FUZZY SET THEORY IN MEDICINE

Klaus-Peter Adlassnig

June 1984
CP-84-22

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
2361 Laxenburg, Austria

PREFACE

In this paper, Klaus-Peter Adlassnig, a participant in the 1983 Young Scientists' Summer Program, shows that fuzzy set theory seems to be a suitable basis for the development of a computerized medical diagnosis and treatment-recommendation system. He describes a medical expert system of this type, CADIAG-2, developed at the University of Vienna, and outlines some results obtained during testing.

Decision making is often characterized by a high degree of fuzziness and uncertainty. This may reside in the imperfect and complex nature of human information processing and/or in the decision systems themselves. It may lie in the generation of possible options, the formation of criteria by which the options are judged, the prediction of the effects of possible decisions, and/or the level of understanding of the underlying processes.

This paper represents a contribution to research in the field of computerized decision support, and was carried out as part of the Interactive Decision Analysis Project.

ANDRZEJ WIERZBICKI
Chairman
System and Decision Sciences

ABSTRACT

Fuzzy set theory has a number of properties that make it suitable for formalizing the uncertain information upon which medical diagnosis and treatment is usually based.

Firstly, it allows us to define inexact medical entities as fuzzy sets. Secondly, it provides a linguistic approach with an excellent approximation to texts. Finally, fuzzy logic offers powerful reasoning methods capable of drawing approximate inferences.

These facts suggest that fuzzy set theory might be a suitable basis for the development of a computerized diagnosis and treatment-recommendation system. This is borne out by trials performed with the medical expert system CADIAG-2, which uses fuzzy set theory to formalize medical relationships.

FUZZY SET THEORY IN MEDICINE

Klaus-Peter Adlassnig

Department of Medical Computer Sciences, University of Vienna,
Garnisongasse 13, A-1090 Vienna, Austria

1 INTRODUCTION

It is widely accepted that the information available to the physician about his patient and about medical relationships in general is inherently uncertain. Nevertheless, the physician is still quite capable of drawing (approximate) conclusions from this information. This paper describes an attempt to provide a formal model of this process using fuzzy set theory, and implement it in the form of a computerized diagnosis and treatment-recommendation system.

In medicine, the principle of "Measuring everything measurable and trying to make measurable that which has not been measurable so far" (Galileo) is still practiced, although its fundamental limitations have been recognized during the course of this century. We now know that all real-world knowledge is characterized by:

- incompleteness (implying that the human process of cognition is infinite)
- inaccuracy (as stated in Heisenberg's Uncertainty Principle)
- inconsistency (anticipated by Gödel's Theorem).

Fuzzy set theory, which was developed by Zadeh [1], makes it possible to define inexact medical entities as fuzzy sets. It offers a linguistic approach which represents an excellent approximation to medical texts [2,3]. In addition, fuzzy logic provides powerful reasoning methods capable of making approximate inferences [4,5]. These facts suggest that fuzzy set theory might be a suitable basis for the development of a computerized diagnosis and treatment-recommendation system [6]. Tests carried out with the medical expert system CADIAG-2 [7–9] are described which show that this is indeed the case.

2 REAL-WORLD KNOWLEDGE

Precision exists only through abstraction. Abstraction may be defined as the ability of human beings to recognize and select the relevant properties of real-world phenomena and objects. This leads to the construction of conceptual models defining abstract classes of phenomena and objects. However, in actual fact every real-world phenomenon and object is of course unique.

Abstract models of real-world phenomena and objects such as mathematical structures (circle, point, etc.), equalities ($a = b + c$) and propositions (yes, no) are artificial constructs. They represent ideal structures, ideal equalities and ideal propositions.

Nevertheless, despite these caveats, abstraction forms the basis of human thought, and human knowledge is its result.

2.1 Incompleteness

Abstraction, however, is not a static concept. The process of abstraction is continuous and is constantly producing new results. The set of properties of real-world phenomena and objects under consideration is continually being enlarged and changed. Knowledge is therefore always and necessarily incomplete.

2.2 Inaccuracy

Unlimited precision is impossible in the real world. Anything said to be "precise" can only be considered as "precise to a certain extent".

The pursuit of maximum precision is still an important aim in science. Galileo, who is often credited with being the father of the quantitative scientific experiment, was certainly responsible for many scientific advances through his philosophy of "Measuring everything measurable and trying to make measurable that which has not been measured so far", although the limitations of this approach should be recognized.

Heisenberg's Uncertainty Principle [10] states the limits to accurate measurement very clearly. Of course, the Principle applies only to the world of microphenomena and microobjects, but its philosophical implications go further. It shows that nature is fundamentally indeterministic. And it seems meaningless to ask whether nature inherently lacks determinism or whether uncertainty stems only from experimentation.

2.3 Inconsistency

Abstraction does not always lead to the same results, which in turn are not always interpreted in the same way. "Knowledge" may differ according to nation, culture, religion, social status, education, etc., and information from different sources may therefore be inconsistent. To eliminate inconsistency from the information system is only possible in limited systems, and Gödel's Theorem [11] clearly demonstrates that contradictions within a system cannot be eliminated by the system itself.

3 MEDICAL INFORMATION

In medicine, it is not necessary to deal with microphenomena and microobjects to run into the problems of incompleteness, uncertainty and inconsistency. The lack of information, and its imprecise and sometimes contradictory nature, is much more a fact of life in medicine than in, say, the physical sciences. These problems have to be taken into account in every medical decision, where they may have important, even vital consequences for the object of medical attention, the patient.

3.1 Information about the patient

Data about the patient can be divided into a number of different categories that are all characterized by an inherent lack of certainty.

1. *Medical history of the patient*

   The medical history of the patient is given by the patient himself. It is highly subjective and may include simulated, exaggerated or understated symptoms. Ignorance of previous diseases in himself or his family, failure to mention previous operations and general poor recollection often raise doubts about a patient's medical history in the mind of the doctor. On the other hand, however, the information that finally leads to the correct diagnosis is very often found here.

2. *Physical examination*

   The physician subjects the patient to a physical examination from which he obtains more or less objective data. But of course, physicians can make mistakes, overlook important indications or fail to carry out a complete examination. Furthermore, they may misinterpret other indications because the boundary between normal and pathological status is not always clearly defined.

3. *Results of laboratory tests*

   The results of laboratory tests are considered to be objective data. However, measurement errors, organizational problems (mislabelling samples, sending them to the wrong laboratory, etc.) or improper behavior on the part of the patients prior to examinations can lead to imprecise and sometimes even totally incorrect data. Again, the boundaries between normal and pathological results are generally not strict: there are always borderline values that cannot be said to be either normal or pathological.

4. *Results obtained by histological, X-ray, ultrasonic examinations, etc.*

   These results again depend on correct interpretation by medical or other staff. Such findings are often crucial because they frequently indicate invasive therapy. In many cases, consideration of uncertainty is part of the evaluation procedure, for example in cell counts, cell determination, picture analysis, etc.

3.2 Information about medical relationships

Medical knowledge consists of medical descriptions and assertions that are incomplete and uncertain. It has been built up step by step, and is based partly on theoretical studies (in areas such as anatomy and physiology) and partly on almost purely empirical observations (made in the course of surgery, for example). Medical knowledge may be said to comprise knowledge about causal relationships based in theory, statistical information, pure definitions and personal judgement.

To add to the problem, the elements considered to form medical relationships differ according to place and time, vary between medical schools and in some cases have not been studied to any significant extent.

3.3 Medical inference

This is the process by which the physician uses his medical knowledge to infer a diagnosis from the symptoms displayed by the patient, his lab test results and medical history. It is a complex and almost uninvestigated process in which the physician is obviously able to work with uncertain and imprecise sets of data. To some extent it is a subconscious activity, which is why it is often called an art.

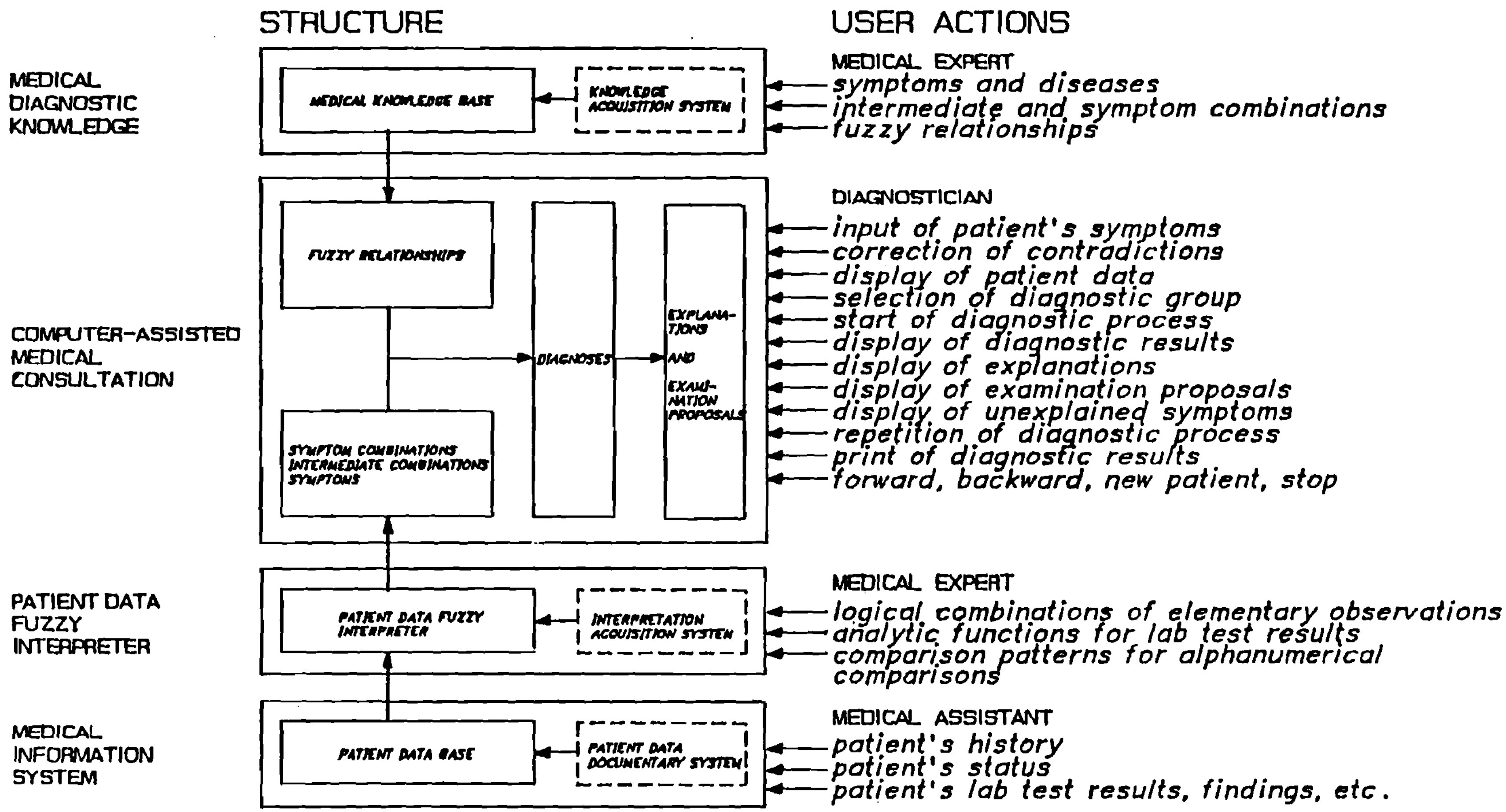


FIGURE 1: Structure of CADIAG-2 with connection to a medical information system
(dashed lines mark components effective before starting the consultation)

4 MEDICAL EXPERT SYSTEM CADIAG-2

CADIAG-2 (a Computer-Assisted DIAGnosis system) is intended to be an active assistant to the physician in diagnostic situations. In this way the experience, creativeness and intuition of the physician may be supplemented by the information-based computational power of the computer. The general structure of CADIAG-2 is shown in Figure 1.

4.1 Representation of medical information

CADIAG-2 considers four classes of medical entities:

- symptoms, indications, test results, findings (S_i)
- diseases, diagnoses (D_j)
- intermediate combinations (IC_k)
- symptom combinations (SC_l).

Symptoms S_i take values μ_{S_i} in $[0,1] \cup \phi$. The value μ_{S_i} indicates the degree to which the patient exhibits symptom S_i (a value of ϕ implies that symptom S_i has not yet been studied). In the language of fuzzy set theory, μ_{S_i} expresses the grade of membership of the patient's symptom manifestation S_i. An example of this mode of representation is given in Table 1.

A binary fuzzy relationship $R_{PS} \subset \Pi \times \Sigma$ is then established, defined by $\mu_{R_{PS}}(P_q, S_i) = \mu_{S_i}$ for patient P_q, where $P_q \in \Pi = \{P_1, \ldots, P_r\}$ and $S_i \in \Sigma = \{S_1, \ldots, S_m\}$.

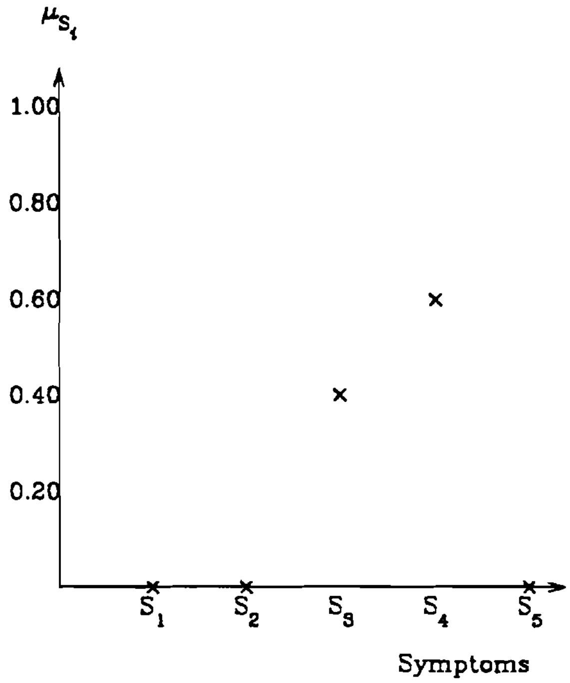
Diseases or diagnoses also take values in $[0,1] \cup \phi$. Fuzzy values $0.00 < \mu_{D_j} < 1.00$ represent possible diagnoses while the values $\mu_{D_j} = 1.00$ and $\mu_{D_j} = 0.00$ correspond to confirmed and excluded diagnoses, respectively. Diagnoses which have not yet been considered take the value $\mu_{D_j} = \phi$. Formally, a relationship $R_{PD} \subset \Pi \times \Delta$ is established, defined by $\mu_{R_{PD}}(P_q, D_j) = \mu_{D_j}$ for patient P_q, where $D_j \in \Delta = \{D_1, \ldots, D_n\}$.

Intermediate combinations (fuzzy logical combinations of symptoms and diseases) were introduced to model the pathophysiological states of patients; symptom combinations are combinations of symptoms, diseases and intermediate combinations. Both entities take their values μ_{IC_k} and μ_{SC_l} (respectively) in $[0,1] \cup \phi$, where ϕ implies that the actual value has not yet been determined.

Table 1. An example of the representation of medical knowledge.

Quantitative value		Symptom	Fuzzy value
		Potassium, greatly decreased	$\mu_{S_1} = 0.00$
		Potassium, decreased	$\mu_{S_2} = 0.00$
Measured potassium level of 5.3 mmol/l	Fuzzy interpreter →	Potassium, normal	$\mu_{S_3} = 0.40$
		Potassium, increased	$\mu_{S_4} = 0.60$
		Potassium, greatly increased	$\mu_{S_5} = 0.00$

5.3 — Fuzzy interpreter →



The relationship $R_{PSC} \subset \Pi \times K$ is defined by $\mu_{R_{PSC}}(P_q, SC_l) = \mu_{SC_l}$ for patient P_q, where $SC_l \in K = \{SC_1, \ldots, SC_t\}$ formally describes the symptom combinations observed in the patient (both the presence and absence of symptoms are regarded as observations).

The fuzzy logical connectives are defined as follows:

Conjunction:

$$x_1 \wedge x_2 = \begin{cases} \min\ (x_1, x_2) & \text{if } x_1 \in [0,1] \text{ and } x_2 \in [0,1] \\ \phi & \text{if } x_1 = \phi \text{ and/or } x_2 = \phi \end{cases}$$

Disjunction:

$$x_1 \vee x_2 = \begin{cases} \max\ (x_1, x_2) & \text{if } x_1 \in [0,1] \text{ and } x_2 \in [0,1] \\ x_1 & \text{if } x_1 \in [0,1] \text{ and } x_2 = \phi \\ x_2 & \text{if } x_1 = \phi \text{ and } x_2 \in [0,1] \\ \phi & \text{if } x_1 = \phi \text{ and } x_2 = \phi \end{cases}$$

Negation:

$$\bar{x}_1 = \begin{cases} 1 - x_1 & \text{if } x_1 \in [0,1] \\ \phi & \text{if } x_1 = \phi \end{cases}$$

The following relationships between medical entities are considered in CADIAG-2:

- symptom–disease relationships ($S_i D_j$)
- symptom combination–disease relationships ($SC_l D_j$)
- symptom–symptom relationships ($S_i S_j$)
- disease–disease relationships ($D_i D_j$).

These relationships are characterized by two parameters:

- frequency of occurrence (o)
- strength of confirmation (c).

For a relationship between medical entities X and Y (where X and Y may be symptoms, diseases or symptom combinations), the frequency of occurrence describes the frequency with which X occurs when Y is present. Similarly, the strength of confirmation reflects the degree to which the presence of X implies the presence of Y.

The relationships between medical entities are given in the form of relationship rules with associated relationship tupels. The general formulation of these rules is:

IF (premise) THEN (conclusion) WITH (o, c) .

The relationship tupels (o,c) contain either numerical fuzzy values μ_o and μ_c or linguistic fuzzy values λ_o and λ_c, or both [3].

The definitions of the linguistic values λ_o and λ_c, the fuzzy intervals that they cover and their representative numerical values are given in Table 2. Representative numerical values are necessary in order to make fuzzy inferences possible (see Section 4.2). The way in which the linguistic fuzzy values, the fuzzy numerical intervals and their representative numerical values were chosen is described in more detail in refs. 8 and 9. Some examples of relationship rules are given below.

Table 2. Linguistic fuzzy values, numerical intervals and representative numerical values describing frequency of occurrence and strength of confirmation.

Frequency of occurrence			Strength of confirmation		
Value λ_o	Interval	Representative value μ_o	Value λ_c	Interval	Representative value μ_c
Always	[1.00,1.00]	1.00	Always	[1.00,1.00]	1.00
Almost always	[0.99,0.98]	0.99	Almost always	[0.99,0.98]	0.99
Very often	[0.97,0.83]	0.90	Very strong	[0.97.0.83]	0.90
Often	[0.82,0.68]	0.75	Strong	[0.82,0.68]	0.75
Medium	[0.67,0.33]	0.50	Medium	[0.67,0.33]	0.50
Seldom	[0.32,0.18]	0.25	Weak	[0.32,0.18]	0.25
Very seldom	[0.17,0.03]	0.10	Very weak	[0.17,0.03]	0.10
Almost never	[0.02,0.01]	0.01	Almost never	[0.02,0.01]	0.01
Never	[0.00,0.00]	0.00	Never	[0.00,0.00]	0.00
Unknown	ϕ	ϕ	Unknown	ϕ	ϕ

Example 1

IF (ultrasonic of pancreas is pathological)

THEN (pancreatic carcinoma)

WITH (0.75 = often, 0.25 = weak)

Example 2

IF (tophi)

THEN (gout)

WITH (0.25 = seldom, 1.00 = always)

Example 3

IF (lower back pain $\wedge$ limitation of motion of the lumbar spine $\wedge$ diminished chest expansion $\wedge$ male patient $\wedge$ age between 20 and 40 years)

THEN (ankylosing spondylitis)

WITH (–, 0.90 = very strong)

The values μ_o and μ_c are interpreted as the values of the fuzzy relationships between premises and conclusions:

$S_i D_j$ (occurrence relationship)	$R^o_{SD} \subset \Sigma \times \Delta$
$S_i D_j$ (confirmation relationship)	$R^c_{SD} \subset \Sigma \times \Delta$
$SC_l D_j$ (occurrence relationship)	$R^o_{SCD} \subset K \times \Delta$
$SC_l D_j$ (confirmation relationship)	$R^c_{SCD} \subset K \times \Delta$
$S_i S_j$ (occurrence relationship)	$R^o_{SS} \subset \Sigma \times \Sigma$
$S_i S_j$ (confirmation relationship)	$R^c_{SS} \subset \Sigma \times \Sigma$
$D_i D_j$ (occurrence relationship)	$R^o_{DD} \subset \Delta \times \Delta$
$D_i D_j$ (confirmation relationship)	$R^c_{DD} \subset \Delta \times \Delta$

4.2 Fuzzy logical inference

The compositional inference rule proposed by Zadeh [4] and introduced into medical diagnosis by Sanchez [12,13] is adopted as an inference mechanism. It accepts fuzzy descriptions of the patient's symptoms and infers fuzzy descriptions of the patient's condition by means of the fuzzy relationships described in the previous section.

Three such inference rules (compositions) are used to deduce the diseases D_j suffered by patient P_q from the observed symptoms S_i:

1. Composition for $S_i D_j$ confirmation:

$$R^1_{PD} = R_{PS} \circ R^c_{SD} \qquad (1)$$

defined by

$$\mu_{R^1_{PD}}(P_q, D_j) = \max_{S_i} \min \left[\mu_{R_{PS}}(P_q, S_i); \mu_{R^c_{SD}}(S_i, D_j)\right] .$$

2. Composition for $S_i D_j$ non-confirmation:

$$R^2_{PD} = R_{PS} \circ (1 - R^c_{SD}) \tag{2}$$

defined by

$$\mu_{R^2_{PD}}(P_q, D_j) = \max_{S_i} \min \left[\mu_{R_{PS}}(P_q, S_i);\ 1 - \mu_{R^c_{SD}}(S_i, D_j)\right]$$

3. Composition for $S_i D_j$ without symptoms:

$$R^3_{PD} = (1 - R_{PS}) \circ R^o_{SD} \tag{3}$$

defined by

$$\mu_{R^3_{PD}}(P_q, D_j) = \max_{S_i} \min \left[1 - \mu_{R_{PS}}(P_q, S_i);\ \mu_{R^o_{SD}}(S_i, D_j)\right]$$

The following diagnostic results are obtained:

- a diagnosis is confirmed if

$$\mu_{R^1_{PD}}(P_q, D_j) = 1.00 \tag{4}$$

- a diagnosis is possible if

$$0.10 \leq \mu_{R^1_{PD}}(P_q, D_j) \leq 0.99 \tag{5}$$

The boundary value 0.10 is a heuristic value which rejects diagnoses with very low evidence.

- a diagnosis is excluded if

$$\mu_{R^2_{PD}}(P_q, D_j) = 1.00 \tag{6}$$

or

$$\mu_{R^3_{PD}}(P_q, D_j) = 1.00 \tag{7}$$

Symptom combination–disease inferences (compositions 4,5 and 6) are carried out and interpreted in an analogous way. Symptom–symptom inferences (compositions 7, 8 and 9) are computed in order to complete the patient's symptom patterns. Disease–disease inferences (compositions 10, 11 and 12) are also performed in order to confirm the underlying disease from the presence of the secondary complaints or to exclude entire areas of secondary complaints if a particular primary disease is absent.

4.3 Acquisition of medical knowledge

The knowledge acquisition system is capable of acquiring information on medical entities and the relationships between them. In CADIAG-2, relationships are stored as numerical fuzzy values in the range [0,1]. Medical information can be acquired in two ways:

- through linguistic evaluation by medical experts
- by statistical evaluation of a data base containing medical data on patients with confirmed diagnoses.

Information on relationships can be gathered linguistically using predefined linguistic values to determine parameters such as frequency of occurrence o and strength of confirmation c (cf. Table 2). Empirical, judgemental and definitive knowledge may be acquired in this way.

CADIAG-2 relationships have the important property that they may be interpreted statistically. The values of the frequency of occurrence μ_o and the strength of confirmation μ_c may be defined as follows:

$$\mu_o = \frac{F(S_i \cap D_j)}{F(D_j)} = F(S_i / D_j) \tag{8}$$

$$\mu_c = \frac{F(S_i \cap D_j)}{F(S_i)} = F(D_j / S_i) \quad , \tag{9}$$

where

$F(S_i \cap D_j)$ – absolute frequency of occurrence of S_i and D_j

$F(D_j)$ – absolute frequency of occurrence of D_j

$F(S_i)$ – absolute frequency of occurrence of S_i

$F(S_i / D_j)$ – conditional frequency of S_i given D_j

$F(D_j / S_i)$ – conditional frequency of D_j given S_i.

With definitions (8) and (9), extended statistical evaluations of known medical relationships or as yet unidentified relationships can be carried out using data on patients with confirmed diagnoses.

4.4 The diagnostic process

4.4.1 Symptoms

The symptoms of the patient can be entered into CADIAG-2 in three ways (described in detail in [9]):

(i) by natural language input of symptoms S_i

(ii) by natural language input of keywords that trigger whole groups of symptoms S_i

(iii) by accessing a data base containing the patient's data and transferring information via a fuzzy interpreter.

Natural language input of symptoms S_i such as "high fever", "increased GOT" or "blood stool positive" is achieved by a symptom search algorithm with an embedded word segmentation algorithm that allows the use of synonyms and abbreviations, orthographic variants and different parts of speech.

Input of keywords such as "present complaints", "previous complaints", "blood count" and "ultrasonic" causes whole sections of the symptom thesaurus to be displayed. Subsequently, fuzzy values can be linked with these symptoms by the physician.

The existence of a data base which already contains the patient's symptoms suggests the automatic transfer of information from the data base to CADIAG-2. During this transfer, the data is passed through a fuzzy interpreter which contains instructions about the assignment of fuzzy values to observations, lab test results and even simple alphanumeric texts.

After the patient's symptoms have been collected, symptom—symptom inferences are performed. The symptom list contains all necessary items of data, including fuzzy value, origin (measured; inferred), predefined symptom class (routine; specially requested; invasive or expensive), numerical value, units and date of observation. The list of symptoms is then checked for contradictions.

4.4.2 Symptom combinations

Intermediate combinations of symptoms are evaluated in the next step. Having passed the consistency check, fuzzy values for all symptom combinations are computed. The resulting lists are now as complete as possible and do not contain any contradictions.

4.4.3 Confirmed diagnoses

The fuzzy values $\mu_{D_j} = 1.00$, i.e., confirmed diagnoses D_j for patient P_q, are identified using the following equation:

$$\mu_{D_j} = 1.00 \text{ if } \begin{cases} \mu_{R^1_{PD}}(P_q, D_j) = 1.00 \\ \text{or} \\ \mu_{R^4_{PD}}(P_q, D_j) = 1.00 \end{cases} \tag{10}$$

4.4.4 Excluded diagnoses

The fuzzy values $\mu_{D_j} = 0.00$, i.e., excluded diagnoses D_j for patient P_q, are identified using:

$$\mu_{D_j} = 0.00 \text{ if } \begin{cases} \mu_{R^2_{PD}}(P_q, D_j) = 1.00 \\ \text{or} \\ \mu_{R^3_{PD}}(P_q, D_j) = 1.00 \\ \text{or} \\ \mu_{R^5_{PD}}(P_q, D_j) = 1.00 \\ \text{or} \\ \mu_{R^6_{PD}}(P_q, D_j) = 1.00 \end{cases} \tag{11}$$

Disease–disease relationships now allow the inference of further diagnoses (confirmed or excluded):

$$\mu_{D_j} = \begin{cases} 1.00 \text{ if } \mu_{R^{10}_{PD}}(P_q, D_j) = 1.00 \\ 0.00 \text{ if } \begin{cases} \mu_{R^{11}_{PD}}(P_q, D_j) = 1.00 \\ \text{or} \\ \mu_{R^{12}_{PD}}(P_q, D_j) = 1.00 \end{cases} \end{cases} \tag{12}$$

4.4.5 Possible diagnoses

Method 1. Fuzzy values μ_{D_j} such that $0.10 \le \mu_{D_j} \le 0.99$ indicate possible diagnoses. These are determined as follows:

$$\mu_{D_j} = \max[\mu_{R^1_{PD}}(P_q,D_j); \mu_{R^4_{PD}}(P_q,D_j); \mu_{R^{10}_{PD}}(P_q,D_j)] \text{ if } \begin{cases} 0.10 \le \mu_{R^1_{PD}}(P_q,D_j) \le 0.99 \\ \text{and/ or} \\ 0.10 \le \mu_{R^4_{PD}}(P_q,D_j) \le 0.99 \\ \text{and/ or} \\ 0.10 \le \mu_{R^{10}_{PD}}(P_q,D_j) \le 0.99 \end{cases} \quad (13)$$

Method 2. Because the value μ_{D_j} calculated by (13) is independent of the rules that can be used to define D_j, a powerful heuristic function is introduced which considers the number of criteria present which suggest but do not confirm disease D_j, and then calculates the corresponding number of points PN_{D_j}. The values of PN_{D_j} are helpful in judging between the various possible diagnoses, although the ultimate aim should be to obtain a confirmed diagnosis. The number of points PN_{D_j} is calculated as follows:

$$PN_{D_j} = \sum_{i=1}^{m^*} [\alpha \mu_{R^o_{SD}}(S_i,D_j) + \beta \mu_{R^c_{SD}}(S_i,D_j)] \quad . \qquad (14)$$

where m^* is the number of symptoms exhibited by the patient that occur in the definition of D_j, and $\alpha + \beta = 1.00$. We generally take $\alpha = 0.09$ and $\beta = 0.91$, i.e., the strength of confirmation has ten times more influence than the frequency of occurrence on the value of PN_{D_j}.

4.4.6 Explanation of diagnostic results

The physician's acceptance of CADIAG's diagnoses depends strongly on the ability of CADIAG-2 to explain its diagnostic output. On request, the information supporting confirmed diagnoses, excluded diagnoses and possible diagnoses is presented; this takes the form of the names of the medical entities, their definitions, their measured and fuzzy values, and their relationships to the diagnostic output.

4.4.7 Proposals for further examination of the patient

One of the main objectives of CADIAG-2 is to provide iterative consultations, starting with simple, easy-to-examine and cheap data. A number of possible diagnoses can usually be inferred from these data, and further examinations are then necessary to confirm or exclude these hypotheses. CADIAG-2 uses the medical information stored in its data bank to propose what form these further examinations should take. The symptoms selected for further study are clearly those which would confirm or exclude a particular diagnosis.

4.4.8 Unexplained symptoms

The confirmed diagnoses and any remaining possible diagnoses should together explain any pathological symptom, indication or lab test result of the patient. Unexplained data (usually) indicates further diseases that should be investigated.

5 RESULTS

5.1 Rheumatic diseases

CADIAG-2/RHEUMA has undergone partial tests with data from patients at a rheumatological hospital. A study of 169 patients with rheumatoid arthritis, Sjögren's disease, systemic lupus erythematodes, Reiter's disease or sclerodermia showed that CADIAG-2 obtained the correct diagnosis in 77.16% of the cases considered. This figure was calculated by comparing the clinical diagnoses established by the consultant at the rheumatological hospital (assumed to be correct) with the confirmed diagnoses made by CADIAG-2. Most of the cases in which clinical diagnoses could not be confirmed fell into two classes:

(i) The patient was in hospital only temporarily to check the efficacy of drugs already administered

(ii) The patient was in the early stages of one of the rheumatic diseases considered; in almost all of these cases a possible diagnosis was suggested.

5.2 Pancreatic diseases

CADIAG-2/PANCREAS was tested with data from 31 patients. The final clinical diagnoses of these patients had not been confirmed by histological examination, but were nevertheless assumed to be correct.

Pancreatic carcinoma was confirmed twice. Confirmation was aided by the existence of a result "Specific abnormal pancreatic biopsy", which has a strength of confirmation μ_c = 1.00 for pancreatic carcinoma.

Possible hypotheses were generated for the other cases, and the heuristically determined number of points was taken as the basis for evaluation. The results are given in Table 3.

Table 3. Comparison of CADIAG-2 possible diagnoses with the clinical diagnoses.

Clinical diagnosis	Percentage of cases
CADIAG diagnosis with highest number of points	50.0
CADIAG diagnosis with second highest number of points	21.4
CADIAG diagnosis with third highest number of points	10.8
CADIAG diagnosis with fourth highest number of points	7.0
No CADIAG diagnosis	10.8

ACKNOWLEDGEMENTS

The author gratefully acknowledges the contributions of G. Kolarz, M.D., and W. Scheithauer, M.D., in the medical documentation of rheumatic and pancreatic diseases.

REFERENCES

1. L.A. Zadeh. Fuzzy sets. *Information and Control* 8 (1965) 338–353.
2. L.A. Zadeh. A fuzzy-algorithmic approach to the definition of complex or imprecise concepts. In H. Bossel, S. Klaczko and N. Müller (Eds.), *Systems Theory in the Social Sciences*. Birkhäuser Verlag, Basel and Stuttgart, 1976, pp. 202–282.
3. L.A. Zadeh. Linguistic variables, approximate reasoning and dispositions. *Med. Inform.* 8 (1983) 173–186.
4. L.A. Zadeh. Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Transactions on Systems, Man, and Cybernetics* 3 (1973) 28–44.
5. R.E. Bellman and L.A. Zadeh. *Local and Fuzzy Logics*. Memorandum No. ERL-M584, Electronics Research Laboratory, College of Engineering, University of California, Berkeley 94720, 11 May 1976.
6. K.-P. Adlassnig. A survey on medical diagnosis and fuzzy subsets. In M.M. Gupta and E. Sanchez (Eds.), *Approximate Reasoning in Decision Analysis*. North-Holland Publishing Company, Amsterdam, New York and Oxford, 1982, pp. 203–217.
7. K.-P. Adlassnig. A fuzzy logical model of computer-assisted medical diagnosis. *Meth. Inform. Med.* 19 (1980) 141–148.

8. K.-P. Adlassnig and G. Kolarz. CADIAG-2: computer-assisted medical diagnosis using fuzzy subsets. In M.M. Gupta and E. Sanchez (Eds.), *Approximate Reasoning in Decision Analysis*. North-Holland Publishing Company, Amsterdam, New York and Oxford, 1982, pp. 219–247.

9. K.-P. Adlassnig. *Ein Computerunterstütztes Medizinisches Diagnosesystem unter Verwendung von Fuzzy Teilmengen*. Ph.D. Dissertation, Technical University of Vienna, Vienna, 1983.

10. *Encyclopedia Britannica*. Encyclopedia Britannica Inc., Chicago, 1975.

11. *Lexicon der Kybernetik*. Akademie-Verlag, Berlin, 1981.

12. E. Sanchez. Compositions of fuzzy relations. In M.M. Gupta, R.K. Ragade and R.R. Yager (Eds.), *Advances in Fuzzy Set Theory and Applications*. North-Holland Publishing Company, Amsterdam, New York and Oxford, 1979, pp. 421–433.

13. E. Sanchez. Medical diagnosis and composite fuzzy relations. In M.M. Gupta, R.K. Ragade and R.R. Yager (Eds.), *Advances in Fuzzy Set Theory and Applications*. North-Holland Publishing Company, Amsterdam, New York and Oxford, 1979, pp. 437–444.