



ELSEVIER

Fuzzy Sets and Systems 129 (2002) 203–228

**FUZZY**  
sets and systems

www.elsevier.com/locate/fss

## A fuzzy-based methodology for the analysis of diabetic neuropathy

Luigi Di Lascio<sup>a</sup>, Antonio Gisolfi<sup>a,\*</sup>, Antonio Alburnia<sup>a</sup>, Giuseppe Galardi<sup>b</sup>,  
Franco Meschi<sup>c</sup>

<sup>a</sup> *Dipartimento di Matematica e Informatica, Università di Salerno, Via Salvador Allende 64, 84081 Baronissi (SA) Italy*

<sup>b</sup> *Neurophysiological Department, Istituto Scientifico San Raffaele, Via Olgettina 60, 20132 Milano, Italy*

<sup>c</sup> *Pediatric Department, Istituto Scientifico San Raffaele, Via Olgettina 60, 20132 Milano, Italy*

Received 29 February 2000; received in revised form 19 February 2001; accepted 6 March 2001

### Abstract

A new model for the fuzzy-based analysis of diabetic neuropathy is illustrated, whose pathogenesis so far is not well known. The underlying algebraic structure is a commutative l-monoid, whose support is a set of classifications based on the concept of linguistic variable introduced by Zadeh. The analysis is carried out by means of patient's anagraphical and clinical data, e.g. age, sex, duration of the disease, insulinic needs, severity of diabetes, possible presence of complications. The results obtained by us are identical with medical diagnoses. Moreover, analyzing suitable relevance factors one gets reasonable information about the etiology of the disease, our results agree with most credited clinical hypotheses. © 2002 Elsevier Science B.V. All rights reserved.

**Keywords:** Medical diagnosis; Commutative l-monoid; Fuzzy algebraic structure; Fuzzy triangular number; Linguistic label; Diabetic neuropathy

### 1. Introduction

Fuzzy-based analysis in order to carry out medical diagnoses is widely used nowadays. We just quote CADIAG-2 [1] for internal medicine, ARC [2] using case-based techniques, and IBIS [4] for analyzing pathological images MR.

CADIAG-2—*computer-assisted diagnosis*—is an expert system that utilizes rules of fuzzy inference, suitably adapted to the field of medical diagnosis. The system consists of several functional modules. One of them has peculiar features in order to update the knowledge representation related to the clinical data of a disease and the relationships between disease and symptoms. This system is currently used for diagnosing pancreas and rheumatic diseases and its reliability is over 90%.

\* Corresponding author. Tel.: +39-089965225; fax: +39-089965438.

E-mail addresses: [gisolfi@unisa.it](mailto:gisolfi@unisa.it) (A. Gisolfi), [galardi@hstr.it](mailto:galardi@hstr.it) (G. Galardi).

The system ARC, instead, utilises production rules whereas the inferential process is CBR-type (case based reasoning). ARC furnishes a diagnosis by taking into account diagnoses carried out in similar situations.

Interpretation of biomedical, industrial and remote sensing images (IBIS) is a fuzzy expert system able to explain the images obtained by means of different techniques (nuclear magnetic resonance, computerised axial tomography). By comparing the image under examination with a set of sample images, the acknowledgement process assigns to each elementary part of the image a value that evaluates the degree of reliability of the interpretation in that specific part.

Other recent results are presented in [8,14,15,17,19,21,22].

In this paper we use a fuzzy algebraic structure in order to deal with the etiological uncertainty affecting one of the most severe complications of diabetes, i.e. the diabetical neuropathy.

For each patient several data have been taken into account, such as age, sex, duration and severity of the disease, insulimic needs, and several others. It is worth noting that pathogenesis so far is not well known at all. Moreover, the problem of diagnosing neuropathies is a difficult one facing present physicians. We aim at developing a tool which can give, for each patient under examination, the risk factor related to this disease. The problem modelling is based on two samples related to individuals who have been clinically diagnosed. The results obtained by our methodology fully agree with those achieved clinically, also as regards possible origins of the disease.

The paper is organised as follows. In order to make clear the relationship between the mathematical model and clinical data, some clinical elements concerning diabetic neuropathies are described in Sections 2 and 3. Section 4 illustrates triangular fuzzy numbers, they are used to represent medical expertise about the disease. These numbers are fuzzy sets associated with linguistic variables related to clinical data [26]. The algebraic structure underlying our methodology is presented in Section 5 and its basic features are briefly discussed. Next section deals with the computational aspects of the methodology and Section 7 faces the problem of singling out the clinical factors inducing the disease. In Section 8 it is shown how pathogenic factors can be discriminated against those not directly relevant for the insurgence of the neuropathy. Critical factors are equal to those singled out in [3,18]. Last section briefly discusses the results so far obtained and problems currently under investigation are sketched.

## 2. Pathogenesis of diabetic neuropathies

It is well known that diabetes is a metabolic disease characterised by pancreas' inability to produce insulin, protein necessary in order to glucose could cross the cellular membrane and subsequently be utilised by the cell itself. Consequently one has hyperglycemia which leads to serious complications, such as neuropathies, nephropathies and retinopathies [3,8,18,23]. In particular, neuropathies affect peripheral nerves which are different according to the type of signal carried by them: sensations directed to central nervous system, orders to muscles arising from central system, involuntary control of internal organs (e.g. cardiac rhythm, arterial pressure). According to the type of nervous fibres one has sensitive neuropathy, motorial neuropathy, vegetative neuropathy. The sensitive neuropathy is the most frequent and about a third of people affected by diabetes suffers from distal polyneuropathy. The diagnosis of polyneuropathy is carried out by means of electroneurography (ENG): this test is based on the electro-stimulation of a nerve and the subsequent measurement of parameters related to the amount of nervous fibres and the conduction velocity of nervous impulses. More specifically, one has the motor parameters motor action potential (MAP) and motor conduction velocity (MCV) and the sensitive parameters sensory action potential (SAP) and sensory conduction velocity (SCV). On the grounds of these parameters polyneuropathy is diagnosed.

### 3. Clinical data about diabetic neuropathy

We consider two samples of patients, hospitalised in different times and diagnosed with diabetic neuropathy, their relevant clinical data are reported in Tables 1 and 2:

- *HbA1c*: denotes the glycosilate haemoglobin and measures the patient's hemoglicidic state, in terms of percentage of haemoglobin affected by glycosilation. Normal values are lower than 6%.

Table 1

No.	Age/sex	Duration diabetes	AER	HbA1c	BPs	BPd	BPm	Insu	GFR	Fluor	BdB	Fundus
Age	years	years	mcg/min	%	mm/Hg	mm/Hg	mm/Hg	U/Kg	60/135			
1	16.8/M	15.6	2.1	8.3	125	90	116	1	127	1	1	1
2	22.8/F	13.7	64.8	9	140	90	106	0.9	106	1	1	2
3	25.2/M	12.3	14.6	9.6	140	90	106	1.3	119	1	1	1
4	23/M	13.7	8.5	8	120	70	86	0.7	116	1	1	1
5	23.5/M	19.1	12.4	7.6	120	70	86	1.1	120	2	1	1
6	28.3/M	26.9	7.4	7.4	120	75	90	0.7	120	1	1	1
7	22.2/F	10.1	12.6	9.4	130	80	96	0.5	166	3	1	3
8	28.5/M	24.1	16.3	9.3	120	80	93	0.8	108	3	1	3
9	25.8/M	17.1	25.3	10.9	110	70	93	0.9	112	3	1	3
10	24.7/F	13.8	3	9.3	110	75	86	0.5	113	3	1	3
11	21.8/M	10.4	6.3	14.2	125	75	90	1	141	1	1	1
12	22.2/M	11.4	65.8	6.6	130	70	90	1.2	120	1	1	1
13	22.5/M	10.2	3.9	7.8	125	75	90	0.6	114	2	1	2
14	22.7/M	11.2	7.3	10.1	140	80	100	1	152	1	1	1
15	18.8/F	11.8	29.2	12.4	140	80	100	0.8	116	2	1	2
16	20.2/F	15.1	16.6	9	130	75	63	0.8	124	2	1	1
17	26.4/M	18.2	4.2	8.6	130	80	96	0.7	91	1	1	1

Table 2

No.	Duration diabetes	AER	HbA1c	Insu	BdB	Fundus
years	years	mcg/min	%	U/Kg		
1	5	12.6	7.5	1.2	1	2
2	11	57	6.7	0.8	1	2
3	0.9	1	8.6	0.7	1	1
4	13	22.9	10.6	1.2	1	1
5	4.1	41	7.3	0.6	1	1
6	11.4	4.3	8.5	0.9	1	1
7	8.6	4.5	8.1	0.8	1	1
8	11.3	8.8	7.4	0.6	1	1
9	10.61	7.2	8.9	0.8	1	1
10	11.8	5.9	8	0.7	1	1
11	13.8	6.2	6.9	0.5	1	1
12	5.6	4.1	7.9	0.9	1	1
13	9.2	6.5	10	0.8	1	1
14	15.2	3.6	11.6	0.7	1	2
15	5.5	8	9.1	0.6	1	2
16	7.4	4.4	10	0.9	1	2

- *AER*: this parameter measures the amount of albumin in urine. High values are symptoms of nephropathy and signals that diabetes starts damaging the organism.
- *BP<sub>s</sub>*, *BP<sub>d</sub>*: denote, respectively, the values of systolic and diastolic pressure.
- *I<sub>ins</sub>*: denotes the amount of insulin administered to the patient.
- *GFR*: measures the amount of plasma filtered by renal glomerule per unit of time. This parameter gives further information about renal functionality. Normal values range from 90 to 130, greater values may denote early unbalance.
- *Fluor*: denotes fluorangiography, namely the eyeball test devoted to ascertain the state of vessels.
- *BdB*: stands for beat to beat and denotes the degree of cardiac enervation. Possible values are: 1 = normal, 2 = borderline, 3 = pathological.
- *Fundus*: denotes the state of ocular fundus.

The importance of these parameters is widely accepted [3]. In particular the parameter *HbA1c* gives most reliable information about the severity of the disease.

The parameter *AER* is also useful to evaluate the patient's renal damage. Steady high values of this parameter lead, in a few years, to the renal insufficiency. Also the index *GFR* measures renal functionality as regards glomeruli involved in the nephropathy. The index has a peculiar behaviour: in early stages of the disease it increases, whereas in late stages decreases.

The retinopathy is another severe complication of diabetes, characterised by a gradual degradation of the patient's retina. Main tests to check the state of the retina include a direct test of ocular fundus and retinal fluorangiography by which captions of the retina can be analysed in detail.

#### 4. Representation of clinical and diagnostic data

In order to model the uncertainty inherently present in clinical data we have used triangular fuzzy numbers [16,20], each one representing values of linguistic variables. The fuzzification of uncertain data allows to map an interval of real numbers onto a fuzzy number.

##### 4.1. Triangular fuzzy numbers

A fuzzy set  $S$  [27] on the set  $R$  of real numbers is *convex* if and only if  $\forall x_1, x_2 \in R$ , and  $\forall p \in [0, 1]$  one has:  $\mu_S(px_1 + (1-p)x_2) \geq \min[\mu_S(x_1), \mu_S(x_2)]$ .

$A$  is a fuzzy number if and only if

- (1)  $\mu_S : x \in R \rightarrow \mu_S(x) \in [0, 1]$ ,
- (2)  $S$  is convex,
- (3) there is one  $x_0 \in R$ , called mean value, such that  $\mu_S(x_0) = 1$ ,
- (4)  $\mu_S$  is continuous, possibly stepwise.

A class of fuzzy numbers very useful in practice is the class of triangular fuzzy numbers. For this class the membership function is characterised by a triple of reals  $l \leq c \leq r$  and has the form

$$\mu_S = \begin{cases} (x-1)/(c-l) & \text{if } x \in [l, c], \\ (r-x)/(x-c) & \text{if } x \in [c, r]. \end{cases}$$

The importance of triangular numbers [20] stems from the ease in representing them. In fact, a triangular number is characterised by the triple  $[l, c, r]$  with no loss of information and operations are consequently greatly simplified.

Moreover, fuzzy numbers play an important role due to the following reasons:

(a) by the *extension principle*, usual operations for reals can be extended to fuzzy numbers. For example, as regards triangular numbers one has:

- (1)  $[a, b, c] + [d, e, f] = [a + d, b + e, c + f]$ .
- (2)  $\lambda * [a, b, c] = [\lambda * a, \lambda * b, \lambda * c]$ , where  $\lambda \in R$ .

(b) fuzzy numbers qualify as a very good model for representing linguistic labels, i.e. the variables of fuzzy logic and approximate reasoning [24–26].

#### 4.2. Linguistic labels

We aim at classifying each patient according to the severity of the symptoms. Suitable linguistic labels denote a different severity grade of the symptom. Each label, in turn, is represented by a triangular number. Their values are assigned partly by consultancy of an expert in the field, and partly by activating the phase of initial training. The labels for each attribute are the following:

*Labels for diabetes-age [Di-age]*: Choosing labels for this attribute takes into account juvenile diabetes, whose diagnosis takes place generally at about 12. Consequently the following labels are singled out:

- *Very early*: if diagnosis occurs before 3 years old, i.e. the value  $Di\text{-age} \in [0, 3]$ . This label is represented by the fuzzy number  $ve = [0.8, 1.0, 1.0]$ .
- *Early*: if  $Di\text{-age} \in [3, 10] \Rightarrow e = [0.6, 0.7, 0.9]$ .
- *Average*: if  $Di\text{-age} \in (10, 25] \Rightarrow a = [0.1, 0.5, 0.7]$ .
- *Late*: if  $Di\text{-age} \in (25, \infty) \Rightarrow l = [0.0, 0.0, 0.2]$ .

*Labels for HbA1c*: For healthy people this index takes value zero, however values below 6% are considered as acceptable.

- *Serious pathology*: if the value is greater than 15%, i.e.  $HbA1c \in (15, \infty]$ . The corresponding fuzzy number is  $sp = [0.8, 1.0, 1.0]$ .
- *High value*: if  $HbA1c \in [9, 15] \Rightarrow hv = [0.5, 0.75, 0.9]$ .
- *Altered value*: if  $HbA1c \in (6, 9) \Rightarrow av = [0.2, 0.4, 0.6]$ .
- *Normal*: if  $HbA1c \in [0, 6] \Rightarrow n = [0.0, 0.0, 0.3]$ .

*Labels for AER*:

- *Serious pathology*: if the value is greater than 30, i.e. index  $AER \in (30, \infty) \Rightarrow sp = [0.5, 0.7, 1.0]$ .
- *High value*: If  $AER \in (20, 30] \Rightarrow hv = [0.4, 0.6, 1.0]$ .
- *Altered value*: if  $AER \in (15, 20] \Rightarrow av = [0.2, 0.5, 0.6]$ .
- *Normal*: if  $AER \in [0, 15] \Rightarrow n = [0.0, 0.0, 0.3]$ .

*Labels for insulin*: The values are assigned considering that daily quantity is usually 1 U/kg.

- *High*: if the value  $Insu \in (1.1, \infty) \Rightarrow h = [0.9, 1.0, 1.0]$ .
- *Normal*: if  $Insu \in [0.9, 1.1] \Rightarrow n = [0.5, 0.7, 1]$ .
- *Low*: if  $Insu \in [0.0, 0.9] \Rightarrow l = [0.2, 0.5, 0.6]$ .

*Labels for GFR*: Regular values lie between 90 and 130, lower are viewed as pathological and upper ones deserve further investigation.

- *Serious pathology*: if index  $GFR \in [0, 40] \Rightarrow sp = [0.7, 1.0, 1.0]$ .
- *High*: if  $GFR \in [40, 60] \Rightarrow h = [0.4, 0.6, 0.8]$ .

- **Borderline:** if  $GFR \in (60, 90] \cup (130, \infty) \Rightarrow b = [0.2, 0.4, 0.5]$ .
- **Normal:** if  $GFR \in (90, 130] \Rightarrow n = [0.0, 0.0, 0.4]$ .

**Labels for Fluor:** As regards fluorangiography there are standard values representing pathological categories, thus we just attach a label to each category.

- **Serious pathology:** if the value  $Fluor = 3 \Rightarrow sp = [0.7, 1.0, 1.0]$ .
- **Pathological:** if  $Fluor = 2 \Rightarrow p = [0.4, 0.6, 0.8]$ .
- **Normal:** if  $Fluor = 1 \Rightarrow n = [0.2, 0.4, 0.5]$ .

#### Labels for BdB:

- **Pathological:** if the index  $BdB = 3 \Rightarrow p = [0.7, 1.0, 1.0]$ .
- **Borderline:** if  $BdB = 2 \Rightarrow b = [0.3, 0.5, 0.7]$ .
- **Normal:** if  $BdB = 1 \Rightarrow n = [0.2, 0.4, 0.5]$ .

#### Labels for Fundus:

- **Serious pathology:** if the index  $Fundus = 3 \Rightarrow sp = [0.7, 1.0, 1.0]$ .
- **Pathological:** if  $Fundus = 2 \Rightarrow p = [0.3, 0.5, 0.7]$ .
- **Normal:** if  $Fundus = 1 \Rightarrow n = [0.1, 0.3, 0.5]$ .

### 4.3. Attribute strings

The set of parameters and the related labels induce the following classification on the set of patients (Table 1):

$$\begin{aligned}
 \text{Diab-age} &= [1, 6]^{pe} [2, 4, 5, 8, 9, 15, 16, 17]^p [3, 7, 10, 11, 12, 13, 14]^p [-]^f, \\
 \text{HbA1c} &= [-]^{ph} [3, 7, 8, 9, 10, 11, 14, 15]^b [1, 2, 4, 5, 6, 12, 13, 16, 17]^n [-]^n, \\
 \text{AER} &= [2, 12]^{wh} [9, 15]^b [8, 16]^n [1, 3, 4, 5, 6, 7, 10, 11, 13, 14, 17]^n, \\
 \text{Insu} &= [3, 12]^b [1, 2, 5, 9, 11, 14]^n [4, 6, 7, 8, 10, 13, 15, 16, 17]^f, \\
 \text{GFR} &= [-]^{sp} [-]^p [7, 11, 14]^b [1, 2, 3, 4, 5, 6, 8, 9, 10, 12, 13, 15, 16, 17]^n, \\
 \text{Fluor} &= [7, 8, 9, 10]^{sp} [5, 13, 15, 16]^p [1, 2, 3, 4, 6, 11, 12, 14, 17]^n, \\
 \text{BdB} &= [-]^p [-]^b [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17]^n, \\
 \text{Fundus} &= [7, 8, 9, 10]^{sp} [2, 13, 15]^p [1, 3, 4, 5, 6, 11, 12, 14, 16, 17]^n.
 \end{aligned}$$

Each attribute string denotes the grade, expressed via linguistic labels, patients are affected by the parameters. For example, *Diab-age* informs us that patients 1 and 6 have been suffering diabetes since early years, whereas for patients 2, 4, 5, 8, 9, 15, 16, 17 the disease appeared when less than 10 and for patients 3, 7, 10, 11, 12, 13 and 14 when youngster. No patient has suffered from diabetes after 25.

### 4.4. Labels for the linguistic approximation of results

The diagnostic process should combine previous strings in order to get the final string which classifies patients according to their risk factor as regards neuropathy.

We divide the interval  $[0, 1]$  in two sub-intervals: the first contains the class of triangular fuzzy numbers whose mean values lie in the range  $[0, 0.5]$ , the second contains the class of triangular fuzzy numbers whose mean values lie in the range  $[0.5, 1]$ .

If we consider the labels previously chosen to represent the clinical data of the patients, the following labels are finally chosen for a further splitting of the above-mentioned sub-intervals:

$$\text{high risk} \Rightarrow HR = [0.7, 1, 1]; \quad \text{moderate risk} \Rightarrow R = [0.5, 0.55, 0.75];$$

$$\text{low risk} \Rightarrow LR = [0.2, 0.4, 0.5]; \quad \text{no at risk} \Rightarrow NR = [0, 0, 0.25].$$

The patients whose labels belong to the interval  $[0.5, 1]$  will be considered as affected by the disease. Thus the relevant features of 17 patients have been embodied in a fuzzy model. Now we take into account the way strings can be manipulated.

### 5. The algebraic fuzzy structure

In this section we are going to illustrate the basic features of an algebraic fuzzy structure that allows to develop suitable computations on the above-mentioned strings. Initially the structure has been considered as a tool for approximate reasoning [9]. Then it has proved useful for classifying with fuzzy attributes [10] and suitable operations have been introduced to deal with several application fields [11–13]. More specifically, we have investigated: classification of bacteria regarding their danger for human beings [11], diets for breast-feeding mothers [11], learning evaluation [6], financial investments [12], user modelling in hypermedia systems [5], young growing diseases [7].

#### 5.1. Basic definitions

Suppose that the crisp set  $U$  is the universe of discourse  $U = \{u_1, u_2, \dots, u_r\}$  and that each element is represented by the  $k$ -ple  $(A_1(u), A_2(u), \dots, A_k(u))$ , where the quantities  $A_i(u)$  are fuzzy terms (represented by triangular numbers) with respect to the attribute  $A_i$ . Each  $A_i(u)$  is a linguistic label denoting the compatibility degree of  $u$  with respect to  $A_i$ .

In turn, each attribute  $A_i$ , taking values on a set of linguistic labels  $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$ , induces on  $U$  a classification that is called *attribute string*, represented as follows:

$$a_n^{\alpha_n} a_{n-1}^{\alpha_{n-1}} \dots a_2^{\alpha_2} a_1^{\alpha_1},$$

where  $a_j = A_j^{-1}(\alpha_j)$  are the subsets of  $U$  for which the  $j$ th attribute takes the value  $\alpha_j$ .

**Definition 1.** An attribute  $A$  on a set  $U$  is a string of the type

$$a_n^{\alpha_n} a_{n-1}^{\alpha_{n-1}} \dots a_2^{\alpha_2} a_1^{\alpha_1}$$

together with a  $n + 1$ -ple  $(k; d_n, d_{n-1}, \dots, d_2, d_1)$  said outfit string such that:

- (1)  $a_i \subseteq U \quad \forall i = 1 \dots n \quad (a_i = A^{-1}(\alpha_i))$ .
- (2)  $\{a_n, a_{n-1}, \dots, a_2, a_1\}$  is a partition of  $U$ .
- (3)  $\alpha_j$ , called linguistic label, is a triangular fuzzy number lying in  $[0, 1] \quad \forall j = 1 \dots n$ .
- (4)  $\alpha_n > \alpha_{n-1} > \dots > \alpha_2 > \alpha_1$ .
- (5)  $k, d_n, d_{n-1}, \dots, d_2, d_1$  are positive integers.

In the following the elements  $a_n, a_{n-1}, \dots, a_2, a_1$  will be referred to as *first parts* whereas the term *second parts* will be used for the quantities  $\alpha_n, \alpha_{n-1}, \dots, \alpha_2, \alpha_1$ .

The outfit string is useful to keep track of the computational evolution of the string and its coefficients are such that the operations involved in the computation are associative. The attribute strings whose outfits consist exclusively of “1” are called *native* or *initial strings*.

5.2. Composing attribute strings

Given two attribute strings  $A$  and  $B$ , the operation of composition  $\diamond$  induces a new classification of  $U$  with respect to an attribute  $C$  obtained from  $A \diamond B$  via a suitable semantic rule. The operation is defined as follows:

**Definition 2.** Let  $\diamond : S(U) \times S(U) \rightarrow S(U)$  be the binary composition of attribute strings defined as follows: if  $A = a_n^{\alpha_n} a_{n-1}^{\alpha_{n-1}} \dots a_2^{\alpha_2} a_1^{\alpha_1}$  ( $k_A; d_{A,n}, \dots, d_{A,2}, d_{A,1}$ ) and

$$B = b_m^{\beta_m} b_{m-1}^{\beta_{m-1}} \dots b_2^{\beta_2} b_1^{\beta_1} \quad (k_B, d_{B,m}, \dots, d_{B,2}, d_{B,1})$$

are attribute strings, then

$$C = A \diamond B = c_{m+n-1}^{\gamma_{m+n-1}} \dots c_2^{\gamma_2} c_1^{\gamma_1} \quad (k_A + k_B; d_{C,m+n-1}, \dots, d_{C,2}, d_{C,1}),$$

where

- for  $n \geq m$  for the first parts one has:

$$c_i = \begin{cases} \bigcup_{j=1}^i (a_{i-j+1} \cap b_j) & 1 \leq i \leq m-1, \\ \bigcup_{j=1}^i (a_{i-j+1} \cap b_j) & m \leq i \leq n-1, \\ \bigcup_{j=i-n+1}^m (a_{i-j+1} \cap b_j) & n \leq i \leq m+n-1 \end{cases}$$

for the second parts:

$$\gamma_i = \begin{cases} \frac{1}{(k_A + k_B) * d_{C,i}} \sum_{j=1}^i d_{B,j} * d_{A,i-j+1} * (k_A * \alpha_{i-j+1} + k_B * \beta_j) & 1 \leq i \leq m-1, \\ \frac{1}{(k_A + k_B) * d_{C,i}} \sum_{j=1}^m d_{B,j} * d_{A,i-j+1} * (k_A * \alpha_{i-j+1} + k_B * \beta_j) & m \leq i \leq n-1, \\ \frac{1}{(k_A + k_B) * d_{C,i}} \sum_{j=i-n+1}^m d_{B,j} * d_{A,i-j+1} * (k_A * \alpha_{i-j+1} + k_B * \beta_j) & n \leq i \leq m+n-1, \end{cases}$$

where the values  $d_{C,i}$  are given by

$$d_{C,i} = \begin{cases} \sum_{j=1}^i d_{B,j} * d_{A,i-j+1} & 1 \leq i \leq m-1, \\ \sum_{j=i-n+1}^i d_{B,j} * d_{A,i-j+1} & m \leq i \leq n-1, \\ \sum_{j=i-n+1}^m d_{B,j} * d_{A,i-j+1} & m \leq i \leq m+n-1. \end{cases}$$

• for  $n \leq m$  for the first parts one has:

$$c_i = \begin{cases} \bigcup_{j=1}^i (a_{i-j+1} \cap b_j) & 1 \leq i \leq n-1, \\ \bigcup_{j=i-n+1}^i (a_{i-j+1} \cap b_j) & n \leq i \leq m-1, \\ \bigcup_{j=i-n+1}^m (a_{i-j+1} \cap b_j) & m \leq i \leq m+n-1 \end{cases}$$

for the second parts:

$$y_i = \begin{cases} \frac{1}{(k_A + k_B) * d_{C,i}} \sum_{j=1}^i d_{B,j} * d_{A,i-j+1} * (k_A * \alpha_{i-j+1} + k_B * \beta_j) & 1 \leq i \leq n-1, \\ \frac{1}{(k_A + k_B) * d_{C,i}} \sum_{j=i-n+1}^i d_{B,j} * d_{A,i-j+1} * (k_A * \alpha_{i-j+1} + k_B * \beta_j) & n \leq i \leq m-1, \\ \frac{1}{(k_A + k_B) * d_{C,i}} \sum_{j=i-n+1}^m d_{B,j} * d_{A,i-j+1} * (k_A * \alpha_{i-j+1} + k_B * \beta_j) & m \leq i \leq m+n-1, \end{cases}$$

where the values  $d_{C,i}$  are given by

$$d_{C,i} = \begin{cases} \sum_{j=1}^i d_{B,j} * d_{A,i-j+1} & 1 \leq i \leq n-1, \\ \sum_{j=i-n+1}^i d_{B,j} * d_{A,i-j+1} & n \leq i \leq m-1, \\ \sum_{j=i-n+1}^m d_{B,j} * d_{A,i-j+1} & m \leq i \leq m+n-1. \end{cases}$$

Moreover, for each attribute string  $A$  one has  $A \diamond U^{NI} = U^{NI} \diamond A = U^{NI}$  and  $A \diamond U^{NC} = U^{NC} \diamond A = A$ .

The idea of such an operation can be intuitively understood if we consider the two-fold meaning of a digit in a number, i.e., the roles of value and position. In the case of our strings, the absolute value of the generic

Table 3

Multiplication of integers		Composition $A \diamond B$	
	First parts	Additional factors	Second parts
Multiplication with carry	Intersection of sets	No carry multiplication	$d_{Bj} * d_{A_i-j+1} * (k_A * \alpha_{i-j+1} + k_B * \beta_j)$
Addition with carry	Union of sets	$d_{Bj} * d_{A_i-j+1}$ No carry addition	$(\sum_{j=1, \dots, i} d_{Bj} * d_{A_i-j+1} * (k_A * \alpha_{i-j+1} + k_B * \beta_j))$

element  $a_i$  is the set of the elements of  $U$  which  $a_i$  represents, whereas the value corresponding to the position is given by the linguistic label  $\alpha_i$ .

In the definition of the composition operator a vital role is played by the additional information present in each string. Suppose that the strings  $A$  and  $B$  are not initial, then the indices  $d_{A_i}$  and  $d_{B_i}$  represent the number of sets whose union have generated the  $i$ th class of  $A$  and  $B$ , respectively, whereas the indices  $k_A$  and  $k_B$  represent, in turn, the number of classes whose intersection have generated the classes of  $A$  and  $B$ , respectively. Thus these indices include the history of our string.

Moreover the quantity  $(\sum_{j=1, \dots, i} d_{Bj} * d_{A_i-j+1} * (k_A * \alpha_{i-j+1} + k_B * \beta_j))$  represents essentially a mean among the labels where each label takes a weight in some way related to its evolution. These weights allow to preserve the associativity for the operation  $\diamond$ , this would not be possible if we considered just the mean values of the labels.

The similarities between our composition and the conventional arithmetical operations, can be sketched by following Table 3.

Although the mechanism can appear complex, the following examples show that two strings can be easily composed in a way that vaguely resembles the ordinary multiplication of integers.

### 5.2.1. Example

Consider two strings that classify the set of colours  $U = \{\text{white, red, sky-blue, violet, yellow, brown}\} = \{w, r, s, v, y, b\}$  according to their degree of brightness and warmth

$$\text{Brightness} = \{w, s, y\}^{\text{high}} \{r\}^{\text{average}} \{v, b\}^{\text{low}}, \quad \text{Warmth} = \{r, y\}^{\text{high}} \{b, v\}^{\text{average}} \{s, w\}^{\text{low}},$$

where the labels high, average, low are represented by the following numbers:

$$\alpha_3 = \beta_3 = \text{high} = [0.7, 1, 1], \quad \alpha_2 = \beta_2 = \text{average} = [0.4, 0.5, 0.8], \quad \alpha_1 = \beta_1 = \text{low} = [0, 0.3, 0.5].$$

For the sake of simplicity, let  $A$  and  $B$  denote the strings *Brightness* and *Warmth*, respectively. Then computing the first parts for  $A \diamond B$ , one gets

$$\frac{\begin{array}{c} \{w, s, y\} \\ \{r, y\} \end{array} \quad \begin{array}{c} \{r\} \\ \{b, v\} \end{array} \quad \begin{array}{c} \{v, b\} * \\ \{s, w\} = \end{array}}{\begin{array}{c} \{s, w\} \cap \{w, s, y\} \\ \cup \\ \{b, v\} \cap \{w, s, y\} \\ \cup \\ \{r, y\} \cap \{w, s, y\} \end{array} \quad \begin{array}{c} \{s, w\} \cap \{r\} \\ \cup \\ \{b, v\} \cap \{r\} \\ \cup \\ \{r, y\} \cap \{r\} \end{array} \quad \begin{array}{c} \{s, w\} \cap \{v, b\} \\ \cup \\ \{b, v\} \cap \{v, b\} \\ \cup \\ \{r, y\} \cap \{v, b\} \end{array} \quad \begin{array}{c} \emptyset_1 \\ \{u, b\} \\ \{s, w\} \\ \{r\} \end{array}}$$

In order to compute the second parts we note that  $A$  and  $B$  are native strings, thus one can write:  $A = \{w, s, \gamma\}^{[0.7, 1, 1]} \{r\}^{[0.4, 0.5, 0.8]} \{v, b\}^{[0, 0.3, 0.5]} (1; 1, 1, 1)$ ,  $B = \{r, y\}^{[0.7, 1, 1]} \{b, v\}^{[0.4, 0.5, 0.8]} \{s, w\}^{[0, 0.3, 0.5]} (1; 1, 1, 1)$ . Then namely  $k_A = 1$ ,  $d_{A,3} = d_{A,2} = d_{A,1} = 1$ ,  $k_B = 1$ ,  $d_{B,3} = d_{B,2} = d_{B,1} = 1$ .

The factors  $d_{C,i}$  are computed as follows:

$$\begin{aligned} d_{C,1} &= d_{A,1} * d_{B,1} = 1, \\ d_{C,2} &= d_{A,1} * d_{B,2} + d_{A,2} * d_{B,1} = 1 * 1 + 1 * 1 = 2, \\ d_{C,3} &= d_{A,1} * d_{B,3} + d_{A,2} * d_{B,2} + d_{A,3} * d_{B,1} = 1 * 1 + 1 * 1 + 1 * 1 = 3, \\ d_{C,4} &= 2, d_{C,5} = 1. \end{aligned}$$

These factors can be computed also in the following way:

$$\begin{array}{c} d_{A,3} \quad d_{A,2} \quad d_{A,1} \quad \otimes \\ d_{B,3} \quad d_{B,2} \quad d_{B,1} \quad = \\ \hline \begin{array}{cccc} d_{3,1} & d_{2,1} & d_{1,1} & \\ \oplus & \oplus & \oplus & \\ d_{3,2} & d_{2,2} & d_{1,2} & - \\ \oplus & \oplus & \oplus & \\ d_{3,3} & d_{2,3} & d_{1,3} & - \\ \hline d_{C,5} & d_{C,4} & d_{C,3} & d_{C,2} \quad d_{C,1}, \end{array} \end{array}$$

where the symbol  $\otimes$  denotes the product with no carry and  $\oplus$  denotes the sum with no carry. One has the previous results.

Then the operation for the second parts produces

$$\begin{aligned} \gamma_1 &= 1 / ((k_A + k_B) * d_{C,1}) * (d_{B1} * d_{A1}) * (k_A * \alpha_1 + k_B * \beta_1) \\ &= 1/2 * 1 * (1 * [0, 0.3, 0.5] + 1 * [0, 0.3, 0.5]) = [0, 0.3, 0.5], \\ \gamma_2 &= 1 / ((k_A + k_B) * d_{C,2}) * \sum_{j=1..2} (d_{Bj} * d_{A,2-j+1}) * (k_A * \alpha_{2-j+1} + k_B * \beta_j) \\ &= 1/4 * \{(d_{B1} * d_{A2}) * (1 * \alpha_2 + 1 * \beta_1) + (d_{B2} * d_{A1}) * \\ &\quad * (1 * \alpha_1 + 1 * \beta_2)\} = 1/4 * \{1 * 1 * ([0.4, 0.5, 0.8] + [0, 0.3, 0.5]) \\ &\quad + 1 * 1 * ([0, 0.3, 0.5] + [0.4, 0.5, 0.8])\} = [0.2, 0.4, 0.65], \\ \gamma_3 &= 1 / ((k_A + k_B) * d_{C,3}) * \sum_{j=1..3} (d_{Bj} * d_{A,2-j+1}) * (k_A * \alpha_{2-j+1} + k_B * \beta_j) \\ &= 1/6 * \{(d_{B1} * d_{A3}) * (1 * \alpha_3 + 1 * \beta_1) + (d_{B2} * d_{A2}) * (1 * \alpha_2 + 1 * \beta_2) + (d_{B3} * d_{A1}) \\ &\quad * (1 * \alpha_1 + 1 * \beta_3)\} = 1/6 * \{1 * 1 * (1 * [0.7, 1, 1] + 1 * [0, 0.3, 0.5]) + 1 * 1 * (1 * [0.4, 0.5, 0.8] \\ &\quad + 1 * [0.4, 0.5, 0.8]) + 1 * 1 * (1 * [0, 0.3, 0.5] + 1 * [0.7, 1, 1])\} = [0.37, 0.6, 0.77], \\ \gamma_4 &= [0.55, 0.75, 0.9], \\ \gamma_5 &= [0.7, 1, 1]. \end{aligned}$$

By juxtaposing the results from both parts, we obtain:

$$\text{Brightness} \diamond \text{Warmth} = \{y\}^{[0.7, 1, 1]} \{r\}^{[0.55, 0.75, 0.9]} \{s, w\}^{[0.36, 0.6, 0.76]} \{v, b\}^{[0.2, 0.4, 0.65]} \emptyset^{[0, 0.3, 0.5]}.$$

### 5.3. The linguistic approximation

In order to improve readability it is worth, given a basic set of linguistic expressions, translating the fuzzy numbers of the second parts into suitable expressions [10]. Thus we have to devise a procedure coping with this problem of linguistic approximation.

Given a set of linguistic labels  $E = \{e_1, e_2, \dots, e_k\}$ , as represented by the triangular numbers  $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_k\}$ , where  $\alpha_i = [l_i, m_i, r_i]$ , and given a fuzzy number  $\beta = [l, m, r]$ , the problem of approximating  $\beta$  in  $E$  consists in singling out a suitable linguistic expression for  $\beta$  in terms of  $e_1, e_2, \dots, e_k$ . This problem can be tackled as follows:

Suppose that the mean value of  $\beta$  lies between those of  $\alpha_i$  and  $\alpha_{i+1}$ , i.e.  $m \in [m_i, m_{i+1}]$ , and let  $d = m_{i+1} - m_i$ , then the following approximation can be carried out:

- (i) if  $m \in [m_i, m_i + d/10]$  then  $\beta$  is approximated by the expression  $e_i$ ;
- (ii) if  $m \in [m_i + (d/10), m_i + (3/10) * d]$  then  $\beta$  is “next to”  $\alpha_i$  and  $\beta$  is approximated by  $n[e_i]$ ;
- (iii) if  $m \in [m_i + (3/10) * d, m_i + (7/10) * d]$  then  $\beta$  is “included between”  $\alpha_i$  and  $\alpha_{i+1}$  and  $\beta$  is approximated by  $ib[e_i, e_{i+1}]$ ;
- (iv) if  $m \in [m_i + (7/10) * d, m_i + (9/10) * d]$  then  $\beta$  is “before”  $\alpha_i + 1$  and  $\beta$  is approximated by  $b[e_{i+1}]$ ;
- (v) if  $m \in [m_i + (9/10) * d, m_i + 1]$  then  $\beta$  is approximated by  $e_{i+1}$ .

It is worth noting that this approximation procedure gives an upper bound to the number of possible labels: namely their number cannot exceed the value  $4k - 3$ , where  $k$  is the number of labels in  $E$ .

Suppose we aim to build the linguistic approximation for  $[0.55 \ 0.75 \ 0.9]$ . For  $m = 0.75 < 1$  then we consider  $\alpha_2$  (average) and  $\alpha_3$  (high). We have  $m_2 = 0.5$  and  $m_3 = 1$  with  $d = m_3 - m_2 = 0.5$ . For  $0.75 \in [0.5 + 3/10 * 0.5, 0.5 + 7/10 * 0.5] = [0.65, 0.85]$ , by (iii) we have  $[0.55, 0.75, 0.9] \approx ib(\text{average}, \text{high})$ . The final result is

$$\text{Brightness} \diamond \text{Warmth} = \{y\}^{\text{high}} \{r\}^{\text{ib(average, high)}} \{s, w\}^{\text{nt(average)}} \{v, b\}^{\text{ib(low, average)}} \emptyset^{\text{low}}.$$

### 5.4. Algebraic properties

We briefly recall basic properties of the algebraic structure, for further details the reader is referred to [7]. It is worth noting that these properties hold for the set  $S(U)$  without considering the linguistic approximation.

**Proposition 1.** *The couple  $\langle S(U), \diamond \rangle$  is a commutative semigroup.*

A total ordering can be introduced on  $S(U)$ , based on the information contained within each string. For a specific string  $A$ , the higher the value of  $k_A$ , the higher the number of attributes whose composition has given the string. Thus we can arrange strings in decreasing order as regards the amount of information. Hence one proves that the structure  $\langle S(U), \diamond, \leq, \text{min}, \text{max} \rangle$  is a totally ordered commutative semigroup and is endowed with the related modular distributive lattice. Finally, using the labels  $NC$  and  $NI$  whose meanings are “not classifiable” and “lack of information”, one can prove our main result:

**Proposition 2.** *The quadruple  $\langle S(U), \diamond, NI, NC, \leq \rangle$  is a commutative l-monoid.*

### 5.5. The weight of attributes

Usually, during classification procedures, each element is more or less relevant as compared with other ones. Thus it would be useful to devise a weighting mechanism in order to assign different relevance degrees to the attributes. The position of the elements in the string is affected by their values as regards the attributes.

From a formal point of view, let  $w$  be a positive number denoting the weight and let  $A$  be the attribute to be weighted:  $A = a_n^{a_n} a_{n-1}^{a_{n-1}} \dots a_2^{a_2} a_1^{a_1}$ .

Then for each class  $a_i$  with label  $\alpha_i$ :

- Construct  $N_{\alpha_i}$  new empty classes, with  $N_{\alpha_i} = \text{Int}(m_{\alpha_i} * w)$ , where  $m_{\alpha_i}$  is the mean value of the fuzzy number represented by the label  $\alpha_i$ , and  $\text{Int}$  denotes the function which furnishes the integer part of the argument.
- Move the set  $a_i$ , along with its label  $\alpha_i$ ,  $N_{\alpha_i}$  positions at left.
- For each  $i \in \{2, 3, \dots, n\}$ , increasing labels are attached to the  $N_{\alpha_i}$  new classes. These labels take values in the range  $(\alpha_{i-1}, \alpha_i)$  according to the formula:

$$s_{ij} = \alpha_{i-1} + \left\{ \frac{j * (\alpha_i - \alpha_{i-1})}{N_{\alpha_i} + 1} \right\}, \quad \text{where } j = 1, \dots, N_{\alpha_i}.$$

#### 5.2.2. Example

We consider the attributes *Brightness* and *Warmth* previously described, respectively, by the strings  $[w, s, y]^{\text{High}} [r]^{\text{Average}} [v, b]^{\text{Low}}$  and  $[r, y]^{\text{High}} [b, v]^{\text{Average}} [s, w]^{\text{Low}}$ . Our goal is to paint a picture very bright. Thus we could assign weight 3 to Brightness attribute.

By applying the weighting mechanism one gets with  $w = 3$ :

$$\begin{aligned} N_{\alpha_3} &= \text{Int}(1 * 3) = 3, \\ s_{31} &= \alpha_2 + 1 / (N_{\alpha_3} + 1) * (\alpha_3 - \alpha_2) = [0.47, 0.69, 0.85], \\ s_{32} &= \alpha_2 + 2 / (N_{\alpha_3} + 1) * (\alpha_3 - \alpha_2) = [0.55, 0.75, 0.9], \\ s_{33} &= \alpha_2 + 3 / (N_{\alpha_3} + 1) * (\alpha_3 - \alpha_2) = [0.62, 0.87, 0.95], \\ N_{\alpha_2} &= \text{Int}(0.5 * 3) = 1, \\ s_{21} &= \alpha_1 + 1 / (N_{\alpha_2} + 1) * (\alpha_2 - \alpha_1) = [0.47, 0.69, 0.85], \\ N_{\alpha_1} &= \text{Int}(0.3 * 3) = 0. \end{aligned}$$

Now the attribute *Brightness* is described by the new string:

$$\text{Brightness}' = [y, s, w]^{\text{High}} [-]^{\text{333}} [-]^{\text{332}} [-]^{\text{331}} [r]^{\text{Average}} [-]^{\text{21}} [b, v]^{\text{Low}}.$$

By composing these strings one obtains

$$\text{Brightness}' \diamond \text{Warmth} = [y]^{\text{High}} [s, w]^{\text{thAverage, High}} [r]^{\text{nt[Average, High]}} [v, b]^{\text{th[Low, Average]}}.$$

It is worth noting that changing the weight for *Brightness*, brightest colours (sky-blue and white) have improved their ranking to the prejudice of less bright ones.

## 6. Fuzzy diagnostics

### 6.1. Analysing the first sample

First, we aim at deducing a relationship between clinical data and neuropathy. Table 1 reports the first sample of patients. The starting point are the attribute strings, obtained from the information about 17 patients (see Section 4.3).

Strings should be suitably composed so that the risk factor for each patient can be determined. Medical expertise is useful to attach the relevance to each diagnostic element with respect to neuropathy. In such way, weights are associated with each attribute string.

More specifically, it is generally accepted by physicians that:

- (i) Effectively diabetes treating prevents in some degree symptoms of neuropathy.
- (ii) Early diabetes deeply affects the insurgence of neuropathy.

To take into account (i) we have to construct a string representing the effectiveness of diabetic treatment beginning from the dose of insulin and the patient's answer as measured by *HbA1c*: the higher the value of *HbA1c* the higher the risk factor for neuropathy. This fact induces to give *HbA1c* a weight greater than that attached to *Insu*. Thus we have to compute the following string:  $Effec = 4 * HbA1c \diamond Insu$ .

The computation yields:

$$Effec = [3]^{[0.54, 0.76, 0.89]} [9, 11, 14]^{[0.51, 0.73, 0.87]} [7, 8, 10, 15]^{[0.47, 0.69, 0.84]} [12]^{[0.4, 0.6, 0.77]} [1, 2, 5]^{[0.363, 0.548, 0.721]} [4, 6, 13, 16, 17]^{[0.31, 0.46, 0.65]}$$

Next step is to take into account point (ii) that asserts the relevance of the attribute *Diab-Age*. The weights  $w_E$  and  $w_D$  associated with *Effec* and *Diab-Age*, respectively, are to be singled out, the training step has suggested the values  $w_E = 2$  and  $w_D = 4$ .

Thus we can write:  $Diag = 2 * Effec \diamond 4 * Diab-Age \diamond AER \diamond GFR \diamond Fluor \diamond BdB \diamond Fundus$ .

By composing these strings one obtains

$$\begin{aligned} Diag = & [-] \cdot \dots [-] [9]^{[0.409, 0.617, 0.766]} [-] \cdot \dots [-] [8]^{[0.393, 0.600, 0.754]} [15]^{[0.387, 0.593, 0.748]} [-] \\ & \times [7]^{[0.373, 0.575, 0.735]} [3, 10]^{[0.364, 0.565, 0.727]} [11, 14]^{[0.355, 0.555, 0.719]} [1]^{[0.345, 0.544, 0.710]} \\ & \times [2]^{[0.335, 0.532, 0.701]} [6]^{[0.324, 0.520, 0.691]} [12]^{[0.313, 0.507, 0.680]} [5]^{[0.300, 0.492, 0.667]} \\ & \times [16]^{[0.287, 0.475, 0.652]} [-] [4, 17]^{[0.256, 0.434, 0.617]} [13]^{[0.238, 0.410, 0.596]} [-] \cdot \dots [-]. \end{aligned}$$

By applying the linguistic approximation to fuzzy numbers corresponding to non-empty classes, we get the following labels:

$$\begin{aligned} \gamma_8 &= [0.238, 0.410, 0.596] \approx LR; & \gamma_9 &= [0.256, 0.434, 0.617] \approx m[LR], \\ \gamma_{11} &= [0.287, 0.475, 0.652] \approx ib[LR, R]; & \gamma_{12} &= [0.300, 0.492, 0.667] \approx ib[LR, R], \\ \gamma_{13} &= [0.313, 0.507, 0.680] \approx b[R]; & \gamma_{14} &= [0.324, 0.520, 0.691] \approx b[R], \\ \gamma_{15} &= [0.335, 0.532, 0.701] \approx b[R]; & \gamma_{16} &= [0.345, 0.544, 0.710] \approx R, \\ \gamma_{17} &= [0.355, 0.555, 0.719] \approx R; & \gamma_{18} &= [0.364, 0.565, 0.727] \approx R, \\ \gamma_{19} &= [0.373, 0.575, 0.735] \approx R, \\ \gamma_{21} &= [0.387, 0.593, 0.748] \approx R, \end{aligned}$$

Table 4  
Comparing diagnoses for the first sample. Yes stands for “ill”,  
no for “healthy”.

No.	Clinical diagnosis	Fuzzy-based results
1	Yes	R
2	No	b[R]
3	Yes	R
4	Yes	ni[LR]
5	No	ib[LR,R]
6	No	b[R]
7	Yes	R
8	Yes	ni[R]
9	Yes	ni[R]
10	Yes	R
11	Yes	R
12	No	b[R]
13	No	LR
14	Yes	R
15	Yes	R
16	Yes	ib[LR,R]
17	No	ni[LR]

$$\gamma_{22} = [0.393, 0.600, 0.754] \approx ni[R],$$

$$\gamma_{25} = [0.409, 0.617, 0.766] \approx ni[R],$$

and the final string is  $Diag = [8, 9]^{ni[R]}[1, 3, 7, 10, 11, 14, 15]^R[2, 6, 12]^{b[R]}[5, 16]^{ib[LR,R]}[4, 17]^{ni[LR]}[13]^{LR}$ .

We choose the following interpretation for the approximated labels:

$ni[R]$  = more than risk,

$b[R]$  = almost at risk,

$ib[LR,R]$  = intermediate between “small risk” and “at risk”,

$ni[LR]$  = more than “small risk”.

Then the interpretation of the string is as follows: *the risk factor is: very high for patients 8 and 9, high for 1, 3, 7, 10, 11, 14, 15, low for patients 2, 5, 6, 12, 16, very low for 4, 13, 17.*

Comparing our results with those reported in Table 4 we note that medical diagnosis and fuzzy-based diagnosis essentially coincide.

There are only two cases where fuzzy-based results and clinical ones disagree: patients 4 and 16. However it is worth noting that:

(a) patient 16 has clinical values very close to patient 5, so both are classified as “healthy”. It is possible that other factors have escaped the clinical analysis.

(b) patient 4 is ill, however only few clinical values are pathological and this fact has caused the wrong attribution.

## 6.2. Analysing the second sample

We have considered a second sample of 16 patients, the following attribute strings describe the situation:

Table 5  
Comparing diagnoses for the second sample

No.	Medical diagnosis	Risk factor	No.	Medical diagnosis	Risk factor
1	Yes	$m[R]$	9	No	$b[LR]$
2	No	$ib[LR, R]$	10	No	$b[LR]$
3	No	$b[R]$	11	No	$b[LR]$
4	No	$m[R]$	12	No	$m[LR]$
5	No	$ib[LR, R]$	13	No	$ib[LR, R]$
6	No	$b[LR]$	14	No	$ib[LR, R]$
7	No	$m[LR]$	15	No	$b[R]$
8	No	$b[LR]$	16	No	$b[R]$

$$Diab-Age = [3]^{ve}[1, 5, 7, 12, 13, 15, 16]^c[2, 4, 6, 8, 9, 10, 11, 14]^g[-]^i,$$

$$HbA1c = [-]^{ch}[4, 13, 14, 15, 16]^h[1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12]^q[-]^n,$$

$$AER = [2, 5]^{sh}[4]^h[-]^q[1, 3, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16]^n,$$

$$Insu = [1, 4]^h[-]^m[2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16]^l,$$

$$BdB = [-]^p[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16]^n,$$

$$Fundus = [-]^p[1, 2, 14, 15, 16]^p[3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13]^n.$$

The computation yields

$$\begin{aligned}
 Diag &= 2 * Effec \diamond 3 * Diab-Age \diamond AER \diamond BdB \diamond Fundus \\
 &= [4]^{[0.433, 0.636, 0.790]}[1]^{[0.406, 0.609, 0.768]}[3, 15, 16]^{[0.313, 0.512, 0.685]}[5, 13]^{[0.297, 0.494, 0.670]}[2, 14]^{[0.261, 0.455, 0.634]} \\
 &\quad \times [7, 12]^{[0.241, 0.431, 0.611]}[6, 8, 9, 10, 11]^{[0.171, 0.336, 0.517]} \\
 &= [1, 4]^{m[R]}[-]^g[3, 15, 16]^h[2, 5, 13, 14]^h[LR, R][7, 12]^m[LR][6, 8, 9, 10, 11]^b[LR].
 \end{aligned}$$

Comparing our results with medical response we have the following results given in Table 5.

Patient 4 is the one for which diagnoses disagree: in fact we consider him at risk, whereas physicians exclude this possibility.

### 7. The importance of symptoms

#### 7.1. The relevance function

It is sometimes important to grasp an idea about the way a string attribute can affect the generation of a string, the concept of *relevance* is useful to this aim [13]. In particular this concept will be useful in order to single out the medical parameters more relevant for diagnosing the disease.

From the intuitive point of view, when two strings *A* and *B* are composed into  $A \diamond B$  in order to get *C* essentially we “shake up” the information in *A* and *B* and put it into *C*.

The relevance measures the ability of *A* in influencing *B* (and vice-versa) in order to get *C*. We give two different measures of relevance showing that each attribute’s importance is double-faced: an absolute relevance

and a relevance related to the composed string, as the operation of composition affects both first and second parts.

From the formal point of view one has:

**Definition 3.** Let  $A = a_1^{a_m} \dots a_2^{a_2} a_1^{a_1}$  be an attribute string and let  $C = c_1^{c_m} \dots c_2^{c_2} c_1^{c_1}$  be the result of the composition of  $A$  with other attributes (e.g.  $C = A \diamond B \diamond D$ ), then the relevance of  $A$  is:

$$R_A^C = \mu_A^C * \rho_A,$$

where

- $\mu_A^C$  is the relative relevance of  $A$  with regard to  $C$  and is computed as follows:

$$\mu_A^C = \sum_{j=1, \dots, n} \sum_{k=1, \dots, m} (p_{jk} - p_j)^2,$$

where  $p_{jk} = \#[A^{-1}(\alpha_j) \cap c_k]/m$  and  $p_j = \#[A^{-1}(\alpha_j)]/m$ ;

- $\rho_A$  is the absolute relevance and is computed as follows:

$$\rho_A = \begin{cases} 0 & \text{if } A = U^{NC}, \\ 1 & \text{if } A = U^{NI}, \\ 2/\pi * \arctg \left\{ \frac{n * (n - 1) * (k_A + \sum_{j=1, \dots, n} d_{A,i})}{2 * \sum_{i=1, \dots, n-1} \sum_{j=i+1, \dots, n} |\alpha_i - \alpha_j|} \right\} & \text{otherwise.} \end{cases}$$

where if  $\alpha_i = [\alpha_{i1}, \alpha_{i2}, \alpha_{i3}]$  and  $\alpha_j = [\alpha_{j1}, \alpha_{j2}, \alpha_{j3}]$ , then  $|\alpha_i - \alpha_j| = |\alpha_{i1} - \alpha_{j1}| + |\alpha_{i2} - \alpha_{j2}| + |\alpha_{i3} - \alpha_{j3}|$ .

It is worth noting that the relative relevance of  $A$  with respect to  $C$  is calculated by using only the first parts of the strings. This relevance represents the quantity of information induced by the attribute  $A$  on  $C$ . In some way it resembles the concept of variance in statistics and, in fact, by  $\mu_A^C$  we measure, in a certain sense, the deviation of the distribution of the first parts of  $A$  in  $C$  from their uniform distribution.

On the other hand, the absolute relevance of  $A$  does not include any element of  $C$ . It is a quantity related to only the information contained in  $A$ . We remember that the operation for second parts is based on the mean among fuzzy numbers, and in case a string consists of many clusters their labels will appear often in the computation of means.

Indeed in the calculus of a label  $\gamma_i$  (see Section 5.2), more the label  $\alpha_i$  of  $A$  is relevant more the coefficients  $k_A$  and  $d_{A,i}$  are high. Thus the relevance of a string is to be directly proportional to the quantity  $(k_A + \sum_{i=1, \dots, n} d_{A,i})$  and inversely proportional to the dispersion of the labels of  $A$  in  $[0, 1]$ , representable as follows:

$$\left( \sum_{i=1, \dots, n-1} \sum_{j=i+1, \dots, n} |\alpha_i - \alpha_j| \right) / (n * (n - 1) / 2).$$

The sum is divided by  $n * (n - 1) / 2$  to find the mean of all differences  $|\alpha_i - \alpha_j|$ .

Moreover the relevance of  $U^{NC}$  equals zero because this string does not affect the string to which it is applied ( $A \diamond U^{NC} = A$ , for every  $A$ ), whereas the relevance for  $U^{NI}$  equals 1 because “adsorbs” any string ( $A \diamond U^{NI} = U^{NI}$ , for every  $A$ ).

### 7.1.1. Example

Consider the previous strings

Brightness =  $\{w, s, y\}^{High} \{r\}^{Average} \{v, b\}^{Low}$  and Warmth =  $\{r, y\}^{High} \{b, v\}^{Average} \{s, w\}^{Low}$ , and let  $C =$

Brightness  $\diamond$  Warmth =  $\{y\}^{High} \{r\}^{ib[Average, High]} \{s, w\}^{ni[Average, High]} \{v, b\}^{ib[Low, Average]} \{s, w\}^{Average} \{s, w\}^{Low}$ .

Then the relative relevance of *Brightness* with respect to *C* is given by

$$\mu_{\text{BRIGHTNESS}}^C = \sum_{j=1, \dots, 3} \sum_{k=1, \dots, 5} (p_{j,k} - p_j)^2,$$

where  $m = 5$ ;  $p_1 = \#\{\{v, b\}\}/m = 2/6 = 1/3$ ;  $p_2 = \#\{\{r\}\}/m = 1/6$ ;  $p_3 = \#\{\{w, s, y\}\}/m = 1/2$  and

$$p_{1,1} = \#\{\{v, m\} \cap \emptyset\} = 0, p_{2,1} = 0, p_{3,1} = 0, p_{1,2} = \#\{\{v, m\} \cap \{v, m\}\} = 2, p_{2,2} = 0, p_{3,2} = 0,$$

$$p_{1,3} = \#\{\{v, m\} \cap \{c, b\}\} = 0, p_{2,3} = 0, p_{3,3} = 2, p_{1,4} = \#\{\{v, m\} \cap \{r\}\} = 0, p_{2,4} = 1, p_{3,4} = 0,$$

$$p_{1,5} = \#\{\{v, m\} \cap \{g\}\} = 0, p_{2,5} = 0, p_{3,5} = 1$$

thus

$$\begin{aligned} \mu_{\text{BRIGHTNESS}}^C &= (0 - 1/3)^2 + (2 - 1/3)^2 + (0 - 1/3)^2 + (0 - 1/3)^2 + (0 - 1/3)^2 + (0 - 1/6)^2 \\ &\quad + (0 - 1/6)^2 + (0 - 1/6)^2 + (1 - 1/6)^2 + (0 - 1/6)^2 + (0 - 1/2)^2 \\ &\quad + (0 - 1/2)^2 + (2 - 1/2)^2 + (0 - 1/2)^2 + (1 - 1/2)^2 = 7.2. \end{aligned}$$

The absolute relevance of *Brightness* is

$$\sum_{i=1, \dots, 2} \sum_{j=i+1, \dots, 3} |\alpha_i - \alpha_j| = (|0.5 - 0.8| + |0.3 - 0.5| + |0 - 0.4|) + (|0.5 - 1| + |0.3 - 1| + |0 - 0.7|)$$

$$+ (|0.8 - 1| + |0.5 - 1| + |0.4 - 0.7|) = 3.8$$

and

$$\rho_{\text{BRIGHTNESS}} = 2/\pi \arctg \left\{ \frac{(3 * 2) * (1 + 1 + 1 + 1)}{2 * 3.8} \right\} = 0.805.$$

Thus the relevance of *Brightness* is  $R_{\text{BRIGHTNESS}}^C = \mu_{\text{BRIGHTNESS}}^C * \rho_{\text{BRIGHTNESS}} = 7.2 * 0.805 = 5.796$ .

For the string *Warmth* one gets

$$\mu_{\text{WARMTH}}^C = 7.6$$

$$\rho_{\text{WARMTH}} = 2/\pi \arctg \left\{ \frac{(3 * 2) * (1 + 1 + 1 + 1)}{2 * 3.8} \right\} = 0.805.$$

$$R_{\text{WARMTH}}^C = \mu_{\text{WARMTH}}^C * \rho_{\text{WARMTH}} = 7.6 * 0.805 = 6.118.$$

## 7.2. Relevance of symptoms for the first sample

In the Section (6.1) we have got for the first sample the following final diagnosis:

$$\begin{aligned} \text{Diag} &= 2 * \text{Effec} \diamond 4 * \text{Diab-Age} \diamond \text{AER} \diamond \text{GFR} \diamond \text{Fluor} \diamond \text{BdB} \diamond \text{Fundus} \\ &= [8, 9]^{m[R]} [1, 3, 7, 10, 11, 14, 15]^R [2, 6, 12]^{pl[R]} [5, 16]^{pl[L,R,R]} [4, 17]^{pl[L,R]} [13]^{LR}. \end{aligned}$$

It is worth noting that the attributes *HbA1c* and *Insu* do not appear directly in the composition but are strictly coupled in the string *Effec*. This means that the parameters are to be considered as strictly related in order to get the correct diagnosis. Thus also computing their relevance is to be carried out indirectly, i.e., we extend the results obtained for the relevance of *Effec*.

Using Definition 3 we are able to compute the relative, absolute and overall relevance for each of the attributes present in *Diag*:

- Relevance for *Effec*:  $\mu_{Effec}^{Diag} = 42$ ;  $\rho_{Effec} = 0.995$ ;  $R_{Effec}^{Diag} = 41.782$ .
- Relevance for *Diab-Age*:  $\mu_{Diab-Age}^{Diag} = 32.360$ ;  $\rho_{Diab-Age} = 0.950$ ;  $R_{Diab-Age}^{Diag} = 30.754$ .
- Relevance for *AER*:  $\mu_{AER}^{Diag} = 38.909$ ;  $\rho_{AER} = 0.866$ ;  $R_{AER}^{Diag} = 33.703$ .
- Relevance for *GFR*:  $\mu_{GFR}^{Diag} = 28.364$ ;  $\rho_{GFR} = 0.842$ ;  $R_{GFR}^{Diag} = 23.833$ .
- Relevance for *Fluor*:  $\mu_{Fluor}^{Diag} = 32.727$ ;  $\rho_{Fluor} = 0.834$ ;  $R_{Fluor}^{Diag} = 27.298$ .
- Relevance for *BdB*:  $\mu_{BdB}^{Diag} = 44.727$ ;  $\rho_{BdB} = 0.834$ ;  $R_{BdB}^{Diag} = 37.307$ .
- Relevance for *Fundus*:  $\mu_{Fundus}^{Diag} = 27.636$ ;  $\rho_{Fundus} = 0.814$ ;  $R_{Fundus}^{Diag} = 22.509$ .

We expect that the parameter *Effec* might affect the insurgence of neuropathy at the highest degree, whereas the parameters *Fluor*, *GFR* and *Fundus* are less relevant for the disease. More specifically the parameter *Fundus* barely affects this pathology [3,18].

### 7.3. Relevance of symptoms for the second sample

Computing the relative, absolute and overall relevance for each of the attributes concerning the second sample (Table 2), one has:

- Relevance for *Effec*:  $\mu_{Effec}^{DIAG} = 34$ ;  $\rho_{EFFEC} = 0.995$ ;  $R_{EFFEC}^{DIAG} = 33.82$ .
- Relevance for *Diab-Age*:  $\mu_{Diab-Age}^{DIAG} = 32.6$ ;  $\rho_{Diab-Age} = 0.942$ ;  $R_{Diab-Age}^{DIAG} = 30.708$ .
- Relevance for *AER*:  $\mu_{AER}^{DIAG} = 30.6$ ;  $\rho_{AER} = 0.866$ ;  $R_{AER}^{DIAG} = 26.506$ .
- Relevance for *BdB*:  $\mu_{BdB}^{DIAG} = 32.4$ ;  $\rho_{BdB} = 0.834$ ;  $R_{BdB}^{DIAG} = 27.025$ .
- Relevance for *Fundus*:  $\mu_{Fundus}^{DIAG} = 29.4$ ;  $\rho_{Fundus} = 0.814$ ;  $R_{Fundus}^{DIAG} = 23.945$ .

We can infer that *Effec* and *Diab-Age* are most relevant for the disease, *BdB* and *AER* are quite relevant, whereas *Fundus* is not relevant at all.

## 8. Experimentation results

In this section we aim to assess the robustness of our model. In particular, we aim to verify the consistence of data results related to the insurgence of pathology according to the relevance of the attributes. More specifically, we carry out the following case studies:

*Case study 1*: Take into account patients really ill and classified by us as “at risk”. We wonder whether giving “normal” values to most relevant attributes, ill patients become “not at risk”.

*Case study 2*: Consider the same sub-sample as before of ill patients classified as “at risk” and suppose that less relevant attributes are set to normal values. It is interesting to check whether in this case ill patients are still classified as “at risk”.

*Case study 3*: Take into account healthy patients classified as “not at risk”. Suppose that the most relevant attributes are given values meaning serious illness, then healthy patients, along with ill ones, should become “at risk”.

**Case study 4:** Take into account healthy patients classified as “not at risk”. Suppose that high values are given to the least relevant attributes, then healthy patients should remain “not at risk”.

### 8.1. Changing values in the first sample

#### 8.1.1. Case study 1: Giving normal values to most relevant attributes (ill patients)

We take into account patients really ill and classified by us as “at risk”. We show that giving normal values to most relevant attributes, ill patients become “not at risk”. To accomplish this we lower the importance of symptoms related to most relevant attributes.

Owing to the relevance analysis carried out on the first sample we realised that most relevant attributes are:

$$Effec (R = 41.782), AER (R = 33.703), Diab-Age (R = 30.754), BdB (R = 37.307).$$

Suppose that ill patients take normal values for these attributes. Of course, as *Effec* is composed by *HbA1c* and *Insu*, it is necessary that changes be made to both elements. Attribute strings are rewritten as follows (underlining is used for ill patients in order to single out them immediately).

$$(a) \quad HbA1c = [-]^{wh} [3, 8, 9, 10, \underline{11}, \underline{14}, \underline{15}]^h [1, 2, 4, 5, 6, 12, 13, \underline{16}, 17]^a [-]^{wh} [-]^{wh} [2, 5, 6, 12, 13, 17]^a [1, 3, 4, 7, 8, 9, 10, \underline{11}, \underline{14}, \underline{15}, \underline{16}]^{wh} \text{ and } Insu = [3, 12]^h [1, 2, 5, 9, \underline{11}, \underline{14}]^{wh} [4, 6, 7, 8, \underline{10}, 13, \underline{15}, \underline{16}, 17]^l \text{ becomes } Ins-1a = [12]^h [2, 5]^w [1, 3, 4, 6, 7, 8, 9, \underline{10}, \underline{11}, 13, \underline{14}, \underline{15}, \underline{16}, 17]^l.$$

$$(b) \quad AER = [2, 12]^{wh} [9, 15]^{wh} [8, 16]^{wh} [1, 3, 4, 5, 6, 7, 10, \underline{11}, 13, \underline{14}, 17]^w \text{ becomes } AER-1b = [2, 12]^{wh} [-]^{wh} [-]^{wh} [1, 3, 4, 5, 6, 7, 8, 9, 10, \underline{11}, 13, \underline{14}, \underline{15}, \underline{16}, 17]^w.$$

$$(c) \quad Diab-Age = [1, 6]^{wh} [2, 4, 5, 8, 9, 15, 16, 17]^e [3, 7, 10, \underline{11}, 12, 13, \underline{14}]^a [-]^{wh} \text{ becomes } Dia-1c = [6]^{wh} [2, 5, 17]^e [12, 13]^a [1, 3, 4, 7, 8, 9, 10, \underline{11}, 14, \underline{15}, \underline{16}]^l.$$

(d) As regards the string

$$BdB = [-]^{wh} [-]^{wh} [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, \underline{11}, 12, 13, 14, \underline{15}, \underline{16}, 17]^w$$

we note that the lowest values are already present thus it is affected by no changes.

Thus the new composition for effectiveness is

$$E-1a = 4 * Hb-1^o \diamond Ins-1a \\ = [12]^{[0,4,0,6,1,0,77]} [2, 5]^{[0,3,6,0,54,0,72]} [6, 13, 17]^{[0,3,1,0,4,6,0,65]} \\ [1, 3, 4, 7, 8, 9, 10, \underline{11}, \underline{14}, \underline{15}, \underline{16}]^{[0,1,0,25,0,45]}.$$

The analysis is now carried out as usual, with the only provision that most relevant string are different. The new composition is

$$DIAG-1 = 2 * E-1a \diamond 4 * Dia-1c \diamond AER-1b \diamond Fluor \diamond GFR \diamond BdB \diamond Fundus,$$

whose value is:

$$DIAG-1 = [-]^{wh[R]} [-]^{wh} [2, 6, 12]^{wh[R]} [5]^{wh[L,R,R,R]} [17]^{wh[L,R,R]} [13]^{LR} [7, 8, 9, 10]^{wh[L,R]} [1, 3, 4, 11, 14, 15, 16]^{wh[NR, LR]}.$$

We note that the original diagnosis was:

$$DIAG = [8, 9]^{wh[R]} [1, 3, 7, 10, \underline{11}, \underline{14}, \underline{15}]^R [2, 6, 12]^{wh[R]} [5, 16]^{wh[L,R,R]} [4, 17]^{wh[L,R]} [13]^{LR} [-]^{wh[L,R]} [-]^{wh[NR, LR]}.$$

We see that all ill patients by originally classified as “at risk” change their status if most relevant attributes are given by normal values.

that this result occurs when all relevant attributes are changed. If we are given normal values, we would expect that the risk factor is reduced proportionally to the attribute.

If the attribute *Effec* is given normal values. Then we have:

$$\diamond 4 * Diab-Age \diamond AER \diamond Fluor \diamond GFR \diamond BdB \diamond Fundus$$

$$= [2, 6, 9, 12]^{b[R]} [1, 5, 8, 15]^{b[LR,R]} [7, 16, 17]^{m[LR]} [4, 10, 13]^{LR} [3, 11, 14]^{b(LR)}.$$

patient with normal values of *HbA1c* and *Insu* (and consequently *Effec*) is not at risk, since *Effec* is the most relevant attribute. What happens in case when only the attribute *AER* takes normal values:

$$\diamond 4 * Diab-Age \diamond AER-1b \diamond Fluor \diamond GFR \diamond BdB \diamond Fundus$$

$$= [3, 7, 8, 10, 11, 14, 15]^{R} [2, 6, 12]^{b[R]} [5, 16]^{b[LR,R]} [4, 17]^{m[LR]} [13]^{LR}.$$

is result is that decreasing the value given to the attribute *AER* does not induce a change in the risk factor. We note that the relevance of *AER* is high, however its influence is not significant.

In the case that only the attribute *Diab-Age* takes normal values. We have:  $DIAG-1C = Fluor \diamond GFR \diamond BdB \diamond Fundus = [-]^{m[R]} [7, 8, 9]^{R} [2, 3, 6, 10, 11, 12, 14, 15]^{b(R)} [16]^{b[LR]} [4]^{b[LR, LR]}$ .

As a result, only patients 7, 8, 9 are still classified as ill and the risk factor vanishes for all other patients. The relevance of *Diab-Age* is not very high. Thus we can deduce that the attribute *Diab-Age* prevents almost completely the resurgence of the disease.

*Normal values to least relevant attributes (ill patients)*

If we are given the same sub-sample as before (ill patients) classified as “at risk” and less relevant attributes are given normal values. We show that ill patients are still classified as “at risk”.

We assume the following:  $GFR (R = 23.833)$ ,  $Fluor (R = 27.298)$  and  $Fundus (R = 27.298)$ . We assume that all ill patients are characterised by not serious values for  $GFR$ ,  $Fluor$  and  $Fundus$ . The following re-classification:

$$[11, 14]^{b} [1, 2, 3, 4, 5, 6, 8, 9, 10, 12, 13, 15, 16, 17]^{m} \text{ becomes } GFR-2a = [-]^{p} [-]^{p} [-]^{b} [0, 11, 12, 13, 14, 15, 16, 17]^{m}.$$

$$[5, 13, 15, 16]^{p} [1, 2, 3, 4, 6, 11, 12, 14, 17]^{m} \text{ becomes } FI-2b = [-]^{p} [5, 13]^{p} [1, 2, 3, 4, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17]^{m}.$$

$$[2, 13, 15]^{p} [1, 3, 4, 5, 6, 11, 12, 14, 16, 17]^{m} \text{ becomes } FI-2c = [-]^{p} [2, 13]^{p} [1, 3, 4, 5, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17]^{m}.$$

we have

$$\diamond 4 * Diab-Age \diamond AER \diamond FI-2b \diamond GFR-2a \diamond BdB \diamond Fluor-2c$$

Almost all patients initially classified as “at risk”, in case less relevant indices become normal, are still classified as “at risk”.

This claim can be checked by investigating the changes induced by changes in just one of them.

(A) In case only *GFR* is modified one has

$$\begin{aligned} DIAG-2A &= 2 * Effec \diamond 4 * Diab-Age \diamond AER \diamond Fluor \diamond GFR-2a \diamond BdB \diamond Fundus \\ &= [8, 9]^{m[R]} [1, 3, 7, 10, 11, 14, 15]^R [2, 6, 12]^{b[R]} [5, 16]^{b[LR,R]} [4, 17]^{m[LR]} [13]^{LR} \approx DIAG. \end{aligned}$$

We see that F-DIAGNOSI-2A is identical with the original one. Then:

all patients originally classified as “at risk”, in case their *GFR* values move towards normal levels, are still considered as “at risk”.

(B) In case *Fluor* takes the place of *Fluor* we get:

$$\begin{aligned} DIAG-2B &= 2 * Effec \diamond 4 * Diab-Age \diamond AER \diamond Fl-2b \diamond GFR \diamond BdB \diamond Fundus \\ &= [9]^{m[R]} [1, 3, 7, 8, 10, 11, 14, 15]^R [2, 6, 12]^{b[R]} [5, 16]^{b[LR,R]} [4, 17]^{m[LR]} [13]^{LR} \approx DIAG. \end{aligned}$$

We conclude that:

all patients originally classified as “at risk”, in case *Fluor* values move towards normal levels, are still classified as “at risk”.

(C) In case only *Fundus* is changed:

$$\begin{aligned} DIAG-2C &= 2 * Effec \diamond 4 * Diab-Age \diamond AER \diamond Fluor \diamond GFR \diamond BdB \diamond Fu-2c \\ &= [9]^{m[R]} [1, 3, 7, 8, 10, 11, 14, 15]^R [2, 6, 12]^{b[R]} [5, 16]^{b[LR,R]} [4, 17]^{m[LR]} [13]^{LR} \approx DIAG. \end{aligned}$$

This is not surprising as Fundus is the least relevant attribute. Thus also in this case all patients initially classified as “at risk” remain in this state if the Fundus index moves towards normal levels.

### 8.1.3. Case study 3: Giving high values to most relevant attributes (healthy patients)

We consider healthy patients classified as “not at risk”. Suppose that most relevant attributes are given values meaning serious illness, then we see that healthy patients, along with ill ones, are classified as “at risk”.

In fact, assume that healthy patients take pathological values for the relevant attributes. The new strings are:

- (a)  $HbA1c = [-]^{ph} [3, 7, 8, 9, 10, 11, 14, 15]^h [1, 2, 4, 5, 6, 12, 13, 16, 17]^q [-]^{n}$  becomes,  $Hb-3a = [2, 5, 6, 12, 13, 17]^{ph} [3, 7, 8, 9, 10, 11, 14, 15]^h [1, 4, 16]^q [-]^{nm}$  and  $Insu = [3, 12]^h [1, 2, 5, 9, 11, 14]^m [4, 6, 7, 8, 10, 13, 15, 16, 17]^l$  becomes,  $Ins-3a = [2, 3, 5, 6, 12, 13, 17]^h [1, 9, 11, 14]^m [4, 7, 8, 10, 15, 16]^l$ .
- (b)  $AER = [2, 12]^{ph} [9, 15]^h [8, 16]^q [1, 3, 4, 5, 6, 7, 10, 11, 13, 14, 17]^n$  becomes  $AER-3b = [2, 5, 6, 12, 13, 17]^{ph} [9, 15]^h [8, 16]^q [1, 3, 4, 7, 10, 11, 14]^n$ .
- (c)  $Diab-Age = [1, 6]^{ve} [2, 4, 5, 8, 9, 15, 16, 17]^q [3, 7, 10, 11, 12, 13, 14]^q [-]^{l}$  becomes  $Diab-3c = [1, 2, 5, 6, 12, 13, 17]^{ve} [4, 8, 9, 15, 16]^q [3, 7, 10, 11, 14]^q [-]^{l}$ .
- (d)  $BdB = [-]^{pl} [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17]^n$  becomes  $B-3d = [2, 5, 6, 12, 13, 17]^{pl} [-]^{b} [1, 3, 4, 7, 8, 9, 10, 11, 14, 15, 16]^n$ .

The effectiveness is now described by

$$\begin{aligned} E-3a &= 4 * Hb-3a \diamond Ins-3a = [2, 5, 6, 12, 13, 17]^{[0.85, 1, 1]} [3]^{[0.54, 0.76, 0.89]} [9, 11, 14]^{[0.51, 0.73, 0.87]} \\ &= [7, 8, 10, 15]^{[0.47, 0.69, 0.84]} [1]^{[0.36, 0.54, 0.72]} [4, 16]^{[0.31, 0.46, 0.65]} \end{aligned}$$

and we get the diagnosis

$$DIAG-3 = 2 * E-3a \diamond 4 * Dia-3c \diamond AER-3b \diamond Fluor \diamond GFR \diamond B-3d \diamond Fundus = [2, 5, 6, 12, 13, 17]^{ib(LR,HR)} [8, 9]^{ml(LR)} [1, 3, 7, 10, 11, 14, 15]^{R} [-]^{ml(LR,R)} [4]^{ml(LR)} [-]^{LR}$$

The result confirms our expectations: all healthy patients are now classified as “at risk”.

In case just one attribute is changed we have:

(A) In case just the attribute *Effec* is modified we get:

$$DIAG-3A = 2 * E-3a \diamond 4 * Diab-Age \diamond AER \diamond Fluor \diamond GFR \diamond BdB \diamond Fundus = [2, 6]^{ib(LR,HR)} [5, 8, 9, 12, 13, 17]^{ml(LR)} [1, 3, 7, 10, 11, 14, 15]^{R} [16]^{ib(LR,R)} [4]^{ml(LR)} [-]^{LR}$$

We see that this single modification induces a radical change in the classification: all healthy patients are now “at risk”, this fact stems from the high relevance of this attribute.

(B) In case high values are given only to the attribute *AER*, one has:

$$DIAG-3B = 2 * Effec \diamond 4 * Diab-Age \diamond AER-3b \diamond Fluor \diamond GFR \diamond BdB \diamond Fundus = [8, 9]^{ml(LR)} [1, 3, 6, 7, 10, 11, 14, 15]^{R} [2, 5, 12]^{b(LR)} [13, 16, 17]^{ib(LR,R)} [4]^{ml(LR)} [-]^{LR}$$

In a way similar to what happens for Case study 1, the attribute *AER*, though relevant, is not able to change the final classification. However we see in the following that if also another relevant attribute is given high values, then patients become “at risk”.

(C) In case high values are given only to the attribute *Diab-Age*, we get:

$$DIAG-3C = 2 * Effec \diamond 4 * Dia-3c \diamond AER \diamond Fluor \diamond GFR \diamond BdB \diamond Fundus = [8, 9]^{ml(LR)} [1, 2, 3, 5, 7, 10, 11, 12, 13, 14, 15]^{R} [6, 17]^{b(LR)} [16]^{ib(LR,R)} [4]^{ml(LR)} [-]^{LR}$$

*Diab-Age* is a relevant attribute and in fact most healthy patients originally classified as “not at risk” have changed their status.

Our result confirms that:

most healthy patients originally classified as “not at risk”, in case theory *Di-age* values move towards high levels, change their state to “at risk”.

More precisely only two healthy patients (6 and 17) remain “not at risk”, and for the remaining four the change in *Diab-Age* has induced a corresponding variation in the risk factor.

We are going to show that in case another relevant attribute is also given high values the risk factor becomes correspondingly high.

(D) Finally, if only the attribute *BdB* is given high values, the diagnosis becomes:

$$DIAG-3D = 2 * Effec \diamond 4 * Diab-Age \diamond AER \diamond Fluor \diamond GFR \diamond B-3d \diamond Fundus = [8, 9]^{ml(LR)} [1, 2, 3, 6, 7, 10, 11, 14, 15]^{R} [5, 12]^{b(LR)} [13, 16, 17]^{ib(LR,R)} [4]^{ml(LR)} [-]^{LR}$$

This result means that the attribute *BdB*, though relevant, by itself is almost always unable to induce the disease. In case another attribute is also given high values the situation becomes critical.

(E) In fact, we show that, although relevant attributes such as *AER*, *Diab-Age* and *BdB* are unable to induce the disease in case only one of them is given high values, in case two of them are changed all healthy patients become “at risk”.

- $DIAG-3E = 2 * Effec \diamond 4 * Dia-3c \diamond AER-3b \diamond Fluor \diamond GFR \diamond BdB \diamond Fundus = [8, 9]^{ml(LR)} [1, 2, 3, 5, 6, 7, 10, 11, 12, 13, 14, 15, 17]^{R} [16]^{ib(LR,R)} [4]^{ml(LR)} [-]^{LR}$

- $DIAG-3F = 2 * Effec \diamond 4 * Dia-3c \diamond AER \diamond Fluor \diamond GFR \diamond B-2d \diamond Fundus = [2, 8, 9, 12]^{ml(LR)} [1, 3, 5, 6, 7, 10, 11, 13, 14, 15, 17]^{R} [16]^{ib(LR,R)} [4]^{ml(LR)} [-]^{LR}$

- $DIAG-3G = 2 * Effec \diamond 4 * Diab-Age \diamond AER-3b \diamond Fluor \diamond GFR \diamond B-2d \diamond Fundus = [8, 9]^{ml(LR)} [1, 2, 3, 5, 6, 7, 10, 11, 14, 15]^{R} [12, 13, 17]^{b(LR)} [16]^{ib(LR,R)} [4]^{ml(LR)}$

In particular we note that simultaneous worsening of attributes *AER* and *BdB* makes very likely the insurgence of the disease.

#### 8.1.4. Case study 4: Giving high values to least relevant attributes (healthy patients)

Consider now healthy patients classified as “not at risk”. We show that in case high values are given to less relevant attributes, healthy patients still remain classified as “not at risk”.

Suppose that all healthy patients are given high values for *GFR*, *Fluor* and *Fundus*. Correspondingly, attribute strings change as follows:

- (a)  $GFR = [-]^p[-]^p[7, 11, 14]^b[1, 2, 3, 4, 5, 6, 8, 9, 10, 12, 13, 15, 16, 17]^n$  becomes  $GFR-4a = [-]^p[-]^p[5, 13, 17]^p$   
 $[2, 6, 7, 11, 12, 14]^b[1, 3, 4, 8, 9, 10, 15, 16]^n$   
 (b)  $Fluor = [7, 8, 9, 10]^p[5, 13, 15, 16]^p[1, 2, 3, 4, 6, 11, 12, 14, 17]^n$  becomes  $Fl-4b = [7, 8, 9, 10]^p[2, 5, 6, 12, 13, 15, 16, 17]^p[1, 3, 4, 11, 14]^n$   
 (c)  $Fundus = [7, 8, 9, 10]^p[2, 13, 15]^p[1, 3, 4, 5, 6, 11, 12, 14, 16, 17]^n$  becomes  $Fu-4c = [7, 8, 9, 10]^p[2, 5, 6, 12, 13, 15, 17]^p[1, 3, 4, 11, 14, 16]^n$ .

We see that high values for each irrelevant attribute are unable to affect seriously the diagnosis.

(A) In case the attribute *GFR* is given high values, we get:

$$DIAG-4A = 2 * Effec \diamond 4 * Diab-Age \diamond AER \diamond Fluor \diamond GFR-4a \diamond BdB \diamond Fundus = [8, 9]^{m[R]} [1, 2, 3, 7, 10, 11, 14, 15]^{R} [5, 6, 12]^{b[L,R]} [13, 16, 17]^{b[L,R,R]} [4]^{m[L,R]} [-]^{LR}$$

Only patient 2 has changed his/her status, all others are unaffected by the change in *GFR*.

(B) In case the attribute *Fluor* is change, we have

$$DIAG-4B = 2 * Effec \diamond 4 * Diab-Age \diamond AER \diamond Fl-4b \diamond GFR \diamond BdB \diamond Fundus = [8, 9]^{m[R]} [1, 2, 3, 7, 10, 11, 14, 15]^{R} [6, 12]^{b[R]} [5, 16, 17]^{b[L,R,R]} [4]^{m[L,R]} [13]^{LR}$$

Also in this case, except for patient 2, we can affirm that:

healthy patients originally classified as “not at risk”, in case the *Fluor* index moves towards high levels, remain classified as “not at risk”.

(C) In turn, if only the string *Fundus* is change, one has

$$DIAG-4C = 2 * Effec \diamond 4 * Diab-Age \diamond AER \diamond Fluor \diamond GFR \diamond BdB \diamond Fu-4c = [8, 9]^{m[R]} [1, 3, 7, 10, 11, 14, 15]^{R} [2, 5, 6, 12]^{b[L,R]} [16, 17]^{b[L,R,R]} [4]^{m[L,R]} [13]^{LR}$$

As regards the *Fundus* index we can affirm that all healthy patients originally classified as “not at risk” remain in this state in case the *Fundus* index moves towards pathological levels.

#### 8.2. Changing values in the second sample

Similar results are obtained if we change values for patients in the second sample. Basic results are as follows:

Case study 1: In case patient 1 (ill and “at risk”) is given normal values as regards most relevant attributes he gets classified as “not at risk”. In fact, one has:

$$DIAG-1 = 2 * E-Ia \diamond 3 * Dia-b \diamond AER \diamond BdB \diamond Fundus = [4]^{m[R]} [3, 15, 16]^{b[L,R]} [2, 5, 13, 14]^{b[L,R,R]} [7, 12]^{m[L,R]} [6, 8, 9, 10, 11]^{b[L,R]} [1]^{b[NR,LR]}$$

where

$$Hb-Ia = [-]^{bH} [4, 13, 14, 15, 16]^b [2, 3, 5, 6, 7, 8, 9, 10, 11, 12]^n [1]^{nR}$$

$$Ins-Ia = [4]^n [-]^n [1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16]^n$$

$$Dia-Ib = [3]^{nR} [5, 7, 12, 13, 15, 16]^n [2, 4, 6, 8, 9, 10, 11, 14]^n [1]^{nR}$$

Case study 2:  $DIAG-2 = 2 * Effec \diamond 3 * Diab-Age \diamond AER \diamond BdB \diamond Fu-2a = [4]^{m[R]} [1]^{nR} [3, 15, 16]^{b[L,R]} [2, 5, 13, 14]^{b[L,R,R]} [7, 12]^{m[L,R]} [6, 8, 9, 10, 11]^{b[L,R]} [7, 12]^{m[L,R]} [6, 8, 9, 10, 11]^{b[L,R]}$

where  $Fu-2a = [-]^p [2, 14, 15, 16]^p [1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13]^n$ .

Case study 3:  $DIAG-3 = 2 * E-3a \diamond 4 * Dia-3b \diamond AER-3d \diamond B-3c \diamond Fundus = [2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16]^{b[L,R]} [4]^{b[R,HR]} [1]^{m[R]} [-]^{nR} [-]^{LR}$

where

$$\begin{aligned} Hb-3a &= [2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16]^{oh} [-]^b [1]^{ot} [-]^n, \\ Ins-3a &= [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16]^{ht} [-]^n [-]^l, \\ Dia-3b &= [2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16]^{oe} [1]^{ot} [-]^a [-]^l, \\ B-3c &= [2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16]^{pt} [-]^b [1]^n, \\ AER-3d &= [2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16]^{oh} [4]^{ht} [-]^a [1]^n. \end{aligned}$$

$$\text{Case study 4: } DIAG-4A \ 2 * Effec \diamond 3 * Diab-Age \diamond AER \diamond BdB \diamond Fu-4a = [1, 4]^{ot(R)} [3]^n [5, 13, 15, 16]^{bt(R)} [2, 7, 12, 14]^{bt(L,R)} [-]^n [6, 8, 9, 10, 11]^{LR},$$

where

$$Fu-4c = [2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16]^{tp} [1]^p [-]^n.$$

## 9. Concluding remarks

In this paper we have presented a fuzzy-based technique for diagnosing the diabetic neuropathy, whose pathogenesis so far is not well known. The technique is based on a commutative  $L$ -monoid, whose support is a set of classifications where a basic role is played by linguistic variables.

Given two samples of patients, both hospitalised at the Istituto Scientifico San Raffaele in Milan, our diagnoses agree with the clinical ones. Moreover, taking into account the relevance factors present in the algebraic structure, we have been able to put on reasonable hypotheses about the aetiology of the disease. Also in this case our results are identical with the most credited ones [3,18].

## References

- [1] K.-P. Adlassnig, Fuzzy set theory in medical diagnosis, IEEE Trans. System, Man, Cybernet. SMC-16 (2) (1986) 260–265.
- [2] K.S. Collazos, J.M. Barreto, S.M. Nassar, Fuzzy analogical reasoning for medical diagnosis, <http://www.informatic.uni-dortmund.de/~basu/lepdiaag.html>, 1999.
- [3] G. Coni, G. Galardi, S. Amadio, E. Bianchi, A. Secchi, S. Martinenghi, R. Caldara, G. Pozza, N. Canal, Neurophysiological study of the effect of combined kidney and pancreas transplantation on diabetic neuropathy: a 2-year follow-up evaluation, Diabetologia 34 (Suppl 1) (1991) S103–107.
- [4] S. Dellepiane, G. Venturi, G. Veruazza, A fuzzy model for the processing and recognition of MR pathological images, in: A.C.F. Colchester, D.J. Hawkes (Eds.), Information Processing in Medical Imaging, IPMI'91, Lecture Notes in Computer Science, Vol. 511, Springer, Berlin, 1991, pp. 444–457.
- [5] L. Di Lascio, E. Fischetti, A. Gisolfi, A fuzzy-based approach to stereotype selection in hypermedia, User Modeling User-Adapted Interaction 9 (1999) 285–320.
- [6] L. Di Lascio, E. Fischetti, A. Gisolfi, Fuzzy modeling and evaluation of an ITM user, Computacion and Systemas, 2002, to appear.
- [7] L. Di Lascio, A. Gisolfi, G. Rosa, A commutative  $L$ -monoid for classifications with fuzzy attributes, Int. J. Approximate Reasoning 26 (2001) 1–46.
- [8] E.L. Feldman, M.J. Stevens, P.K. Thomas, M.B. Brown, N. Canal, D.A. Greene, A practical two-step quantitative clinical and electrophysiological assessment for the diagnosis and staging of diabetic neuropathy, Diabetes Care 17 (11) (1994) 1281–1289.
- [9] A. Gisolfi, An algebraic fuzzy structure for the approximate reasoning, Fuzzy Sets and Systems 45 (1992) 37–43.
- [10] A. Gisolfi, Classifying through an algebraic fuzzy structure: the relevance of the attributes, Int. J. Intell. Systems 10 (1995) 715–734.
- [11] A. Gisolfi, F. Cicalese, Classifying through a fuzzy algebraic structure, Fuzzy Sets and Systems 78 (1996) 317–331.
- [12] A. Gisolfi, V. Loia, A complete, flexible fuzzy-based approach to the classification problem, Int. J. Approximate Reasoning 13 (1995) 151–183.
- [13] A. Gisolfi, G. Nunez, An algebraic approximation to the classification with fuzzy attributes, Int. J. Approximate Reasoning 8 (1993) 75–95.
- [14] R. Golubovskij, Petrovska-Boshkova Tatyana, Trajkovski, A software package for EXG analysis, Proceedings EUFIT'99, Aachen, September 13–16, Verlag Mainz, Germany, 1999.
- [15] L.-K. Hyong, S. Yoon-Seon, L. Keon-Myong, Similarity measure between fuzzy sets and between elements, Fuzzy Sets and Systems 62 (1994) 291–293.

- [16] G.J. Klir, B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice-Hall, London, 1995.
- [17] I.K. Ludmila, Fuzzy multi-level classifier for medical applications, *Comput. Biol. Med.* 20 (6) (1990) 421–431.
- [18] S. Martinenghi, G. Comi, G. Galardi, V. Di Carlo, G. Pozza, A. Secchi, Amelioration of nerve conduction velocity following simultaneous kidney/pancreas transplantation is due to the glycaemic control provided by the pancreas, *Diabetologia* 40 (9) (1997) 1110–1112.
- [19] A. Metin, C. Maurice, H. Donna, Fuzzy sets in life science, *Fuzzy Sets and Systems* 90 (1997) 219–224.
- [20] W. Pedrycz, Why triangular membership function?, *Fuzzy Sets and Systems* 64 (1) (1994) 21–30.
- [21] A.P. Rošhtein, E. Loiko, D. Katselnikov, Prognosis of diseases number using fuzzy logic, *Proceedings EUFIT'99*, Aachen, September 13–16, Verlag Mainz, Germany, 1999.
- [22] M. Schurz, K.-P. Adlassnig, C. Lagor, B. Schneider, G. Grabner, Semi-automatic knowledge acquisition of fuzzy symptom–diagnose-relationship with hepatitides, *Proceedings EUFIT'99*, Aachen, September 13–16, Verlag Mainz, Germany, 1999.
- [23] A.A. Sima, P.V. Chertian, Neuropathology of diabetic neuropathy and its correlations with neurophysiology, *Clin. Neurosci.* 4 (6) (1997) 359–364.
- [24] L.-X. Wang, *A Course In Fuzzy Systems And Control*, Prentice-Hall International, 1997.
- [25] L.A. Zadeh, Fuzzy logic—computing with words, *IEEE Trans. Fuzzy Systems* 4 (2) (1996) 103–111.
- [26] L.A. Zadeh, The concept of a linguistic variable and its application to approximate reasoning, I, II, III, *Inform. Sci.* 8 (1975) 199–249, 8 (1975) 301–357, 9 (1975) 43–89.
- [27] L.A. Zadeh, Fuzzy sets, *Inform. Control* 8 (3) (1965) 338–353.