab. 7, respectively. For further details about the collection of the data material, see [9, 10].

For each patient the surgeon estimated the patient's risk of having acute appendicitis in increments of 10% from 0 to 100%. There were nine different surgeons with two to six years of surgical training who participated in this probability estimation. The estimation was done before the result of a blood test was ready. The attributes based on the blood test, and thus not available to the surgeon when he performed the probability estimation, are: ESR, CRP, WBC, and NEUTRO. The probability estimates were, of course, neither a part of the logistic regression analysis nor of the rough set analysis.

The DIAGNOSIS attribute is the (*a posteriori*) decision attribute d in the analysis. It shows which patients actually turned out to have appendicitis. As can be seen in Tab. 6, 98 patients (38%) turned out to have appendicitis and 159 (62%) turned out to have some other disease or non-specific abdominal pain. The final diagnosis of acute appendicitis was based on histological examination of the excised appendix. Other diagnoses were based on routine investigation with repeated clinical examination, biochemical tests, imaging techniques and, if necessary, surgery.

In the analysis, different subsets of all the attributes will be used as A in the decision system $\mathbf{A} = (U, A \cup d)$. This is done in order to make the comparison of the diagnostic ability of the logistic regression model, the rough set model, and the surgeons' probability estimate as fair as possible wrt. to the attributes.

The data set obtained from the medical doctor was already discretized. The numerical attributes in Tab. 7 are discretized in Tab. 8.

We were informed that the discretization of the CRP, WBC, and NEUTRO attributes was done manually, while the ESR attribute was discretized using the same intervals as in [21]. The AGE, DURATION, and TEMP attributes were discretized into three intervals, each containing approximately the same number of objects.

5 Methodology

Hallan et al. [9, 10] applied a logistic regression analysis which resulted in a function that maps each patient to a probability

Attribute	Intervals	Count	Description
AGE	$[-\infty, 17)$	84	Low
	$[17, 31)$	90	Middle
	$[31, \infty)$	83	High
DUR-ATION	$[-\infty, 13)$	96	Short
	$[13, 30)$	71	Middle
	$[30, \infty)$	90	Long
TEMP	$[-\infty, 37.5)$	84	Low
	$[37.5, 38.1)$	89	Middle
	$[38.1, \infty)$	81	High
ESR	$[-\infty, 10)$	126	Normal
	$[10, 25)$	99	Slightly raised
	$[25, \infty)$	32	Considerably raised
CRP	$[-\infty, 6)$	103	Normal
	$[6, 40)$	95	Slightly raised
	$[40, \infty)$	59	Considerably raised
WBC	$[-\infty, 10.0)$	91	Normal
	$[10.0, 14.0)$	80	Slightly raised
	$[14.0, \infty)$	86	Considerably raised
NEU-TRO	$[-\infty, 75)$	87	Normal
	$[75, 85)$	87	Slightly raised
	$[85, \infty)$	83	Considerably raised

Table 8. Discretization of numerical attributes

of disease. They chose a splitting strategy where the original group of objects was split in two approximately equal parts. Logistic regression analysis was then performed on one part and subsequently tested on the other. This was done for 20 random splits, and from each iteration a *Receiver Operator Characteristic* curve[3] (hence abbreviation ROC) curve was generated by varying the cut-off probability of disease. The mean area under the 20 ROC curves was subsequently calculated [11].

We followed the previous study as close as possible and performed rough set analysis in a similar fashion, i.e. with 20 different splits of the original table into equal parts. However, these splits were chosen randomly and are thus not the same splits as those used by Hallan et al. Rulesets were generated from the first parts and tested on the other ones. Points on the ROC curves were generated by systematically varying cut-off thresholds across the output from the previosly describe rule voting procedure. The Area Under the ROC Curve (hence abbreviation AUC) can then be computed using the trapezoidal integration rule, and the mean AUC can be computed over a number of different runs where different splits of the original table are used.

Hallan et al. presented in [10] analyses on three different variable sets. The first variable set, which will be called A for simplicity, consists of the following clinical variables: CLASSIC, REBTEND, SEX, TENDRLQ, COUGHING and

[3] An ROC curve is a plot of *1 - specificity* on the x-axis versus *sensitivity* on the y-axis for different cut-off values. Sensitivity (resp. specificity) is the proportion of the patients with positive (resp. negative) disease status who are correctly identified by the test.

Attribute	Description	Statistics	
		Yes % (count)	No % (count)
SEX	Male sex?	55.3 (142)	44.7 (115)
ANOREXIA	Anorexia?	69.3 (178)	30.7 (79)
NAUSEA	Nausea or vomiting?	70.8 (182)	29.2 (75)
PREVIOUS	Previous surgery?	9.3 (24)	90.7 (233)
MOVEMENT	Aggravation of pain by movement?	61.5 (158)	38.5 (99)
COUGHING	Aggravation of pain by coughing?	59.9 (154)	40.1 (103)
MICTUR	Normal micturition?	87.2 (224)	12.8 (33)
TENDRLQ	Tenderness in right lower quadrant?	86.0 (221)	14.0 (36)
REBTEND	Rebound tenderness in right lower quadrant?	55.3 (142)	44.7 (115)
GUARD	Guarding or rigidity?	30.7 (79)	69.3 (178)
CLASSIC	Classic migration of pain?	49.4 (127)	50.6 (130)
DIAGNOSIS	(Final diagnosis:) acute appendicitis?	38.1 (98)	61.9 (159)

Table 6. Binary attributes

Attribute	Description	Unit	Statistics		
			Mean (SD)	Median	Range
AGE	Age	years	26.8 (17.0)	22	3–86
DURATION	Duration of pain	hours	35.3 (53.8)	22	2–600
TEMP	Rectal temperature	°C	37.8 (0.746)	37.7	36.4–40.3
ESR	Erythrocyte sedimentation rate	mm	14.1 (15.8)	10	1–90
CRP	C–reactive protein concentration	mg/L	32.8 (48.7)	12	0–260
WBC	White blood cell count	$\times 10^9$	12.3 (4.79)	12.1	2.9–31.0
NEUTRO	Neutrophil count	%	77.1 (11.4)	80	38–93

Table 7. Numerical attributes

Variable set	Logistic Regression	Rough Sets	Surgeons
A	0.854 ± 0.0283	0.850 ± 0.0235	
B	0.901 ± 0.0174	0.905 ± 0.0231	
C	0.920 ± 0.0238	0.923 ± 0.0225	
Clinical variables			0.817

Table 9. Results: Average AUC ± SD

GUARD. The second analysis used the variable WBC in addition, let $B = A \cup \{\text{WBC}\}$. The last analysis used variables CRP and NEUTRO in addition. This variable set will be called C, $C = B \cup \{\text{CRP}, \text{NEUTRO}\}$. Hallan et al. found that adding other clinical variables or the ESR did not improve the logistic regression model further.

The ROSETTA software system [18] was used to carry out all computations. The presented rough set analyses were done on the same sets of variables as used by Hallan et al. for reasons of comparison. Dynamic reducts were calculated using an exhaustive algorithm on each sampled subtable. The dynamic reduct sampling strategy was the following: Subtables were sampled on 10 equally spaced levels with 50 samples per level from 5% to 95% of the original table. Then rules were generated from all of the resulting reducts in the following way.

The synthesis was done for two cases: exact rules (i.e. exactness coefficient 1.0) and approximate rules with exactness not less than 0.7. We show a fragment of the rule set for decision system C in Tab. 10. The attributes are: CLASSIC, REBTEND, SEX, TENDRLQ, COUGHING, GUARD, WBC, CRP and NEUTRO. The ninth column named D is the decision. Column S_1 gives the number of objects that support the rule, i.e. the number of objects which match the predecessor and successor of the rule; column S_2 – the number of objects matching the predecessor but not the successor, and

A0	A1	A2	A3	A4	A5	A6	A7	A8	D	S_1	S_2	*Consist*
						(6=0)			(9=0)	83	20	0,81
	(1=0)								(9=0)	79	24	0,77
(0=1)	(1=1)			(4=1)					(9=1)	26	10	0,72
(0=1)	(1=0)	(2=1)	(3=0)	(4=0)			(7=1)	(8=2)	(9=1)	1	0	1,00
(0=0)									(9=0)	89	26	0,77
(0=0)	(1=0)	(2=1)				(6=2)			(9=1)	4	1	0,80
(0=1)	(1=1)	(2=1)	(3=0)					(8=0)	(9=1)	2	0	1,00
(0=1)	(1=0)	(2=1)	(3=0)			(6=1)		(8=1)	(9=1)	2	0	1,00
(0=1)	(1=0)	(2=1)		(4=0)		(6=0)	(7=1)	(8=2)	(9=1)	1	0	1,00
(0=0)	(1=1)	(2=1)				(6=1)	(7=1)	(8=1)	(9=1)	1	0	1,00
(0=1)	(1=1)		(3=1)						(9=1)	45	17	0,73

Table 10. A fragment of the rule set for decision system C.

the last one is a consistency coefficient defined by $S_1/(S_1+S_2)$.

6 Results

The results of the rough set analysis and the logistic regression analysis by Hallan et al. are summarized in Tab. 9. The results are presented as average AUC values ± standard deviation for experiments on 20 different splits of the original table as described in Sect. 5. We see that the logistic regression models and the rough set models perform approximately equally well. Both methods perform somewhat better than the surgeons. Just like the logistic regression model on variable set C, the corresponding rough set model did not improve further when adding clinical variables or the ESR.

In Tab. 11, a comparison of the rough set approach and the surgeons is done using some other common measures of performance (sensitivity, specificity and accuracy), in addition to the mean area under the ROC curves. The two following methods are added. The strategy of the best attribute, CLASS, is: If classic migration of pain is present at a patient, classify as "sick", else classify as "healthy". The baseline strategy is to classify all patients as the majority decision class, which in this data table is "not appendicitis". Any advanced method should at least perform better than those two strategies.

Method	Sens.	Spec.	Acc.	AUC
RS on A	0.684	0.837	0.774	0.850
RS on B	0.801	0.853	0.830	0.905
RS on C	0.876	0.850	0.858	0.923
Surgeons	0.867	0.679	0.751	0.817
Best attribute	0.755	0.667	0.700	0.711
Baseline	0	1	0.619	0.5

Table 11. Comparison between rough set approach and surgeons.

We can see from this table that the rough set approach scores slightly better than the surgeons on accuracy, when only clinical variables are considered. The accuracy rises from 0.774 to 0.858 when CRP, WBC, and NEUTRO are added to the analysis. The high sensitivity and low specificity for the surgeons reflects the usual tendency to overestimate the diagnosis acute appendicitis to avoid perforation. Varying the cut-off thresholds, the rough set models could also have been tuned to overestimate the diagnosis, giving a higher sensitivity and a lower specificity. Such tuning also affects the accuracy, but not the AUC.

7 Analysis and Discussion

Rough set analysis confirms the results obtained in [10] and it adds the possibility of inspecting the discovered diagnostic algorithm. On the other hand, logistic regression is a well-known and extensively used statistical method which will be preferred by most medical doctors. The relative advantage of rough sets needs to be further investigated in order to show substantial gains over the standard approaches, if it is to be widely used in the medical community. This may happen with the help of computer scientists trained in the rough set methods who will be likely to provide services of higher quality and possibly lead to the discovery of new medical knowledge.

We have synthesized rules for set C for two cases: exact and approximate rules. The first approach produced 494 rules while the second one resulted in 95 rules. The second set gave best classification. The rules were not analyzed by a medical doctor at the moment of submitting this article. The results obtained by rough sets are nevertheless remarkable. Using only 257 objects the diagnostic algorithm provides an explicit representation and performs better than a surgeon with a 2 to 6 year training. Thus such algorithms may eventually be very useful decision aids, even for the experienced diagnosticians.

In order to get a better assessment of the relative "goodness" of both techniques, their predictive capability should be assessed on a new independent set of data. There is a risk that both are over-optimistic. If it were possible, we would have preferred to do automatic discretization of the numerical attributes. Unfortunately, the data sets delivered to us were already discretized. The 20-split scheme was chosen in order to compare the results with the study of Hallan et al.

These results are also in line with other experiments performed in our group on predicting acute myocardial infarction [39] and modelling prognostic power of cardiac tests [19].

8 Acknowledgments

Special thanks to Stein Hallan, MD, for making the data available. We appreciate cooperation with Jan Bazan, Rzeszow University, Poland, who independently processed our data sets and could confirm our results. He has also synthesized the rules. Andrzej Skowron, Warsaw University, Poland, was, as usual, a friend and a provider of insightful suggestions and comments.

This research is supported in part by the European Union 4th Framework Telematics project CardiAssist, by the Human Capital and Mobility Norwegian Research Council contract #101341/410, by the Norwegian Research Council grant #74467/410, and by the Norwegian Research Council grant for Cooperation with Central Europe.

REFERENCES

[1] A. Aamodt and J. Komorowski (Eds.) (1995), *Proc. Fifth Scandinavian Conference on Artificial Intelligence*, Trondheim, Norway, May 29–31, Frontiers in Artifical Intelligence and Applications, Vol. 28, IOS Press.

[2] J. Bazan (1998), *A Comparison of Dynamic and non-Dynamic Rough Set Methods for Extracting Laws from Decision Tables.* In: Polkowski and Skowron [30].

[3] J. Bazan, A. Skowron, P. Synak (1994), Dynamic Reducts as a Tool for Extracting Laws from Decision Tables, Proc. Symp. on Methodologies for Intelligent Systems, Charlotte, NC, USA, October 16–19, Lecture Notes in Artificial Intelligence, Springer Verlag, Vol. 869, pp. 346–355.

[4] F. M. Brown (1990), *Boolean Reasoning*, Kluwer Academic Publishers, Dordrecht.

[5] U. Carlin (1998), Mining medical data with rough sets, M.Sc. (siv.ing.) thesis, Department of Computer and Information Science, The Norwegian University of Science and Technology, Trondheim, Norway.

[6] S. Eriksson (1996), Acute Appendicitis – Ways to Improve Diagnostic Accuracy, European Journal of Surgery, Vol. 162, pp. 435–442.

[7] J. M. Grönroos, J. J. Forsström, K. Irjala, T. J. Nevalainen (1994), Phospholipase A_2, C–Reactive Protein, and White Blood Cell Count in the Diagnosis of Acute Appendicitis, Clinical Chemistry, Vol. 40, No. 9, pp. 1757–1760.

[8] J.W. Grzymała-Busse and L.K. Goodwin (1997), *Predicting preterm birth risk using machine learning from data with missing values*, In: S. Tsumoto (Ed.), *Bulletin of International Rough Set Society* **1/2**, pp. 17–21.

[9] S. Hallan, A. Åsberg, T.-H. Edna (1997), Estimating the Probability of Acute Appendicitis Using Clinical Criteria of a Structured Record Sheet: The Physician Against the Computer, European Journal of Surgery, Vol. 163, No 6, pp. 427–432.

[10] S. Hallan, A. Åsberg, T.-H. Edna (1997), Additional Value of Biochemical Tests in Suspected Acute Appendicitis, European Journal of Surgery, Vol. 163, No 7, pp. 533–538.

[11] J. Hanley, B. McNeil (1982), *The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve*, Radiology, 143, pp. 29–36.

[12] T.-K. Jensen, J. Komorowski and A. Øhrn (1998), *Some Heuristics for Default Knowledge Discovery.* In: Polkowski and Skowron [29].

[13] T. Mollestad (1997), *A Rough Set Approach to Data Mining: Extracting a Logic of Default Rules from Data*, PhD thesis, Norwegian University of Science and Technology.

[14] T. Mollestad and J. Komorowski (1998), *A Rough Set Framework for Propositional Default Rules Data Mining.* In: Pal and Skowron [26].

[15] T. Mollestad and A. Skowron (1996), *A Rough Set Framework for Data Mining of Propositional Default Rules* In: Pawlak and Ras [28] pp. 448–457. Full version available at `http://www.idi.ntnu.no`

[16] A. Øhrn, J. Komorowski (1997), Rosetta – A Rough Set Toolkit for Analysis of Data, Proc. Third International Joint Conference on Information Sciences, Durham, NC, USA, March 1–5, Vol. 3, pp. 403–407.

[17] A. Øhrn, J. Komorowski (1998), Analyzing The Prognostic Power of Cardiac Tests Using Rough Sets, Working Notes of the Invited Session on Intelligent Prognostic Methods in Medical Diagnosis and Treatment Planning, Second IMACS Multi-conference on Computational Engineering in Systems Applications, Nabeul-Hammamet, Tunisia, April 1-4, pp. 54-61.

[18] A. Øhrn, J. Komorowski, A. Skowron and P. Synak (1998), *The Design and Implementation of a Knowledge Discovery Toolkit Based on Rough Sets - The* Rosetta *System.* In Polkowski and Skowron [30], 24 pages.

[19] A. Øhrn, S. Vinterbo, P. Szymański and J. Komorowski (1997), *Modelling Cardiac Patient Set Residuals Using Rough Sets.* In: *Proc. AMIA Annual Fall Symposium* (formerly SCAMC), Nashville, TN, USA, Oct. 25–29, pp. 203–207.

[20] H.S. Nguyen and S.H. Nguyen (1998), *Discretization methods in data mining*, In: Polkowski and Skowron [30].

[21] W. P. Oosterhuis, A. H. Zwinderman, M. Teeuwen, et al. (1993), C reactive protein in the diagnosis of acute appendicitis, European Journal of Surgery, Vol. 159, pp. 115–119.

[22] E. Orłowska (Ed.) (1998), *Incomplete Information: Rough Set Analysis.* ISBN 3-7908-1049-5. Physica Verlag, 613 pages.

[23] Z. Pawlak (1982), Rough Sets, International Journal of Computer and Information Sciences, Vol. 11, No 5, pp. 341–356.

[24] Z. Pawlak (1991), Rough Sets – Theoretical Aspects of Reasoning about Data, Kluwer Academic Publishers, Dordrecht.

[25] Z. Pawlak, A. Skowron (1994), *Rough Membership Functions*, Advances in the Dempster-Shafer Theory of Evidence, R. Yager, M. Fedrizzi and J. Kacprzyk (eds.), Wiley, pp. 251–271.

[26] S.K. Pal and A. Skowron (Eds.) (1998), *Rough Sets, Fuzzy Hybridizations and Decision Support Systems*, Springer-Verlag Singapore. To appear.

[27] Z. Pawlak (1991), *Rough Sets – Theoretical Aspects of Reasoning about Data*, Kluwer Academic Publishers, Dordrecht.

[28] Z. Pawlak and Z. Ras, (Eds.), (1996), *Proc. Ninth International Symposium on Methodologies for Intelligent Systems, ISMIS'96*, June, Springer Verlag, LNAI.

[29] L. Polkowski and A. Skowron (Eds.) (1998), *Proc. First International Conference on Rough Sets and Soft Computing* – RSCTC'98, Warszawa, Poland, June 22–27 , Springer-Verlag, LNAI Vol. 1424.

[30] L. Polkowski and A. Skowron, (Eds.) (1998) *Rough Sets in Knowledge Discovery*, Physica Verlag. To appear.

[31] *http://www.idi.ntnu.no/~aleks/rosetta/* – the Rosetta WWW homepage.

[32] A. Skowron (1995), *Synthesis of Adaptive Decision Systems from Experimental Data.* In: Aamodt and Komorowski [1], pp. 220–238.

[33] A. Skowron, L. Polkowski and J. Komorowski (1996), *Learning Tolerance Relations by Boolean Descriptors: Automatic Feature Extraction from Data Tables*, in the Proc. of the 4th Workshop on Rough Sets, Fuzzy Sets and Machine Discovery, pp. 11–17, ISBN 4-947717-01-7, University of Tokyo.

[34] A. Skowron and C. Rauszer (1992), The Discernibility Matrices and Functions in Information Systems, In R. Slowinski (ed.), Intelligent Decision Support – Handbook of Applications and Advances of the Rough Sets Theory, Dordrecht, Kluwer Academic Publishers, pp. 331–362.

[35] K. Slowinski (1988), Rough Set Approach to Analysis of Data from Peritoneal Lavage in Acute Pancreatitis, Medical Informatics, 13, no. 3, pp. 143–159.

[36] K. Slowinski (1992), Rough classification of HSV patients, In R. Slowinski (ed.), Intelligent Decision Support – Handbook of Applications and Advances of the Rough Sets Theory, Dordrecht, Kluwer Academic Publishers, pp. 77–94.

[37] K. Słowiński and J. Stefanowski (1998), *Multistage rough set analysis of therapeutic experience with acute pancreatis.* In: Pal and Skowron [26].

[38] S. Tsumoto (1998), *Modelling Medical Diagnostic Rules based on Rough Sets.* In: Pal and Skowron [26].

[39] S. Vinterbo, L. Ohno-Machado, H. Fraser (1998), Prediction of Acute Myocardial Infarction using Rough Sets, In preparation.

[40] W. Ziarko (1993), *Variable Precision Rough Set Model*, Journal of Computer and System Sciences, 46, pp. 39–59.