

H. Leitich¹, K.-P. Adlassnig¹,
G. Kolarz^{2,3}

Development and Evaluation of Fuzzy Criteria for the Diagnosis of Rheumatoid Arthritis

¹Department of Medical Computer Sciences, University of Vienna,

²Clinic for Rheumatic Diseases of the Social Insurance Company for Trade and Industry, Baden,

³Institute for Rheumatology in Baden, Austria

Abstract: In 1987, the American Rheumatism Association issued a set of criteria for the classification of rheumatoid arthritis (RA) to provide a uniform definition of RA patients. Fuzzy set theory and fuzzy logic were used to transform this set of criteria into a diagnostic tool that offers diagnoses at different levels of confidence: a definite level, which was consistent with the original criteria definition, as well as several possible and superdefinite levels. Two fuzzy models and a reference model which provided results at a definite level only were applied to 292 clinical cases from a hospital for rheumatic diseases. At the definite level, all models yielded a sensitivity rate of 72.6% and a specificity rate of 87.0%. Sensitivity and specificity rates at the possible levels ranged from 73.3% to 85.6% and from 83.6% to 87.0%. At the superdefinite levels, sensitivity rates ranged from 39.0% to 63.7% and specificity rates from 90.4% to 95.2%. Fuzzy techniques were helpful to add flexibility to preexisting diagnostic criteria in order to obtain diagnoses at the desired level of confidence.

Keywords: Fuzzy Set Theory, Expert Systems, Diagnosis, Rheumatoid Arthritis, CADIAG-2

1. Introduction

Rheumatologic diagnoses are often based on a combination of different symptoms and findings, because rheumatic diseases can rarely be characterized by a single pathognomonic feature. Without any proof of a diagnosis, rheumatologists rely on a consensus policy to classify patients and to compare study results. This policy demands strict definitions of symptoms and findings, and it has motivated the development of classification criteria for rheumatic diseases. Most of these criteria have been issued by the American College of Rheumatology (ACR, formerly the American Rheumatism Association, ARA) and are now widely accepted.

This classification problem has also motivated the development of computer consultant systems and the application of artificial intelligence techniques. CADIAG-2/RHEUMA is a medical expert system, which – along with several other systems – has been developed

to support the differential diagnosis of rheumatic diseases [1-15]. Within the system's knowledge base, which currently includes 166 rheumatologic diagnoses, knowledge is represented as a series of disease profiles and diagnostic rules.

In order to be accepted by expert rheumatologists, most of CADIAG-2/RHEUMA's diagnostic rules are based on ACR (ARA) classification criteria. As mentioned above, these criteria were designed to support the selection of more uniform study patients, not to provide diagnostic assistance. A study carried out with the 1958 and 1987 ACR criteria for the classification of rheumatoid arthritis (RA) showed that they could be used for diagnostic purposes – if some limitations were accepted [16-18]. Diagnostic performance was satisfying in RA patients with longer disease course and more advanced disease features. Some RA patient groups, such as seronegative RA patients or those in early stages of the disease,

however, yielded poor performance rates. These patient groups are known to display more subtle disease features.

A characteristic of CADIAG-2 is the use of fuzzy-set theory to model the vagueness of symptom terms. If classification criteria are replaced with fuzzy sets, they are partially fulfilled by patients with borderline findings. A study carried out with (such a fuzzy version of) the 1987 RA classification criteria showed that overall performances rates were especially improved in RA patient groups with borderline findings, such as seronegative RA patients and those in early disease stages [19].

In this article we again present diagnostic rules based on fuzzy versions of the 1987 RA classification criteria (they will be called "fuzzy models"). We aim to show that different values of the rules' conclusion c represent different levels of diagnostic confidence: a definite as well as several possible and superdefinite levels.

The definite level of confidence represents the original criteria definition, which indicates that this definition is used as a definite reference. Possible levels represent criteria which are less rigid than the original definition, and superdefinite levels criteria which are more rigid than the original definition.

If such a fuzzy model is used as a diagnostic tool for real patients, the choice of a threshold ϵ determines the level of confidence on which results are obtained. If ϵ represents the definite level, positive results will only be assigned to patients who fulfill the original criteria definition. A lower choice of ϵ (representing possible levels) will include patients who do not fulfill the original definition, but only a less rigid one (such as patients with borderline findings). A higher choice of ϵ (representing superdefinite levels) will assign positive results only to patients who fulfill more rigid criteria (such as patients with more advanced disease features).

The numbers of positive results which are obtained with different choices of ϵ can be used to calculate sensitivity and specificity rates for each level of confidence. A lower choice of ϵ will normally yield a higher sensitivity and a lower specificity rate. A high sensitivity rate is useful if the detection of pathological features is more important than a definite diagnosis (i.e., if false-positive results are more acceptable than false-negative results). In this way the fuzzy model may serve as a screening or warning tool for rare diseases. A higher choice of ϵ will usually lead to a higher specificity and a lower sensitivity rate. A high specificity rate is useful if the certainty of a diagnosis is most important (i.e., if false-negative results are more acceptable than false-positive results). In this way the fuzzy model may serve as a selection tool for patients who are to be included in a study or treatment regimen (e.g., a treatment with immunosuppressive drugs).

If a fuzzy model is tested with patients from the intended application setting (e.g., an outpatient clinic), a table can be created which assigns fixed sensitivity and specificity rates to each level of confidence. In this setting, the model can be used as a diagnostic tool with variable and predictable sensitivity and specificity. This procedure is compar-

able to studies in which sensitivity and specificity rates are determined for different thresholds of a single finding. The difference is that, in our study, results of different kinds of examinations (patient history data, results of the physical examination, laboratory tests, X-ray findings) are combined to a single fuzzy logical value, and that sensitivity and specificity rates are determined for different thresholds of this logical value.

The purpose of this study is to find out whether the 1987 RA classification criteria can be transformed into a flexible diagnostic tool, providing diagnostic results at many different levels of confidence and yielding high sensitivity rates at lower levels of confidence, and high specificity rates at higher levels.

We compared three models of the 1987 RA classification criteria: (1) a reference model, which represents the

Table 1 The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis: the (shortened) original criteria definition and the formalized definition of the reference model. In order to establish a diagnosis, at least four of the seven criteria must be present in a patient. This condition can be formalized as a rule with a premise which contains the disjunctive normal form of the criteria s_i : $(s_1 \wedge s_2 \wedge s_3 \wedge s_4) \vee (s_1 \wedge s_2 \wedge s_3 \wedge s_5) \vee \dots \vee (s_4 \wedge s_5 \wedge s_6 \wedge s_7) \rightarrow c$.

ARA definition of criterion s_i	Definition of criterion s_i as used in the reference model
s_1 : Morning stiffness in and around the joints, lasting at least one hour before maximum improvement where t is the duration of morning stiffness in minutes	$s_1 = \begin{cases} 0, & \text{if } t < 60; \\ 1, & \text{if } t \geq 60; \\ \nu, & \text{if } t \text{ is unknown.} \end{cases}$
s_2 : Arthritis of 3 or more joint areas (The 14 possible areas are right or left proximal interphalangeal, metacarpophalangeal, wrist, elbow, knee, ankle, and metatarsophalangeal joints) where n is the number of involved joint areas	$s_2 = \begin{cases} 0, & \text{if } n < 3; \\ 1, & \text{if } n \geq 3; \\ \nu, & \text{if } n \text{ is unknown.} \end{cases}$
s_3 : Arthritis of hand joints (wrist, metacarpophalangeal or proximal interphalangeal joints only)	$s_3 = \begin{cases} 0, & \text{if hand joints are not involved;} \\ 1, & \text{if hand joints are involved;} \\ \nu, & \text{if the involvement of hand joints is unknown.} \end{cases}$
s_4 : Symmetric arthritis (simultaneous involvement of the same joint areas on both sides of the body) where the symmetry ratio q is the number of symmetrically involved joint areas divided by the total number of involved joint areas ¹	$s_4 = \begin{cases} 0, & \text{if } q < 0.5; \\ 1, & \text{if } q \geq 0.5; \\ \nu, & \text{if } q \text{ is unknown.} \end{cases}$
s_5 : Rheumatoid nodules	$s_5 = \begin{cases} 0, & \text{if nodules are not observed;} \\ 1, & \text{if nodules are observed;} \\ \nu, & \text{if the presence of nodules is unknown.} \end{cases}$
s_6 : Serum rheumatoid factor positive where $1 : t$ is the titer of the Waaler-Rose test	$s_6 = \begin{cases} 0, & \text{if } t < 64; \\ 1, & \text{if } t \geq 64; \\ \nu, & \text{if } t \text{ is unknown.} \end{cases}$
s_7 : Radiographic changes typical of RA	$s_7 = \begin{cases} 0, & \text{if changes are not observed;} \\ 1, & \text{if changes are observed;} \\ \nu, & \text{if the presence of changes is unknown.} \end{cases}$

¹If no joint area is involved, then $q := 0$.

Table 2 Definitions of the membership functions $\mu_{s_i}(\chi)$ of criteria s_1, s_2, s_4 and s_6 , as used in fuzzy models 1 and 2. Fuzzy sets were applied to only four of the seven criteria. In criteria s_3, s_5 and s_7 the definitions of the reference model, as shown in Table 1, are valid for all models.

Criterion s_i	Membership function $\mu_{s_i}(x)$	
	of fuzzy model 1	of fuzzy model 2
s_1 : Morning stiffness lasting at least one hour	$\mu_{s_1}(t) = \begin{cases} 0, & \text{if } t \leq 15; \\ \frac{t-15}{45}, & \text{if } 15 < t < 60; \\ 1, & \text{if } t \geq 60; \\ \nu, & \text{if } t \text{ is unknown.} \end{cases}$ <p>where t is the duration of morning stiffness in minutes</p>	$\mu_{s_1}(t) = \begin{cases} 0, & \text{if } t \leq 15; \\ \frac{t-15}{90}, & \text{if } 15 < t < 105; \\ 1, & \text{if } t \geq 105; \\ \nu, & \text{if } t \text{ is unknown.} \end{cases}$
s_2 : Arthritis of 3 or more joint areas	$\mu_{s_2}(n) = \begin{cases} 0, & \text{if } n \leq 1; \\ 0.5, & \text{if } n = 2; \\ 1, & \text{if } n \geq 3; \\ \nu, & \text{if } n \text{ is unknown.} \end{cases}$ <p>where n is the number of involved joint areas</p>	$\mu_{s_2}(n) = \begin{cases} 0, & \text{if } n \leq 1; \\ 0.25, & \text{if } n = 2; \\ 0.50, & \text{if } n = 3; \\ 0.75, & \text{if } n = 4; \\ 1, & \text{if } n \geq 5; \\ \nu, & \text{if } n \text{ is unknown.} \end{cases}$
s_4 : Symmetric arthritis	$\mu_{s_4}(q) = \begin{cases} 0, & \text{if } q = 0; \\ 2q, & \text{if } 0 < q < 0.5; \\ 1, & \text{if } q \geq 0.5; \\ \nu, & \text{if } q \text{ is unknown.} \end{cases}$ <p>where the symmetry ratio q is the number of symmetrically involved joint areas divided by the total number of involved joint areas¹</p>	$\mu_{s_4}(q) = \begin{cases} q, & \text{if } 0 \leq q \leq 1; \\ \nu, & \text{if } q \text{ is unknown.} \end{cases}$
s_6 : Serum rheumatoid factor positive	$\mu_{s_6}(t) = \begin{cases} 0, & \text{if } t \leq 16; \\ 0.5, & \text{if } t = 32; \\ 1, & \text{if } t \geq 64; \\ \nu, & \text{if } t \text{ is unknown.} \end{cases}$ <p>where $1 : t$ is the titer of the Waaler-Rose test</p>	$\mu_{s_6}(t) = \begin{cases} 0, & \text{if } t \leq 16; \\ 0.25, & \text{if } t = 32; \\ 0.50, & \text{if } t = 64; \\ 0.75, & \text{if } t = 128; \\ 1, & \text{if } t \geq 256; \\ \nu, & \text{if } t \text{ is unknown.} \end{cases}$

¹If no joint area is involved, then $q := 0$.

original criteria definition and provides results on a definite level only; (2) a fuzzy model, which offers results on a definite and several possible levels (fuzzy model 1); and (3) a fuzzy model, which offers results on a definite, several possible, and several superdefinite levels (fuzzy model 2). Details of the three models and the development guidelines of the fuzzy sets are given in Section 2. Section 3 presents the results obtained by testing the different models

with 292 clinical cases from a hospital for rheumatic diseases. Section 4 discusses the obtained results.

2. Development of the Reference Model and Fuzzy Models 1 and 2

Table 1 presents a shortened definition of the 1987 RA classification crite-

ria. At least four out of the list of seven criteria must be present in a patient to establish a diagnosis. This condition can be formalized as a rule with a premise that contains a disjunctive normal form of the criteria s_i : $(s_1 \cap s_2 \cap s_3 \cap s_4) \cup (s_1 \cap s_2 \cap s_3 \cap s_5) \cup \dots \cup (s_4 \cap s_5 \cap s_6 \cap s_7) \rightarrow c$. The rule's conclusion c is calculated in different ways by the different models described below.

2.1 The Reference Model

The reference model represents the original criteria definition, with an additional logical value for indefinite statements (for criteria that have not been examined). The definitions of present, absent, and unknown criteria are given in Table 1. To provide results with three different logical values, the reference model used Kleene's logic [20] (Table 3, left side). Because there was only one logical value for positive results, diagnoses could be provided on only one (definite) level.

2.2 The Fuzzy Models

In the fuzzy models, criteria s_i were replaced with fuzzy sets, which were characterized by membership functions μ_{s_i} . If a criterion s_i (e.g., s_1 : "morning stiffness lasting at least one hour") could be based on a variable x (in our example: "duration of morning stiffness in minutes"), the membership function $\mu_{s_i}: X \rightarrow [0,1] \cup \nu$ assigned to every possible $x \in X$ a degree of membership of x in s_i . Thus, $\mu_{s_i}(x)$ expressed the degree of membership to which criterion s_i was observed in a patient with a finding x (which allowed the mathematical modeling of borderline findings).

In our fuzzy models, the membership functions $\mu_{s_i}(x)$ were piecewise linear and they could be characterized by an "upper boundary" (which was the smallest value of x , where $\mu_{s_i}(x) = 1$) and a "lower boundary" (which was the largest value of x , where $\mu_{s_i}(x) = 0$). This was in contrast to the reference model, where criteria were characterized by a single boundary.

Four out of the seven criteria were replaced with fuzzy sets (criteria s_1, s_2, s_4 , and s_6 , see Table 2). In criteria s_3, s_5 , and s_7 , variables for the membership functions were not available (which will

Three-Valued Logic of Kleene				Fuzzy Logic		
conjunction:				conjunction:		
$s_x \wedge s_y$	$s_y = 0$	$s_y = 1$	$s_y = \nu$	$s_x \wedge s_y$	$s_y \in [0,1]$	$s_y = \nu$
$s_x = 0$	0	0	0	$s_x \in [0,1]$	$\min(s_x, s_y)$	ν
$s_x = 1$	0	1	ν	$s_x = \nu$	ν	ν
$s_x = \nu$	0	ν	ν	disjunction:		
disjunction:				disjunction:		
$s_x \vee s_y$	$s_y = 0$	$s_y = 1$	$s_y = \nu$	$s_x \vee s_y$	$s_y \in [0,1]$	$s_y = \nu$
$s_x = 0$	0	1	ν	$s_x \in [0,1]$	$\max(s_x, s_y)$	s_x
$s_x = 1$	1	1	1	$s_x = \nu$	s_y	ν
$s_x = \nu$	ν	1	ν	negation:		
negation:				negation:		
	$\neg s_x$				$\neg s_x$	
$s_x = 0$	1			$s_x \in [0,1]$	$1 - s_x$	
$s_x = 1$	0			$s_x = \nu$	ν	
$s_x = \nu$	ν					

Table 3 Operator definitions of Kleene's logic (as used in the reference model) and of fuzzy logic (as used in fuzzy models 1 and 2).

be discussed in section 4). Because a replacement with fuzzy sets was not possible for these criteria, they were defined in the same way as in the reference model.

In the fuzzy models the rule was calculated with the application of fuzzy logic (for a definition of the fuzzy logical connectives see Table 3, right side). As a result, the rule's conclusion c took values in $[0,1] \cup \nu$.

An important design guideline of the fuzzy models was that the definite level of confidence should be consistent with the original criteria definition.

In fuzzy model 1 (which is a modified version of the model described in [19]), the upper boundaries of all membership functions were set equal to the boundary of the reference model. Thus, the definite level of fuzzy model 1 was equivalent to a result $c = 1$. Values $c \in (0,1)$ represented possible levels, whereas $c \in \{0, \nu\}$ indicated negative results.

Note that the definite level of confidence is only consistent with the original criteria definition if negations do not occur in the premise of the diagnostic rule. This restriction is not considered harmful for our purpose, because the use of exclusive criteria was totally abandoned during the development of the ACR classification criteria. This means that negations would never occur in the premises of diagnostic rules for ACR criteria.

To provide a wide range of different possible levels, the lower boundaries of the membership functions were set as low as possible. The limiting factor was that the medical intention of a criterion might be reversed if criteria definitions were not rigid enough. In criterion s_1 ("morning stiffness lasting at least one hour") the lowest boundary that could be accepted was 15 minutes. Stiffness of short duration, especially at the onset of joint motion, is not known as a typical sign of RA, but rather of osteoarthritis, a different rheumatic disease. In criterion s_2 ("arthritis of three or more joint areas"), an involvement of only a single joint (monarthritis) was not satisfying, because monarthritis and polyarthritis are opposite medical concepts. In criterion s_4 ("symmetric arthritis"), a symmetry ratio $q = 0$ mean that the joint involvement was completely asymmetric.

In criterion s_6 ("serum rheumatoid factor positive"), the lower boundary was equal to the threshold that was used in our clinic to classify patients as being seropositive or seronegative.

Fuzzy model 2 was an extension of fuzzy model 1, developed to provide additional superdefinite levels. In this model, the boundaries of the reference model were equal to values x , where $\mu_{s_i}(x) = 0.5$ in criteria s_1, s_2, s_4 , and s_6 . The lower boundaries of the membership functions were set equal to those in fuzzy model 1, because the medical considerations discussed before were valid for both models. In this way, the membership functions $\mu_{s_i}(x)$ were defined within the interval $\mu_{s_i}(x) \in [0, 0.5]$. A linear extension of $\mu_{s_i}(x)$ from the interval $[0,0.5]$ to $[0.5,1]$ led to the upper boundaries of all membership functions. Thus, in this model the definite level was equivalent to a result $c = 0.5$. Values $c \in (0,0.5)$ represented possible levels, values $c \in (0.5,1]$ superdefinite levels. As in fuzzy model 1, results $c \in \{0, \nu\}$ were considered negative.

Note that this is in contrast to the usual definition of a fuzzy set, where $\mu_{s_i}(x) = 1$, if x is fully compatible with the medical concept s_i , i. e., if s_i is definitely present at a value x . In section 4 we discuss an alternative version of fuzzy model 2, which is consistent with this definition.

3. Evaluation of the Reference Model and Fuzzy Models 1 and 2

3.1 Patients

All patients in this study underwent treatment in a 140-bed hospital for rheumatic diseases in Baden, Austria, between 1979 and 1987. Outpatients were not included. For each patient, two completed questionnaires for medical history and the results of the physical examination, as well as laboratory test results including serum rheumatoid factor were available. X-ray examinations were carried out in most patients and all available results were entered into a patient databank. Baseline characteristics of RA and control patients were not identical (RA patients: 77.4%

Table 4 Diagnoses and numbers of control patients with rheumatic diseases other than RA.

diagnosis	number of patients
osteoarthritis	45
gouty arthritis	32
ankylosing spondylitis	30
psoriatic arthritis	20
Reiter's disease	4
systemic lupus erythematosus	4
chronic scleritis	4
chondrocalcinosis	3
polymyositis	2
polymyalgia rheumatica	2
total number of patients	146

female, mean age: 62.1 years; control patients: 38.4% female, mean age: 57.3 years), because ages and sex ratios of patients with different rheumatic diseases tended to be different from each other.

RA patients: All 146 RA patients had a confirmed clinical diagnosis of RA. Diagnoses were validated by at least two experienced rheumatologists. Because of the known lack of sensitivity of the ACR classification criteria in cases with early RA, no criteria were applied to select study participants. If patients had been admitted to the hospital several times, we selected only the patient records of their first admission to include a larger percentage of patients with early stages of the disease in the study. If the RA patient group is additionally subdivided according to a disease staging based on radiographic findings [21], 24 patients were in disease stage 1, 51 in stage 2, 48 in stage 3, and 23 in stage 4.

Control patients: All 146 control patients had a confirmed clinical diagnosis of a rheumatic disease other than RA. The diagnoses (see Table 4) included osteoarthritis and chronic inflammatory rheumatic diseases.

3.2 Methods

For each patient, the conclusion c of the diagnostic rule was calculated three times, once for each model.

In the reference model, the presence or absence of each criterion was determined with the criteria definition in Table 1. Results were obtained for each patient when Kleene's logic was applied to the diagnostic rule.

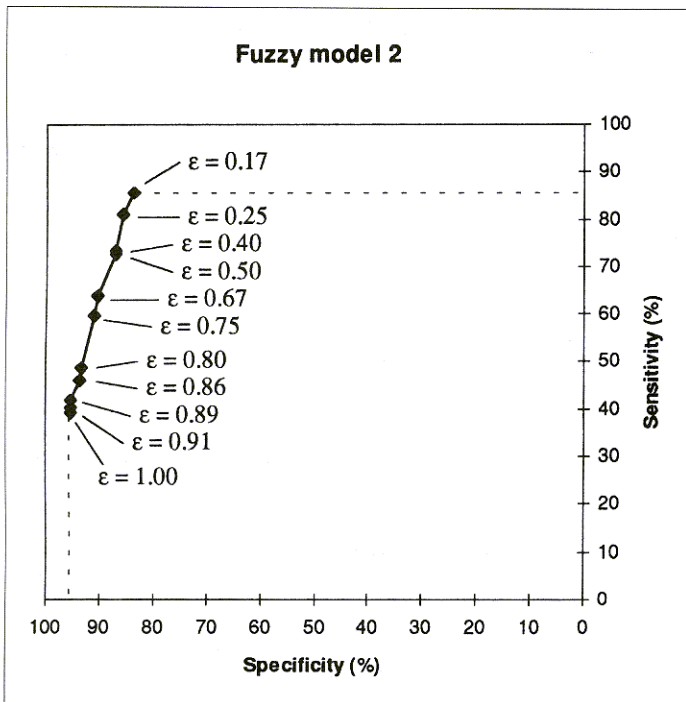
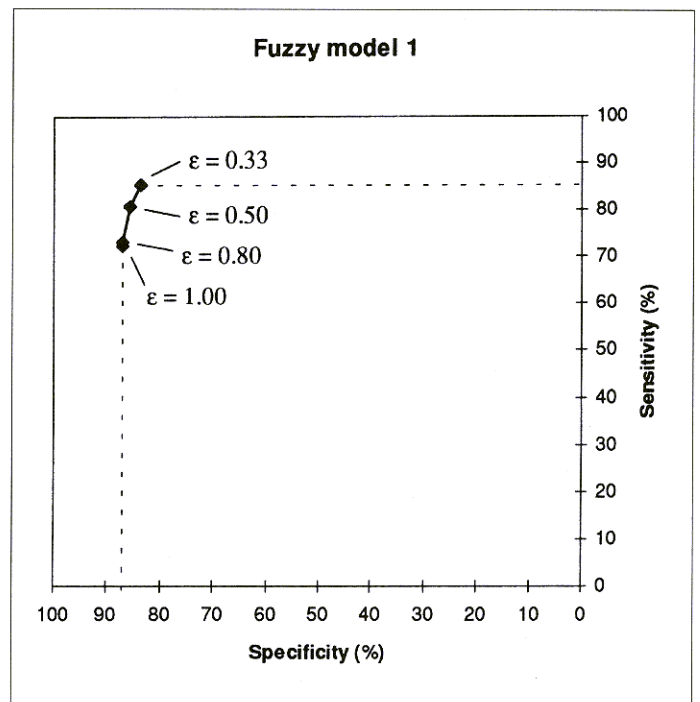
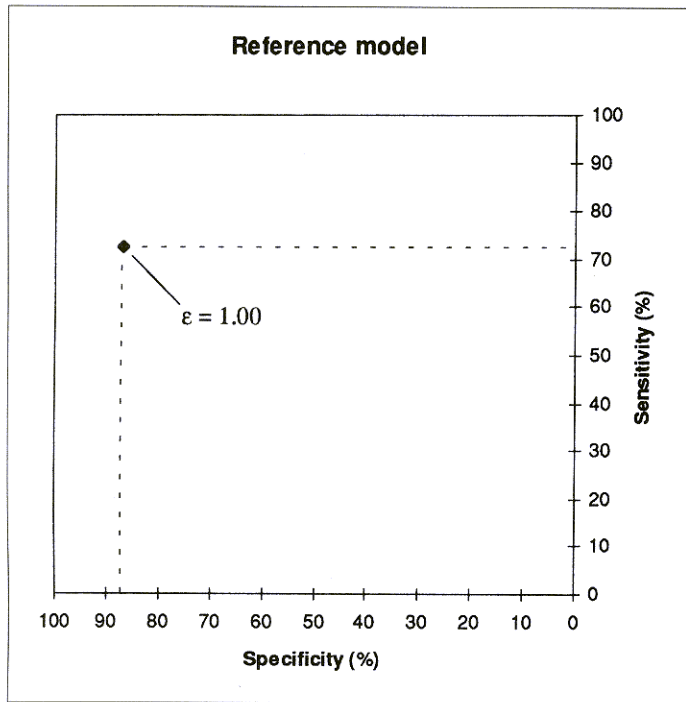


Fig. 1 ROC curves of the reference model and the fuzzy models 1 and 2. Rates of sensitivity and specificity are represented for different values of the threshold ϵ .

Results of the diagnostic rule were compared with the available clinical discharge diagnoses. In this way positive and negative results at each level were subclassified as true-positive, false-positive, true-negative, and false-negative results.

Sensitivity and specificity rates for each level of confidence were obtained with the numbers of patients with true-positive, false-positive, true-negative, and false-negative results.

To calculate sensitivity rates, the number of true-positive results with RA patients was divided by the number of true-positive results with RA patients plus the number of false-negative results with control patients.

To calculate specificity rates, the number of true-negative results with control patients was divided by the number of true-negative results with control patients plus the number of false-positive results with RA patients.

In the fuzzy models, the membership functions of criteria s_1, s_2, s_4 and s_6 were calculated according to the definitions in Table 2. The presence or absence of criteria s_3, s_5 and s_7 was determined in the same way as in the reference model. Results were obtained for each patient when fuzzy logic was applied to the diagnostic rule.

To obtain results at different levels of confidence, the numbers of patients with positive results were determined for each level. Each level was represented by a threshold ϵ , thus positive results c at ϵ were greater than or equal to ϵ ($c \geq \epsilon$ for $0 < \epsilon \leq 1$). Results $c < \epsilon$ and $c = v$ were considered negative results at ϵ .

3.3 Results

Table 5 presents the results of the membership functions $\mu_{s_i}(x)$ that were observed in the study patients. In the set of criteria, which were replaced by fuzzy sets, the numbers of different values in a single criterion were either one (s_2, s_4, s_6) or two (s_1) in fuzzy model 1,

Table 5 Results of the membership functions $\mu_{s_i}(x)$ for each criterion, obtained in RA and control patients with the reference model and fuzzy models 1 and 2.

Reference model							
criterion	s_1	s_2	s_3	s_4	s_5	s_6	s_7
RA patients							
$\mu_{s_i}(x) \in \{0.00, \nu\}$	106	39	19	26	126	91	7
$\mu_{s_i}(x) = 1.00$	40	107	127	120	20	55	139
control patients							
$\mu_{s_i}(x) \in \{0.00, \nu\}$	126	123	109	110	143	145	114
$\mu_{s_i}(x) = 1.00$	20	23	37	36	3	1	32

Fuzzy model 1

criterion	s_1	s_2	s_3	s_4	s_5	s_6	s_7
RA patients							
$\mu_{s_i}(x) \in \{0.00, \nu\}$	53	17	19	24	126	74	7
$\mu_{s_i}(x) = 0.33$	52	0	0	0	0	0	0
$\mu_{s_i}(x) = 0.50$	0	22	0	0	0	17	0
$\mu_{s_i}(x) = 0.66$	1	0	0	0	0	0	0
$\mu_{s_i}(x) = 0.80$	0	0	0	2	0	0	0
$\mu_{s_i}(x) = 1.00$	40	107	127	120	20	55	139
control patients							
$\mu_{s_i}(x) \in \{0.00, \nu\}$	110	102	109	109	143	143	114
$\mu_{s_i}(x) = 0.33$	16	0	0	0	0	0	0
$\mu_{s_i}(x) = 0.50$	0	21	0	0	0	2	0
$\mu_{s_i}(x) = 0.66$	0	0	0	0	0	0	0
$\mu_{s_i}(x) = 0.80$	0	0	0	1	0	0	0
$\mu_{s_i}(x) = 1.00$	20	23	37	36	3	1	32

Fuzzy model 2

criterion	s_1	s_2	s_3	s_4	s_5	s_6	s_7
RA patients							
$\mu_{s_i}(x) \in \{0.00, \nu\}$	53	17	19	24	126	74	7
$\mu_{s_i}(x) = 0.17$	52	0	0	0	0	0	0
$\mu_{s_i}(x) = 0.25$	0	22	0	0	0	17	0
$\mu_{s_i}(x) = 0.33$	1	0	0	0	0	0	0
$\mu_{s_i}(x) = 0.40$	0	0	0	2	0	0	0
$\mu_{s_i}(x) = 0.50$	18	18	0	3	0	15	0
$\mu_{s_i}(x) = 0.67$	0	0	0	19	0	0	0
$\mu_{s_i}(x) = 0.75$	0	21	0	3	0	15	0
$\mu_{s_i}(x) = 0.80$	0	0	0	9	0	0	0
$\mu_{s_i}(x) = 0.83$	2	0	0	0	0	0	0
$\mu_{s_i}(x) = 0.86$	0	0	0	10	0	0	0
$\mu_{s_i}(x) = 0.89$	0	0	0	4	0	0	0
$\mu_{s_i}(x) = 0.91$	0	0	0	2	0	0	0
$\mu_{s_i}(x) = 1.00$	20	68	127	70	20	25	139
control patients							
$\mu_{s_i}(x) \in \{0.00, \nu\}$	110	102	109	109	143	143	114
$\mu_{s_i}(x) = 0.17$	16	0	0	0	0	0	0
$\mu_{s_i}(x) = 0.25$	0	21	0	0	0	2	0
$\mu_{s_i}(x) = 0.33$	0	0	0	0	0	0	0
$\mu_{s_i}(x) = 0.40$	0	0	0	1	0	0	0
$\mu_{s_i}(x) = 0.50$	10	5	0	3	0	1	0
$\mu_{s_i}(x) = 0.67$	0	0	0	6	0	0	0
$\mu_{s_i}(x) = 0.75$	0	7	0	0	0	0	0
$\mu_{s_i}(x) = 0.80$	0	0	0	2	0	0	0
$\mu_{s_i}(x) = 0.83$	0	0	0	0	0	0	0
$\mu_{s_i}(x) = 0.86$	0	0	0	3	0	0	0
$\mu_{s_i}(x) = 0.89$	0	0	0	0	0	0	0
$\mu_{s_i}(x) = 0.91$	0	0	0	0	0	0	0
$\mu_{s_i}(x) = 1.00$	10	11	37	22	3	0	32

Table 6 Results obtained in RA and control patients with the reference model and fuzzy models 1 and 2. Table entries represent numbers of patients with a given value of the rule's conclusion c . (This Table is continued on the next page.)

Reference model

diagnosis	total number $c \in \{0.00, \nu\}$ $c = 1.00$ of patients		
RA patients	146	40	106
control patients (total)	146	127	19
osteoarthritis	45	45	0
gouty arthritis	32	28	4
ankylosing spondylitis	30	29	1
psoriatic arthritis	20	10	10
Reiter's disease	4	4	0
systemic lupus erythematosus	4	1	3
systemic sclerosis	4	4	0
chondrocalcinosis	3	3	0
polymyositis	2	1	1
polymyalgia rheumatica	2	2	0

Fuzzy model 1

diagnosis	total number $c \in \{0.00, \nu\}$ $c = 0.33$ $c = 0.50$ $c = 0.80$ $c = 1.00$ of patients					
RA patients	146	21	7	11	1	106
control patients (total)	146	122	3	2	0	19
osteoarthritis	45	44	0	1	0	0
gouty arthritis	32	28	0	0	0	4
ankylosing spondylitis	30	28	0	1	0	1
psoriatic arthritis	20	7	3	0	0	10
Reiter's disease	4	4	0	0	0	0
systemic lupus erythematosus	4	1	0	0	0	3
systemic sclerosis	4	4	0	0	0	0
chondrocalcinosis	3	3	0	0	0	0
polymyositis	2	1	0	0	0	1
polymyalgia rheumatica	2	2	0	0	0	0

and either three (s_2, s_6), four (s_1), or eight (s_4) in fuzzy model 2.

Table 6 shows the numbers of RA and control patients within the different result categories. In our study results $c \in \{0, \nu\}$ were always considered negative and they were placed in the same category. The numbers of different result categories were two in the reference model, five in fuzzy model 1, and 12 in fuzzy model 2.

The high percentage of false-positive results in patients with psoriatic arthritis and systemic lupus erythematosus shows that patients with these diseases posed most difficulties for classification. Even at the highest level of confidence ($\epsilon = 1.00$) of fuzzy model 2, some of these patients were still misclassified as RA patients. In contrast,

patients with other rheumatic diseases were much less likely to be misclassified.

Table 7 presents the numbers of patients with true- and false-positive results and the rates of sensitivity and specificity for each level of confidence. A different level (and a new choice of ϵ) was presented for each different result c which occurred in the study patients.

The reference model provided results at one definite level (at $\epsilon = 1.00$) and yielded a sensitivity rate of 72.6% and a specificity rate of 87.0%. Fuzzy model 1 offered results at one definite (at $\epsilon = 1.00$) and three possible levels (at $\epsilon = 0.33, 0.50, 0.80$). Sensitivity and specificity rates ranged from 72.6% to 85.6% and from 83.6% to 87.0%. Fuzzy model 2 provided results at one definite

Table 6 continued

Fuzzy model 2

diagnosis	total number of patients	$c \in \{0.00, \nu\}$	$c = 0.17$	$c = 0.25$	$c = 0.40$	$c = 0.50$	$c = 0.67$
RA patients	146	21	7	11	1	13	6
control patients (total)	146	122	3	2	0	5	1
osteoarthritis	45	44	0	1	0	0	0
gouty arthritis	32	28	0	0	0	2	0
ankylosing spondylitis	30	28	0	1	0	0	0
psoriatic arthritis	20	7	3	0	0	3	1
Reiter's disease	4	4	0	0	0	0	0
systemic lupus erythematosus	4	1	0	0	0	0	0
systemic sclerosis	4	4	0	0	0	0	0
chondrocalcinosis	3	3	0	0	0	0	0
polymyositis	2	1	0	0	0	0	0
polymyalgia rheumatica	2	2	0	0	0	0	0

diagnosis	$c = 0.75$	$c = 0.80$	$c = 0.86$	$c = 0.89$	$c = 0.91$	$c = 1.00$
RA patients	16	4	6	2	2	57
control patients (total)	3	1	2	0	0	7
osteoarthritis	0	0	0	0	0	0
gouty arthritis	1	0	1	0	0	0
ankylosing spondylitis	1	0	0	0	0	0
psoriatic arthritis	1	1	0	0	0	4
Reiter's disease	0	0	0	0	0	0
systemic lupus erythematosus	0	0	0	0	0	3
systemic sclerosis	0	0	0	0	0	0
chondrocalcinosis	0	0	0	0	0	0
polymyositis	0	0	1	0	0	0
polymyalgia rheumatica	0	0	0	0	0	0

(at $\epsilon = 0.50$), three possible (at $\epsilon = 0.17, 0.25, 0.40$), and seven superdefinite levels (at $\epsilon = 0.67, 0.75, 0.80, 0.86, 0.89, 0.91, 1.00$). Sensitivity and specificity rates ranged from 39.0% to 85.6% and from 83.6% to 95.2%.

To make the results of the models more comparable to each other, the rates of sensitivity and specificity were represented as ROC curves (Fig. 1).

Fuzzy model 1 can be seen as an extension of the reference model. It provided three additional result categories, which represented three additional possible levels of confidence. On the lowest possible level (at $\epsilon = 0.33$), 19 additional patients were included as true-positive diagnostic hypotheses and a maximum sensitivity rate of 85.6% was obtained.

Fuzzy model 2 can be seen as an extension of fuzzy model 1. It provided seven additional result categories, which represented seven additional superdefinite levels of confidence. On the highest superdefinite levels (at $\epsilon = 0.89, 0.91, 1.00$) the number of patients

with false-positive results was reduced to seven and a maximum specificity rate of 95.2% was obtained. The maximum sensitivity rate at the lowest possible level (at $\epsilon = 0.17$) was 85.6%, which was equal to the rate in fuzzy model 1.

4. Discussion

The purpose of this study was to establish whether a set of classification criteria could be transformed into a flexible diagnostic tool, providing diagnostic results at many different levels of confidence, and yielding high sensitivity rates at the possible levels and high specificity rates at the superdefinite levels.

The limiting factors that restricted the number of observed different levels of confidence and the maximum performance rates at these levels can be divided into: (1) those resulting from the design guidelines of the fuzzy sets, and (2) those resulting from patient characteristics.

As a first design guideline, we decided that criteria would only be replaced with fuzzy sets if variables for the membership functions were available. For this reason, fuzzy sets were applied to only four of the seven criteria. New variables could possibly have been introduced to define a degree of presence of the other three criteria (such as variables for the progression of rheumatoid nodules or the progression of X-ray changes). However, it would have been difficult to find a definition of the criteria consistent with the original definition (and which was accepted by the rheumatic community).

Another decision made during the design of the fuzzy sets was the choice of the lower and upper boundaries of the membership functions. They were chosen with regard to both logical and medical considerations. The lower boundaries were set as low as possible without affecting the medical intention of the criteria. The choice of the upper boundaries was determined by the definition of the definite levels of confidence. Both of these design guidelines restricted the degree of variability of the fuzzy results and the degree of variability of the obtained performance rates.

The number of different result categories and the number of different levels of confidence were limited by membership functions, which were based on discrete variables. As shown in Table 2, $\mu_{s_2}(n)$ and $\mu_{s_6}(t)$ were defined as discrete functions. $\mu_{s_4}(q)$ was defined as a continuous function, but it took only discrete values, because the symmetry ratio q itself was a discrete function. Discrete functions were especially restrictive if the ranges between lower and upper boundaries were small. In two of the criteria of fuzzy model 1 (s_2, s_6) only a single value of the variables n, t could result in a value of the membership functions $\mu_{s_2}(n)$ and $\mu_{s_6}(t)$ that was unequal 0 or 1.

Even if variables were continuous and the ranges between both boundaries were sufficiently large, patient characteristics could still be a limiting factor if variable values occurred in a discrete rather than in a continuous manner. In criterion s_1 ("morning stiffness lasting at least one hour"), the duration of morning stiffness was reported in multi-

ples of 15 minutes in most of our patients.

The reported sensitivity and specificity rates for the different levels of confidence can only be generalized if the models are used in a patient setting similar to our setting. Otherwise, the models have to be tested with a learning set of patients from the intended application setting first. With results obtained in this manner, the models can be used in two ways: (1) as a diagnostic tool with variable sensitivity and specificity, or (2) as a diagnostic tool used at a fixed level of confidence, which was best suited for the application setting during testing. This, for example, is the case when a maximum rate of accuracy is required. (Accuracy rates are means of sensitivity and specificity rates, if the numbers of RA and control patients are equal.)

A problem which has to be addressed is that the definition of fuzzy model 2, as presented in this study, is not consistent with the common definition of a fuzzy set. Normally, the value of a membership function $\mu_{s_i}(x)$ is equal to 1, if the criterion s_i is definitely present at a value x . This is true for fuzzy model 1, but not for fuzzy model 2, where $\mu_{s_i}(x) = 0.5$, if s_i is definitely present at x . Fuzzy model 2 was developed as a demonstrative model for this study, but it can be readjusted to be consistent with the common definition. This adjustment is also necessary, if the model is to be used in the CADIAG-2/RHEUMA system.

An example of such an adjustment is the division of fuzzy model 2 into fuzzy model 1 and an additional superdefinite model. Fuzzy model 1 provides results on definite and possible levels, and the superdefinite model results on superdefinite levels. In order to divide fuzzy model 2, all the membership functions of fuzzy model 2 must be divided. Table 8 shows an example where the membership function $\mu_{s_1}(t)$ of fuzzy model 2 is divided into $\mu_{s_1}(t)$ of fuzzy model 1 and $\mu_{s_1}(t)$ of the superdefinite model. The parameter $\mu_{s_1}(t)$ of the superdefinite model is characterized by a lower boundary, which is equal to the upper boundary of $\mu_{s_1}(t)$ of fuzzy model 1 and an upper boundary, which is equal to the upper boundary of $\mu_{s_1}(t)$ of fuzzy model 2. The superdefinite model

Table 7 Results obtained in 146 RA patients and 146 control patients with the reference model and fuzzy models 1 and 2. The numbers of true- and false-positive results and the sensitivity and specificity rates are shown for each level of confidence (represented by the threshold ϵ). A different level (and a new choice of ϵ) was presented for each different result c which occurred in our study patients.

Reference model					
threshold ϵ	number of true-positive results	sensitivity rate	number of false-positive results	specificity rate	level of confidence
$\epsilon = 1.00$	106	72.6%	19	87.0%	definite
Fuzzy model 1					
threshold ϵ	number of true-positive results	sensitivity rate	number of false-positive results	specificity rate	level of confidence
$\epsilon = 0.33$	125	85.6%	24	83.6%	possible
$\epsilon = 0.50$	118	80.8%	21	85.6%	possible
$\epsilon = 0.80$	107	73.3%	19	87.0%	possible
$\epsilon = 1.00$	106	72.6%	19	87.0%	definite
Fuzzy model 2					
threshold ϵ	number of true-positive results	sensitivity rate	number of false-positive results	specificity rate	level of confidence
$\epsilon = 0.17$	125	85.6%	24	83.6%	possible
$\epsilon = 0.25$	118	80.8%	21	85.6%	possible
$\epsilon = 0.40$	107	73.3%	19	87.0%	possible
$\epsilon = 0.50$	106	72.6%	19	87.0%	definite
$\epsilon = 0.67$	93	63.7%	14	90.4%	superdefinite
$\epsilon = 0.75$	87	59.6%	13	91.1%	superdefinite
$\epsilon = 0.80$	71	48.6%	10	93.2%	superdefinite
$\epsilon = 0.86$	67	45.9%	9	93.8%	superdefinite
$\epsilon = 0.89$	61	41.8%	7	95.2%	superdefinite
$\epsilon = 0.91$	59	40.4%	7	95.2%	superdefinite
$\epsilon = 1.00$	57	39.0%	7	95.2%	superdefinite

Table 8 To be consistent with the common definition of a fuzzy set, fuzzy model 2 can be separated into fuzzy model 1 and an additional superdefinite model. For this purpose, all the membership functions of fuzzy model 2 must be separated. Here, the membership function $\mu_{s_1}(t)$ of fuzzy model 2 is separated into $\mu_{s_1}(t)$ of fuzzy model 1 and $\mu_{s_1}(t)$ of the superdefinite model.

Membership function $\mu_{s_1}(t)$ of fuzzy model 2	Membership function $\mu_{s_1}(t)$ of fuzzy model 1	Membership function $\mu_{s_1}(t)$ of the superdefinite model
$\mu_{s_1}(t) = \begin{cases} 0, & \text{if } t \leq 15; \\ \frac{t-15}{30}, & \text{if } 15 < t < 105; \\ 1, & \text{if } t \geq 105; \\ \nu, & \text{if } t \text{ is unknown.} \end{cases}$	$\mu_{s_1}(t) = \begin{cases} 0, & \text{if } t \leq 15; \\ \frac{t-15}{45}, & \text{if } 15 < t < 60; \\ 1, & \text{if } t \geq 60; \\ \nu, & \text{if } t \text{ is unknown.} \end{cases}$	$\mu_{s_1}(t) = \begin{cases} 0, & \text{if } t \leq 60; \\ \frac{t-60}{45}, & \text{if } 60 < t < 105; \\ 1, & \text{if } t \geq 105; \\ \nu, & \text{if } t \text{ is unknown.} \end{cases}$

is based on criteria which are more rigid than in the original definition. It is handled separately and used as a supplement to fuzzy model 1 to provide additional results on superdefinite levels.

The design guidelines proposed in this paper are, perhaps, too restrictive to take real advantage of fuzzy set theory and fuzzy logic. The main reason for following the guidelines was to preserve the original definition of the ACR criteria on the one hand and to provide them with additional flexibility on the other.

We felt that this approach would be more accepted within the rheumatic community than a proposal of completely new criteria. Strict guidelines are also useful for applying fuzzy logic to more than one set of classification criteria. Results of different sets of criteria would then be directly comparable with each other, allowing a differential diagnosis of rheumatic diseases. This approach is also well suited as a method to develop new diagnostic rules for the CADIAG-2/RHEUMA system.

Acknowledgments

We thank I. Gröger and G. Sedivy, who helped us prepare the patient records and CADIAG-2's knowledge base; C. Schuh, F. Fischler and W. Temsch for their programming work for CADIAG-2; G. Kolousek for technical advice; and F. Steimann for helpful discussions on the presented subject. This research was supported by the *Hochschuljubiläumsstiftung 1994 der Stadt Wien*.

REFERENCES

1. Adlassnig K-P. A fuzzy logical model of computer-assisted medical diagnosis. *Meth Inform Med* 1980; 19: 141-8.
2. Adlassnig K-P. Fuzzy set theory in medical diagnosis. *IEEE T Syst Man Cyb* 1986; 16: 260-5.
3. Adlassnig K-P. Uniform representation of vagueness and imprecision in patient's medical findings using fuzzy sets. In: Trappl R (ed). *Proceedings of Cybernetics and Systems '88*. Dordrecht: Kluwer Academic Publishers 1988; 685-92.
4. Adlassnig K-P, Kolarz G. Representation and semiautomatic acquisition of medical knowledge in CADIAG-1 and CADIAG-2. *Comput Biomed Res* 1986; 19: 63-79.
5. Adlassnig K-P, Kolarz G, Scheithauer W, Effenberger H, Grabner G. CADIAG: Approaches to computer-assisted medical diagnosis. *Comput Biol Med* 1985; 15: 315-35.
6. Adlassnig K-P, Scheithauer W. Performance evaluation of medical expert systems using ROC curves. *Comput Biomed Res* 1989; 22: 297-313.
7. Bernelot Moens HJ, Van der Korst JK. Development and validation of a computer program using Bayes' theorem to support diagnosis of rheumatic diseases. *Ann Rheum Dis* 1992; 51: 266-71.
8. Horn W. MESICAR – A medical expert system integrating causal and associative reasoning. *Appl Artif Intell* 1989; 3: 305-56.
9. Kingsland III LC, Lindberg DAB. The criteria form of knowledge representation in medical artificial intelligence. In: Salamon R, Blum B, Jorgensen M, eds. *Proceedings of MEDINFO '86*. Elsevier Science Publ 1986; 12-16.
10. Kingsland III LC, Lindberg DAB, Sharp GC. Anatomy of a knowledge-based consultant system: AI/RHEUM. *MD Comput* 1986; 3: 18-26.
11. Kolarz G, Adlassnig K-P. Problems in establishing the medical expert systems CADIAG-1 and CADIAG-2 in rheumatology. *J Med Syst* 1986; 10: 395-405.
12. Mathew B, Norris D, Hendry D, Waddell G. Artificial intelligence in the diagnosis of low back pain and sciatica. *Spine* 1988; 13: 168-72.
13. McCrea JD, McCredie MR, McSherry DM, Brooks PM. A controlled evaluation of diagnostic criteria in the development of a rheumatology expert system. *Brit J Rheumatol* 1989; 28: 13-7.
14. Schewe S, Herzer P, Krüger K. Prospective application of an expert system for the medical history of joint pain. *Klin Wochenschr* 1990; 68: 466-71.
15. Schewe S, Scherrmann W, Gierl L. Evaluation and measuring of benefit of an expert system for differential diagnosis in rheumatology. In: Rienhoff O, Piccolo U, Schneider B, eds. *Expert Systems and Decision Support in Medicine*. Berlin: Springer-Verlag 1988; 351-4.
16. Adlassnig K-P, Leitich H, Kolarz G. On the applicability of diagnostic criteria for the diagnosis of rheumatoid arthritis in an expert system. *Expert Syst Appl* 1993; 6: 441-8.
17. Arnett FC, Edworthy SM, Bloch DA, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 1988; 33: 315-24.
18. Ropes MW, Bennett GA, Cobb S, Jacox R, Jassar RA. Revision of diagnostic criteria for rheumatoid arthritis. *B Rheum Dis* 1958; 9: 175-7.
19. Leitich H, Adlassnig K-P, Schuh C, Kolarz G. Improving diagnostic performance with fuzzy logic: A demonstration with the medical expert system CADIAG-2. In: Trappl R, ed. *Proceedings of Cybernetics and Systems '94*. Singapore: World Scientific Publishing 1994; 263-70.
20. Bolc L, Borowik P. *Many-Valued Logics I – Theoretical Foundations*. New York: Springer-Verlag 1992.
21. Steinbrocker O, Traeger CH, Batterman RC. Therapeutic criteria in rheumatoid arthritis. *JAMA* 1949; 140: 659-62.

Address of the authors:
Dr. Harald Leitich,
Albertgasse 6/9,
1080 Vienna,
Austria