UDC 681.3 : 616—07

A Fuzzy Logical Model of Computer-Assisted Medical Diagnosis

(From the Department of Medical Computer Science (Director: Prof. Dr. G. Grabner), University of Vienna, Austria)

K.-P. Adlassnig

A model of a computer-assisted diagnostic system using fuzzy subsets has been developed. The physician documents symptom—diagnosis presence relationships and symptom—diagnosis conclusiveness relationships by means of labels of the fuzzy subsets *never*, *almost never*, *very very seldom*, *very seldom*, *seldom*, *more or less seldom*, *not known*, *more or less often*, *often*, *very often*, *very very often*, *almost always*, *always*.

Symptoms are regarded as fuzzy subsets of reference sets. The reference set includes all values the symptom may assume. The degree of membership of a value in the fuzzy subset of a symptom is calculated when the patient's symptom pattern is available. By means of compositions of fuzzy relations, four different diagnostic indications are determined for every diagnosis under consideration: *presence indication*, *conclusiveness indication*, *non-presence indication* and *non-symptom presence indication*.

By performing the diagnostic process, the system provides the physician with *proven diagnoses*, *excluded diagnoses* and *diagnostic hints*, including *reasons* for the diagnoses displayed. *Proposals* for further investigations may also be requested.

Key-Words: Computer-assisted Diagnosis, Fuzzy Subsets, Medical Documentation, Diagnostic Hints, Symptom—Diagnosis Relationships, Proven and Excluded Diagnoses

EIN MODELL ZUR COMPUTERUNTERSTÜTZTEN DIAGNOSE MIT FUZZY LOGIK

Es wurde ein computerunterstütztes medizinisches Diagnosemodell unter Verwendung von Fuzzy-Teilmengen entwickelt. Die medizinische Vorarbeit besteht in der Dokumentation des Vorhandenseins eines Symptoms bei einer Diagnose und der Beweiskraft eines Symptoms für eine Diagnose. Zur Dokumentation verwendet der Mediziner die Bezeichnung der Fuzzy-Teilmengen *nie*, *fast nie*, *sehr sehr selten*, *sehr selten*, *selten*, *mehr oder weniger selten*, *unbekannt*, *mehr oder weniger oft*, *oft*, *sehr oft*, *sehr sehr oft*, *fast immer*, *immer*.

Die Symptome werden als Fuzzy-Teilmengen von Bezugsmengen, die alle Werte, die das Symptom annehmen kann, enthalten, betrachtet. Der Grad der Zugehörigkeit eines Wertes zu diesem Symptom wird dann ermittelt, wenn das Symptomenmuster des Patienten verfügbar ist. Mittels der Komposition von Fuzzy-Relationen werden vier verschiedene Fuzzy-Hinweise für jede betrachtete Diagnose errechnet: *Vorhandensein-Hinweis*, *Beweiskraft-Hinweis*, *Nichtvorhandensein-Hinweis*, *Symptom-Nichtvorhandensein-Hinweis*.

Nach Ablauf des Diagnoseprozesses liefert das System dem Arzt eine Liste von *bewiesenen Diagnosen*, *ausgeschlossenen Diagnosen* und *Diagnosehinweisen*. Für die ausgegebenen Diagnosen werden Begründungen geliefert, und zur Sicherung der Diagnosen werden Untersuchungsvorschläge angeboten.

Schlüssel-Wörter: Computerunterstützte Diagnose, Fuzzy-Teilmengen, Medizinische Dokumentation, Symptom—Diagnose-Beziehungen, Bewiesene und ausgeschlossene Diagnosen, Diagnosenhinweise

Introduction

The high grades of precision which prevail in sciences like mathematics, physics, chemistry and civil engineering stand in sharp contrast to the imprecision of definitions which can be found in sociology, psychology, medicine, linguistics, literature, science of art, philosophy, etc. Although the traditional mathematical techniques can still be applied very usefully to the above-mentioned sciences, a concept seems necessary that considers the huge complexity and imprecision of definitions in these fields.

L. A. Zadeh's theory of *fuzzy subsets* [2, 8—12, 15, 17, 27, 28) is an attempt at a mathematical theory of vagueness and imprecision. If one does not know exactly the probability of an event but is able to say *more or less probable*, *not very probable*, or, if the question of the age of a person can only be answered with *old*, *very young* or *not very young*, one can establish fuzzy subsets A of a reference set*) $\mathfrak{U}$ labelled *probable*, *old* or *young* which are characterized by a *membership function* $\mu_A: \mathfrak{U} \rightarrow [0, 1]$. The membership function μ_A associates with each element x of $\mathfrak{U}$ a number $\mu_A(x)$ in the interval [0, 1] which represents the degree of membership of x in A. The theory of ordinary sets is a particular case of the theory of fuzzy subsets with $\mu_A(x) = 1$ in the case of full membership of x in A and $\mu_A(x) = 0$ for no membership of x in A.

Bibliographies of the theory of fuzzy subsets may be found in [5, 7].

In medical science, it is rarely possible to work with exact definitions, descriptions or assertions. In medical diagnosis, there is rarely a sharp boundary between diseases. The appearance of more than one disease in the patient at the same time destroys the symptom pattern of the disease and makes the diagnostic and therapeutic decision more difficult. The assignment of laboratory test

*) L. A. Zadeh [27, 28] and others call $\mathfrak{U}$ universe of discourse.

results to normal or pathological ranges is arbitrary in borderline cases. The intensity of pain can only be described verbally and depends on the subjective estimation of the patient. In descriptions of diseases, precise relationships between symptoms and diseases can very seldom be found.

Thus, one can find in [21] (page 676ff.) the following statements about acute pancreatitis:

—There is *overweight* of about 62% and *abuse of alcohol* could be found in *40—70%* of the cases of acute pancreatitis (depends on the authors who published results of their medical research).

—*Frequent* historical information is related to the *ulcus ventriculi* and *duodeni*, *pregnancy*, etc.

—*Typically*, *acute pancreatitis* begins with *sudden pain of the abdomen* which may *increase rapidly*.

—*The intensity of ache* which is *not obligatory* can vary *enormously*.

—*Acute pancreatitis* is *almost always* connected with *sickness*, *vomiting*, etc.

—*Very high levels of amylase* (more than five times) are *almost proving* for *acute pancreatitis*.

From these examples (there are similar descriptions in [16, 20]) one is able to see that fuzzy terms like *frequent*, *typically*, *not obligatory*, *40—70%*, *almost always*, *seldom*, *almost proving*, etc. are common medical descriptions.

Strong relationships between symptoms and diagnoses like *obligatory*, *proving* or *excluding* may appear and it is absolutely necessary to use them, but they are rare [1, 6, 13, 24].

Due to these considerations, a model of a computer-assisted diagnostic system based on fuzzy subsets is presented. The system is able to work with assertions like

—symptom S is *often* present and *seldom* proving for diagnosis D,

—*strongly decreased* symptom S is *almost always* present and *always* proving for diagnosis D,

and to determine logical relationships between the patient's symptom pattern and diagnoses under consideration.

Review of Published Methods

A review of some published approaches using fuzzy subsets follows.

Description of symptoms, signs and test results

Esogbue et al. [3] and Moon et al. [47] propose the application of fuzzy subsets to determine the severity level of non-binary symptoms. The membership function may reflect the painfulness or blueness of symptoms such as *headache* and *cyanosis* or represents the degree of abnormality of possible clinical or diagnostic test results. In [3], for example, an appropriate linear function of the fuzzy subset of *abnormal cholesterol* C expressed in milligrams per 100 milliliters of serum would be

$$\mu_C(x) = \begin{cases} 0, & \text{for } x < 260 \\ \dfrac{x}{340} - \dfrac{26}{34}, & \text{for } 260 \leq x \leq 600 \\ 1, & \text{for } x > 600 \end{cases} \quad (1)$$

This must be done for all the patient information obtained from the physician--patient interaction during the initial interview as well as for the clinical or diagnostic test results gained by laboratory examination.

Moon et al. [14] use linguistic modifiers [27, 28] to compute the degree of membership of a test result in the fuzzy subset S_2 out of the degree of membership of the test result in S_1. In this way, it is possible to compute the degree of membership of the result x of the urine sodium concentration test in the fuzzy subset *very high urine sodium concentration* $\mu_{S_2}(x)$ out of the degree of membership in the fuzzy subset *high urine sodium concentration* $\mu_{S_1}(x)$ by performing

$$\mu_{S_2}(x) = \mu_{S_1}(x)^2. \quad (2)$$

In [3], Esogbue et al. introduce fuzzy subsets to determine diseases in the patient's past history which might be missed or remain undiagnosed. The presence or absence of such histories might then be determined from past undiagnosed symptoms, but the physician only considers the *prominent* symptoms of past sicknesses. The fuzzy subset S_j, containing the *prominent* symptoms of past disease j, is defined as

$$S_j = \{(x_{ij}, \mu_{S_j}(x_{ij}) \mid x_{ij} \in \beta_j, \mu_{S_j} \leqslant \alpha\} \quad (3)$$

where $\mu_{S_j}(x_{ij}) \in [0, 1]$ and $0 \leqslant \alpha \leqslant 1$. The set of possible symptoms for past undiagnosed disease j is β_j where $1 \leqslant j \leqslant p$ and p is the number of undiagnosed diseases in the patient's past history. Whenever the physician assigns the membership function $\mu_{S_j}(x_{ij})$ of symptom x_{ij} for disease j over a specified level α, the symptom becomes a *prominent* symptom.

Description of Diseases and Diagnoses

Smets et al. [23] mention the fuzziness of diagnostic terms like *arteriosclerosis* or *angina pectoris*. The diseases are not clearly defined entities. It is often impossible to determine precisely the symptoms related to the disease. The use of fuzzy subsets represents a mathematical method of handling imprecise diagnostic entities. Diagnoses are defined as fuzzy subsets with symptoms as elements. The symptoms are combined with a degree of membership which characterizes the intensity of belonging to the fuzzy subset that represents the disease under consideration.

Fuzzy Relations between Symptoms and Diagnoses

Sanchez [18, 19] and Moon et al. [14] start with the consideration of introducing a fuzzy relation $R \subset S \times D$ between symptoms S and diagnoses D. Sanchez calls it *medical knowledge* expressing associations between symptoms and diagnoses. Let, further, A be a fuzzy subset of S related to a patient, then the computation of the MAX-MIN-composition $B = A \circ R$ is assumed to describe the state of the patient in terms of diagnoses as a fuzzy subset B of D, characterized by its membership function

$$\mu_B(d) = \underset{S}{\text{MAX}} \text{ MIN } \{\mu_A(s); \mu_R(s,d)\} \quad (4)$$

where $s \in S$ and $d \in D$. If we consider several patients belonging to a set P and define a relation $Q \subset P \times S$, (4) becomes:

$$\mu_T(p,d) = \underset{S}{\text{MAX}} \text{ MIN } \{\mu_Q(p,s); \mu_R(s,d)\} \quad (5)$$

where $p \in P$ and $T \subset Q \times R$.

In fuzzy logic (4) or (5) is called *compositional rule of inference* (Zadeh [28]) or a fuzzy *meta-implication* (Kaufmann [8]).

In Sanchez [18, 19] the fuzzy relation R is assumed to be given by a physician. The physician translates his own knowledge and experience into degrees of association between symptoms and diagnoses.

Symptom Combinations

An attempt is made by Moon et al. [14] to present symptom combinations as AND or OR functions. This implies that the degree of membership of a patient p in the fuzzy subset of a diagnosis is related to the degrees of membership of p in the fuzzy subsets of the symptoms. The calculation of the symptom combinations is performed using the fuzzy intersection for AND and the algebraic product for OR. A different way of presenting OR function is mentioned. Moon et al. [14] use the convex combination

$$\mu_D (s_1 \ldots s_k) = \frac{\sum_i w_i \mu_{S_i}}{\sum_i w_i \lceil \mu_{S_i}} \qquad (6)$$

where $\lceil \mu_{S_i}$ is the ceiling function of μ_{S_i} and the weights w_i are set equal to the posterior probabilities $p(D/S_i)$ calculated for the single symptom.

Cluster Analysis

Fordon et al. [4] present a very interesting application of fuzzy clustering techniques to the diagnosis of hypertensive patients. In one example, a fuzzy k-means algorithm was applied to 218 essential and renovascular hypertensive patients described by four continuous parameters. The overall efficiency of the algorithm was 80,7%.

Computer-assisted Medical Diagnosis using Fuzzy Logic

Objectives

The following requirements are made of computer-assisted diagnostic system bases on fuzzy subsets:

a) Medical knowledge should be stored as logical relationships between symptoms and diagnoses.
b) The logical relationships might be fuzzy. They are not obliged to correspond to Boolean logic [1, 6, 13, 24].
c) Frequent as well as rare diseases are offered after analysing the patient's symptom pattern.
d) The diagnostic process can be performed iteratively.
e) Both proposals for further investigations of the patient and reasons for all diagnostic results are put out on request.

Definitions

General Considerations

Let $\Sigma = \{S_1, S_2, \ldots S_m\}$ be the set of symptoms and $\Delta = \{D_1, D_2, \ldots D_n\}$ the set of diagnoses taken into account and m and n the cardinality of Σ and Δ. Σ and Δ are non-fuzzy sets.

Every $S_i \in \Sigma$, $1 \leqslant i \leqslant m$, is a fuzzy subset of a reference set $\mathfrak{G} = \{x_1, x_2, \ldots\}$ characterized by a membership function $\mu_{S_i}(x)$.

The set $\mathfrak{G}$ contains any possible values which can be assumed by S_i. The membership function $\mu_{S_i}(x)$ defines the strength of affiliation of $x \in \mathfrak{G}$ in S_i.

Every $D_i \in \Delta$ is a fuzzy subset of the reference set $\mathfrak{P} = \{p_1, p_2, \ldots\}$ characterized by a membership function $\mu_{D_i}(p)$.

The set $\mathfrak{P}$ has as elements all patients under consideration and $\mu_{D_i}(p)$ assigns to every patient p a degree of membership of $p \in \mathfrak{P}$ in D_i.

Symptom—Diagnosis Fuzzy Relationships

Two aspects of a symptom S_i are of essential value in order to find out its relation to a diagnosis D_j:

1. *presence* of S_i in the case of D_j
2. *conclusiveness* of S_i for D_j

where $1 \leqslant i \leqslant m$ and $1 \leqslant j \leqslant n$. *Presence* comprises the frequency of occurrence of the symptom S_i in regard to a diagnosis D_j. *Conclusiveness* includes the importance of a symptom S_i to make a certain diagnosis D_j.

These two terms seem to be of great importance in making diagnoses because they provide two different assertions about the relationship between a symptom and a diagnosis.

In practice, the physician does not make his diagnosis only by taking into account the frequencies of symptoms regarding a certain diagnosis, as when applying statistical methods (see [2]). Further, the physician cannot subsume all his knowledge and experience in only one degree of association between a symptom and a diagnosis (see [18, 19]).

Thus, it seems necessary to take a step forward by utilizing two different aspects of a symptom in order to get more specification in describing the relationships between symptoms and diagnoses.

Now, the *presence* of S_i at D_j is defined as a fuzzy subset P of the reference set $\mathfrak{U}_p(x) = \{0, 1, 2, \ldots, 100\}$, where x means the occurrence of S_i in x of one hundred cases of D_j. Similary, the *conclusiveness* of S_i for D_j is determined as a fuzzy subset C of the reference set $\mathfrak{U}_c(x) = \{0, 1, 2, \ldots, 100\}$, i.e. S_i has been proved D_j in x of one hundred cases.

The membership functions of P and C are very simple. They are

$$\mu_p(x) = f_1(x; 2, 50, 98),\ x \in \mathfrak{U}_p \qquad (7)$$

and

$$\mu_c(x) = f_1(x; 2, 50, 98),\ x \in \mathfrak{U}_c \qquad (8)$$

where $f_1(x)$ is the following standardized function (see [28]):

$$f_1(x; \alpha, \beta, \gamma) = \begin{cases} 0 & \text{for } x \leqslant \alpha \\ 2\left(\dfrac{x-\alpha}{\gamma-\alpha}\right)^2 & \text{for } \alpha < x \leqslant \beta \\ 1-2\left(\dfrac{x-\gamma}{\gamma-\alpha}\right)^2 & \text{for } \beta < x \leqslant \gamma \\ 1 & \text{for } x > \gamma \end{cases} \qquad (9)$$

The membership functions $\mu_p(x)$ and $\mu_c(x)$ are shown in Fig. 1.

Determining Symptom—Diagnosis Relationships

There are two different ways of determining a certain $\mu_p(x)$ and $\mu_c(x)$ for a specific S_iD_j relation:

1. Evaluating a medical data base.
   One obtains the basis for calculating $\mu_p(x)$ by counting the number of occurrences of S_i in D_j.
   Similary, one gets the basis for calculating $\mu_c(x)$ by summing up the number of times D_j occurs in the case of S_i.
2. Documentation work by a physician.
   A physician documents the S_iD_j *presence* relationship as well as the S_iD_j *conclusiveness* relationship by answering the following questions:
   presence: »How often does S_i occur in D_j?«
   conclusiveness: »How strongly does S_i prove D_j?«

The possible set of physician's answers to specify *presence* or *conclusiveness* might be the fuzzy subsets P_i, $1 \leqslant i \leqslant 13$, or C_i, $1 \leqslant i \leqslant 13$, in Table 1.

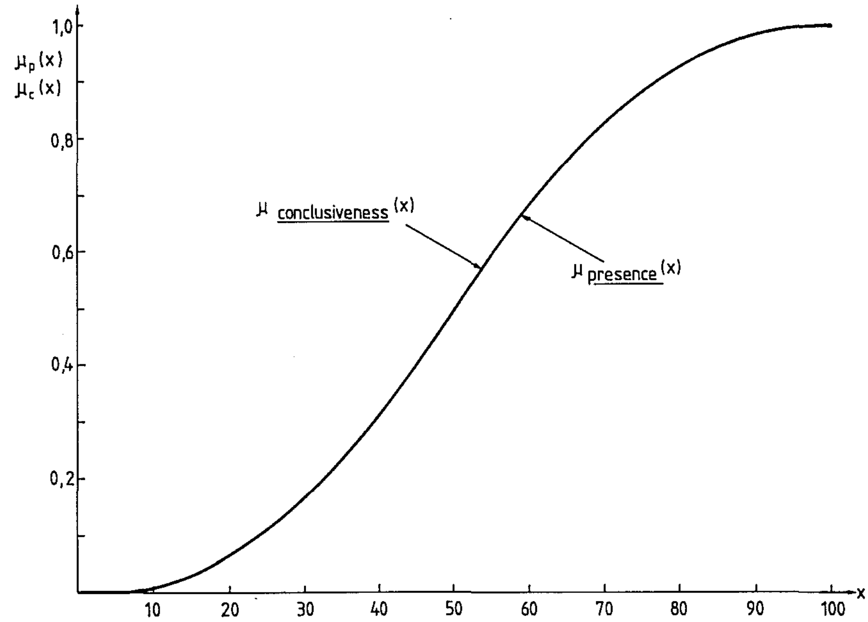

Fig. 1: Membership Function $\mu_p(x)$ for *Presence* of S_i at D_j and $\mu_c(x)$ for *Conclusiveness* of S_i for D_j.

Table 1: Fuzzy Subsets *Presence*$_i$ and *Conclusiveness*$_i$ for Physician's Documentation of S_iD_j Relationships

i	*Presence*$_i$ or *Conclusiveness*$_i$
1	*never*
2	*almost never*
3	*very very seldom*
4	*very seldom*
5	*seldom*
6	*more or less seldom*
7	*not known*
8	*more or less often*
9	*often*
10	*very often*
11	*very very often*
12	*almost always*
13	*always*

The fuzzy subsets P_i or C_i have the same reference set $\mathfrak{U}_{p_i}(x)$ or $\mathfrak{U}_{c_i}(x)$ as the reference set of *presence* $\mathfrak{U}_p(x)$ or *conclusiveness* $\mathfrak{U}_c(x)$ where $1 \leqslant i \leqslant 13$.

The fuzzy subsets P_1, P_2, P_5, P_7, P_9, P_{12} and P_{13} are characterized by the following membership functions:

$$\mu_{\textit{always}}(x) = f_1(x; 96, 97, 98) \qquad (10)$$

$$\mu_{\textit{almost always}}(x) = f_1(x; 85, 90, 95) \qquad (11)$$

$$\mu_{\textit{often}}(x) = f_1(x; 40, 60, 80) \qquad (12)$$

$$\mu_{\textit{not known}}(x) = f_2(x; 8, 50) \qquad (13)$$

$$\mu_{\textit{seldom}}(x) = 1—f_1(20, 40, 60) \qquad (14)$$

$$\mu_{\textit{almost never}}(x) = 1—f_1(5, 10, 15) \qquad (15)$$

$$\mu_{\textit{never}}(x) = 1—f_1(2, 3, 4) \qquad (16)$$

where $f_2(x)$ is a also a standardized function [28] (see (17)).

$$f_2(x; \alpha, \beta) = \begin{cases} f_1\left(x; \beta—\alpha, \beta—\frac{\alpha}{2}, \beta\right) & \text{for } x \leqslant \beta \\ 1—f_1\left(x; \beta, \beta+\frac{\alpha}{2}, \beta+\alpha\right) & \text{for } x > \beta \end{cases} \qquad (17)$$

By using the following operations on fuzzy subsets (see [12]), also called linguistic modifiers [28],

$$\text{concentration: } \mu_{\textit{very}}F(x) = (\mu_F(x))^2 \qquad (18)$$

$$\text{dilatation: } \mu_{\textit{more or less}}F(x) = (\mu_F(x))^{0.5} \qquad (19)$$

the determination of P_3, P_4, P_6, P_8, P_{10}, P_{11}

yields

$$\mu_{\textit{very very often}}(x) = (\mu_{\textit{often}}(x))^4 \qquad (20)$$

$$\mu_{\textit{very often}}(x) = (\mu_{\textit{often}}(x))^2 \qquad (21)$$

$$\mu_{\textit{more or less often}}(x) = (\mu_{\textit{often}}(x))^{0.5} \qquad (22)$$

$$\mu_{\textit{more vr less seldom}}(x) = (\mu_{\textit{seldom}}(x))^{0.5} \qquad (23)$$

$$\mu_{\textit{very seldom}}(x) = (\mu_{\textit{seldom}}(x))^2 \qquad (24)$$

$$\mu_{\textit{very very seldom}}(x) = (\mu_{\textit{seldom}}(x))^4 \qquad (25)$$

The definition of $C_1, C_2, \ldots, C_{13}$ will analogously be performed as for $P_1, P_2, \ldots, P_{13}$. The membership functions for P_i, $1 \leqslant i \leqslant 13$, as well as for C_i, $1 \leqslant i \leqslant 13$, are illustrated in Fig. 2.

But now, after defining the fuzzy subsets P_i, $1 \leqslant i \leqslant 13$, and C_i, $1 \leqslant i \leqslant 13$, and after the physician has specified the relation between an S_i, $1 \leqslant i \leqslant m$, and a D_j, $1 \leqslant i \leqslant n$, the question is raised as to which numerical degree of membership in the fuzzy subsets *presence* and *conclusiveness* might be assumed.

It seems understandable that, at the moment of documentation of the certain S_iD_j relation, i.e. to choose a P_i and C_i, the physician moves in the range of occurrences of S_i at D_j or D_j at S_i where the memberships function $\mu_{p_i}(x) \geqslant 0.5$ or $\mu_{c_i}(x) \geqslant 0.5$.

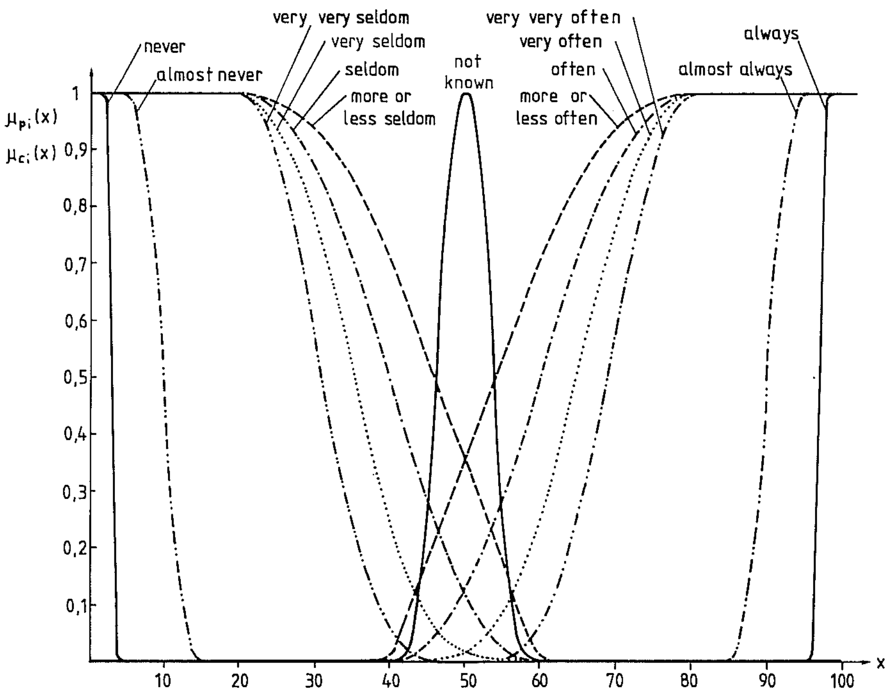

Fig. 2: Membership Functions of P_i, $1 \leqslant i \leqslant 13$, and $C_i, 1 \leqslant i \leqslant 13$

In the case of »S_i occurs *often* at D_j«, i.e. the physician has chosen P_5 = *often* and »S_i has *almost never* proved D_j«, i.e. the physician has documented C_{12} = *almost never*, the ranges would be

P_{often}: 60 to 100 occurrences and

$C_{almost\ never}$: 90 to 100 occurrences.

From Fig. 2 it can be seen that

$$P_1 \subset P_2 \subset P_3 \subset P_4 \subset P_5 \subset P_6$$

and

$$P_{13} \subset P_{12} \subset P_{11} \subset P_{10} \subset P_9 \subset P_8 \quad (26)$$

with $\subset$ as the containment of fuzzy subsets which is defined (see [12] and others) as

$$A \subset B \text{ if } \mu_A(x) \leqslant \mu_B(x), \ \forall x \in \mathfrak{X} \quad (27)$$

where A and B are fuzzy subsets of the reference set $\mathfrak{X}$.

Because of (26) the range r_{p_i} where $\mu_{p_i}(x) \geqslant 0.5$ is included in the range of P_{i-1} or P_{i+1}. We eliminate this by calculating new ranges for every P_i which is not included by another.

Thus

$$r_{p_i} = \begin{cases} r_{p_i} \setminus r_{p_{i-1}} & \text{for } 2 \leqslant i \leqslant 6 \\ r_{p_i} & \text{for } \quad i = 7 \\ r_{p_i} \setminus r_{p_{i+1}} & \text{for } 8 \leqslant i \leqslant 12 \end{cases} \quad (28)$$

The analoguous way is valid to define the ranges of C_i.

By establishing these ranges we have determined an interval of occurrences for every P_i or C_i which corresponds to the intention of the physician when he documents an S_iD_j relationship. In order to determine a numerical degree of membership for *presence* or for *conclusiveness*, we calculate the means of the suitable r_{p_i} or r_{c_i}. Now, we are able to compute $\mu_{presence}(x)$ and $\mu_{conclusiveness}(x)$.

The ranges for P_i and C_i, the means and the attached degrees of membership in P or C are shown in Table 2.

Table 2: Determination of $\mu_{presence}(x)$ and $\mu_{conclusiveness}(x)$ from the documented P_i or C_i, $1 \leqslant i \leqslant 13$

Fuzzy Subset P_i or C_i	Range	Mean	$\mu_p(x)$ or $\mu_c(x)$
never	0 to 3	1.5	0
almost never	3 to 10	6.5	0.004
very very seldom	10 to ~31	20.5	0.074
very seldom	~31 to ~35	33	0.209
seldom	~35 to 40	37.5	0.273
more or less seldom	40 to ~46	43	0.365
not known	~46 to ~54	50	0.5
more or less often	~54 to 60	57	0.635
often	60 to ~65	62.5	0.727
very often	~65 to ~69	67	0.791
very very often	~69 to 90	79.5	0.926
almost always	90 to 97	93.5	0.996
always	97 to 100	98.5	1

Symptoms as Fuzzy Subsets

All observable properties of a patient are applied in medical diagnosis. These can be symptoms, signs, laboratory test results and complaints related to the patient (also to his family and his living circumstances) at any observation time. Let us term every property as a symptom.

As stated above, every symptom $S_i \in \Sigma$, $1 \leqslant i \leqslant m$, is a fuzzy subset of a reference set $\mathfrak{S} = \{x_1, x_2, \ldots\}$ characterized by a membership function $\mu_{S_i}(x)$.

Let us now distinguish two different kinds of reference sets $\mathfrak{S}$.

1. $\mathfrak{S} = \{x_1, x_2, \ldots\}$ contains continuous values, e.g. as $\mathfrak{S}$ for S = *increased alcaline phosphatase*.
2. $\mathfrak{S} = \{0,1\}$, i.e. the symptom S_i is a binary symptom, which can be absent (0) or present (1) (see also [3, 25]), e.g. as $\mathfrak{S}$ for S = *drug anamnesis, medicine against rheumatism* (see Fig. 3).

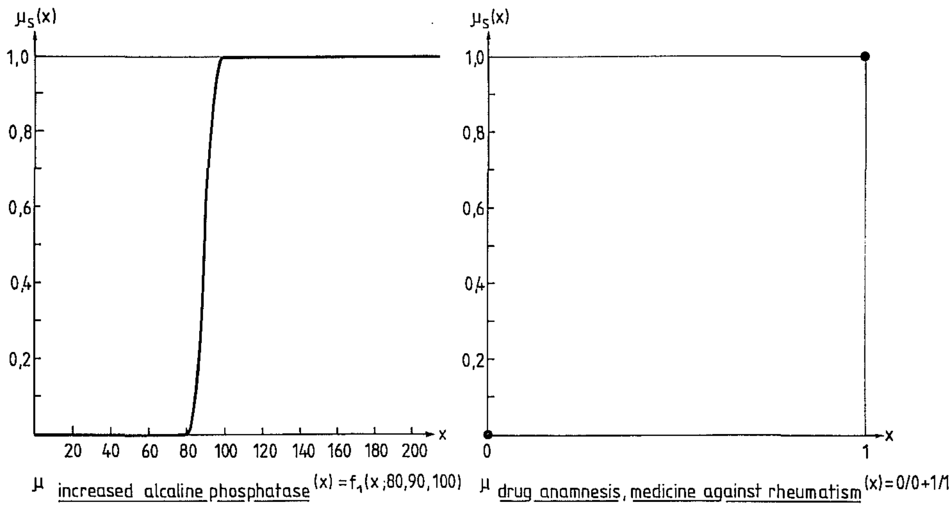

Fig. 3: Symptom Membership Functions

Most of the laboratory test results and some of the signs observed on physical examination (e.g. SCHOBER distance, normal $\geqslant$ 4 cm, pathological $<$ 4 cm) provide continuous values as a result of the investigation, and most of symptoms of patient's history, complaints, evidence of infection, X-ray findings etc. are binary symptoms. Some of the binary symptoms are real ones, e.g. *drug anamnesis, medicine against rheumatism*; others are not, e.g. *headache*. But we do not want to establish fuzzy subsets to define the severity level of symptoms, because so far it does not seem practical to calculate numerical values for, e.g., *severe headache*.

Now, all symptoms taken into account may be defined to have one of the two kinds of reference sets. Uninvestigated symptoms are not considered (in contrast to [1]).

Diagnostic Fuzzy Indications

Let us introduce a binary fuzzy relation R_p which is a fuzzy subset of $\Sigma \times \Delta$ and characterized by the two parameter membership functions $\mu_{R_p}(S_i, D_j)$ in the interval [0, 1]. The $\mu_{R_p}(S_i, D_j)$ are identical to $\mu_p(x)$ which has been documented for S_i with regard to D_j as S_iD_j presence relationships.

The S_iD_j conclusiveness relationships form the elements of the binary fuzzy relation $R_c \subset \Sigma \times \Delta$. Thus, R_c is defined by $\mu_{R_c}(S_i, D_j)$ which is equal to $\mu_c(x)$ considering the symptom S_i and diagnosis D_j.

Furthermore, we introduce the binary fuzzy relation $R_S \subset \mathfrak{P} \times \Sigma$. The sets $\mathfrak{P}$ and Σ are given when the symptom patterns of the patients under consideration are available. Symptoms which have not been investigated but are considered in the documentation work of the physician have to be dropped from the relation matrices R_p and R_c. Now, the functions $\mu_{R_S}(p, S_i)$ which characterize the fuzzy relation R_S are identical with $\mu_{S_i}(x)$ calculated for the patient corresponding to the defined fuzzy subsets S_i.

Finally, four different fuzzy indications are calculated by means of the composition of fuzzy relations:

1. ΣD_j presence indication $R_1 = R_S \circ R_p$

$$\mu_{R_1}(p, D_j) = \underset{S_i}{\mathrm{MAX}} \ \mathrm{MIN}\,\{\mu_{R_S}(p, S_i); \mu_{R_p}(S_i, D_j)\} \quad (29)$$

2. ΣD_j conclusiveness indication $R_2 = R_S \circ R_c$

$$\mu_{R_2}(p, D_j) = \underset{S_i}{\mathrm{MAX}} \ \mathrm{MIN}\,\{\mu_{R_S}(p, S_i); \mu_{R_c}(S_i, D_j)\} \quad (30)$$

3. ΣD_j non-presence indication $R_3 = R_S \circ (1 - R_p)$

$$\mu_{R_3}(p, D_j) = \underset{S_i}{\mathrm{MAX}} \ \mathrm{MIN}\,\{\mu_{R_S}(p, S_i); 1 - \mu_{R_p}(S_i, D_j)\} \quad (31)$$

4. Non ΣD_j presence indication $R_4 = (1 - R_S) \circ R_p$

$$\mu_{R_4}(p, D_j) = \underset{S_i}{\mathrm{MAX}} \ \mathrm{MIN}\,\{1 - \mu_{R_S}(p, S_i); \mu_{R_p}(S_j, D_j)\} \quad (32)$$

where $p \in \mathfrak{P}$, $S_i \in \Sigma$, $1 \leqslant i \leqslant m$ and $D_j \in \Delta$, $1 \leqslant i \leqslant n$.

Computer-assisted Diagnostic Process

The procedure of the computerized diagnostic process may be divided into several steps:

1. Any symptoms observed in patient p under consideration are entered either as numerical values or as 0 or 1 corresponding to the kind of reference set. The definition of the symptom fuzzy subsets allows the calculation of the degree of membership $\mu_{R_S}(p, S_i)$.
2. Now, $\mu_{R_1}(p, D_j)$, $\mu_{R_2}(p, D_j)$, $\mu_{R_3}(p, D_j)$ and $\mu_{R_4}(p, D_j)$ are calculated for patient p. The system then provides the physician with the following results.
3. All D_j are put out as *proven diagnoses* (DP) if
$$0.98 \leqslant \mu_{R_2}(p, D_j) \leqslant 1. \quad (33)$$
4. All D_j are displayed as *excluded diagnoses* (DE) if either
$$0.98 \leqslant \mu_{R_3}(p, D_j) \leqslant 1 \text{ or} \quad (34)$$
$$0.98 \leqslant \mu_{R_4}(p, D_j) \leqslant 1. \quad (35)$$
5. All diagnoses D_j where a degree of membership $\mu_H(p, D_j)$ of
$$0.5 < \mu_H(p, D_j) = \mathrm{MIN}\{\mu_{R_1}(p, D_j); \mu_{R_2}(p, D_j)\} \quad (36)$$
is calculated are put out as *diagnostic hints* (DH).
6. On request, reasons for the diagnoses displayed as proven or excluded diagnoses or as diagnostic hints are put out. The reasons are the degrees of membership $\mu_{R_S}(p, S_i)$ and the labels of the fuzzy subsets P_i and C_i which have been documented by the physician

considering the symptoms S_i that have caused the D_j to be a DP, DE or DH.

7. The physician gets a list of *suggested investigations.* All symptoms which prove or exclude a DH-diagnosis are displayed.
8. If the physician wishes to enter further symptoms and to repeat the whole process with an enlarged symptom pattern, he may continue with step 1 by adding symptoms.

At the end of procedure the physician has extensive information to enable him to make the differential diagnosis. The procedure guarantees that both clearly defined symptom—diagnosis relationships like *never*, *always* (obligatory, proving, excluding in the sense of [1, 13, 24] and fuzzy relationships like *almost never*, *very very seldom*, *very seldom*, *seldom*, *more or less seldom*, *more or less often*, *often*, *very often*, *very very often*, *almost always* (facultative, not proving in the sense of [1, 13, 24]) are used to support the physician's diagnostic decision.

Application

Suppose, for example, a small application with

$\mathfrak{P} = \{p_1, p_2\}$, $\Sigma = \{S_1, S_2, S_3, S_4, S_5, S_6\}$ and $\Delta = \{D_1, D_2\}$

to illustrate the computer-assisted diagnostic system. A physician documented Table 3:

Table 3: Documented Symptom—Diagnosis *Presence* and *Conclusiveness* Relationships

Symptom	D_1		D_2	
	Presence	*Conclusiveness*	*Presence*	*Conclusiveness*
S_1	*always*	*always*	*seldom*	*often*
S_2	*always*	*often*	*often*	*not known*
S_3	*never*	*never*	*not known*	*not known*
S_4	*often*	*not known*	*not known*	*not known*
S_5	*very often*	*always*	*almost never*	*not known*
S_6	*seldom*	*not known*	*often*	*very very often*

R_p as well as R_c are determined from the documentation. The patient's symptom pattern is given and R_S (see Table 4) is calculated.

Table 4: Patient-Symptom Relation Matrix R_S

R_S	S_1	S_2	S_3	S_4	S_5	S_6
p_1	1	0.6	0	1	1	0.2
p_2	0.8	1	0.5	1	0	1

After computing the compositions of fuzzy relations, the following results are obtained (Table 5):

Table 5: Results of the Diagnostic Process for Patient p_1 and p_2

Patient	Category	Diagnosis	Reasons		
			Symptom Pattern	*Presence*	*Conclusiveness*
p_1	proven diagnoses	D_1	$\mu_{R_S}(p_1, S_1) = 1$ $\mu_{R_S}(p_1, S_5) = 1$	*always* *very often*	*always* *always*
	excluded diagnoses	D_2	$\mu_{R_S}(p_1, S_5) = 1$	*almost never*	*not known*
	diagnostic hints	/			
p_2	proven diagnoses	/			
	excluded diagnoses	/			
	diagnostic hints	D_1	$\mu_{R_S}(p_2, S_2) = 1$ $\mu_{R_S}(p_2, S_1) = 0.8$	*always* *always*	*often* *always*
		D_2	$\mu_{R_S}(p_2, S_6) = 1$ $\mu_{R_S}(p_2, S_2) = 1$	*often* *often*	*very very often* *not known*

Critical Remarks

So far, the present computer-assisted diagnostic system using fuzzy subsets has been developed as a theoretical concept.

It is also possible to use the algebraic product [8, 12] instead of the intersection in (36). Also (7) and (8) may be changed if necessary. Further, the most left-hand values of the inequalities (33), (34), (35) and (36) have not been proven in practical applications.

Perhaps it will also be possible to calculate diagnostic hints to exclude diagnoses from $\mu_{R_3}(p, D_j)$ and $\mu_{R_4}(p, D_j)$.

References

[1] Adlassnig, K.-P., Gergely, T., Grabner, H., Grabner, G.: A Computer Assisted System for Diagnostic Decision Making. Online Usage of the Database of the Medical Information System WAMIS. In D. B. Shires, H. Wolf (Eds): MEDINFO 77, pp. 213—218. (Amsterdam: North-Holland 1977).

[2] Bellman, R. E., Zadeh, L. A.: Local and Fuzzy Logics. Memorandum No. ERL-M584, 11. May 1976. Electronics Research Laboratory, College of Engineering, Univ. of California, Berkeley 94720.

[3] Esogbue, A. O., Elder, R. C.: Fuzzy Sets and Modelling of Physician Decision Processes. Part I: The Initial Interview-Information Gathering Session. Fuzzy Sets Syst. *2* (1979) 279—291.

[4] Fordon, W. A., Bezdek, J. C.: The Application of Fuzzy Set Theory to Medical Diagnosis. In M. M. Gupta, R. K. Ragade, R. R. Yager (Eds): Advances in Fuzzy Set Theory and Applications, pp. 445—461. (Amsterdam: North-Holland 1979).

[5] Gaines, B. R., Kohout, L. J.: The Fuzzy Decade: A Bibliography of Fuzzy Systems and Closely Related Topics. Int. J. Man-Machine Stud. *9* (1977) 1—68.

[6] Grabner, H., Grabner, G.: Aims and Structure of the Vienna General Medical Information System WAMIS. In J. Anderson, J. M. Forsythe (Eds): MEDINFO 74, pp. 375—379. (Amsterdam: North-Holland 1974).

[7] Kandel, A., Yager, R. R.: A 1979 Bibliography on Fuzzy Sets, their Applications, and Related Topics. In M. M. Gupta, R. K. Ragade, R. R. Yager (Eds): Advances in Fuzzy Set Theory and Applications, pp. 621—744. (Amsterdam: North-Holland 1979).

[8] Kaufmann, A.: Introduction to the Theory of Fuzzy Subsets. Vol. I. Fundamental Theoretical Elements. (New York: Academic Press 1975).

[9] Kaufmann, A.: Introduction à la théorie des sous-ensembles flous. Tome II. Application à la linguistique et à la sémantique. (Paris: Masson 1975).

[10] Kaufmann, A.: Introduction à la théorie des sous-ensembles flous. Tome III. Application à la classification et à la reconnaissance des formes, aux automates et aux systèmes, ou choix des critères. (Paris: Masson 1975).

[11] Kaufmann, A.: Introduction à la théorie des sous-ensembles flous. Tome IV. Compléments et nouvelles applications. (Paris: Masson 1977).

[12] Kickert, W. J. M.: Fuzzy Theories on Decision-making. (Leiden: Martinus Nijhoff 1978).

[13] Ledley, R. S.: Use of Computers in Biology and Medicine. (New York: McGraw-Hill 1965).

[14] Moon, R. E., Jordanov, S., Perez, A., Turksen, I. B.: Medical Diagnostic System with Human-like Reasoning Cabability. In D. B. Shires, H. Wolf (Eds): MEDINFO 77, pp. 115—119. (Amsterdam: North-Holland 1977).

[15] Negoita, C. V.: Fuzzy Models for Social Processes. In H. Bossel, S. Klaczko, N. Müller (Eds): System Theory in the Social Sciences, pp. 283—291. (Basel: Birkhäuser 1976).

[16] Overzier, C.: Systematik der inneren Medizin: Daten, Fakten, Übersichten. (Stuttgart: Thieme 1975).

[17] Reisinger, L.: On Fuzzy Thesauri. In G. Bruckmann, F. Ferschl, L. Schmetterer (Eds): COMPSTAT 1974, pp. 119—127. (Wien: Physica 1974).

[18] Sanchez, E.: Inverses of Fuzzy Relations. Application to Possibility Distributions and Medical Diagnosis. Fuzzy Sets Syst. *2* (1979) 75—86.

[19] Sanchez, E.: Medical Diagnosis and Composite Fuzzy Relations. In M. M. Gupta, R. K. Ragade, R. R. Yager (Eds.): Advances in Fuzzy Set Theory and Applications, pp. 437—444. (Amsterdam: North-Holland 1979).

[20] Schaldach, H. (Hrsg.): Wörterbuch der Medizin. (Berlin: VEB Verlag Volk und Gesundheit 1974).

[21] Schwiegk, H. (Hrsg.): Handbuch der inneren Medizin. 3. Band: Verdauungsorgane. (Berlin: Springer 1976).

[22] Skala, H. J.: Fuzzy Concepts: Logic, Motivation, Application. In H. Bossel, S. Klaczko, N. Müller (Eds): System Theory in the Social Sciences, pp. 292—306. (Basel: Birkhäuser 1976).

[23] Smets, P., Vainsel, H., Bernard, R., Kornreich, F.: Bayesian Probability of Fuzzy Diagnosis. In D. B. Shires, H. Wolf (Eds): MEDINFO 77, pp. 213—218. (Amsterdam: North-Holland 1977).

[24] Spindelberger, W., Grabner, G.: Ein Computerverfahren zur diagnostischen Hilfestellung. In K. Fellinger (Hrsg.): Computer in der Medizin — Probleme, Erfahrungen, Projekte, S. 189—221. (Wien: Hollinek 1968).

[25] Tautu, P., Wagner, G.: The Process of Medical Diagnosis: Routes of Mathematical Investigations. Meth. Inform. Med. *17* (1978) 1—10.

[26] Wagner, G., Tautu, P., Wolber, U.: Problems of Medical Diagnosis — A Bibliography. Meth. Inform. Med. *17* (1978) 55—74.

[27] Zadeh, L. A.: Outline of a New Approach to the Analysis of Complex Systems and Decision Processes. IEEE Transact. Syst. Man Cybern. *SMC-3* (1973) 28—44.

[28] Zadeh, L. A.: A Fuzzy-algorithmic Approach to the Definition of Complex or Imprecise Concepts. In H. Bossel, S. Klaczko, N. Müller (Eds): System Theory in the Social Sciences, pp. 202—282. (Basel: Birkhäuser 1976).

Address of the author: Dipl.-Ing. Klaus-Peter Adlassnig, Institut für Medizinische Computerwissenschaften der Medizinischen Fakultät Wien, Garnisongasse 13, A-1090 Wien, Österreich.