Expert Systems With Applications, Vol 6, pp 441–448, 1993
Printed in the USA

0957-4174/93 $6.00 + .00

On the Applicability of Diagnostic Criteria for the Diagnosis of Rheumatoid Arthritis in an Expert System

KLAUS-PETER ADLASSNIG AND HARALD LEITICH

University of Vienna, Vienna, Austria

GERNOT KOLARZ

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

Abstract—*CADIAG-2/RHEUMA is a medical expert system developed to assist in the differential diagnosis of rheumatic diseases. Based on fuzzy set theory and fuzzy logic, it supports the formalization of vague and uncertain medical information (i.e., medical entities and relationships between them) and draws justifiable conclusions from these imprecise data. Given a patient's finding pattern, CADIAG-2 provides confirmed and excluded diagnoses, diagnostic hypotheses, and suggestions for further examinations. The knowledge base of CADIAG-2 has been designed to contain simple finding/disease relationships as well as diagnostic rules of high complexity to confirm or hypothesize diseases. We shall present results obtained with 300 clinical cases from a hospital for rheumatic diseases. Different rules for the diagnosis of rheumatoid arthritis based upon classification criteria issued by the American Rheumatism Association were tested against each other. That diagnostic rule which had shown the best results was then further improved by a rheumatology expert, which finally yielded a sensitivity of 83.3% and a specificity of 95.3%.*

1. INTRODUCTION

THE CENTRAL AIM of the CADIAG-2 project is the development of a medical consultation system for general internal medicine. The underlying clinical issue is to assist in the differential diagnostic process (a) by indicating all possible diseases which might be the cause of a patient's pathological findings, with special emphasis on rare diseases; (b) by offering further useful examinations to confirm or to exclude diagnostic hypotheses gained or to find stronger support for them; and (c) by indicating a patients' pathological findings not yet accounted for by the expert system's proposed diagnoses.

After gaining experience with the medical expert system CADIAG-1 which was formally based on first-order predicate logic and pattern matching (Adlassnig et al., 1985), the successor system CADIAG-2 was developed and implemented (Adlassnig, 1980; 1986). This system applies fuzzy set theory to model inherent vagueness of medical concepts and fuzzy logic to infer diagnostic conclusions.

At present, the knowledge base of CADIAG-2 contains disease profiles and diagnostic rules for 267 diseases, among them 185 rheumatic diseases (diagnostic group CADIAG-2/RHEUMA, with 69 joint diseases, 12 diseases of the spinal column, 38 diseases of the soft tissue and connective tissue systems, 45 diseases of cartilage and bone, and 21 regional pain syndromes (Kolarz & Adlassnig, 1986)) and 82 gastroenterological diseases (diagnostic group CADIAG-2/GALL with 35 gall bladder and bile duct diseases (Adlassnig & Akhavan-Heidari, 1989), diagnostic group CADIAG-2/PANCREAS with 10 pancreatic diseases (Adlassnig & Scheithauer, 1989; Adlassnig, Scheithauer, & Grabner, 1984), and diagnostic group CADIAG-2/COLON with 37 colon diseases). All these diagnostic groups reside in the CADIAG-2 shell and are self-contained to allow differential diagnosis.

The CADIAG-2 system is integrated into the medical information system WAMIS (the German acronym for *W*iener *A*llgemeines *M*edizinisches *I*nformations-*S*ystem, Vienna General Medical Information System of the Vienna General Hospital (Adlassnig, Kolarz,

Requests for reprints should be sent to Klaus-Peter Adlassnig, Department of Medical Computer Sciences, University of Vienna, Wahringer Gurtel 18-20, A-1090, Vienna, Austria

Scheithauer, & Grabner, 1986). This integration allows collecting a patient's findings for CADIAG-2 via the routine medical documentation and laboratory system of WAMIS. Through a data abstraction and aggregation process (Adlassnig, 1988), patient data are made available to the CADIAG-2 system which tries to infer diagnoses from these abstracted findings in a data-driven manner. In addition, patient data not routinely collected in WAMIS can be added to CADIAG-2 through a man-machine interface which processes medical terms given in natural language. A word segmentation algorithm allows the usage of medical synonyms and abbreviations; moreover, it accepts various orthographic variants and takes into account different medical suffixes (Adlassnig & Grabner, 1985).

The diagnostic process of CADIAG-2 is based on both stored disease profiles and diagnostic rules (usually very complex ones such as the ARA criteria for rheumatic diseases (Arnett et al., 1988; Ropes et al., 1958). Two relationships define the association between findings and diseases in these disease profiles: (a) the necessity of occurrence of a certain finding with a given disease (frequency of occurrence degree) and (b) its sufficiency to infer that disease (strength of confirmation degree). The same relationships are applicable to define the associations between the antecedents and consequents of diagnostic rules.

The inference process of CADIAG-2 aims at generating one or more differential diagnoses and—at the same time—at excluding some or all remaining diagnoses. A diagnosis is either established as definitely confirmed or proposed as a diagnostic hypothesis to be confirmed or excluded after additional examinations are performed.

Diagnoses are indicated as definitely confirmed if pathognomonic findings were found in the patient or if confirming rules were triggered by the patient's findings. Because of the hierarchical relationships among diseases in CADIAG-2, diagnoses at a higher level in the disease hierarchy are confirmed as well if subdiagnoses are indicated as being confirmed.

Excluded diagnoses are established either by present excluding criteria or by absent obligatory criteria. Excluding criteria may be single excluding findings, excluding rules or other, already established diagnoses that exclude other diagnoses. Findings and rule criteria which are defined to be obligatorily present in the patient to establish a certain diagnosis but which are definitely absent consequently exclude the respective diagnosis. Definitely excluded disease categories in the disease hierarchy also cause the exclusion of the respective entire set of existing subdiagnoses.

Diagnoses confirmed and excluded at the same time—which might happen due to contradictory patient data and/or knowledge base errors—are termed as diagnostic contradictions. They are displayed separately, with the inference process made transparent.

Diagnostic hypotheses are generated (a) if a diagnosis is neither confirmed nor excluded nor a contradictory result and (b) if the strength of confirmation of at least one present finding, one triggered rule, or one already established subdiagnosis is equal to, or higher than, a given threshold ϵ ($0 < \epsilon < 1$). Since the application of fuzzy set theory allows for mathematical modeling of borderline findings, the degree of presence of a finding (degree of membership in a fuzzy set) is combined with its strength of confirmation. If the resulting value, which is a measure of confirmation for the concluded disease, lies between threshold ϵ and unity (unity means full confirmation), the respective disease has to be taken into consideration as a diagnostic hypothesis. In addition, diagnostic hypotheses are ranked according to a score of support. This score is calculated on the basis of (a) the number of single findings present, or present to a certain degree, and having a relationship to the disease under consideration; (b) the degree of presence of these findings; and (c) the degrees for frequency of occurrence and strength of confirmation between these findings and the respective disease.

Diagnoses which are neither confirmed nor excluded nor diagnostic hypotheses nor contradictory results are put into a category termed "diagnoses not generated." This allows the physician to obtain a complete survey of all diseases included in the knowledge base of CADIAG-2.

In CADIAG-2, two forms of knowledge acquisition have been applied: (a) acquisition of knowledge from medical experts and (b) semiautomatic acquisition of medical knowledge from a patient data base. Medical experts provide definitional and judgmental knowledge from textbooks and their own practical experience. The estimation of appropriate values for the frequency of occurrence and strength of confirmation degrees is assisted by an automatic procedure which calculates the respective values from stored records of patients with established diagnoses (Adlassnig & Kolarz, 1986).

CADIAG-2/RHEUMA, being part of the CADIAG-2 host system, was developed to support differential diagnosis of rheumatic diseases. In a prior evaluation, the diagnostic accuracy of CADIAG-2/RHEUMA attained a sensitivity of 93.7%, evaluating 426 patients with rheumatoid arthritis, gout, ankylosing spondylitis, psoriatic arthritis, Sjögren's disease, systemic lupus erythematosus, Reiter's disease, and systemic sclerosis (Adlassnig & Grabner, 1985).

This study is focused on the diagnosis of rheumatoid arthritis (RA) only. Three different sets of RA diagnostic criteria were used and applied to both RA patients and a control group with non-RA rheumatic diseases. All three sets were based on RA classification criteria issued by The American Rheumatism Association (ARA, now The American College of Rheumatology) (Arnett et al., 1988; Ropes et al., 1958). In CADIAG-2/RHEUMA, each set of diagnostic criteria was

TABLE 1
RA Staging According to Steibrocker et al. (1949)

Stage 1	No signs of destruction of cartilage or bone; osteoporosis may be present
Stage 2	Erosions and bony decalcification adjacent to the joint involved without joint deformity
Stage 3	As in Stage 2 plus subluxation.
Stage 4	As in Stage 3 plus ankylosis

implemented as one IF-THEN rule. The three established IF-THEN rules are:

- rule 1 was implemented according to the revised 1958 ARA criteria for the classification of definite RA (Ropes et al., 1958);
- rule 2 was implemented according to the revised 1987 ARA criteria for the classification of (definite) RA (Arnett et al., 1988);
- rule 3 combined both literature definition and specific clinical experience of a rheumatology expert: Like rule 1, it was implemented according to the revised 1958 ARA criteria for the classification of definite RA (Ropes et al., 1958); however, several criteria were changed and redefined by the rheumatology expert.

The aims of the study were to find out: (a) which of the two initial sets of RA diagnostic criteria performed best; (b) whether the best performing set of criteria could still be improved; and (c) whether splitting both RA patients and control subjects into subsets (according to disease stages, disease characteristics, and concomitant diseases) would give a more detailed picture of the accuracy obtained with the expert system.

2. PATIENT DATA

All 150 RA patients and 150 control subjects of this study underwent treatment in a 140-bed hospital for rheumatic diseases in Baden/Austria. Only adults with disease onset after age 16 were included in this study. The mean ages of RA and non-RA patients were similar, whereas sex percentages differed due to a higher number of female RA patients (RA group: 79.3% females, control group: 38.7% females). Two completed questionnaires for medical history and the results of the physical examination, as well as X-ray findings of the hands and feet and laboratory test results including serum rheumatoid factor were available for each patient.

RA patients: All 150 RA patients had a confirmed clinical diagnosis of RA (no classification criteria were applied to select study participants). They were additionally subdivided according to the following conditions: (a) disease stage (the disease staging used in this study is based on radiographic findings and was introduced by Steinbrocker, Traeger, & Batterman (1949) (Table 1); (b) presence or absence of rheumatoid factor (seropositivity/seronegativity), determined by Waaler-Rose test and defined as being positive with a titer of at least 1:24; and (c) concomitant rheumatic diseases with no relation to RA according to the following categorization: (a) concomitant diseases of the vertebral column; and (b) concomitant osteoporosis.

Control subjects: Any patient with a rheumatic disease other than RA was designated a control subject. The clinical diagnoses represent a cross-section of patients that were treated in the hospital mentioned above.

3. METHOD

The results shown in Tables 2–6 were obtained by comparing the diagnostic results of CADIAG-2/RHEUMA with the available confirmed clinical diagnoses. It should be mentioned that a CADIAG-2/RHEUMA diagnosis was taken to be established if it was either a confirmed diagnosis or a diagnostic hypothesis with a degree of confirmation of at least 0.5 (cf., Adlassnig, 1986; Adlassnig et al., 1986). Moreover, only a partial diagnostic inference process was carried out. Simple finding/disease relationships were left out of

TABLE 2
True Positive Results (= Sensitivity Rates) Obtained by the ARA Criteria-Based Rules 1 and 2 in Patients With Rheumatoid Arthritis (RA)

Diagnosis	Total Number of Patients	Rule 1 (ARA 1958 Definite)	Rule 2 (ARA 1987 Definite)
Seropositive RA, Stage 1	9	7	6
Seronegative RA, Stage 1	19	10	10
Seropositive RA, Stage 2	25	22	21
Seronegative RA, Stage 2	26	17	16
Seropositive RA, Stage 3	28	23	25
Seronegative RA, Stage 3	20	18	15
Seropositive RA, Stage 4	17	13	12
Seronegative RA, Stage 4	6	5	5
Total number of diagnoses	150	115	110
Sensitivity rates		76.7%	73.3%

TABLE 3
True Positive Results (= Sensitivity Rates) Obtained by the ARA Criteria-Based Rules 1 and 2 in RA Patients, Disease Stages 1–4

Diagnosis	Total Number of Patients	Rule 1 (ARA 1958 Definite)	Rule 2 (ARA 1987 Definite)
RA, Stage 1	28	17 (60.7%)	16 (57.1%)
RA, Stage 2	51	39 (76 5%)	37 (72.6%)
RA, Stage 3	48	41 (85 4%)	40 (83.3%)
RA, Stage 4	23	18 (78.2%)	17 (73.9%)

consideration in this study; only the respective diagnostic rules provided the diagnostic results. The primary goal was to test the available ARA criteria for applicability in CADIAG-2/RHEUMA. This implies, however, that a complete inference process of CADIAG-2/RHEUMA might improve the reported sensitivity and specificity results achieved with the single ARA criteria.

4. RESULTS

4.1. Results Obtained with Rules 1 and 2 (Based on ARA Criteria Definition Only)

True positive results (= sensitivity rates) with RA patients As shown in Table 2, rule 1 performed better than rule 2 reaching a sensitivity of 76.7%. Tables 3–5 show the different diagnostic results obtained in the RA subgroups. Cases with early disease stage (stage 1), cases with seronegative RA, and cases with concomitant diseases of the vertebral column tended to cause a lower sensitivity.

False positive results (= 100% − specificity rates) *with control subjects* As shown in Table 6, rule 1 performed better than rule 2 reaching a specificity of 88.0%. A substantial number of patients with psoriatic arthritis and systemic lupus erythematosus yielded incorrect results, a fact which led to further developments, as described in Section 4.2.

4.2. Development of an Improved Rule 3 (Based on ARA Criteria Definition and Clinical Experience)

To improve the diagnostic accuracy of CADIAG-2/RHEUMA, clinical experience of a rheumatology expert was needed to modify the established rules. Rule 1 was selected for further improvement because of the higher sensitivities and specificities obtained with it. Its criteria were consecutively changed to reach higher rates of both sensitivity and specificity. The problem was successfully approached in two different ways:

1. Redefinition of diagnostic criteria:
   The symptom "morning stiffness" was redefined; it had to last for only 30 minutes instead of 60 minutes.
   The sign "symmetrical joint swelling," which had to be observed by a physician, was remodelled to "symmetrical joint involvement," observed by a physician or reported in the patient history.
2. Addition of further exclusion criteria:
   To avoid false positive results in cases of psoriatic arthritis, an exclusion in case of present psoriasis was added to rule 1. This exclusion prevents the diagnosis of definite RA if there is sufficient evidence that a patient might actually suffer from psoriatic arthritis.

A comparison of all diagnostic and excluding criteria included in rules 1–3 is shown in Tables 7 and 8. In rules 1 and 3, a patient must fit at least 5 out of 11 criteria and none of the exclusions to establish RA as the confirmed diagnosis, whereas in rule 2 at least 4 out of 7 criteria must be present; no exclusion criteria are contained in rule 2.

4.3. Diagnostic Results Obtained with the Improved Rule 3 (Based on ARA Criteria Definition and Clinical Experience)

All improvements led to a definite (confirming) rule for RA which showed a sensitivity of 83.3% and a specificity of 95.3%, thus reaching a total accuracy of 89.3% (the mean of sensitivity and specificity rates) as shown in Table 9.

TABLE 4
True Positive Results (= Sensitivity Rates) Obtained by the ARA Criteria-Based Rules 1 and 2 in Seropositive and Seronegative RA Patients

Diagnosis	Total Number of Patients	Rule 1 (ARA 1958 Definite)	Rule 2 (ARA 1987 Definite)
Seropositive RA	79	65 (82 3%)	64 (81 0%)
Seronegative RA	71	50 (70.4%)	46 (64 8%)

TABLE 5
True Positive Results (= Sensitivity Rates) Obtained by the ARA Criteria Based Rules 1 and 2 in RA Patients With and Without Concomitant Spinal Diseases

Diagnosis	Total Number of Patients	Rule 1 (ARA 1958 Definite)	Rule 2 (ARA 1987 Definite)
No concomitant diseases	74	59 (79.7%)	54 (73.0%)
Diseases of the vertebral column	50	36 (72.0%)	34 (68.0%)
Osteoporosis	26	20 (76.9%)	22 (84.6%)

5. DISCUSSION

Computer-assisted differential diagnoses for rheumatic diseases has been an issue of clinical interest for more than 20 years.

A very early approach is already described by Horak, et al. (1968). This system was based on a method for computer-assisted differential diagnosis in internal medicine that employs Boolean logic and heuristic pattern matching (Spindelberger & Grabner, 1968). It was later redesigned and termed CADIAG-1 (Adlassnig & Grabner, 1985) that finally initiated the work on CADIAG-2 (Adlassnig, 1980, 1986). The earliest versions of these programs were not tested systematically with real cases. Evaluation studies of CADIAG-1/RHEUMA and CADIAG-2/RHEUMA including 282 clinically verified RA cases were described first in Adlassnig (1980). For these RA cases, sensitivity rates of 99.3% and 93.7%, respectively, could be obtained. It should be mentioned, however, that prompting RA as either a confirmed diagnosis, diagnostic hypothesis, or possible diagnosis (the latter in CADIAG-1/RHEUMA only) was judged to be a correct outcome. The presently described evaluation study is based on a more rigid definition of a correct diagnosis (hypothesis threshold $\epsilon \geq 0.5$, cf., Section 3).

Similar performance studies in the area of rheumatology including RA cases were carried out with the AI/RHEUMA consultant system (Kingsland & Lindberg, 1986; Kingsland, Lindberg, & Sharp, 1986; Porter et al., 1988). In Porter et al. (1988), its diagnostic accuracy was evaluated using information that was supplied by Japanese rheumatologists on 59 patients with connective tissue diseases (two RA cases overlapping with other diseases were included). The diagnoses of the AI/RHEUMA model were in full or partial agreement with those of the Japanese rheumatologists in 54 of 59 cases (91.5%). Preliminary evaluation of the criteria used by this model to diagnose mixed connective tissue disease showed a sensitivity of 90% and a specificity of 96%.

An earlier study on the performance of AI/RHEUMA that included a total of 384 cases (254 connective tissue diseases, 34 spondyloarthropathies, 19 crystal-induced arthritides, 30 infection-induced arthritides, 17 juvenile rheumatoid arthritis, and 30 other rheumatic disorders) is described in Kingsland et al. (1986). Here, an overall sensitivity rate of 94% could be obtained.

A further, extended evaluation of AI/RHEUMA with 1,570 consecutive outpatients of a Dutch rheumatological clinic was carried out and described in Bernelot Moens and Kingsland (1990). After a complete revision of the knowledge base of AI/RHEUMA, 93 definite RA cases, 95 possible RA cases, and 1,392 non-RA cases were tested. The definite RA cases yielded a sensitivity of 99%, the possible RA cases 62%, and the obtained specificity rate amounted to 98%.

The prospective application of the +RHEUMA expert

TABLE 6
False Positive Results (= 100%-Specificity Rates) Obtained by ARA Criteria-Based Rules 1 and 2 in Control Subjects

Diagnosis	Total Number of Patients	Rule 1 (ARA 1958 Definite)	Rule 2 (ARA 1987 Definite)
Osteoarthrosis	44	2	0
Gouty arthritis	32	0	4
Ankylosing spondylitis	30	2	0
Psoriatic arthritis	20	10	10
Bacterial arthritis	4	1	1
Reiter's disease	4	0	0
Systemic lupus erythematosus	4	2	3
Systemic sclerosis	4	0	0
Polymyositis	3	0	1
Chondrocalcinosis	3	1	0
Polymyalgia rheumatica	2	0	0
Total number of diagnoses	150	18	19
Specificity rates		88.0%	87.3%

TABLE 7
Comparison of All Diagnostic Criteria to be Fulfilled in Rules 1-3

Diagnostic Criteria	Rule 1 (ARA 1958 Definite)	Rule 2 (ARA 1987 Definite)	Rule 3 (Improved Rule 1)
Morning stiffness 60 min	yes	yes	—
Morning stiffness 30 min.	—	—	yes
Pain on motion or tenderness in at least one joint	yes	—	yes
Swelling in at least one joint	yes	—	yes
Swelling in at least one other joint	yes	—	yes
Arthritis of three or more joint areas	—	yes	—
Arthritis of hand joints	—	yes	—
Symmetrical joint swelling	yes	yes	—
Symmetrical joint involvement	—	—	yes
Subcutaneous nodules	yes	yes	yes
X-ray changes typical of RA	yes	yes	yes
Positive serum rheumatoid factor	yes	yes	yes
Poor mucin precipitate from synovial fluid	yes	—	yes
Characteristic histologic changes in synovium	yes	—	yes
Characteristic histologic changes in nodules	yes	—	yes
Total number of criteria	5 out of 11 and excluding criteria in Table 8	4 out of 7	5 out of 11 and excluding criteria in Table 8

system for the medical history of joint pain (Schewe, Herzer, & Krüger, 1990; Schewe, Krüger, Herzer, & Schattenkirchner, 1991; Schewe, Sherrman, & Gierl, 1988) is reported in Schewe et al. (1990). Sixty-seven cases of RA yielded a sensitivity rate of 89.5%. The hit rates of the other rheumatic diseases included in this study (lupus erythematosus, ankylosing spondylitis, psoriatic arthritis, osteoarthritis of greater joints, spinal osteoarthritis, systemic sclerosis, infectious arthritis, polymyalgia rheumatica, and osteoarthritis of the fingers) range from 58.2% for osteoarthritis of the greater joints to 100% for systemic sclerosis.

Both CADIAG-2/RHEUMA and AI/RHEUMA employ criteria sets for the formal representation of RA defi-

TABLE 8
Comparison of all Excluding Criteria in Rules 1-3 (Some Excluding Criteria Given in Ropes et al. (1958) and Arnett et al. (1988) Were Simplified and Substituted by Their Respective Disease Terms)

Excluding Criteria	Rule 1 (ARA 1958 Definite)	Rule 2 (ARA 1987 Definite)	Rule 3 (Improved Rule 1)
Systemic lupus erythematosus	yes	—	yes
Panarteritis nodosa	yes	—	yes
Polymyositis/dermatomyositis	yes	—	yes
Systemic sclerosis	yes	—	yes
Rheumatic fever	yes	—	yes
Gouty arthritis	yes	—	yes
Infectious arthritis	yes	—	yes
Joint tuberculosis	yes	—	yes
Reiter's disease	yes	—	yes
Shoulder-hand syndrome	yes	—	yes
Hypertrophic pulmonary osteoarthropathy	yes	—	yes
Neuroarthropathy	yes	—	yes
Ochronosis	yes	—	yes
Sarcoidosis	yes	—	yes
Multiple myeloma	yes	—	yes
Erythema nodosum	yes	—	yes
Leukemia or lymphoma	yes	—	yes
Psoriatic arthritis	—	—	yes
Total number of criteria	17	—	18

TABLE 9
Sensitivity, Specificity, and Accuracy Rates Obtained by Rules 1 and 2 and Improved Rule 3

	Rule 1 (ARA 1958 Definite)	Rule 2 (ARA 1987 Definite)	Rule 3 (Improved Rule 1)
Sensitivity	76.7%	73.3%	83.3%
Specificity	88.0%	87.3%	95.3%
Accuracy	82.4%	80.3%	89 3%

nitions. The +RHEUMA system contains a large rule base for representing medical knowledge on rheumatic diseases.

Our study especially analyzes the applicability of the different ARA criteria for representing knowledge on RA. Historically, a committee of the American Rheumatism Association proposed the (revised) 11 diagnostic criteria for RA with 20 exclusions in 1958. A diagnosis of classical RA required at least seven criteria, definite RA required at least five criteria, probable RA required at least three criteria and in all disease categories none of the exclusions was allowed to be fulfilled.

The 1958 criteria have been applied for more than 30 years; they assisted in a more uniform definition of RA, while over the same period several new forms of arthritis such as HLA B-27 associated spondylarthropathies were classified separately.

In 1987, a new ARA committee reviewed the 1958 RA criteria and issued an updated set of criteria allowing only one single (definite) disease category and requiring neither invasive diagnostic procedures (which were performed rarely) nor exclusionary criteria (which were difficult to handle in practice). Instead of formulating diagnostic criteria as had been done in 1958, the 1987 criteria were formed to facilitate the classification of RA (i.e., the selection of more uniform patient cohorts for RA studies).

Though it seems meaningful to use classification criteria for diagnostic purposes, there are limitations—limitations with respect to the achievable sensitivity and specificity rates. Since classification criteria contain features of typical RA cases, sensitivity rates may only be as good as the frequency rates of typical RA cases within the whole disease spectrum. RA patients who do not fit in the typical disease picture of RA might be easily misdiagnosed. Two RA patient groups with such features are shown in Table 3 and Table 4. Both seronegative RA patients and those with early disease stage yield lower sensitivity rates. But since the majority of RA patients exhibits typical disease features, an expert system can still perform successfully even if it were purely based on classification criteria. (However, the sensitivity and specificity rates obtained in our study are not as high as those published by the ARA committee in Arnett et al. (1988).)

Moreover, it was not surprising to find out that the 1958 set of criteria (rule 1) performed better than the 1987 set (rule 2) because the former was especially designed to support diagnosis. Even though a list of 20 exclusions might not be handled easily in clinical practice, computer expert systems with direct connection to a central patient data base containing all necessary information can manage this without problems. The inclusion of psoriatic arthritis as a further exclusion criterion, an arthritis form not yet classified separately in 1958, was one of the reasons why rule 3 showed higher rates of specificity.

Furthermore, the obtained results were shown to be highly sensitive to symptom definitions. In our study, the redefinition of two symptoms led to improved rates of sensitivity and specificity, a goal which usually is not easily accomplished due to the mutual trade-off between sensitivity and specificity. The increase in performance of rule 3 compared to rule 1 (which it is based upon) proved that the subjective clinical experience of a medical specialist is helpful in achieving a better expert system's performance.

Acknowledgements—The authors thank I Gröger for secretarial assistance, G. Sedivy for medical data input, C. Schuh, F. Fischler, and W. Temsch for extensive programming work. This research was partly supported by IBM Austria.

REFERENCES

Adlassnig, K.-P (1980). A fuzzy logical model of computer-assisted medical diagnosis. *Methods of Information in Medicine*, **19**, 141–148

Adlassnig, K.-P. (1986) Fuzzy set theory in medical diagnosis. *IEEE Transactions on Systems, Man, and Cybernetics*, **SMC-16**, 260–265.

Adlassnig, K.-P (1988). Uniform representation of vagueness and imprecision in patient's medical findings using fuzzy sets. In *Proceedings of Cybernetics and Systems* '88 (pp. 685–692). Dordrecht Kluwer Academic Publishers.

Adlassnig, K.-P , & Akhavan-Heidari, M (1989) CADIAG-2/GALL: An experimental diagnostic expert system for gallbladder and biliary tract diseases. *Artificial Intelligence in Medicine*, **1**, 71–77.

Adlassnig, K.-P., & Grabner, H. (1985). Verarbeitung natürlichsprachiger medizinischer Begriffe. In G Grabner (Ed.), *WAMIS—Wiener Allgemeines Medizinisches Informations-System*, (pp. 162–189). Berlin: Springer-Verlag

Adlassnig, K -P , & Kolarz, G. (1986). Representation and semiautomatic acquisition of medical knowledge in CADIAG-1 and CADIAG-2. *Computers and Biomedical Research*, **19**, 63–79.

Adlassnig, K.-P., Kolarz, G., Scheithauer, W , Effenberger, H., & Grabner, G (1985) CADIAG· Approaches to computer-assisted

medical diagnosis. *Computers in Biology and Medicine*, **15**, 315–335

Adlassnig, K.-P., Kolarz, G., Scheithauer, W., & Grabner, H. (1986) Approach to a hospital-based application of the medical expert system CADIAG-2. *Medical Informatics*, **11**, 205–223

Adlassnig, K.-P., & Scheithauer, W. (1989) Performance evaluation of medical expert systems using ROC curves. *Computers and Biomedical Research*, **22**, 297–313

Adlassnig, K.-P., Scheithauer, W., & Grabner, G. (1984) Computerunterstutzte Medizinische Diagnostik und ihr Einsatz bei Pankreaserkrankungen *Acta Med Austriaca*, **11**, 125–134.

Arnett, F. C., Edworthy, St. M., Bloch, D. A., McShane, D. J., Fries, J. F., Cooper, N. S., Healey, L. A., Kaplan, St. R., Liang, M. H., Luthra, H. S., Medsger, Jr., Th. A., Mitchell, D. A., Neustadt, D. A., Pinals, R. S., Schaller, J. G., Sharp, J. T., Wilder, R. L., & Hunder, G. G. (1988) The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis *Arthritis and Rheumatism*, **33**, 315–324

Bernelot Moens, H. J., & Kingsland L. C., III (1990) Evaluation and further development of AI/RHEUM with data from 1570 consecutive outpatients of a Dutch rheumatological clinic In G. Cooper, M. Musen, R. Patil, & L. Widman (Eds.), *Working notes of the AAAI Spring Symposium Series Symposium "AI in Medicine"* (pp. 141–144). Stanford, CA: American Association for Artificial Intelligence

Horak, W., Michalek, P., Richter, H., & Thumb, N. (1968) Computerdiagnostik in der Rheumatologie In K. Fellinger (Eds.), *Computer in der Medizin—Probleme, Erfahrungen, Projekte* (pp. 232–237) Verlag Bruder Hollinek, Wien

Kingsland L. C. III, & Lindberg, D. A. B. (1986) The criteria form of knowledge representation in medical artificial intelligence In *Proceedings MEDINFO 86* (pp. 12–16) North Holland. Elsevier Science Publishers B. V.

Kingsland L. C., III, Lindberg, D. A. B., & Sharp, G. C. (1986). Anatomy of a knowledge-based consultant system AI/RHEUM *M. D. Computing*, **3**, 18–26

Kolarz, G., & Adlassnig, K.-P. (1986) Problems in establishing the medical expert systems CADIAG-1 and CADIAG-2 in rheumatology *Journal of Medical Systems*, **10**, 395–405.

Porter, J. F., Kingsland L. C., III, Lindberg, D. A. B., Shah, I., Benge, J. M., Hazelwood, S. E., Kay, D. R., Homma, M., Akizuki, M., Takano, M., & Sharp, G. C. (1988). The AI/RHEUM knowledge-based computer consultant system in rheumatology *Arthritis and Rheumatism*, **31**, 219–226

Ropes, M. W., Bennett, G. A., Cobb, S., Jacox, R., & Jassar, R. A. (1958) Revision of diagnostic criteria for rheumatoid arthritis *Bulletin on Rheumatic Diseases*, **9**, 175–177

Schewe, S., Herzer, P., & Kruger, K. (1990) Prospective application of an expert system for the medical history of joint pain *Klinische Wochenschrift*, **68**, 466–471.

Schewe, S., Kruger, K., Herzer, P., & Schattenkirchner, M. (1991). Der Arztcomputer als Diagnoseunterstutzung bei Gelenkschmerzen—Patientenmeinung und Ergebnisse *Zeitschrift fur Rheumatologie*, **50**, 211–215

Schewe, S., Scherrmann, W., & Gierl, L. (1988). Evaluation and measuring of benefit of an expert system for differential diagnosis in rheumatology In O. Rienhoff, U. Piccolo, & B. Schneider (Eds.) *Expert systems and decision support in medicine* (pp. 351–354) Berlin Springer-Verlag

Spindelberger, W., & Grabner, G. (1968). Ein Computerverfahren zur Diagnostischen Hilfestellung. In K. Fellinger (Eds.) *Computer in der Medizin—Probleme, Erfahrungen, Projekte* (pp. 189–221) Verlag Bruder Hollinek, Wien

Steinbrocker, O., Traeger, C. H., & Batterman, R. C. (1949) Therapeutic criteria in rheumatoid arthritis *The Journal of the American Medical Association*, **140**, 659–662