

Article - Biological and Applied Sciences

# A Fuzzy Approach for Diabetes Mellitus Type 2 Classification

**Glaucia Maria Bressan<sup>1\*</sup>**

<https://orcid.org/0000-0001-6996-3129>

**Beatriz Cristina Flávia de Azevedo<sup>1</sup>**

<https://orcid.org/0000-0002-8527-7409>

**Roberto Molina de Souza<sup>1</sup>**

<https://orcid.org/0000-0002-0638-3650>

<sup>1</sup>Universidade Tecnológica do Paraná (UTFPR), Campus Cornélio Procópio, Paraná, Brazil.

Received: 2018.12.14; Accepted: 2019.11.26.

\*Correspondence: [glauciabressan@utfpr.edu.br](mailto:glauciabressan@utfpr.edu.br); Tel.: +55-43-3933-3813 (F.L.)

## HIGHLIGHTS

- Automatic fuzzy classification system is proposed for glycemic index.
- Neuro-fuzzy systems are used to obtain the parameters of the functions.
- Decision Trees are used to obtain the linguistic based rules.
- Results indicate the best treatment for each patient

**Abstract:** This paper proposes an automatic fuzzy classification system for glycemic index, which indicates the level of Diabetes Mellitus type 2. Diabetes is a chronic disease occurred when there is deficiency in insulin production or in its action, or both, causing complications. Neuro-fuzzy systems and Decision Trees are used to obtain, respectively, the numerical parameters of the membership functions and the linguistic based rules of the fuzzy classification system. The results goal to categorize the glycemic index into 4 classes: decrease a lot, decrease, stable and increase. Real database from [1] is used and the input attributes of the system are defined. In addition, the proposed automatic fuzzy classification system is compared with an “expert” fuzzy classification system, which is totally modeled using expert knowledge. From linguistic based rules obtained from fuzzy inference process, new scenarios are simulated in order to obtain a larger data set which provides a better evaluation of the classification systems. Results are promising, since they indicate the best treatment - intervention or comparative - for each patient, assisting in the decision-making process of the health care professional.

**Keywords:** Glycemic Index, Fuzzy Classification, Decision Tree, Neuro-fuzzy.

## INTRODUCTION

Chronic diseases are considered the major cause of death and disability worldwide. The Diabetes Mellitus is then considered one of the major problems of the 21 century [2-3]. According to the International Diabetes Federation, there are more than 415 million adults aged 20 to 79 with Diabetes and this number is expected to increase to 642 million in 2040. It is important to note this estimation does not include people older than 80 [3].

The Diabetes could be defined as a set of metabolic disease that developed when the pancreas stops producing insulin (hormone that regulates the sugar (glucose) concentration in the blood, glycemia) enough or when the organism is no longer able to use correctly the insulin produced [4]. The lack of insulin or the inability of the cells to respond to insulin leads to high levels of blood glucose, or hyperglycemia, which cause damage to several organs [3].

According to ADA (American Diabetes Association - 2017), diabetes can be classified into four general categories: the type 1 diabetes is categorized by the destruction of the autoimmune  $\beta$ -cell, consequently, the body produces none or insufficient quantity of insulin. Diabetes type 1 is more common in children and adolescents. The Diabetes type 2 is a progressive loss of  $\beta$ -cell insulin secretion frequently on the background of insulin resistance. This is the most common type of Diabetes diagnosis in adults, however due to the change lifestyle, sedentary life, and poor diet, it is also increasing in children. Other category is the Gestational Diabetes Mellitus which is diagnosed during the pregnancy due to hormone alteration. Finally, the last category is composed by specific types of diabetes, normally caused by genetic mutation, coming from other specific causes as diseases of exocrine pancreas and drug-or chemical-induced diabetes [2].

The Diabetes, regardless of the category, can cause serious complication for the patients, as convulsions, loss of consciousness, damage for the eyes, heart, vessels blood and nervous system, amputation of the body's member, as well as the increased risk of cardiac diseases and cerebrovascular accidents. In some cases lead to death. However, frequently, in the death certificate, the Diabetes is not declared as cause of the death, but its complication [5]. For this reason, the number of death caused by Diabetes is not precise. Although, many of these deaths could be avoided if there were an early diagnosis that result in glycemic control and change of life habits.

Several effective actions could be taken to improve the health of people with Diabetes; not only using medications, but also through food changes, physical activities, reduction of vices such as cigarettes and alcohol, for example. In addition, the early diagnosis and the previous onset of treatment are essential factors to contain the progression of the disease and avoid secondary complications.

Due to the large incidence of diagnosis, forms of control and early diagnosis of Diabetes Mellitus began to be very studied in the scientific community. Particularly, the fuzzy methodology, which is the focus of this work, has been widely used in the Diabetes study, since it is able to work with imprecise and vague information, commonly present in Diabetes database.

Lukmanto and Irwansyah discuss the application of fuzzy logic with hierarchical model as implementation of computational intelligence techniques for identify and determine the potential for someone related to Diabetes Mellitus. This system could contribute to the prevention and early detection of Diabetes [6]. A system for diabetes classification applying fuzzy logic was also studied in [7]. Authors used the technique of cuckoo search optimization algorithm for the rules generation used in the fuzzy model [7]. Osgouie and Azizi present a fuzzy controller to maintain the normoglycaemic average of plasma glucose concentration and other model variables concentration within a certain desired interval, in patients who presented diabetes type 1 [8].

Other interesting approaches are studied by authors that investigate the insulin dosing. Lalka and Jain propose a fuzzy system to work the dynamics of diabetes diagnosis and recommend the best insulin dosage according to the patient's parameters. Authors conclude that fuzzy based systems are robust in nature and can handle dynamism of the nature of a disease [9].

Erroneous dosing of insulin may worsen the clinical situation of patients inducing to hypoglycemia (decreased blood glucose level) or hyperglycemia (increased blood glucose level). Khan and coauthors studied how to calibrate the total daily insulin dosage for type 1 diabetes patients undergoing insulin treatment regimens using a fuzzy logic based system [10]. The fuzzy based insulin dosing system dispensed precise insulin dose for individual patients and hence demonstrated substantial control of blood sugar regulation [10].

Moreira observed the influence of medical intervention and nursing management in the glycemic control of Diabetes Mellitus type 2. Glycemia control is the main target for the prevention of chronic complications, with fasting plasmatic glycemia and glycated hemoglobin as indicators. The aim is to evaluate the effect of case management method on the glycemia control of people with Type 2 Diabetes Mellitus and on its chronic complications risk factors. The author concluded that case management method affected the glycemia control. However, in this work, the variation of the indicators considers only the numerical difference between the measures [1].

In this scenario, this paper proposes an automatic fuzzy classification system [11] for the glycemic index, which indicates the Diabetes Mellitus type 2, using pattern recognition and supervised learning approaches. Neuro-fuzzy systems [12] and Decision Trees [13] are used to obtain, respectively, the numerical parameters

of the functions and the linguistic based rules. This proposed classification system aims to categorize the glycemic index into 4 classes: decrease a lot, decrease, stable and increase. In order to do this classification task, real database from Moreira [1] is analyzed and the input attributes for the classification system - the ones with the most influence in glycemic index - are selected. In addition, the proposed automatic fuzzy classification system is compared with an “expert” fuzzy classification system, which is totally modeled using expert knowledge. From linguistic based rules obtained from fuzzy inference process, new scenarios are simulated in order to obtain a larger data set which provides a better evaluation of the classification systems.

### Fuzzy classification systems proposed

Fuzzy theory was introduced in 1965 by Lotfi A. Zadeh [14]. The fuzzy theory enables the incorporation of inaccurate and uncertain information as part of the system. This ability to work with uncertainties and overlaps between borders of the classes, makes the model more consistent and mathematically easier to work.

In this paper, an automatic and an expert fuzzy classification systems are developed. Besides, for each one of these systems, two fuzzy systems are developed: one system for the called “intervention group” and the second system for the called “comparative group”. Thus, 4 fuzzy classification systems are developed.

In a schematic way, a fuzzy system could be constructed in 5 steps: Input, Fuzzify, Inference, Defuzzify and Output. Figure 1 shows the fuzzy inference process used in this paper to produce the automatic classification system. Firstly, the input, which are crisp (non-fuzzy) numbers limited to a specific range, provide numeric information about the data for the fuzzy system. In the Fuzzify phase, the system receive the inputs information and determine the degree to which they belong to each of the appropriate fuzzy sets via membership functions. In this paper, this process is done by Neuro-fuzzy system, described in Section 2.1. After the inputs are fuzzified, the system has the degree to which each part of the antecedent is satisfied for each rule. In sequence, all rules, provided by Decision Tree, are evaluated using fuzzy reasoning (inference). In this paper, the linguistic based rules is obtained automatically from Decision Trees. The results of the rules are combined and distilled (defuzzified). The input for the defuzzification process is a fuzzy set and the output is a crisp (non-fuzzy) number. As shown in Figure 1, the system input is a numeric information about the data, after passing through the system, these information are processed and the output present the knowledge about the data patterns, that could be used to infer about new data.

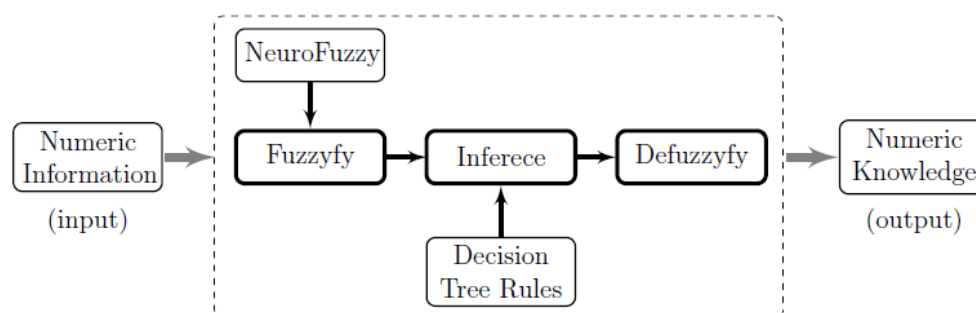


Figure 1. Fuzzy System Proposed.

The linguistic based rules is described as “IF-THEN” rules, for example: “IF attribute A is  $y_1$  AND attribute B is  $y_2$  THEN the class is Z”. If the antecedent of a given rule has more than one part, the fuzzy operator is applied to obtain one number that represents the result of the antecedent for that rule. This number will then be applied to the output function. The input to the fuzzy operator is two or more membership values from fuzzified input variables. The output is a single truth value. Basically, AND (minimum) and OR (maximum) methods. This type of rules are closer to how humans interpret information, which facilitates the understanding of the system.

### Membership Functions

A membership function is defined as a curve that defines how each point in the input space is mapped to a membership value (or degree of membership) between 0 and 1 [11]. The input is always a crisp numerical value limited to the universe of discourse of the input variable and the output is a fuzzy degree of membership (the interval between 0 and 1) [11].

**Definition 1.** The formal mathematical representation of a fuzzy set is characterized by a membership function mapping the elements of a domain, space, or universe of discourse  $X$  to the unit interval  $[0,1]$ , that is  $A: X \rightarrow [0,1]$ .

Thus, a fuzzy set  $A$  in  $X$  may be represented as a set of ordered pairs of a generic element  $x \in X$  and its grade of membership:  $A = \{(A(x)/x) | x \in X\}$ . The membership values express the degrees to which object is compatible with the properties or features distinctive to the collection [11]. In principle any function of the form  $A: X \rightarrow [0,1]$  describes a membership function associated with a fuzzy set  $A$ . This fuzzy set depends the concept and the context in which it is used. The shapes and the properties of each membership functions have strong influence in the results of the models generated. The application and the data behavior determined the membership type. The main membership functions are defined by the Equations (1), (2) and (3), where  $m$  is a model value,  $a$  and  $b$  denote the lower and upper bounds, respectively, for nonzero values of  $A(x)$  [11].

Triangular Function:

$$A(x) = \begin{cases} 0 & \text{se } x \leq a \\ \frac{x-a}{m-a} & \text{se } x \in [a, m] \\ \frac{b-x}{b-m} & \text{se } x \in [m, b] \\ 0 & \text{se } x \geq b \end{cases}, \quad (1)$$

Trapezoidal Function:

$$A(x) = \begin{cases} 0 & \text{se } x < a \\ \frac{x-a}{m-a} & \text{se } x \in [a, m] \\ 1 & \text{se } x \in [m, n], \\ \frac{b-x}{b-n} & \text{se } x \in [n, b] \\ 0 & \text{se } x > b \end{cases}, \quad (2)$$

Gaussian Function:

$$A(x) = e^{-k(x-m)^2}; k > 0, \quad (3)$$

In this paper, Triangular Function are used to formulated inputs and outputs membership functions of the classification systems, since it allows non-symmetric sides (as Gaussian Function requires) and it does not have intervals of stability (as Trapezoidal Function). Therefore, triangular membership functions better represent the selected attributes.

The membership function determine the degree of the relation of the data with the fuzzy sets. The adjustment of this relation could be done using hydride systems as neuro-fuzzy, combining backpropagation, as applied in this work. Using a computational support, the adaptive neuro-fuzzy inference system (ANFIS) of Matlab, adjusts the membership function parameters by neuro-adaptive learning, which is similar of the training an Artificial Neural Network.

The main idea of ANFIS is to provide a method for the fuzzy modeling procedure to learn information about a data set, in order to compute the membership function parameters that best allow the associated fuzzy inference system to track the given input/output data. The computation of these parameters (or their adjustment) is facilitated by a gradient vector, which provides a measure of how well the fuzzy inference system is modeling the input/output data for a given set of parameters. Once the gradient vector is obtained, any of several optimization routines could be applied in order to adjust the parameters so as to reduce some error measure. The neuro-fuzzy system is trained using the Sugeno inference system, and this one must be of unity weight for each rule and linear or constant output membership functions. This adaptive neuro-fuzzy inference system can be seen with more details in [12].

## IF-THEN Rules

The base of rule "IF-THEN", are employed to capture the imprecise modes of reasoning that play an essential role in the human ability to make decisions in an environment of uncertainty and imprecision [12]. The "IF-THEN" rules could be provide by a specialist who contains sufficient knowledge to infer about the data patterns (expert system) or obtained by supervised learning, as Decision Tree algorithm [13].

Decision Tree is a data mining technique used to recognize patterns. The Decision Tree model consists of a data set, partitioned into groups known as nodes. The partition is done recursively, according to divider algorithm, which is normally a TDIDT (Top-Down induction Decision Tree) algorithm. The partition process can be interrupted by a stopped criterion pre-determined or when no more improvement is possible [13].

The Decision Tree structure is basically composed by a top node, internal nodes and leaves. The top node is called root node, which is selected using some attribute selection measures. Under the root node are the internal nodes, originating from the division of the data set; they constitute the tree branches. At the end of each branch is the terminal node, designed leaves, which represent the most appropriated class for the rule. The Decision Tree's rules are described as "IF-THEN" rules and they are composed by the root node, the internal nodes that compose a specific branch and one leaf that assign a class.

The strategy used to generate a Decision Tree varies according to the algorithm used and it directly influences in the results. For this reason, some algorithms work better than others in the same database. The construction of a Decision Tree could be divide into two steps: the first, called growing, is responsible for generate the initial model; and the second step, called pruning, optimizes the computational operation and eliminates overfitting, improving the classification.

Due to the best results, compared to other algorithms, in this paper the authors opted for C4.5 algorithm [15], available in free software WEKA. The C4.5 algorithm uses Gain Ratio [13] and the Information Gain [13], as splitting criteria and consequently generate the initial model. The element with highest Gain Ratio is taken as the root node and data set is split based on the root element values [13]. The Information Gain is calculated for all the sub-nodes and the process is repeated until the prediction is completed. In C4.5, the pruning process, is done using the Error-Based Pruning process. C4.5 equates the predicted error rate at a leaf with this upper limit, on the argument that the tree has been constructed to minimize the observed error rate [15]. Rokach and Maimon say that the error rate is estimated using the upper bound of the statistical confidence interval for proportions [13].

## DATABASE

The database used in this paper is obtained from a clinical research, described in Moreira [1]. This research comprises an anterior-posterior prospective cohort. The attributes (or variables) were measured at 3 different times: initial ( $t_1$ ), after 6 months ( $t_2$ ) and after 12 months ( $t_3$ ). The numerical data of the attributes were measured from 77 patients, which were divided into 2 groups: intervention - group 1 and comparative - group 2. The group 1 has 38 patients and the intervention is based on nursing case management, with multi-professional attendance, group educational activities, home-care and telephone contacts. The group 2 contains 39 patients submitted to usual care in Unified Health System.

All the measured attributes are shown in Table 1, which are organized in 5 groups: social-demographic, clinical, lifestyle, glycemic control, risk factors of chronic complications. The social-demographic attributes refer to the environment and social information of the patients; clinical attributes express pathologic and medicines information; the lifestyle describe the life habits, such as eating and physicals activities; the glycemic control informs about the glycemic index; finally, attributes involving risk factors of chronic complications refer to the consequences that contributes for early deaths and physical disabilities of the patients.

From attributes described in Table database, Moreira has measured numerical data, obtaining the database. Decision Trees were trained using all appropriative combinations of attributes. The set of attributes that provided the best relationship in the Decision Tree and, consequently, generated the rule base that provided the best classification results. Therefore, after these tests, 7 attributes are selected: age, triglycerides, body mass index, abdominal circumference, per capita income, education and evolutionary time. This last one refers to the time, in years, since the beginning of the disease until the measure moment. These attributes are used to generate a fuzzy inference system able to infer about the patient profile and help in the diabetes diagnosis. For fuzzy classification system the 7 input attributes are categorized into 3 classes: low, medium and high.

The output attribute refers to the glycemic index. Since the output of the classification system must be categorized in linguistic classes, as well as input attributes, it is calculated, for each patient, as the numerical difference between its measurement  $t_3 - t_1$ .

The data analyze of Moreira concluded about the index variation of Diabetes Mellitus type 2, but this analyze only considered the numerical difference between measures  $t_3 - t_1$ , determining that positive values correspond to increase and negative values to decrease. However, in order to make a more rigorous and realistic analysis, in this paper the glycemic index is categorized in 4 fuzzy classes: decrease a lot, decrease, stable and increase.

Next section describes in details the ranges of the fuzzy classes.

**Table 1.** Available Attributes.

<b>Socio-demographic</b>	<b>Clinical</b>	<b>Lifestyle</b>	<b>Glycemic Control</b>	<b>Risk Factor of Chronic Complications</b>
Gender	Comorbidities	Physical Activities	Plasma Glycemia	Total cholesterol
Age	Evolutionary Time with Diagnosis of Diabetes Mellitus	Smooking	Glycated Hemoglobin	Low Density Lipoprotein
Marital Status	Medications in Use	Alcohol Ingestion		High Density Lipoprotein
Number of Children	Adherence to Medicinal Therapy	Eating Habits		Triglycerides
Education	Number of Previous Hospitalizations			Systolic and Diastolic Blood Pressure
Work				Weight
Number of People Living in Residence				Stature
Family Income				Body Mass Index
Per Capita Income				Abdominal Circumference
				Tracking Diabetic Nephropathy
				Tracking and Identification of Risk Factors for Diabetic Foot

## CLASSIFICATION RESULTS

This section describes how the proposed fuzzy classification systems are developed: an automatic and an expert systems. For each one of these systems, one system for “intervention group” and one system for “comparative group” are also developed. Therefore, there are 4 fuzzy classification systems to be described.

### Automatic System

In the automatic fuzzy classification system, supervised learning (neuro-fuzzy) and Decision Trees are used to obtain, respectively, the numerical parameters of the membership functions and the linguistic based rules (inference). First of all, using MATLAB, the numerical data from groups 1 and 2 are trained by a neuro-fuzzy system (ANFIS), separately. This training determines the degree of the data belong to each of the appropriate fuzzy sets, generating the membership functions parameters. As mentioned before, the Triangular Function is selected to formulate the membership functions, since it is non-symmetric and there is no stability intervals. Then, from ANFIS, the numerical data of the 7 input attributes are discretized in 3 linguistic classes: low, medium and high.

For group 1 (intervention), Table 2 shows the ranges provided by the Neuro-fuzzy system, after data supervised training, for the input attributes classes. And Table 3 shows the ranges for group 2

(comparative). Since ranges refer to fuzzy sets, there is overlap between the bounds. Similarly, Table 4 shows the ranges for the output classes, for both groups. The output membership functions are determined according to the variation of the glycemia index of patients. This index is categorized in 4 classes: decrease a lot, decrease, stable and increase. Since the output are fuzzy functions, there is overlap between the bounds of the ranges, which were fitted in order to obtain the best results of classification (the best accuracy). Firstly, for each patient, the glycemic index is measured at time  $t_1$  and  $t_3$ ; then they are subtract:  $t_3 - t_1$ . The highest and the lowest value are identified, constituting the initial range, which is partitioned into the 4 classes, as follows: decrease a lot  $]-\infty; -2[$ ; decrease  $]-2; -0.5[$ ; stable  $]-0.5; 0.5[$  and increase  $]0.5; \infty[$ . From this initial range, some adjusted are done in the functions to incorporate the overlaps between the classes and then find the membership functions parameters that best represent the data set.

**Table 2.** Input Attributes - Control Group.

Attributes	Intervention Group		
	Low	Medium	High
Age	[19.5;33;46.5]	[33;46.5;60]	[46.5;60;73.5]
Triglycerides	[-174;47;268.5]	[47;268.5;490]	[268.5;490;711]
Evolutionary Time	[-15.2;0.20;15.6]	[0.2;15.6;31]	[15.6;31;46.4]
Body Mass Index	[3.7;16.06;28.4]	[16.1;28.41;40.8]	[28.1;40.8;53.1]
Abdominal Circumference	[45;72;99]	[72;99;126]	[99;126;153]
Per Capita Income	[-400;4.48e-05;400]	[6.1e-05;400;800]	[400;800;1200]
Time of Schooling	[-7.5;-1.1e-03;7.5]	[1.1e-03;7.5;15]	[7.5;15;22.5]

**Table 3.** Input Attributes - Comparative Group.

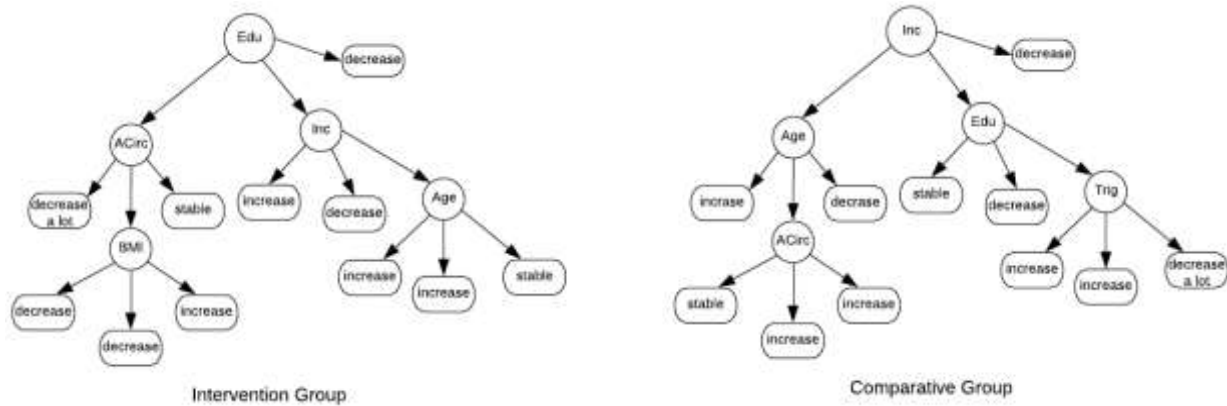
Attributes	Comparative Group		
	Low	Medium	High
Age	[3;22;41]	[22;41;59.9]	[41.0;60;79]
Triglycerides	[-135;45;225]	[45;225;405]	[225;405;585]
Evolutionary Time	[-11.63;0.25;12.1]	[0.27;12.1;24]	[12.1;24;35.9]
Body Mass Index	[2.2;21.7;41.2]	[21.7;41.2;60.7]	[41.2;60.7;80.2]
Abdominal Circumference	[52.8;81;109.2]	[81;109.2;137.5]	[109.3;137.5;165.8]
Per Capita Income	[-500;-9.2e-05;500]	[8.4e-05;500;1000]	[500;1000;1500]
Time of Schooling	[-7.5;-2.6e-05;7.5]	[6.4e-03;7.5;14.9]	[7.5;15;22.5]

**Table 4.** Output Parameters - Automatic System.

Output	Intervention Group	Comparative Group
Decrease a lot	[-6.69 ; -3.84 ; -1.87]	[-4 ; -2.47 ; -1.08]
Decrease	[-3.33 ; -1.959 ; -0.354]	[-2.014 ; -1.15 ; -0.2857]
Stable	[-0.836 ; 0.03296 ; 0.9019]	[-0.7452 ; 1.388e-17;0.745]
Increase	[0.3609 ; 1.71 ; 2.56]	[0.194 ; 1.881 ; 3.24]

The ANFIS system works with Sugeno inference system [11], which the output is linear or constant. Thus, the membership function parameters obtained are transported to the Mandani inference system [11], which the outputs are fuzzy sets. After that, the linguistic "IF-THEN" rules base is generated from Decision Tree, using C4.5 algorithm and computational support of the free software WEKA. Figure 2 illustrates the Decision Trees, generated using C4.5 algorithm [13], for groups 1 (intervention) and 2 (comparative) and the relationship between the input attributes. In order to construct Decision Trees using C4.5 algorithm, the numerical data also need to be discretized. However, unlike the fuzzy functions, Decision Trees do not allow

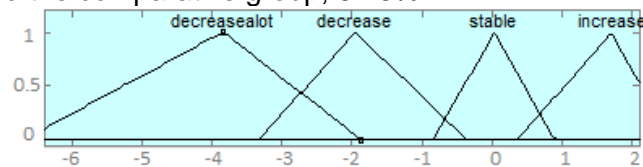
overlaps between the bounds of the ranges. Then, numerical data are split in ranges using WEKA, specifically for decision trees development. And the output classes (reduced a lot, reduce, stable and increase) are discretized considering the initial range, described above, so as to achieve the results that could be incorporated in the fuzzy systems.



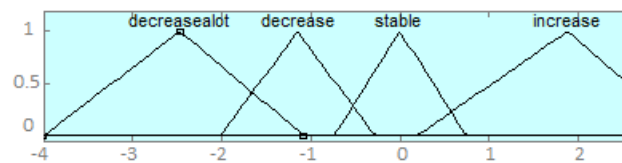
**Figure 2.** Decision Trees for Intervention Group and for Comparative Group.

Observing Figure 2, an example of “IF-THEN” rule extracted from the intervention group could be described as “IF education is medium AND income is high AND age is medium THEN the glycemic index increase”; and an example of rule from the comparative group could be “IF income in low AND age is medium AND abdominal circumference is low THEN the glycemic index is stable”. The obtained “IF-THEN” rules base are then inserted into the fuzzy system, of each group, in order to do the inference process.

The evaluation of the classification systems has been done by cross-validation, which the data is randomly split into  $n$  mutually exclusive subsets of approximately equal size. An inducer is trained and tested  $n$  times; each time it is tested on one of the  $k$  folds and trained using the remaining  $n - 1$  folds. Every data point gets to be in a test set exactly once, and gets to be in a training set  $n - 1$  times. At the end, the average of data correctly classified, determines the accuracy of the model. Figures 3 and 4 show the output membership functions for the intervention and comparative groups, respectively. The intervention group presented 70% of accuracy and the comparative group, 62.5%.



**Figure 3.** Output Membership Function of Intervention Group.



**Figure 4.** Output Membership Function of Intervention Group.

## Expert System

In order to compare the obtained results from automatic system, a system based exclusively on literature and expert knowledge, is developed. Then, the called expert system is totally based on human decision and on literature to provide the input membership functions parameters and the rules base. These input membership functions parameters are the same for the groups 1 and 2, and it are shown in Table 5. The “IF-THEN” rules base is generated separately for each group, according to the patients characteristics, for example: patients with less schooling will have a larger reduction of glycemia if they are assisted in the Group 1, since World Health Organization - WHO relates low schooling with negative sanitary conditions [16], which affects the perception of health status [17]. Thus, case management provides regular information to patients and family members. By the other hand, patients with lower income will have a stability or a increase of



glycemia if they are assisted in the Group 2, since WHO considers that individuals with lower income have more difficulty in accessing health services and materials/resources acquisition that are not offered by Unified Health System [16]. Table 6 shows the output parameters for intervention group (left side) and comparative group (right side). To get them, the initial range was considered, and similarly as done in the automatic system, some adjusted are done to incorporated the overlaps between the classes and find the membership functions parameters that best represent data set.

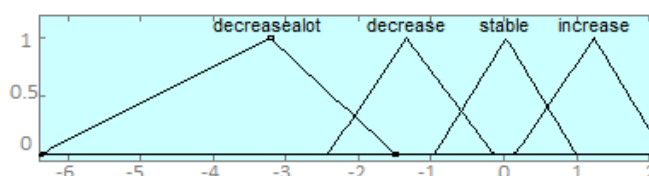
**Table 5.** Input Parameters - Expert System.

Attributes	Intervention and Control Group		
	Low	Medium	High
Age	[31.9;37.02;41.6]	[37.9;45.04;51.5]	[47.9;54.11;60]
Triglycerides	[36.8;105.3;166]	[133;177.8;220]	[197.343.9;490]
Evolutionary Time	[0.221;153;6.02]	[3.99;7.63;11]	[9.29;20.81;31.6]
Body Mass Index	[16.1;21.05;26.1]	[23.9;27.34;31.2]	[29.2;34.8;41]
Abdominal Circumference	[71.1;87.9;105]	[86.25;103;124]	[105;118.7;126]
Per Capita Income	[-0.43;43.77;99.2]	[60.2;180;299.5]	[253.5;487;802]
Time of Schooling	[-0.77;2.74;6.09]	[3.21;7.5;11.79]	7.52;12.12;15.01]

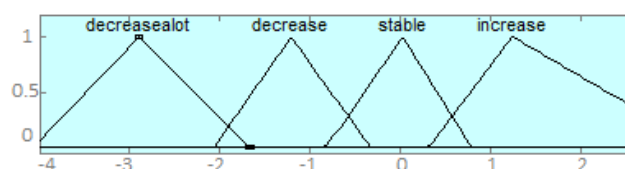
**Table 6.** Output Parameters - Expert System.

Output	Intervention Group	Comparative Group
Decrease a lot	[-6.38 ; -3.209 ; -1.5]	[-4.08 ; -2.9 ; -1.67]
Decrease	[-2.42 ; -1.35 ; -0.1424]	[-2.064 ; -1.2 ; -0.3358]
Stable	[-0.962 ; 0.033 ; 1]	[-0.822 ; 0.0222 ; 0.794]
Increase	[0.152 ; 1.248 ; 2.22]	[0.316 ; 1.245 ; 3.36]

Figures 5 and 6 show the output membership functions for the intervention and comparative groups of the expert system, respectively. The intervention group presented 63.15% of accuracy and the comparative group 55.26%.



**Figure 5:** Output Membership Function of Intervention Group.



**Figure 6:** Output Membership Function of Comparative Group.

## CONCLUSIONS

In this paper, an automatic fuzzy classification system for the glycemic index, which indicates the Diabetes Mellitus type 2, is proposed. Neuro-fuzzy systems and Decision Trees are used to obtain, respectively, the numerical parameters of the membership functions and the linguistic based rules. Database from Moreira was used to categorize the glycemic index into 4 classes: decrease a lot, decrease, stable and increase [1]. Data are divided into 2 groups: intervention and comparative. For this reason, a fuzzy

classification is developed for each one of the groups. These mentioned automatic systems are compared with “expert” fuzzy classification systems, using expert knowledge.

For practical situations, a health care professional can measure the input attributes of his patient and, using the linguistic based rules obtained from the inference fuzzy system, the professional can forward the appropriate treatment. Therefore, from linguistic based rules obtained from the fuzzy inference process, new scenarios are simulated in order to obtain a larger data set: 10.000 patients are simulated, which present the input attributes according to the based rules of the automatic classification system. This larger data provides a better evaluation of the classification systems. For these simulated patients, Table 7 describes, in percentage, the estimates of data classification into the output classes of the automatic fuzzy classification system, considering groups 1 and 2. Table 7 shows that “decrease” and “decrease a lot” classes occur mostly in Group 1, in comparison with Group 2. Consequently, increases occur mostly in Group 2. These results are in concordance with Moreira [1].

**Table 7.** Output Classification Percentage of 10.000 Simulated Patients.

Group	Increase	Stable	Decrease	Decrease a lot
1	0.1518	0.1162	0.5399	0.1921
2	0.2228	0.2510	0.5185	0.0080

Although Group 1 presents a higher proportion of decrease, not all individuals should necessarily be forwarded to this treatment, since some individuals also decreased in Group 2. The methodology proposed in this paper provides the classification of a particular patient, which will allow the health care professional to measure the patient’s input attributes and, using the based rules, forward the patient to the group that provided the best results. Table 8 presents the cross table for Groups 1 and 2. Numbers refer to the percentage of the 10.000 simulated patients. Main diagonal shows that approximately 35% of simulated patients were classified equally, both for Group 1 and Group 2. In these cases, for costs issues, these patients should be forwarded to Group 2. For the values below the main diagonal, approximately 39% of simulated patients present best results for Group 1. Finally, approximately 26% should be forwarded to Group 2.

**Table 8.** Output Cross Table: Percentage of 10.000 Simulated Patients.

Group 1	Group 2			
	Increase	Stable	Decrease	Decrease a lot
Increase	0.0393	0.0214	0.0911	0.0000
Stable	0.0233	0.0320	0.0320	0.0609
Decrease	0.1509	0.1035	0.2775	0.0000
Decrease a lot	0.0090	0.0941	0.0890	0.0000

Results are promising, since they indicate the best treatment - intervention or comparative - for each patient, assisting in the decision-making process of the health care professional. This paper contributes greatly to prioritize actions to patients who really have a greater benefit compared to what they would if assisted by the usual care (Group 2).

**Acknowledgments:** Fundação Araucária, Brazil.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results”.

## REFERENCES

1. Moreira RC. The effect of case management method on the glycemia control of people with type 2 diabetes mellitus, [Doctorate Thesis]. Curitiba (PR): Universidade Tecnológica Federal do Paraná; 2013. 227 p.
2. ADA, Association, Standards of Medical Care in Diabetes 2017, Vol. 40. In: Diabetes Care: The Journal of Clinical and Applied Research and Education. American Diabetes Association, 2017.
3. IDF, Diabetes atlas estimates, International Diabetes Federation. Atlas 8 ed., 2017.
4. WHO, World Health Organization. Global Report on diabetes, 2016.
5. Milech A, et al. Diretrizes da Sociedade Brasileira de Diabetes 2015 - 2016, Organização José Egídio Paulo de Oliveira, Sérgio Vencio, São Paulo: A.C. Farmacêutica, 2016.
6. Lukmanto RB, Irwansyah E. The early detection of diabetes mellitus (DM) using fuzzy hierarchical model. *Procedia Compute Science*. 2015, 59: 312-319.

7. Sahu N, Verma T, Reddy, GT. Diabetes classification using fuzzy logic and adaptive cuckoo search optimization techniques. *International Journal on Future Revolution in Computer Science & Communication Engineering*. 2017; 3(9): 252-255.
8. Osgouie KG, Azizi A. Optimizing fuzzy logic controller for diabetes type I by genetic algorithm. *Proceedings of the 2nd International Conference on Computer and Automation Engineering (ICCAE)*; 2010: 2:4-8.
9. Lalka N, Jain S. Fuzzy based expert system for diabetes diagnosis and insulin dosage control. *International Conference on Computing, Communication and Automation*. 2015: 262-267.
10. Khan RI, Nirzhor SSR, Chowdhury AM, Shishir TA, Khan AI. A fuzzy logic based approach for the adjustment of insulin dosage for type 1 diabetes patients. *J. Innov. Pharm. Bio. Sci.* 2017;4(4): 145-152.
11. Pedrycz WE, Gomide F. *An Introduction to Fuzzy Sets*, MIT PRESS; 1998.
12. Jang JR. Anfis: Adaptive - network - based fuzzy inference system. *IEEE Trans Sys Man Cyberns.* 1993;23(3): 665-685.
13. Rokach L, Maimon O. *Data mining with decision trees: theory and applications*, Vol. 69, World Scientific; 2008.
14. Zadeh LA. Fuzzy sets. *Inf. Control.* 1965;8(3): 338-353.
15. Quinlan JR C4.5: *Programs for machine learning*, Morgan Kaufmann; 1993.
16. WHO, World Health Organization. *Global Report on diabetes*, 2003.
17. Medes EV. *As redes de atenção a saúde, Organização Pan-Americana da Saúde*, 2011.



© 2020 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY NC) license (<https://creativecommons.org/licenses/by-nc/4.0/>).