H. Leitich[1], H. P. Kiener[2], G. Kolarz[3,4], C. Schuh[1], W. Graninger[2], K.-P. Adlassnig[1]

A Prospective Evaluation of the Medical Consultation System CADIAG-II/RHEUMA in a Rheumatological Outpatient Clinic

[1]Department of Medical Computer Sciences, Section of Medical Expert and Knowledge-Based Systems, University of Vienna Medical School,
[2]Department of Internal Medicine III, Section of Rheumatology, University of Vienna Medical School
[3]Clinic for Rheumatic Diseases of the Social Insurance Company for Trade and Industry, Baden, Austria
[4]Institute for Rheumatology in Baden, Austria

Abstract: To evaluate the performance of CADIAG-II/RHEUMA as consultant in the primary evaluation of patients visiting a rheumatological outpatient clinic, a CADIAG-II/RHEUMA consultation was done for 54 patients and the list of generated diagnostic hypotheses was compared to each clinical discharge diagnosis. For 26 of a total of 126 rheumatological discharge diagnoses, no matching CADIAG-II/RHEUMA diagnosis was available. 94% of all other discharge diagnoses were found in the list of CADIAG-II/RHEUMA hypotheses, 82% among the first third of the list of hypotheses and 48% among the first five hypotheses. We identified the following factors limiting the ability of CADIAG-II/RHEUMA to generate a comprehensive and correctly ranked list of diagnostic hypotheses: (1) a large percentage of patients with early stages of not clearly identified rheumatological conditions; (2) the limited number of CADIAG-II/RHEUMA diagnoses compared to the large number of known rheumatological conditions; (3) the fact that rheumatological diseases are rarely characterized by a single pathognomonic feature but are usually diagnosed by combinations of rather unspecific findings.

Keywords: Fuzzy Set Theory, Expert Systems, Rheumatology, Evaluation, CADIAG-II

1. Introduction

CADIAG-II is a consultation system in internal medicine which was developed at the Department of Medical Computer Sciences, University of Vienna Medical School. The CADIAG project was started with a system based on Boolean logic in 1968 [1]. This was followed by the system CADIAG-I, in which first-order predicate calculus formulas were used to define relationships between symptoms and diagnoses [2]. In 1980, Adlassnig described a new concept for a successor system based on fuzzy set theory to formalize symptoms and diagnoses and on fuzzy logic as the processing mechanism [3]. Named CADIAG-II, the new system was implemented in 1982 and was completely incorporated into the medical information system WAMIS in 1984 [4]. It was continuously improved in the following years [5-9].

CADIAG-II is the predecessor of a series of new approaches to a more generalized medical consultation system. The current development of MEDFRAME, the latest successor of the CADIAG project, includes a highly-structured object-oriented patient and knowledge base [10, 11], a more generalized use of fuzzy set theory in symptom generation [12], a broader definition of relationships between medical entities [13], and a more generalized use of fuzzy logic as inference mechanism [14].

In CADIAG-II, symptoms and diagnoses are formalized as fuzzy sets, which are characterized by fuzzy membership functions [9, 15]. If a symptom is a symbolic description of a medical datum x (e.g., the result of a laboratory test), the membership function μ: $X \rightarrow [0,1]$ assigns to every possible $x \epsilon X$ a degree of membership of x in S. Thus, μ expresses the degree of compatibility of a measured or observed finding x with linguistically expressed symptoms. In a given patient with the finding x, symptoms may either be definitely present ($\mu = 1$), partially present ($0<\mu<1$) or definitely absent ($\mu = 0$).

Relationships between symptoms and diagnoses are characterized by two aspects: (1) the frequency of occurrence degree; (2) the strength of confirmation degree, both of which also take fuzzy values in [0,1]. Frequency of occurrence degrees of 1 or 0 are assigned to symptoms which definitely must be present or absent in order to establish a diagnosis, and the interval]0,1[is used to describe the frequency of occurrence of a symptom in a diagnosis. Strength of confirmation degrees of 1 or 0 are assig-

Table 1 CADIAG-2/RHEUMA's knowledge base – total numbers of diagnoses according to a current classification of rheumatic diseases and numbers of diagnoses documented in CADIAG-2/RHEUMA's knowledge base.

Diagnoses	Total numbers of diagnoses	Numbers of diagnoses documented in CADIAG-2/RHEUMA´s knowledge base
1) Joint diseases		
Inflammatory joint diseases	87	60
Degenerative joint diseases	17	8
Other joint diseases	24	4
2) Diseases of the vertebral column		
Inflammatory diseases of the vertebral column	21	3
Degenerative diseases of the vertebral column	9	8
Other diseases of the vertebral column	14	0
3) Other diseases of cartilage and bone	76	35
4) Developmental rheumatic disorders	42	1
5) Diffuse connective tissue diseases	39	12
6) Soft tissue diseases	46	18
7) Regional pain syndromes	28	21
Total	**403**	**170**

ned to symptoms which definitely confirm or exclude a diagnosis and the interval]0,1[is used to describe the extent to which a symptom confirms a diagnosis.

Relationships between symptoms and diagnoses are formalized as diagnostic rules, where symptoms are antecedents and diagnoses are conclusions. A diagnostic rule may have only one variable in the antecedent (simple S-D relationship) or a logical combination of several variables forming the antecedent (complex S-D relationship). Logical combinations of variables and rules are calculated with the use of fuzzy operators. If a specific diagnosis D is supported by different rules with the conclusions $c_i \in [0,1]$, i = 1,..,n, the degree of presence μ_{PD} of the diagnosis D in a patient P is calculated as $\mu_{PD} = \max c_i$ (except that if $c_i = 0$ in any of the rules, $\mu_{PD} := 0$). Thus, μ_{PD} also takes values in [0,1] and diagnoses can therefore be definitely confirmed ($\mu_{PD} = 1$), excluded ($\mu_{PD} = 0$), or diagnostic hypotheses made ($0<\mu_{PD}<1$).

In a given patient, different diagnostic hypotheses may be ranked by using two different ranking procedures. With the first procedure, the diagnosis with the maximum degree of presence μ_{PD}, is ranked first. With the second procedure, diagnoses are ranked according to a ranking score, which takes the number of symptoms supporting the diagnosis and their frequency of occurrence and strength of confirmation degrees into account [4].

The list of diagnostic hypotheses can be shortened by using a hypothesis threshold. If a threshold $t \in [0,1]$ is used, diagnostic hypotheses are displayed only if $\mu_{PD} \geq t$.

The rheumatological knowledge base of CADIAG-II was first developed by Kolarz as a knowledge base for the system CADIAG-I and was later modified and expanded for the CADIAG-II system [16]. It currently contains 1126 symptoms (261 symptoms of patient history, 519 signs from the general and rheumatological physical examination, 183 laboratory test results, 89 X-ray findings, 56 biopsy and histological findings, and 18 other test results) and 170 documented diagnoses. In Table 1, the numbers of documented diagnoses are compared to total numbers of known diagnoses in different rheumatological classes according to an unpublished classification of rheumatic diseases developed by one of the authors (G.K.). The total numbers of simple and complex S-D relationships are 16,040 and 60, respectively. To evaluate the system CADIAG-II/RHEUMA, a large database of patient charts from a 140-bed rheumatological hospital in Baden/Austria was subsequently created. The database currently contains the data of more than 3500 patients with a large variety of rheumatological conditions.

CADIAG-II/RHEUMA was first evaluated in 1986, in 322 patients who were treated for rheumatoid arthritis (RA), Sjögren's syndrome, ankylosing spondylitis, systemic lupus erythematodes, gout, and progressive sclerosis [17]. Depending on the patients' diagnoses, the percentage of correctly diagnosed cases ranged from 38.3% (gout) to 90.6% (ankylosing spondylitis). The overall performance of 81.7% correct diagnoses was satisfying, but several limitations of the system were identified, including incomplete or incorrect patient histories, the difficulty to define borderlines between normal and abnormal findings, limitations in terms of covering all symptoms and diagnoses that may occur, overlapping and concomitant diseases, and especially the effect of therapy in modifying the clinical appearance of a disease.

In three studies [18-20], complex diagnostic rules based on the 1958 diagnostic criteria for RA and the 1987 revised criteria for the classification of RA, both of which were published by the American College of Rheumatology, were evaluated.

In one study [18], which included 150 RA patients and 150 patients with other rheumatic diseases, we found that performance rates were better with the 1958 criteria. The latter are more suitable for diagnostic purposes than the 1987 criteria, which are explicit classification criteria. The 1987 criteria were developed to select more uniform patient cohorts for RA studies, thus including features of typical RA cases. In our study it appeared that the usefulness of classification criteria as a diagnostic tool is limited because patients with less pronounced or atypical symptoms, such as those with early-stage disease or with seronegative RA, might easily be misdiagnosed.

In order to improve the performance of CADIAG-II/RHEUMA in these patient groups, new fuzzy versions of the 1987 criteria were tested in two subsequent studies comprising 146 RA and 146 control patients [19, 20]. It appeared that the application of fuzzy sets to four of the seven classification criteria transformed the set of criteria into a diagnostic tool that offered diagnoses at different levels of confidence. Both patients with non-typical and typical features of RA were correctly diagnosed at lower levels (with higher sensitivity and lower specificity rates) or higher levels (with lower sensitivity and higher specificity rates) of diagnostic confidence.

Apart from well-known computer consultant systems in internal medicine such as INTERNIST-I/QMR [21, 22] or ILIAD [23, 24], several other systems have been specifically designed to support the differential diagnosis of rheumatic diseases, including AI/RHEUM [25], RENOIR [26, 27], MESICAR [28, 29], RHEUMA [30, 31], and the systems designed by Bernelot Moens et al. [32, 33], McCrea et al. [34] and Mathew et al. [35]. These systems use different forms of knowledge representation and inference such as criteria tables [25], modified Bayes's theorem [32], decision trees [34], detailed anatomical and functional knowledge [29], discrimination and connectivity analysis [35], simple if-then rules [30] or fuzzy logic [26]. In the latter, fuzzy set theory and fuzzy logic are used to define fuzzy facts and to allow rules to have certainty values. The numbers of different rheumatological diagnoses included in the systems' knowledge bases range from 4 [35] to 67 [31]. All systems were tested with real patient data, but only two [31, 36] were evaluated in a prospective fashion using data of real patients from outside the hospital in which the systems were originally developed and tested. In both studies, performance rates differed significantly from those obtained in the original setting and, as a result of the evaluations, major changes were introduced in the knowledge bases of the systems.

In the preceding studies, CADIAG-II/RHEUMA was evaluated in patients from our database only and we felt that it was necessary to evaluate the system in a different patient setting. We planned to conduct this study in a rheumatological out-patient clinic of a university hospital, where patients were expected to be different from those previously evaluated. Patients are usually referred to this clinic by other hospital departments or private physicians when the first signs of a more severe rheumatological condition occur. Thus, we expected to encounter a larger number of patients with early stages of inflammatory rheumatic diseases requiring immediate evaluation and well-planned management.

Although, mainly for organizational reasons, we did not intend to use CADIAG-II/RHEUMA as an online consultation system, we planned to evaluate the performance of CADIAG-II/RHEUMA as consultant in the early stages of patient evaluation and work-up. The main tasks of CADIAG-II/RHEUMA in this process were to offer a list of diagnostic hypotheses which may explain a patient's condition. This list of hypotheses was intended to include both a) all possible explanatory diagnoses; and b) the diagnoses most closely related to the patient's condition. The first task was thought to be important, namely to remind the physician to keep rare diseases in mind because these are prone to be misdiagnosed more easily. The second task was to make the physician aware of diagnoses that most probably account for the patient's condition and thus require primary attention – both from a practical and financial point of view. Therefore, the aim of the study was to determine whether CADIAG-II/RHEUMA was able to offer both comprehensive *and* correctly ranked lists of diagnostic hypotheses.

2. Material and Methods

The present study was designed to include 50 consecutive patients visiting the outpatient clinic of the Department of Internal Medicine III of the University of Vienna Medical School, Section of Rheumatology. The primary examination was carried out by one of the authors (H. K.) according to the following predefined plan:

At the first visit, both a structured patient history form and two structured physical examination forms were completed. These forms were specifically designed for the CADIAG-II/RHEUMA system, to ensure a more detailed and comprehensive clinical database for the consultation process. The patient history form was completed by the patient prior to the first examination and contained questions regarding current and previous complaints, diseases, and medication. The physical examination forms were completed by H. K. and contained detailed information about the general and rheumatological physical examination.

Apart from completing these forms, all patients were managed according to the common clinical practice and guidelines of the department. For the purpose of the study, no additional diagnostic procedures were carried out.

The results of all diagnostic procedures of the first and the follow-up visits were included. If both normal and pathological results of the same diagnostic procedure were available, only pathological test results were included.

For each patient, all information of the patient history and physical examination forms and other symptoms that could be matched to CADIAG-II/RHEUMA symptoms were manually entered into the CADIAG-II/RHEUMA patient database. Most laboratory test results were directly transferred to the CADIAG-II/RHEUMA database via the local hospital information system and were interpreted by the CADIAG-II/RHEUMA system as the first step of the consultation process. All other test results (including X-rays, cultures, and Schirmer's test) had already been classified as normal or pathological by the referring departments.

For each patient, four types of diagnoses were recorded: *Referral diagnoses:* the suspected conditions which led to the referral of the patient to the outpatient clinic. *Discharge diagnoses:* the diagnoses which were printed on the final report and were the result of medical evaluation at the first and the follow-up visits. *Suspected diagnoses:* Other diagnoses which were considered by H.K. as possible conditions to explain the patient's symptoms were recorded specifically for the purpose of this study. *Last available clinical diagnoses:* We planned to review all study patients charts at least one year after finishing the patient data collection in order to check whether the discharge diagnoses were still valid or would have to be altered on the basis of additional information obtained after the study period. Diagnoses were only recorded as last available clinical diagnoses if they were regarded as new and clinically significant insights into the patient's rheumatological condition.

Before starting the CADIAG-II/RHEUMA consultation process, we tried to match all patient diagnoses, including referral, discharge, suspected,

and last available diagnoses, to the list of available CADIAG-II/RHEUMA diagnoses. If a specific CADIAG-II/ RHEUMA diagnosis was not available, but only a more unspecific one, the latter was chosen. For instance, the discharge diagnosis "osteoarthrosis of the carpal joint", which is not included in the list of CADIAG-II/RHEUMA diagnoses, would be matched to the CADIAG-II/RHEUMA diagnosis "osteoarthrosis, unspecified".

After patient data had been collected, the CADIAG-II/RHEUMA consultation was done consecutively for all patients as a batch job. We then recorded the numbers of confirmed and excluded diagnoses and the numbers of diagnostic hypotheses for each patient.

Each patient diagnosis and the list of CADIAG-II/RHEUMA diagnoses were compared and the rank of each matching CADIAG-II/RHEUMA diagnosis was calculated.

To compare the two main ranking procedures, results were calculated both according to the hypothesis score and the maximum degree of confirmation.

To obtain a smaller and more specific list of CADIAG-II/RHEUMA diagnoses, we also calculated the rank of each matching CADIAG-II/ RHEUMA diagnosis without and with the use of hypothesis thresholds at levels of 0.2 and 0.3. Using these thresholds, only diagnostic hypotheses with a degree of presence of 0.2 or above and 0.3 or above would be included in the list of hypotheses.

3. Results

A total of 54 patients visiting the outpatient clinic on the days specified by H.K. to evaluate patients according to the study plan were included. The patients were not preselected and represented a typical mixture of those visiting the clinic.

At the first visit, history and physical examination forms were obtained from all patients. The number of follow-up visits for each individual ranged between 0 and 6. At the end of the study period, all final reports including the results of diagnostic procedures were available for analysis. Two years after completing the initial study, all patients charts were again reviewed to check whether the current diagnoses still matched those in the final reports.

The numbers of discharge diagnoses in our patients ranged from 0-9 (median: 3), of which 0-7 (median: 1.5) were rheumatological discharge diagnoses. Three patients had non-rheumatological problems and did not get a rheumatological discharge diagnosis. For a total of 26 rheumatological discharge diagnoses in 20 other patients, no matching CADIAG-II/RHEUMA diagnosis was found. Most of them were either functional vertebral disorders (e.g., hypermobility syndrome or scoliosis) or unspecific disorders (e.g., monarthritis, oligoarthritis, or arthralgia), which, due to the absence of specific symptoms, were not included in CADIAG-II/ RHEUMA's knowledge base. In 12 of the 20 patients, we could not match a CADIAG-II/RHEUMA diagnosis to any of the patients' discharge diagnoses.

As a result of the CADIAG-II/ RHEUMA consultation process, the number of confirmed and excluded diagnoses in individual patients ranged between 0-2 (median: 0) and 7-20 (median: 9), respectively. The number of diagnostic hypotheses without or with the use of hypothesis thresholds of 0.2 or 0.3 ranged between 29-149 (median: 134), 2-35 (median: 19), or 0-20 (median: 9), respectively.

In two patients, the CADIAG-II/ RHEUMA consultation resulted in a total of three confirmed diagnoses, all of which were true positive results. In both patients, the diagnoses "prosthesis of the hip joint", "prosthesis, unspecified", and "spondylitis, unspecified" were confirmed by the presence of definite radiological signs.

All excluded diagnoses were true negative results, i.e. they did not match any referral, discharge, suspected, or last available diagnosis, with the exception of one false negative result. In this patient with the discharge diagnosis of "osteoarthrosis of the knee joint", the corresponding CADIAG-II/ RHEUMA diagnosis was excluded because the obligatory CADIAG-II/ RHEUMA symptom "present complaints, affection of one or both knee joints" was not present.

The number of generated CADIAG-II/RHEUMA hypotheses (including the three confirmed diagnoses) are shown in Table 2. Numbers and corresponding sensitivity rates are shown for both discharge or last available and for referral or suspected diagnoses, for both ranking procedures, with and without the use of hypothesis thresholds.

In seven patients, a total of 8 last available diagnoses provided new and

Total number of diagnoses	Without hypothesis threshold and ranked according to score (or maximum degree of presence)			Using a hypothesis threshold of 0.2 and ranked according to score (or maximum degree of presence)			Using a hypothesis threshold of 0.3 and ranked according to score (or maximum degree of presence)		
	occurring in the list of hypotheses	among the first five hypotheses	among the first third of the list of hypotheses	Occurring in the list of hypotheses	Among the first five hypotheses	among the first third of the list of hypotheses	occurring in the list of hypotheses	among the first five hypotheses	among the first third of the list of hypotheses
Discharge and last available confirmed diagnoses									
100	**94 (94)**	**48 (53)**	**82 (83)**	**78 (78)**	**47 (52)**	**53 (58)**	**67 (67)**	**53 (52)**	**41 (38)**
%	94 (94)	48 (53)	82 (83)	78 (78)	47 (52)	53 (58)	67 (67)	53 (52)	41 (38)
Referral and suspected diagnoses									
75	**70 (70)**	**37 (24)**	**64 (60)**	**46 (46)**	**37 (24)**	**40 (32)**	**36 (36)**	**34 (24)**	**31 (21)**
%	93 (93)	49 (32)	85 (80)	61 (61)	49 (32)	53 (43)	48 (48)	45 (32)	41 (28)

Table 2 Numbers of generated hypotheses (including confirmed diagnoses) among discharge and last available confirmed diagnoses or among referral and suspected diagnoses using different hypothesis thresholds and with ranking according to hypothesis score or maximum degree of presence μ_{PD}.

Table 3 Results ranked according to hypothesis score (or maximum degree of presence μ_{PD}) for patients in whom the last available diagnosis was significantly different from the discharge diagnosis.

Patient	Discharge diagnosis	Matching CADIAG-II diagnosis	Rank of CADIAG-II diagnosis	Last available diagnosis	Matching CADIAG-II diagnosis	Rank of CADIAG-II diagnosis
A.E., 26 yr., m	Seronegative spondarthritis	Seronegative spondarthritis	12 (11)	Ankylosing spondylitis	Ankylosing spondylitis	29 (2)
F.E., 40 yr., f	Interphalangeal osteoarthrosis	Interphalangeal osteoarthrosis	3 (6)	Rheumatoid arthritis	Rheumatoid arthritis	1 (2)
G.J., 35 yr., f	Reactive arthritis (with *U. urealyticum* infection)	Reactive arthritis	2 (1)	Seronegative spondarthritis	Seronegative spondarthritis	13 (28)
H.G., 37 yr., m	Oligoarthritis, unspecified	No matching CADIAG-II diagnosis	-	Seronegative spondarthritis	Seronegative spondarthritis	11 (27)
N.O., 19 yr., m	Unilateral sacroileitis	Seronegative spondarthritis	13 (19)	Reactive arthritis (with *S. enteritidis* infection)	Reactive arthritis	8 (10)
U.J., 22 yr., m	Discitis Th12/L1	Spondylitis, unspecified	Confirmed diagnosis	Discitis Th12/L1 occurring as part of reactive arthritis (with *U. urealyticum* infection), seronegative spondarthritis	Reactive arthritis	8 (5)
					Seronegative spondarthritis	5 (20)
W.A., 57 yr., m	Oligoarthritis, unspecified	No matching CADIAG-II diagnosis	-	Activated osteoarthrosis	Osteoarthrosis, unspecified	2 (3)
					Rheumatoid arthritis	1 (1)

clinically significant insights into the patients' rheumatological conditions. In Table 3, the results for both discharge and last available diagnoses are shown.

4. Discussion

The results of this study indicate that the two aims of the study, namely the generation of a comprehensive *and* correctly ranked list of hypotheses were successfully met, but the results also show the limitations of a computer consultant system in a patient setting as that used in the study.

The numbers of confirmed and excluded diagnoses – only a total of three confirmed diagnoses in two patients and a median of nine excluded diagnoses per patient – were quite small. The reason for this result, which was also observed in a preceding evaluation [17], is that rheumatological diseases are rarely characterized by a single pathognomonic feature, but have to be diagnosed by combinations of more or less unspecific findings. This fact is also reflected in our knowledge base, in which only a minority of symptoms have strength of confirmation degrees of 1 or 0 and thus are able to confirm or exclude a diagnosis.

The median number of generated hypotheses, which was 134 without using a hypothesis threshold, was sharply reduced to a median of 19 or 9 hypotheses with a threshold of 0.2 or 0.3. Thus, a comprehensive list of hypotheses (with nearly 80% of all documented diagnoses) including rare rheumatological conditions, as well as two shorter and more practical lists of hypotheses were generated.

The main factor limiting the first aim of this study – to generate a comprehensive list of diagnoses – was the single fact that the number of different known rheumatological conditions is vast.

Even though CADIAG-II/RHEUMA's knowledge base contains 170 documented diagnoses, this is only a minority compared to published classifications of rheumatic diseases with more than 1000 diagnoses [37] or even compared to the 403 diagnoses included in the classification we are currently using (see Table 1). Documentation is nearly complete for the group of inflammatory joint diseases, which from a clinical standpoint require primary attention, but documentation is still incomplete for other groups of rheumatic diseases. As a result, 21% of the discharge diagnoses of our study patients could not be matched to any documented CADIAG-II/RHEUMA diagnosis. When discharge diagnoses could be matched to CADIAG-II/RHEUMA diagnoses, the system was successful in including them in the list of hypotheses, with no difference between either referral and suspected or discharge and last available diagnoses (93% or 94%, respectively, see Table 2).

The second aim of this study – to generate a correctly ranked list of diagnoses – was only partially met. About 82% of discharge and last available diagnoses were ranked among the first third of the list of hypotheses, but only 48% were among the first five hypotheses. We identified two major limitations to the second task: the characteristics of patients included in the study, and the manner of representing knowledge in CADIAG-II/RHEUMA.

Three patients were referred to the out-patient clinic because of signs that may have been a part of a rheumatological condition, which was not confirmed after the initial evaluation, and therefore they were discharged without a rheumatological diagnosis. The majority of patients were referred with symptoms suggesting an incipient inflammatory rheumatic disease and during the initial evaluation several different diagnoses were suspected. More than 70% of referral and suspected diagnoses were inflammatory joint disea-

Diagnosis	Total number of diagnoses	Without hypothesis threshold and ranked according to score (or maximum degree of presence)			Using a hypothesis threshold of 0.2 and ranked according to score (or maximum degree of presence)			Using a hypothesis threshold of 0.3 and ranked according to score (or maximum degree of presence)		
		occurring in the list of hypotheses	among the first five hypotheses	among the first third of the list of hypotheses	occurring in the list of hypotheses	among the first five hypotheses	among the first third of the list of hypotheses	occurring in the list of hypotheses	among the first five hypotheses	among the first third of the list of hypotheses
Rheumatoid arthritis	4	4 (4)	4 (4)	4 (4)	4 (4)	4 (4)	4 (4)	4 (4)	4 (4)	4 (4)
Seronegative spondarthritis	5	5 (5)	1 (0)	5 (5)	1 (1)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Psoriatic arthritis	2	2 (2)	2 (2)	2 (2)	2 (2)	2 (2)	2 (2)	2 (2)	2 (2)	1 (1)
Osteoarthrosis of the knee	5	4 (4)	3 (4)	4 (4)	4 (4)	3 (4)	4 (4)	4 (4)	3 (4)	2 (3)
Interphalangeal osteoarthrosis	10	10 (10)	7 (4)	10 (10)	10 (10)	7 (4)	8 (6)	10 (10)	7 (4)	5 (0)
Spondyloarthrosis	7	7 (7)	5 (6)	7 (7)	7 (7)	5 (6)	5 (7)	7 (7)	6 (6)	5 (6)
Spondylosis	9	9 (9)	0 (6)	9 (9)	9 (9)	0 (6)	2 (6)	6 (6)	2 (6)	0 (6)
Myalgia, unspecified	8	8 (8)	0 (0)	1 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Carpal tunnel syndrome	4	1 (1)	1 (1)	1 (1)	1 (1)	1 (1)	1 (1)	1 (1)	1 (1)	0 (1)

Table 4 Numbers of generated hypotheses (including confirmed diagnoses) among selected discharge and last available confirmed diagnoses using different hypothesis thresholds and with ranking according to hypothesis score or maximum degree of presence μ_{PD}.

ses; only 13% were degenerative rheumatic diseases and 4%, soft tissue diseases. After the patient work-up, these numbers changed significantly, with only 18% of patients being discharged with inflammatory joint diseases, but 55% with degenerative rheumatic diseases and 18% with soft tissue diseases. This indicates that most patients who were included in the study had no typical features of a single rheumatic disease, but unclear conditions requiring expert evaluation. As a result, discharge diagnoses were less likely to be correctly ranked than would be expected if CADIAG-II/RHEUMA were being evaluated in patients presenting with typical rheumatological conditions.

The majority of symptoms in CADIAG-II/RHEUMA's knowledge base have low strength of confirmation degrees, most of them in the range of 0.01 to 0.17. As mentioned earlier, this is due to the fact that most rheumatological diagnoses are rarely characterized by a single pathognomonic feature and must be diagnosed by combinations of rather unspecific findings. Without symptoms with high specificity and high discriminative power between different diagnoses, differential diagnosis is difficult. As a result, rheumatological discharge diagnoses were less likely to be correctly ranked than would be expected if CADIAG-II/RHEUMA had been evaluated in a different medical domain with more specific disease profiles.

In this study, two ranking procedures were compared with each other: the maximum degree of the presence of a diagnosis, which is mainly determined by a single symptom with a high strength of confirmation degree, and the hypothesis score, which is determined by the number of supporting symptoms and their frequency of occurrence and strength of confirmation degrees. As shown in Table 2, the results of both ranking procedures are similar in discharge and last available diagnoses, but the results are better with the hypothesis score in referral and suspected diagnoses. Thus, combining the evidence of all supporting symptoms seems to be advantageous for generating diagnostic hypotheses, but not for establishing a correct final diagnosis.

The results were also analyzed according to different discharge and last available diagnoses.

In the main groups of diagnoses, 100% or 50% of inflammatory joint diseases, 97% or 82% of degenerative joint diseases, 100% or 29% of degenerative diseases of the vertebral column and 28% or 0% of soft tissue diseases were ranked among the first third or among the first five hypotheses when hypotheses were ranked according to the hypothesis score. Again, the main factor accounting for the differences between the results of these groups of diagnoses is the manner in which knowledge is represented in CADIAG-II/RHEUMA. A single diagnosis was more likely to be highly ranked according to the maximum degree of presence

Diagnoses	Total number of simple S-D relationships with defined strength of confirmation	Total number of complex S-D relationships with defined strength of confirmation	Number of S-D relationships with strength of confirmation within specified range								
			1.00	0.98-0.99	0.83-0.97	0.68-0.82	0.33-0.67	0.18-0.32	0.03-0.17	0.01-0.02	0.00
Rheumatoid arthritis	323	6	0	0	0	2	10	59	191	60	1
Seronegative spondarthritis	122	0	0	0	0	0	0	3	106	11	2
Psoriatic arthritis	218	2	0	0	0	2	7	23	137	39	10
Osteoarthrosis of the knee	91	0	0	0	0	0	10	30	42	9	0
Interphalangeal osteoarthrosis	76	0	0	0	0	0	14	12	45	3	2
Spondyloarthrosis	70	0	0	0	0	0	13	25	14	17	1
Spondylosis	66	0	0	0	0	1	0	6	29	29	1
Myalgia, unspecified	66	0	0	0	0	0	0	2	15	48	1
Carpal tunnel syndrome	20	0	0	0	1	0	1	1	3	14	0

Table 5 Numbers of S-D relationships in CADIAG-II/RHEUMA's knowledge base in selected diagnoses.

if the knowledge base contained more supporting symptoms with higher strength of confirmation degrees and was more likely to be highly ranked with the hypothesis score if both the number of S-D relationships and the strength of confirmation and frequency of occurrence degrees were higher.

Without using a hypothesis threshold, the performance of CADIAG-II/RHEUMA was high in diagnoses with both, a large number of symptoms and a high proportion of symptoms with higher degrees of confirmation such as "rheumatoid arthritis" or "psoriatic arthritis", or with fewer but more specific symptoms like the group of degenerative joint diseases ("osteoarthrosis of the knee", "interphalangeal osteoarthrosis") or "spondyloarthrosis", as seen in Tables 4 and 5. Performance was poor in diagnoses with both few and unspecific symptoms such as "myalgia, unspecified" or with very few symptoms like "carpal tunnel syndrome". Differences in the results between both ranking procedures became evident when the discharge and last available diagnoses were expected to be among the first five diagnostic hypotheses. Results were better in diagnoses with more specific symptoms (e.g., the group of degenerative joint diseases) than in diagnoses with less specific symptoms (e.g., "spondylosis").

The use of hypothesis thresholds did not influence the good results of diagnoses with a high number of more specific findings such as "rheumatoid arthritis" or "psoriatic arthritis" or with a higher proportion of more specific findings like the groups of degenerative joint diseases. Again, hypothesis generation was strongly diminished in diagnoses with unspecific symptoms such as "myalgia, unspecified" or "seronegative spondarthritis". With the use of hypothesis thresholds, the percentage of correct results among the first five or among the first third of the list of hypotheses was improved with ranking according to the hypothesis score in diagnoses with more (e.g., the group of degenerative joint diseases) than with less specific findings (e.g., "spondyloarthrosis").

The results of this study show that the main limitations of the diagnostic performance of CADIAG-II/RHEUMA in a patient setting as that used in this study were the large number of known rheumatological diagnoses, the lack of specific signs for rheumatological diagnoses and patients presenting with early stages of unclear conditions.

As the characteristics of patients and rheumatological diagnoses cannot be changed, the performance of CADIAG-II/RHEUMA can be best improved by extending the number of documented diagnoses and continuously updating the knowledge base in terms of new diagnostic tests.

CADIAG-II/RHEUMA's performance will also be improved by the efforts currently being made to develop an improved medical consultation system (MEDFRAME). These include the definition of negative evidence, in order to use both positive evidence for and negative evidence against diagnoses in hypothesis generation, and the development and evaluation of new knowledge processing and hypothesis ranking procedures. All of these efforts may bring us closer to the vision of an expert computer consultant. Considering the good results of CADIAG-II/RHEUMA in predicting the final diagnosis of the study patients (as shown in Table 3), this is a promising vision.

REFERENCES

1. Spindelberger W, Grabner G. Ein Computerverfahren zur diagnostischen Hilfestellung. In: Fellinger K, ed. Computer in der Medizin – Probleme, Erfahrungen, Projekte. Vienna, Verlag Brüder Hollinek, 1968; 189-221.
2. Moser W, Adlassnig K-P. Consistency checking of binary relationships in a medical knowledge base. Artif Intell Med 1992; 4: 389-407.
3. Adlassnig K-P. A fuzzy logical model of computer-assisted medical diagnosis. Method Inform Med 1980; 19: 141-8.
4. Adlassnig K-P, Kolarz G, Scheithauer W. Approach to a hospital-based application of a medical expert system. Med Inform 1986; 11: 205-23.
5. Adlassnig K-P, Kolarz G, Scheithauer W, Effenberger H, Grabner G. Computer-assisted medical diagnosis. Comp Biol Med 1985; 15: 315-35.
6. Adlassnig K-P, Kolarz G, Scheithauer W. Present state of the medical expert system CADIAG-2. Method Inform Med 1985; 24: 13-20.
7. Adlassnig K-P. Fuzzy set theory in medical diagnosis. IEEE Transactions on Systems, Man, Cybernetics SMC 1986; 16: 260-5.
8. Adlassnig K-P, Kolarz G. Representation and semiautomatic acquisition of medical knowledge in CADIAG-1 and CADIAG-2. Comp Biomed Res 1986; 19: 63-79.
9. 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.
10. Kolousek G. The system architecture of an integrated medical consultation system and its implementation based on fuzzy technology. Ph.D. Thesis, Technical University of Vienna, Austria, 1997.
11. Bögl K. Design and implementation of a web-based knowledge acquisition toolkit for medical expert consultation systems. Ph. D. Thesis, Technical University of Vienna, Austria, 1997.
12. Bögl K, Leitich H, Kolousek G, Rothenfluh TE, Adlassnig K-P. Clinical data interpretation in MEDFRAME/CADIAG-4 using fuzzy sets. Biomed Eng Appl, Basis Communications 1996; 8: 488-95.
13. Rothenfluh TE, Bögl K, Adlassnig K-P. Representation and acquisition of knowledge for a fuzzy medical consultation system. In: Szczepaniak PS, Lisboa PJ, Tsumoto S, eds. Fuzzy Systems in Medicine. Heidelberg, New York: Springer, 2000: 636-51.
14. Brein L. Das medizinische Expertensystem CADIAG-IV: Inferenzprozeß, Erklärungen und Untersuchungsvorschläge. Master Thesis, Technical University of Vienna, Austria, 1997.
15. Zadeh LA. Fuzzy Sets. Inform Contr 1965; 8: 338-53.
16. Adlassnig K-P, Kolarz G. CADIAG-2: Computer-assisted medical diagnosis using fuzzy subsets. In: Gupta MM, Sanchez E, eds. Approximate Reasoning in Decision Analysis. Amsterdam: North-Holland Publishing Company, 1982; 219-47.
17. 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.
18. Adlassnig K-P, Leitich H, Kolarz G. On the applicability of diagnostic criteria for the diagnosis of rheumatoid arthritis in an expert system. Exp Syst Appl 1993; 6: 441-8.
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. Leitich H, Adlassnig K-P, Kolarz G. Development and evaluation of fuzzy criteria for the diagnosis of rheumatoid arthritis. Method Inform Med 1996; 35: 334-42.
21. Miller RA, Pople HE, Myers JD. INTERNIST-I, an experimental computer-based diagnostic consultant for general internal medicine. N Engl J Med 1982; 307: 468-76.
22. Bankowitz RA, McNeil MA, Challinor SM, Parker RC, Kapoor WN, Miller RA. A computer-assisted medical diagnostic consultation service. Implementation and

prospective evaluation of a prototype. Ann Intern Med 1989; 110: 824-32.
23. Warner HR. ILIAD: Moving medical decision making into new frontiers. Method Inform Med 1989; 28: 370-2.
24. Bergeron B. ILIAD: A diagnostic consultant and patient simulator. MD Computing 1991; 8: 46-53.
25. Kingsland LC III, Lindberg DAB, Sharp GC. AI/RHEUM: a consultant system for rheumatology. J Med Syst 1983; 7: 221-7.
26. Belmonte-Serrano M, Sierra C, Lopez de Mantaras R. RENOIR: An expert system using fuzzy logic for rheumatology diagnosis. Int J Intell Syst 1994; 9: 985-1000.
27. Hernandez C, Sancho JJ, Belmonte MA, Sierra C, Sanz F. Validation of the medical expert system RENOIR. Comput Biomed Res 1994; 27: 456-71.
28. Horn W. MESICAR – A medical expert system integrating causal and associative reasoning. Appl Artif Intell 1989; 3: 305-56.
29. Horn W. Utilizing detailed anatomical knowledge for hypothesis formation and hypothesis testing in rheumatological decision support. Artif Intell Med 1991; 3: 21-39.
30. 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.
31. Schewe S, Schreiber MA. Stepwise development of a clinical expert system in rheumatology. Clin Investig 1993; 71: 139-44.
32. Bernelot Moens HJ, Van der Korst JK. Development and validation of a computer program using Bayes's theorem to support diagnosis of rheumatic disorders. Ann Rheum Dis 1992; 51: 266-71.
33. Bernelot Moens HJ, Van der Korst JK. Comparison of rheumatological diagnoses by a Bayesian program and by physicians. Method Inform Med 1991; 30: 187-93.
34. McCrea JD, McCredie MR, McSherry DM, Brooks PM. A controlled evaluation of diagnostic criteria in the development of a rheumatology expert system. Br J Rheumatol 1989; 28: 13-7.
35. Mathew B, Norris D, Hendry D, Waddell G. Artificial intelligence in the diagnosis of low back pain and sciatica. Spine 1988; 13: 168-72.
36. Bernelot Moens HJ. Validation of the AI/RHEUM knowledge base with data from consecutive rheumatological outpatients. Method Inform Med 1992; 31: 175-81.
37. Mathies H, Otte P, Villiaumey J, St. Dixon A, eds. Classification of diseases of the locomotor apparatus. Compendia Rheumatologica 4. Basel: Eular Publishers, 1978.

Address of the authors:
Harald Leitich, MD, MSc,
Department of Medical Computer Sciences,
Section of Medical Expert
and Knowledge-Based Systems,
University of Vienna Medical School,
Spitalgasse 23,
1090 Vienna, Austria
E-mail: harald.leitich@univie.ac.at