

Evaluation of two different models of semi-automatic knowledge acquisition for the medical consultant system CADIAG-II/RHEUMA

Harald Leitich^{a,b,c,*}, Klaus-Peter Adlassnig^{a,b,c}, Gernot Kolarz^{d,e}

^aDepartment of Medical Computer Sciences, Section of Medical Expert and Knowledge-Based Systems, University of Vienna Medical School, Spitalgasse 23, A-1090 Vienna, Austria

^bDepartment of Obstetrics and Gynecology, University of Vienna Medical School, Währinger Gürtel 18-20, A-1090 Vienna, Austria

^cLudwig Boltzmann Institute for Expert Systems and Quality Management in Medicine, Spitalgasse 23, A-1090 Vienna, Austria

^dClinic for Rheumatic Disease of the Social Insurance Company for Trade and Industry, Adolfine Malchergasse 1, A-2500 Baden, Austria

^eInstitute for Rheumatology, Marchetstraße 78, A-2500 Baden, Austria

Received 7 July 2001; received in revised form 11 October 2001; accepted 15 October 2001

Abstract

As part of a plan to promote semi-automatic knowledge acquisition for the medical consultant system CADIAG-II/RHEUMA, this study sought to explore and cope with the variability of results that may be anticipated when performing knowledge acquisition with patient data from different patient settings. Patient data were drawn both from a published study for the classification of rheumatoid arthritis (RA) and from a large database of rheumatological patient charts developed for the CADIAG-II/RHEUMA system. An analysis of the relationships between RA and selected CADIAG-II/RHEUMA symptoms was done using two models. In one of them, we controlled for the differences in baseline frequencies of symptoms and diseases in the two study populations as an important factor influencing the results of the calculations. Other factors that were identified included inconsistent definitions of symptoms and diseases, and the different composition of study groups in the two study populations. By eliminating differences in baseline frequencies as the most important bias, the results obtained from the two different knowledge sources became more consistent. All remaining inconsistencies and uncertainties about the contribution and relative importance of the factors were formalized using fuzzy intervals. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: CADIAG-II/RHEUMA; Expert system; Knowledge acquisition; Rheumatoid arthritis

* Corresponding author. Tel.: +43-1-40400-2822; fax: +43-1-40400-2861.
E-mail address: harald.leitich@akh-wien.ac.at (H. Leitich).

1. Introduction

Knowledge acquisition (KA) has consistently been described as a bottleneck in the development of computer consultant systems. Areas of high complexity impose great demands on domain experts to generate knowledge bases that are both consistent and complete in the sense that all major aspects of the application domain are correctly represented in the knowledge base. Automatic KA from databases within the application domain was proposed early to support traditional manual KA. Knowledge acquired within such a database, however, will automatically reflect biases and peculiarities of the specific set of data. In order to generate knowledge that is more consistent among different data sets, it is important to identify and to control for biasing factors.

As part of a plan to promote semi-automatic KA for the medical consultant system CADIAG-II/RHEUMA, this study sought to explore and cope with the variability of results that may be expected during KA with patient data from different patient settings. By identifying and controlling for major influencing factors, we expected the results obtained from different patient data sets to become more consistent. Finally, we planned to use fuzzy techniques to cope with the remaining inconsistencies and uncertainties about the contribution and relative importance of all factors.

The medical consultant system CADIAG-II was developed at the Department of Medical Computer Sciences, University of Vienna Medical School. The predecessors of CADIAG-II include a system based on Boolean logic [34], followed by the system CADIAG-I, in which first-order predicate calculus formulas were used to define relationships between symptoms and diseases [30]. A new concept for a successor system based on fuzzy set theory to formalize symptoms and diseases, and on fuzzy logic as the processing mechanism, was described by Adlassnig in 1980 [1]. The new system, named CADIAG-II, was implemented in 1982, and was completely incorporated into the medical information system WAMIS in 1984 [7]. It was continuously improved over the following years [2,3,5,6,8].

CADIAG-II itself was the starting point of a series of new approaches to a more generalized medical consultant system known as MEDFRAME. Current areas of research and development include a highly structured object-oriented patient data and medical knowledge base [16,24], a more generalized use of fuzzy set theory in symptom generation [17], and a broader definition of relationships between medical entities [31].

Symptoms and diseases in CADIAG-II are formalized as fuzzy sets, which are characterized by fuzzy membership functions [3,36]. Relationships between symptoms and diseases 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 the range [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 disease. Strength of confirmation degrees of 1 or 0 are assigned 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.

The rheumatological knowledge base of CADIAG-II was first developed by Adlassnig and Kolarz as a knowledge base for the CADIAG-I system, and was later modified and expanded for the CADIAG-II system [4]. It currently contains 170 documented diseases and 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). The total numbers of simple and complex symptom–disease (S–D) relationships are 16,040 and 60.

CADIAG-II/RHEUMA has been evaluated in several studies, both as a complete consultant system [23], and as a series of evaluations of diagnostic rules that were based on the 1958 diagnostic criteria for rheumatoid arthritis (RA), and the 1987 revised criteria for the classification of RA, both published by the American College of Rheumatology (ACR) [9,25,26].

Apart from CADIAG-II/RHEUMA, and apart from well-known computer consultant systems in internal medicine such as INTERNIST-I/QMR [11,29] or ILIAD [13,35], several other systems have been specifically designed to support the differential diagnosis of rheumatic diseases, including AI/RHEUM [22], RENOIR [12,19], MESICAR [20,21], RHEUMA [32,33], and the systems designed by Bernelot Moens and Van der Korst [14,15], McCrea et al. [28], and Mathew et al. [27]. In these systems, different forms of knowledge representation and inference were used, such as criteria tables [22], modified Bayes' theorem [15], decision trees [28], detailed anatomical and functional knowledge [21], discrimination and connectivity analysis [27], simple if–then rules [32], or fuzzy logic [12]. In the latter, fuzzy set theory and fuzzy logic are used to define fuzzy facts and to allow rules to have certainty values.

KA for CADIAG-II/RHEUMA, i.e. the definition of 60 complex, and more than 16,000 simple S–D relationships, proved to be a tedious and time-consuming task. Most of the work was done by a single domain expert (GK) using a linguistic approach: among a predefined set of linguistic expressions, the expression that most closely matched the correct relation between a symptom S and a disease D—for both the frequency of occurrence and the strength of confirmation degrees—was chosen. During the development of the knowledge base for CADIAG-II/RHEUMA, manual KA had already been partially supported by semi-automatic KA [6]. Semi-automatic KA was done by calculating all statistical relationships between symptoms and diseases in the patient database of CADIAG-II/RHEUMA. Results of all statistical calculations were then validated by the domain expert as proposals for the fuzzy values of the frequency of occurrence and the strength of confirmation degrees, and were subsequently entered into the knowledge base.

This study sought to demonstrate the variability of results that may be expected when performing automatic KA for a single disease (RA) in two different patient settings. By eliminating some of the influencing factors, a consensus could be reached. All remaining uncertainties about influencing factors that were not yet accounted for, would then be addressed using expressions of fuzzy set theory.

2. Methods

The patient data used in this study were taken from a large rheumatological patient database that had been built up during the development of CADIAG-II/RHEUMA. All patients were treated in a 140-bed rheumatological hospital in Baden, Austria. At the time of this study, it contained computerized records of 154 patients with RA and 3098 control patients with other rheumatological diseases.

The second source of patient data was a publication of the ACR, in which a new set of revised criteria for the classification of RA were introduced [10]. In this paper, a variety of symptoms leading to the diagnosis of RA were evaluated in 262 RA patients and 262 control patients from a wide range of university and private practice settings throughout the US, and all results were displayed in detail.

In both study populations, 2×2 tables were calculated for seven symptoms, which were finally chosen by the ACR as new criteria for the classification of RA, because of their high discriminatory power between RA and other rheumatological diseases. In all calculations, symptoms were assumed to be either present, absent, or unknown in a given patient.

The models of semi-automatic KA developed for the CADIAG-II system were based on the assumption that relations in CADIAG-II could have statistical interpretations: the frequency of occurrence degree can be statistically interpreted as $P(S/D)$ and the strength of confirmation degree as $P(D/S)$. Thus, Bayes' theorem, as displayed in the upper portions of Tables 1 and 2, might be used to calculate the probabilities $P(S/D)$ and $P(D/S)$, which, in turn, may be transformed to fuzzy values.

Bayes' theorem, in the notation in Table 2, is commonly used to calculate an individual probability $P(D/S)$, i.e. the probability that an individual patient with a symptom S has a diagnosis D . $P(D)$, i.e. the prior probability that this individual patient has the diagnosis D is used as an input variable, whereas the probabilities $P(S/D)$ and $P(S/\neg D)$ are assumed to be fixed values and so-called test characteristics. Thus, following Bayes' theorem as a KA tool for CADIAG-II/RHEUMA, a set of different fuzzy values, depending on different prior probabilities $P(D)$, would have to be acquired.

In contrast, knowledge representation in CADIAG-II/RHEUMA followed a different philosophy: acquired knowledge should be independent of individual predispositions. The rationale for this approach was that a consultation process should be possible even if either no, or unreliable, information about the individual patient, for whom the consultation is done, is available. Another reason not to consider prior probabilities was that even if reliable background information about a patient is available, estimates of the prior probability $P(D)$ itself tend to be unreliable.

To emphasize the difference between probability theory and its application in Bayes' theorem on the one hand, and our model of KA based on calculations with data from a patient database on the other, we used the notations $F(S/D)$ and $F(D/S)$, where F stands for frequency, instead of $P(S/D)$ and $P(D/S)$.

Derived from Bayes' theorem, two models to calculate $F(S/D)$ and $F(D/S)$, with two different assumptions about the frequencies $F(S)$ and $F(D)$, were used in this study. Both models were based on the calculation of 2×2 tables (listing true positive, false positive, true negative, and false negative results) to analyze the relationship between a symptom S and a disease D .

The simpler formulae of model 2 at the bottom of Tables 1 and 2 assume that prior probabilities are, and should be, exactly equal to those found in the patient data set; in fact, in those formulae $F(S/D)$ corresponds to the so-called sensitivity rate and $F(D/S)$ to the so-called positive predictive value. In the more complex formulae of model 1 in the center of Tables 1 and 2, the frequencies $F(D)$ and $F(S)$ are normalized, thus $F(S)$ is set equal to $F(\neg S)$, and $F(D)$ is set equal to $F(\neg D)$. By normalizing predispositions, their influence on the results

Table 1

Calculation of proposals for the frequency of occurrence degree μ_o with models 1 and 2, based on numbers of true positive (TP), false positive (FP), true negative (TN), and false negative (FN) results

Statistical background:

$$P(S/D) = \frac{P(D/S) \cdot P(S)}{P(D/S) \cdot P(S) + P(D/\neg S) \cdot P(\neg S)}$$

Model 1: Calculation with normalized baseline frequencies $F(S)$, $F(\neg S)$

Assumption:

$$F(S) = F(\neg S)$$

Calculation:

$$\mu_o = F(S/D) = \frac{F(S \cap D)}{F(D)} = \frac{F(D/S)}{F(D/S) + F(D/\neg S)} = \frac{\frac{TP}{TP + FP}}{\frac{TP}{TP + FP} + \frac{FN}{FN + TN}}$$

Model 2: Calculation with baseline frequencies $F(S)$, $F(\neg S)$ as found in patient data set

Assumptions:

$$F(S) = \frac{TP + FP}{TP + FP + TN + FN}$$

$$F(\neg S) = \frac{TN + FN}{TP + FP + TN + FN}$$

Calculation:

$$\mu_o = F(S/D) = \frac{TP}{TP + FN}$$

of the calculations is eliminated; or, arguing from a probabilistic standpoint, the chances that an individual does, or does not, have a prior probability $P(S)$ or $P(D)$, are set equal.

The calculation of the frequencies $F(S/D)$ and $F(D/S)$ is also influenced by other factors, among which different definitions of symptoms and diseases, different compositions of the groups of patients D and controls $\neg D$, and different compositions of the groups of persons with S or $\neg S$, are, arguably, the most important ones. By normalizing, and thus eliminating, prior probabilities as influencing factors, the influence of different definitions of symptoms and diseases which shift the balances between the groups D and $\neg D$, or S and $\neg S$, will also be controlled for. The selection of patients and controls, or symptomatic and asymptomatic persons, on the other hand, remains the major influencing factor in our model.

Calculations of $F(S/D)$ and $F(D/S)$ were done with both models depicted in [Tables 1 and 2](#). We also calculated $F(\neg S/\neg D)$ and $F(\neg D/\neg S)$ using equivalent formulae (not shown).

Table 2

Calculation of proposals for the confirmation degree μ_c with models 1 and 2, based on numbers of true positive (TP), false positive (FP), true negative (TN), and false negative (FN) results

Statistical background:

$$P(D/S) = \frac{P(S/D) \cdot P(D)}{P(S/D) \cdot P(D) + P(S/\neg D) \cdot P(\neg D)}$$

Model 1: Calculation with normalized baseline frequencies $F(D)$, $F(\neg D)$

Assumption:

$$F(D) = F(\neg D)$$

Calculation:

$$\mu_c = F(D/S) = \frac{F(S \cap D)}{F(S)} = \frac{F(S/D)}{F(S/D) + F(S/\neg D)} = \frac{\frac{TP}{TP + FN}}{\frac{TP}{TP + FN} + \frac{FP}{FP + TN}}$$

Model 2: Calculation with baseline frequencies $F(D)$, $F(\neg D)$ as found in patient data set

Assumptions:

$$F(D) = \frac{TP + FN}{TP + FP + TN + FN}$$

$$F(\neg D) = \frac{FP + TN}{TP + FP + TN + FN}$$

Calculation:

$$\mu_c = F(D/S) = \frac{TP}{TP + FP}$$

Analogous to the frequency of occurrence degree and the strength of confirmation degrees, we defined a frequency of non-occurrence degree, μ_{NO} , which describes the frequency of $\neg S$ in the group of controls and the strength of exclusion degree, μ_E , which describes the frequency of $\neg D$ in the group of asymptomatic individuals. Thus, μ_{NO} can be set equal to $F(\neg S/\neg D)$, and μ_E can be set equal to $F(\neg D/\neg S)$. In its simplest form, $F(\neg S/\neg D)$ is the so-called true negative or specificity rate and $F(\neg D/\neg S)$, the so-called negative predictive value. Baseline frequencies $F(D)$ for RA and $F(S)$ for all seven symptoms were also calculated for both study populations.

We expected sensitivity and specificity rates, and the positive and negative predictive values obtained in the two study populations to be different from each other—even if the results were normalized for different baseline frequencies $F(S)$, $F(D)$. This is because the different composition of patient groups in the two populations were expected to persis-

tently exercise a strong influence on the outcome of our calculations. To cope with these factors that were not accounted for, we decided to use the definition of a fuzzy interval to describe the remaining uncertainty. Thus, the proposals for μ_O , μ_C , μ_{NO} , and μ_E were made by defining fuzzy intervals in which the limits were set equal to the results obtained with the calculations using model 1 in both study populations.

3. Results

In Tables 3 and 4, the rates of sensitivity, specificity, and the positive and negative predictive values, calculated with model 2, and the baseline frequencies $F(S)$, $F(D)$ are shown for both study populations. The lower baseline frequencies of symptoms and diseases in our own patient data base, as compared to the ACR study, led to markedly diminished rates of sensitivity and lower positive predictive values, and to higher

Table 3
Rates of sensitivity and specificity calculated using model 2, and baseline frequencies $F(S)$ in both study populations

Symptoms	ACR study			CADIAG-II study		
	Sensitivity (%)	Specificity (%)	$F(S)$ (%)	Sensitivity (%)	Specificity (%)	$F(S)$ (%)
Morning stiffness for at least 1 h	81.2	57.5	61.9	14.9	96.4	4.1
Swelling of three or more joint areas	90.6	84.2	53.3	70.8	95.4	7.8
Swelling of the PIP, MCP, or wrist joints	96.6	74.7	61.0	85.1	91.6	12.1
Symmetric joint swelling	94.3	74.3	60.0	81.2	90.6	12.8
Rheumatoid nodules	43.5	97.7	22.9	13.0	98.3	2.3
Presence of rheumatoid factor	80.4	87.0	49.9	37.5	98.6	3.1
Radiographic changes typical for arthritis	77.3	93.7	44.4	89.0	97.3	6.8

Table 4
Positive predictive values (PPV) and negative predictive values (NPV) calculated using model 2, and baseline frequencies $F(D)$ in both study populations

Symptoms	ACR study			CADIAG-II study		
	PPV (%)	NPV (%)	$F(D)$ (%)	PPV (%)	NPV (%)	$F(D)$ (%)
Morning stiffness for at least 1 h	65.7	75.3	50.1	16.7	95.9	4.6
Swelling of three or more joint areas	85.2	89.9	50.1	43.1	98.5	4.7
Swelling of the PIP, MCP, or wrist joints	79.3	95.6	50.1	33.4	99.2	4.7
Symmetric joint swelling	78.7	92.8	50.1	30.0	99.0	4.7
Rheumatoid nodules	95.0	63.3	50.1	27.0	95.8	4.7
Presence of rheumatoid factor	88.2	78.6	54.7	57.6	96.9	4.7
Radiographic changes typical for arthritis	93.4	78.1	53.7	62.0	99.4	4.7

Table 5
Normalized rates of sensitivity and specificity calculated using model 1 in both study populations

Symptoms	ACR study		CADIAG-II study	
	Normalized sensitivity (%)	Normalized specificity (%)	Normalized sensitivity (%)	Normalized specificity (%)
Morning stiffness for at least 1 h	72.6	68.7	80.3	53.5
Swelling of three or more joint areas	89.4	85.8	96.6	63.4
Swelling of the PIP, MCP, or wrist joints	94.7	82.2	97.7	59.8
Symmetric joint swelling	91.6	81.3	96.7	58.6
Rheumatoid nodules	72.1	92.6	86.5	56.8
Presence of rheumatoid factor	80.5	86.9	95.0	69.6
Radiographic changes typical for arthritis	81.0	92.2	99.1	72.3

specificity rates and higher negative predictive values. This influence did not appear to be strong in all symptoms, and thus, the question remained as to whether unknown differences in definitions of symptoms and diseases, based on different concepts of symptoms and diseases in different hospitals, or the different compositions of patient groups in the two study populations, were other important factors. In any event, the results obtained with model 2 in both study populations were too inconsistent to be useful as proposals to define CADIAG-II/RHEUMA relationships.

In Tables 5 and 6, normalized sensitivity and specificity rates, and the positive and negative predictive values, calculated with model 1, are shown for both study populations. Normalization of the baseline frequencies done with the calculations in model 1 helped to eliminate the bias of different baseline frequencies of symptoms and diseases, and to control for possible and unknown divergent definitions of symptoms and diseases. Still, differences in results obtained in the two study populations—although much smaller now—persisted. Explanations for the differences can be found by having a closer look at the different study populations. In the ACR study carried out to determine criteria for a homogenous classification of RA patients for clinical trials, only patients with a definite

Table 6
Normalized positive predictive values (PPV) and negative predictive values (NPV) calculated using model 1 in both study populations

Symptoms	ACR study		CADIAG-II study	
	Normalized PPV (%)	Normalized NPV (%)	Normalized PPV (%)	Normalized NPV (%)
Morning stiffness for at least 1 h	65.6	75.3	80.6	53.1
Swelling of three or more joint areas	85.1	89.9	93.8	76.5
Swelling of the PIP, MCP, or wrist joints	79.2	95.6	91.0	86.0
Symmetric joint swelling	78.6	92.8	89.6	82.8
Rheumatoid nodules	94.9	63.3	88.2	53.0
Presence of rheumatoid factor	86.0	81.6	96.5	61.2
Radiographic changes typical for arthritis	92.4	80.5	97.0	89.8

Table 7

Proposals for the frequency of occurrence degree μ_O , the frequency of non-occurrence degree μ_{NO} , the strength of confirmation degree μ_C , and the strength of exclusion degree μ_E , based on calculations using model 1 in both study populations

Symptoms	$\mu_O = \begin{cases} 1, & \forall x \in I \\ 0, & \forall x \notin I \end{cases}$	$\mu_{NO} = \begin{cases} 1, & \forall x \in I \\ 0, & \forall x \notin I \end{cases}$	$\mu_C = \begin{cases} 1, & \forall x \in I \\ 0, & \forall x \notin I \end{cases}$	$\mu_E = \begin{cases} 1, & \forall x \in I \\ 0, & \forall x \notin I \end{cases}$
Morning stiffness for at least one hour	$I = [0.72; 0.80]$	$I = [0.53; 0.68]$	$I = [0.65; 0.80]$	$I = [0.53; 0.71]$
Swelling of three or more joint areas	$I = [0.89; 0.96]$	$I = [0.63; 0.85]$	$I = [0.85; 0.93]$	$I = [0.76; 0.89]$
Swelling of the PIP, MCP, or wrist joints	$I = [0.94; 0.97]$	$I = [0.59; 0.82]$	$I = [0.79; 0.91]$	$I = [0.86; 0.95]$
Symmetric joint swelling	$I = [0.91; 0.96]$	$I = [0.58; 0.81]$	$I = [0.78; 0.89]$	$I = [0.82; 0.92]$
Rheumatoid nodules	$I = [0.72; 0.86]$	$I = [0.56; 0.92]$	$I = [0.88; 0.94]$	$I = [0.53; 0.63]$
Presence of rheumatoid factor	$I = [0.80; 0.95]$	$I = [0.69; 0.86]$	$I = [0.86; 0.96]$	$I = [0.61; 0.81]$
Radiographic changes typical for arthritis	$I = [0.81; 0.99]$	$I = [0.72; 0.92]$	$I = [0.92; 0.97]$	$I = [0.80; 0.89]$

diagnosis of RA were included in the RA group. In our database, however, a large percentage of RA patients were either at an early stage of disease or had already received disease-modifying drugs—in both cases patients had less pronounced disease features. The control group in the ACR study included a much larger percentage of patients with other inflammatory rheumatic disorders compared to the control group of our database, in which the majority of patients were affected by degenerative rheumatic disorders. For some symptoms, as a consequence, sensitivity rates in our population tended to be lower and specificity rates, higher.

In Table 7, fuzzy intervals as proposals for the definition of the frequency of occurrence and non-occurrence, and the strength of confirmation and exclusion degrees, are shown. The limits of the fuzzy intervals were set equal to the results obtained with the calculations using model 1 in both study populations.

4. Discussion

In this study, we showed that characteristics of patient populations, especially the baseline frequencies of symptoms and diseases, are important factors influencing calculations to support automatic KA from computerized patient data.

All influencing factors would play a minor role if KA were done in a patient setting that closely matched the setting in which the consultant system would later be used. This, however, would be detrimental to the philosophy of an all-purpose consultant system. In contrast, to follow this philosophy, KA should be done in as many as possible different patient settings and a consensus between diverging results achieved in these different settings should be obtained.

Although we successfully controlled for differences in baseline frequencies of symptoms and diseases as the most important bias, it became clear that even if we tried to identify, address, and partly control for some influencing factors, the contribution and relative importance of all factors in a given example would remain unknown.

As a consequence, the uncertainty as to what should be the “true” values that should be assigned to the relationships in CADIAG-II/RHEUMA’s knowledge base persisted. The philosophy of CADIAG-II/RHEUMA, of a consultant system that is independent of any prior information, motivated us, and the mathematical models to formalize uncertainty in fuzzy set theory allowed us to follow the solution that all remaining uncertainty can be expressed as a fuzzy interval that incorporates the entire range of results found in both study populations. This is also a flexible approach because new information obtained in other studies can easily be incorporated into the existing knowledge base by adapting the limits of the fuzzy intervals accordingly.

It should be noted that the present version of CADIAG-II/RHEUMA does not allow the assignment of fuzzy intervals, but only fuzzy values to relationships between symptoms and diseases. Also, in its present form, negative evidence, expressed by the frequency of non-occurrence degree and the strength of exclusion degrees, cannot be included in the knowledge base of CADIAG-II/RHEUMA. Both concepts, however, were discussed in an earlier paper [5] and will be part of the development of a KA tool for the ongoing MEDFRAME/CADIAG-IV project [18].

Acknowledgements

We gratefully acknowledge the financial support provided by ARC Seibersdorf Research GmbH, Biomedical Engineering, Seibersdorf, Austria; the corporate research partner of the Ludwig Boltzmann Institute for Expert Systems and Quality Management in Medicine, Vienna, Austria.

References

- [1] Adlassnig K-P. A fuzzy logical model of computer-assisted medical diagnosis. *Methods Inf Med* 1980;19:141–8.
- [2] Adlassnig K-P. Fuzzy set theory in medical diagnosis. *IEEE Trans Syst Man Cybernet SMC* 1986;16:260–5.
- [3] Adlassnig K-P. Uniform representation of vagueness and imprecision in patient’s medical findings using fuzzy sets. In: Trapp R, editor. *Proceedings of Cybernetics and Systems*. Dordrecht: Kluwer Academic Publishers, 1988. p. 685–92.
- [4] Adlassnig K-P, Kolarz G. CADIAG-II: computer-assisted medical diagnosis using fuzzy subsets. In: Gupta MM, Sanchez E, editors. *Approximate reasoning in decision analysis*. Amsterdam: North-Holland, 1982. p. 219–47.
- [5] Adlassnig K-P, Kolarz G. Representation and semi-automatic acquisition of medical knowledge in CADIAG-I and CADIAG-II. *Comp Biomed Res* 1986;19:63–79.
- [6] Adlassnig K-P, Kolarz G, Scheithauer W. Present state of the medical expert system CADIAG-II. *Methods Inf Med* 1985;24:13–20.
- [7] Adlassnig K-P, Kolarz G, Scheithauer W. Approach to a hospital-based application of a medical expert system. *Med Inform* 1986;11:205–23.
- [8] Adlassnig K-P, Kolarz G, Scheithauer W, Effenberger H, Grabner G. Computer-assisted medical diagnosis. *Comp Biol Med* 1985;15:315–35.
- [9] 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.
- [10] 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.

- [11] 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 Int Med* 1989;110:824–32.
- [12] 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.
- [13] Bergeron B. ILIAD: a diagnostic consultant and patient simulator. *MD Comput* 1991;8:46–53.
- [14] Bernelot Moens HJ, Van der Korst JK. Comparison of rheumatological diagnoses by a Bayesian program and by physicians. *Methods Inf Med* 1991;30:187–93.
- [15] Bernelot Moens HJ, Van der Korst JK. Development and validation of a computer program using Bayes' theorem to support diagnosis of rheumatic disorders. *Ann Rheum Dis* 1992;51:266–71.
- [16] Boegl K. Design and implementation of a web-based knowledge acquisition toolkit for medical expert consultation systems, Ph.D. Thesis. Austria: Technical University of Vienna, 1997.
- [17] Boegl K, Leitich H, Kolousek G, Rothenfluh TE, Adlassnig K-P. Clinical data interpretation in MEDFRAME/CADIAG-IV using fuzzy sets. *Biomed Eng Appl Basis Commun* 1996;8:488–95.
- [18] Boegl K, Rothenfluh TE, Leitich H, Adlassnig K-P. Knowledge acquisition in the fuzzy knowledge representation framework of a medical consultation system. *Artif Intell Med*, in press.
- [19] Hernandez C, Sancho JJ, Belmonte MA, Sierra C, Sanz F. Validation of the medical expert system RENOIR. *Comput Biomed Res* 1994;27:456–71.
- [20] Horn W. MESICAR—a medical expert system integrating causal and associative reasoning. *Appl Artif Intell* 1989;3:305–56.
- [21] Horn W. Utilizing detailed anatomical knowledge for hypothesis formation and hypothesis testing in rheumatological decision support. *Artif Intell Med* 1991;3:21–39.
- [22] Kingsland Jr. LC, Lindberg DAB, Sharp GC. AI/RHEUM: a consultant system for rheumatology. *J Med Syst* 1983;7:221–7.
- [23] Kolarz G, Adlassnig K-P. Problems in establishing the medical expert systems CADIAG-I and CADIAG-II in rheumatology. *J Med Syst* 1986;10:395–405.
- [24] Kolousek G. The system architecture of an integrated medical consultation system and its implementation based on fuzzy technology, Ph.D. Thesis. Austria: Technical University of Vienna, 1997.
- [25] Leitich H, Adlassnig K-P, Kolarz G. Development and evaluation of fuzzy criteria for the diagnosis of rheumatoid arthritis. *Methods Inf Med* 1996;35:334–42.
- [26] Leitich H, Adlassnig K-P, Schuh, C, Kolarz, G. Improving diagnostic performance with fuzzy logic: a demonstration with the medical expert system CADIAG-II. In: Trapp R, editor. *Proceedings of Cybernetics and Systems*. Singapore: World Scientific, 1994. p. 263–70.
- [27] Mathew B, Norris D, Hendry D, Waddell G. Artificial intelligence in the diagnosis of low back pain and sciatica. *Spine* 1988;13:168–72.
- [28] 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.
- [29] Miller RA, Pople HE, Myers JD. INTERNIST-I: an experimental computer-based diagnostic consultant for general internal medicine. *New Engl J Med* 1982;307:468–76.
- [30] Moser W, Adlassnig K-P. Consistency checking of binary relationships in a medical knowledge base. *Artif Intell Med* 1992;4:389–407.
- [31] Rothenfluh TE, Boegl K, Adlassnig K-P. Representation and acquisition of knowledge for a fuzzy medical consultation system. In: Szczepaniak PS, Lisboa PJ, Tsumoto S, editors. *Fuzzy systems in medicine*. Heidelberg: Springer, 2000. p. 636–51.
- [32] 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.
- [33] Schewe S, Schreiber MA. Stepwise development of a clinical expert system in rheumatology. *Clin Invest* 1993;71:139–44.
- [34] Spindelberger W, Grabner G. Ein Computerverfahren zur diagnostischen Hilfestellung. In: Fellinger K, editor. *Computer in der Medizin? Probleme, Erfahrungen, Projekte*. Verlag Brüder Hollinek, Wien, 1968. p. 189–221.
- [35] Warner HR. ILIAD: moving medical decision making into new frontiers. *Methods Inf Med* 1989;28:370–2.
- [36] Zadeh LA. Fuzzy Sets. *Inf Contr* 1965;8:338–53.