Journal of Physical Sciences, Vol. 20, 2015, 7-18
ISSN: 2350-0352 (print), www.vidyasagar.ac.in/journal
Published on 24 December 2015

Fuzzy Set Theory in Real-World Knowledge and Medical Diagnosis Process

Utpalendu Dev[1], Abeda Sultana[2] and Nirmal Kanti Mitra[3]

[1]Department of Mathematics, Govt. Mohila College,Narayanganj, Dhaka.
E-mail: utpalendudev@yahoo.com
[2]Department of Mathematics, Jahangirnagar University, Savar, Dhaka.
[3]Department of Mathematics, Bangladesh University of Business & Technology, Dhaka, Bangladesh

Received 31 August 2015; accepted 7 November 2015

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.

Keywords**:** Phenomenon, pathophysiological, premise, lingustic, ultrasonic, pancreas, carcinoma, lumbar, alphanumaric, lupus erythematodes, sclerodermia.

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 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 Godel's Theorem)

These facts suggest that fuzzy set theory might be a suitable basis for the development of a computerized diagnosis and treatment recommendation system [1]. Tests carried out with the medical expert system CADIAG- 2 [2, 3] 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. 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 proposition (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 [4] states the limits to accurate measurement very clearly of course, the principle applies only to the world of micro phenomena and micro objects, but its philosophical implications go further. It shows that nature is fundamentally in deterministic. 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 Godel’s theorem [5] clearly demonstrates that contradictions within a system cannot be eliminated by the system itself.

3. 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.

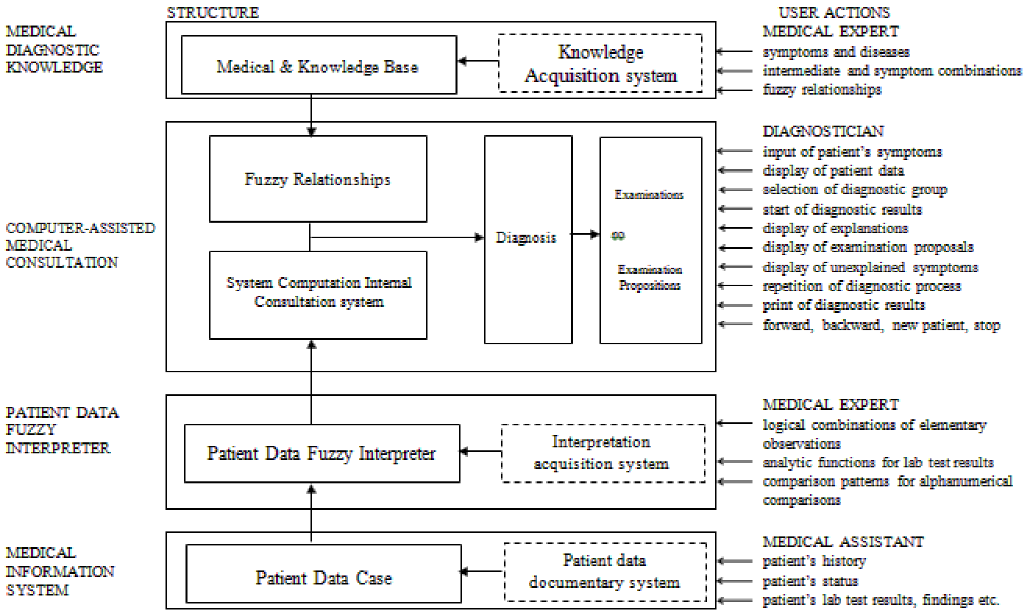


Figure 3 (1): Structure of CADIAG- 2 with connection to a medical information system (dashed lines mark components effective before starting the consultation)

3.1. Representation of medical information

CADIG- 2 Considers four classes of medical entities :

- Symptoms, indications, test results, findings (S_i)
- diseases, diagnoses (D_j)
- intermediate combinations (IC_k)
- System 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.

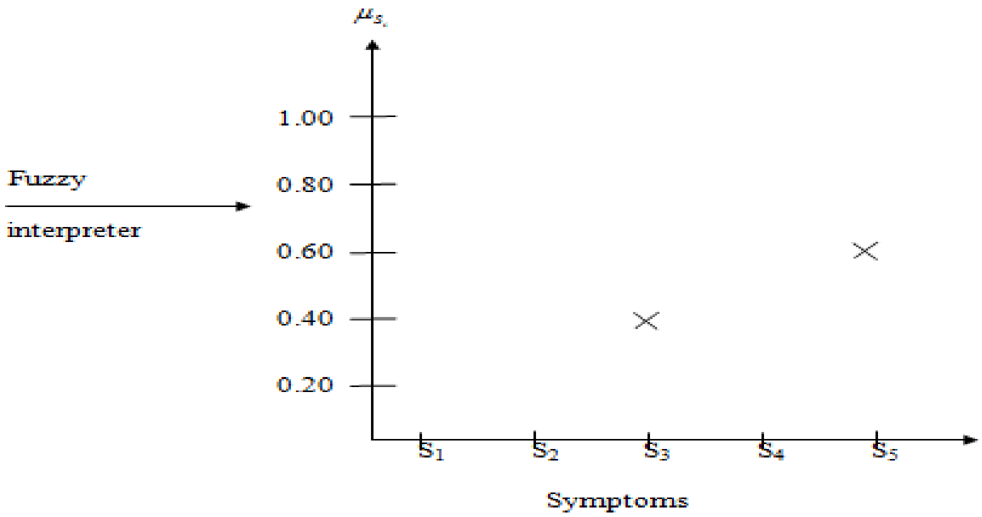
A binary fuzzy relationship $R_{PS} \subset \Pi \times \Sigma$ is then established, defined by $\mu_{R_{PS}}(P_q, S_i) = \mu_{S_i}$ for patient (P_q) where $P_q \in \Pi = \{P_1, \ldots\ldots, P_r\}$ and $S_i \in \Sigma = \{S_1, \ldots\ldots\ldots, S_m\}$.

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 $D_j \in \Delta = \{D_1, \ldots\ldots, D_n\}$.
Intermediate combinations (fuzzy logical combinations of symptoms and diseases) were introduced to model the pathophysiological states of patients: symptoms 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 3.1 (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/1.	Fuzzy → interpreter	Potassium, normal	$\mu_{S_3} = 0.40$
		Potassium, increased	$\mu_{S_4} = 0.60$
		Potassium, greatly increased	$\mu_{S_5} = 0.00$



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 $SC_l \in K = \{SC_1, ..., SC_l\}$ formally describes the symptom combinations observed in 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 :

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

Disjunction :

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

Negation :

$$\overline{x_1} = \begin{cases} 1 - x_1 \; if \;\; x_1 \in [0,1] \\ \phi \; \text{if} \; x_1 = \phi \end{cases}$$

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

- symptom- disease relationships (S_iD_j)
- symptom combination- disease relationships (SC_iD_j)
- symptom- symptom relationships (S_iS_j)
- diseasedisease relationships (D_iD_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 $\mu_\circ$ and μ_c or linguistic fuzzy values λ_o and λ_c, or both [6].

The difinitions of the lingustic 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 (CADIAG- 1). 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. [7, 3]. Some examples of relationship rules are given below.

Table- 3.1 (2).Linguistic fuzzy values, numerical intervals and representative numerical values describing frequency of occurrence and strength of confirmation.

Frequency of occurrence

Value λ_o	**Interval**	**Representative Value** μ_o
Always	[1.00, 1.00]	1.00
Almost Always	[0.99, 0.98]	0.99
Very often	[0.97, 0.83]	0.90
Often	[0.82, 0.68]	0.75
Medium	[0.67, 0.33]	0.50
Seldom	[0.32, 0.18]	0.25
Very Seldom	[0.17, 0.03]	0.10
Almost never	[0.02, 0.01]	0.01
Never	[0.00, 0.00]	0.00
Unknown	ϕ	ϕ

Stregth of Confirmation

Value λ_c	**Interval**	**Representative Value** μ_c
Always	[1.00, 1.00]	1.00
Almost Always	[0.99, 0.98]	0.99
Very strong	[0.97, 0.83]	0.90
Strong	[0.82, 0.68]	0.75
Medium	[0.67, 0.33]	0.50
Weak	[0.32, 0.18]	0.25
Very Weak	[0.17, 0.03]	0.10
Almost never	[0.02, 0.01]	0.01
Never	[0.00, 0.00]	0.00
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 ^ limitation of motion of the lumbar spine ^ diminished chest expansion ^ male patient ^ age between 20 and 40 years)
THEN (ankylosing spondylitis)
WITH (-, 0.90 = very strong)

The values $\mu_{\circ}$ and μ_c are interpreted as the values of the fuzzy relationships between premises and conclusions:

S_iD_j (occurrence relationship) $R^o{}_{SD} \subset \Sigma \times \Delta$
S_iD_j (confirmation relationship) $R^c{}_{SD} \subset \Sigma \times \Delta$
SC_iD_j (occurrence relationship) $R^{\circ}{}_{SCD} \subset k \times \Delta$
SC_iD_j (confirmation relationship) $R^c{}_{SCD} \subset k \times \Delta$
S_iS_j (occurrence relationship) $R^{\circ}{}_{SS} \subset \Sigma \times \Sigma$
S_iS_j (confirmation relationship) $R^c{}_{SS} \subset \Sigma \times \Sigma$
D_iD_j (occurrence relationship) $R^{\circ}{}_{DD} \subset \Delta \times \Delta$
D_iD_j (confirmation relationship) $R^c{}_{DD} \subset \Delta \times \Delta$

3.2. Fuzzy logical inference

The compositional inference rule proposed [8] and introduced into medical diagnosis [9, 10] is adopted as an inference mechanism. It accepts fuzzy descriptions of the patient's symptoms and infers fuzzy descriptions of the fuzzy relationships described in the previous section.

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

1. Composition for $S_iD_j$confirmation:

$$R^1{}_{PD} = R_{PS} \ o \ R^c{}_{SD} \tag{1}$$

defined by

$$\mu_{R^1{}_{PD}}(P_q.D_j) = \overset{\max\min}{S_i} [\mu_{R_{PS}}(P_q, S_i); \mu_{R^c{}_{SD}}(S_i, D_j)]$$

2. Composition for S_iD_j non-confirmation:

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

defined by

$$\mu_{R^2{}_{PD}}(P_q.D_j) = \overset{\max\min}{S_i} [\mu_{R_{PS}}(P_q, S_i); 1 - \mu_{R^c{}_{SD}}(S_i, D_j)]$$

3. Composition for S_iD_j without symptoms :

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

defined by

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

The following diagnostic results obtained:
* a diagnosis is confirmed if

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

* a diagnosis is possible if

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

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.

3.3. 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 relations 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, judgmental 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_0 and the strength of confirmation μ_c may be defined as follows :

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

$$\mu_{c} = \frac{F(S_i \cap D_j)}{F(S_i)} = F(D_j / S_i) \tag{9}$$

where

F $(S_j \cap D_j)$ --- 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 givenD_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.

3.4. The diagnostic process

3.4.1. Symptoms

The symptoms of the patient can be entered into CADIAG- 2 in three ways described in detail in [3]:
(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.

3.4.2. Symptom combinations

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

3.4.3. 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:

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

3.4.4. Excluded diagnoses

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

$$\mu_{D_j} = 0.00 \text{ if } \begin{cases} \mu_{R^2_{PD}}(P_q, D_j) = 1.00 \\ \text{Or} \\ \mu_{R^3_{PD}}(P_q, D_j) = 1.00 \\ \text{Or} \\ \mu_{R^{11}_{PD}}(P_q, D_j) = 1.00 \\ \text{Or} \\ \mu_{R^{12}_{PD}}(P_q, D_j) = 1.00 \end{cases} \tag{11}$$

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

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

3.4.5. 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:

$$\mu_{D_j} = \max[\mu_{R^1_{PD}}(P_q, D_j); \mu_{R^4_{PD}}(P_q, D_j); \mu_{R^{10}_{PD}}(P_q, D_j)]$$

$$\text{if} \quad \begin{cases} 0.10 \le \mu_{R^1{}_{PD}}(Pq, Dj) \le 0.99 \\ and\,/\,or \\ 0.10 \le \mu_{R^4{}_{PD}}(Pq, Dj) \le 0.99 \\ and\,/\,or \\ 0.10 < \mu_{R^{10}{}_{PD}}(Pq, Dj) \le 0.99 \end{cases} \qquad (13)$$

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 suggests 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 confirmed diagnoses. The number of points PN_{D_j} is calculated as follows :

$$PN_{D_j} = \sum_{i=1}^{m} [\alpha \mu_{R^{\circ}{}_{SD}}(S_i, D_j) + \beta \mu_{R^{c}{}_{SD}}(S_i, D_j)] \qquad (14)$$

where m is the number of symptoms exhibited by the patient that occur in the definition of D_j, and $\alpha + \beta = 1.00$ $\alpha + \beta = 1.00$. We generally take $\alpha = 0.9$ 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. Results

4.1. Rheumatic diseases

CADIAG- 2 / RHEUMA has undergone partial tests with data from patients at a rheumatological hospital. A study of 69 patients with rheumatoid arthritis, Sjogren'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 disease considered; in almost all of these cases a possible diagnosis was suggested.

4.2. Pancreatic disease

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 $\mu_c = 1.00$ 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.

5. Conclusion

The use of fuzzy approaches has been successful in our work so far, which has addressed particularly difficult problems in the medical field involving classification and perception by experts of uncertain measured parameters, and visual and linguistic information. We see future directions developing fundamental methods such as supervised learning of type-2 (linguistic) fuzzy sets and exploring their applicability in the very rich and important area of medical diagnosis and analysis.

REFERENCES

1. 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 Comany. Amsterdam, New York and Oxford*, (1982) 203-217.
2. K.P. Adlassnig, A fuzzy logical model of computer-assisted medical diagnosis. *Meth. Infrom. Med.,* 19 (1980) 141-148.
3. K.P. Adlassnig. Ein. *Computerunterstutztes Medizinisches Diagnosesystem unter Verwendung von Fuzzy Teilmengen*, Ph. D. Dissertation. Technical University of Vienna, Vienna, 1983.
4. F.Steimann and K.P.Adlassnig, *Fuzzy medical diagnosis*, 2001. *http: //citeseer.nj.nec.com/160037.html.*
5. L.D. Kybernetik, *Akademie-Verlag*,Berlin, 1981.
6. L.A. Zadeh, Linguistic variables, approximate reasoning and dispositions. *Med. Inform,* 8 (1983) 173-186.
7. 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) 219-247.
8. L.A. Zadeh, Outline of a new approach to the analysis of complex systems and decision processes, *IEEE Transactions on Systems, Man. and Cybernettcs*, 3 (1973) 28-44.
9. 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) 421-433.
10. 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) 437-444.