

Fuzzy Logic Assisted Diagnosis for Atherogenesis Risk

Ana L. Alonso* Oscar A. Rosas-Jaimes**
Juan A. Suárez-Cuenca***

* *Departamento de Asesoría Científica, Amplibio, S.A. de C.V., D.F.,
México (e-mail: analucia@amplibio.com).*

** *Facultad de Ingeniería, Universidad Autónoma del Estado de México
(e-mail: oscar.rosasjaimes@yahoo.com).*

*** *División de Investigación Clínica, CMN "20 de Noviembre"
ISSSTE, DF (e-mail: suarej05@gmail.com)*

Abstract: A fuzzy inference system (FIS) that aids in atherogenesis risk diagnosis is described in this document, taken as point of start data of human lipid levels. This FIS uses Total Cholesterol, Low-Density Lipoproteins, Atherogenic Index and Triglycerides as variables in order to propose a diagnosis method to help in low-cost early detection of atherogenesis risk, in strict agreement to medical convention.

Keywords: Fuzzy Logic, LDL, AI, Triglycerides, Cholesterol.

1. INTRODUCTION

Metabolic syndrome (MS) is a relatively common disorder as a major risk factor for atherosclerosis and cardiovascular disease. Hyperlipidemia, diabetes mellitus, hypertension, obesity are included as part of MS definition. Atherosclerosis is a chronic disease associated to hyperlipidemia, characterized by a lipid unbalance where the vascular intimate thickness increases, constituted by fat cumulation and conjunctive tissue. This phenomenon stays asymptomatic during many years. When arterial obstruction manifestations (myocardial infarct, gangrene, embolism) get present, consequences are mostly irreversible.

Currently, most of the medical professionals agree that main serum lipids can be divided in the intervals shown in table 1, where *TC* stands for Total Cholesterol, *LDL* for Low-Density Lipids, *AI* for Atherogenesis Index and *TG* for Triglycerides.

It has been found by 30-year USA trends analyses that variables in table 1 have a stronger function and should be taken into account to make accurate health diagnosis in any patient related to atherosclerosis or other similar illnesses (Cohen et al. (2010)).

Table 1. Optimal, Borderline, High-Risk and Very High-Risk Serum Lipid Concentrations

	Optimal	Borderline	High-Risk	Very High-Risk
TC [$\frac{mg}{dl}$]	< 200	200 – 240	> 240	–
LDL [$\frac{mg}{dl}$]	< 100	130 – 160	> 160	–
AI	≤ 4.2	–	> 4.2	–
TG [$\frac{mg}{dl}$]	< 150	150 – 200	200 – 500	> 500

2. DIAGNOSIS GENERALITIES OF ATHEROGENESIS RISK

In human plasma, cholesteryl esters and triglycerides can exchange between various lipoprotein fractions through the Cholesteryl Ester Transfer Protein (Lagrost (1994); Tall (1995); Ritsch et al. (2010)). Exchange is a key player in the metabolic interaction amid high density lipoprotein (*HDL*) particles, triglyceride-rich (*TG*) lipoproteins, very low-density lipoproteins (*VLDL*) and low-density lipoproteins (*LDL*).

In individuals with type-2 diabetes, metabolic syndrome and combined dyslipidemia, cardiovascular risk is increased by a clustering of risk factors such as abdominal obesity, impaired fasting glucose, high blood pressure, low HDL-cholesterol (*HDL-C*) levels, increased triglycerides (*TG*), and high levels of small dense LDL particles (Dobiášová (2004)).

On the other hand, since 1993, health national institutes meet in the Consensus Development Conference in order to evaluate the role of triglycerides in cardiovascular risk assessment (NIH Panel (1993)). However the extent to which triglycerides directly promote cardiovascular disease or represent a risk biomarker has been debated for 3 decades (Miller et al. (2011)).

The role of *TG* as regulators of lipoprotein interactions is backed by the evidence related to their increase in plasma, associated directly to a high incidence of coronary artery disease, an increased population of small dense LDL, and an enhanced cholesteryl ester mass transfer from HDL to apolipoprotein B containing lipoproteins. *TG* have also been proposed to be a major determinant of cholesterol esterification/transfer and HDL remodeling in human plasma (Dobiášová (2004)). The ratio of *TG* to *HDL-C* ratio correlates with the plasma level of small dense LDL particles.

Although the diagnosis currently employed permits some certainty, we can only speak of an elevated clinical suspicion of atherosclerosis risk. Without a trustworthy understanding of factors intervening in atherogenesis and its clinical manifestations, it is not possible to establish a totally clear relationship with many parameters considered as risk factors.

3. FUZZY INFERENCE SYSTEM DESIGN

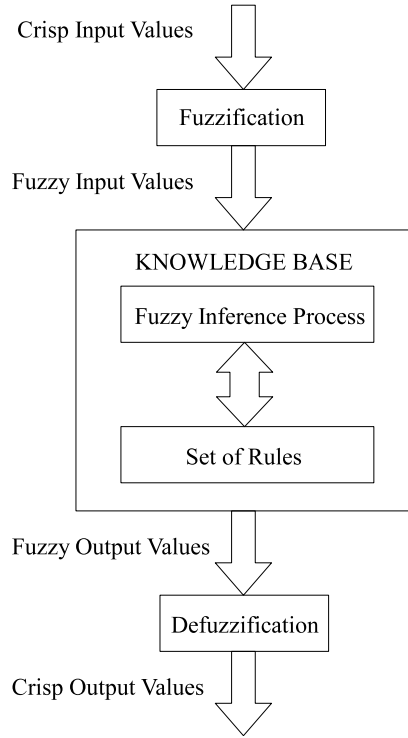


Fig. 1. Fuzzy Inference System Concept.

A fuzzy inference system (FIS) can be designed to emulate to some extent how physicians use data and knowledge to elaborate diagnosis (Hudson and Cohen (1994); Nguyen et al. (2001)). This systems and other hybrids, as neurofuzzy systems (Serhatlioglu et al. (2003)), have been developed for different uses as well as medical applications. However, in many cases they have been used to analyze images or as classifiers (Choi and Krishnapuram (1995); Udupa et al. (1997)). The number of FIS designed to aid diagnosis based on clinical knowledge and analysis is still low (Gorzalczany and McLeish (1992); Wong et al. (2004)).

In this work, we use a fuzzy inference system (Fig. 1) based on rules to determine the state of risk of atherogenesis, in individuals who have had a blood clinical analysis. This FIS has been designed following criteria in which the number, name and shape of membership functions used, along with the inference rules and the way in which results are obtained, emulate the reasoning of a physician.

The same set of four variables shown in Table 1 has been picked up to design such a FIS: Total Cholesterol TC , Low-Density Lipids LDL , Atherogenesis Index AI and Triglycerides TG .

Fig. 2 depicts membership functions for each variable's universe of discourse and (1) shows the mathematical

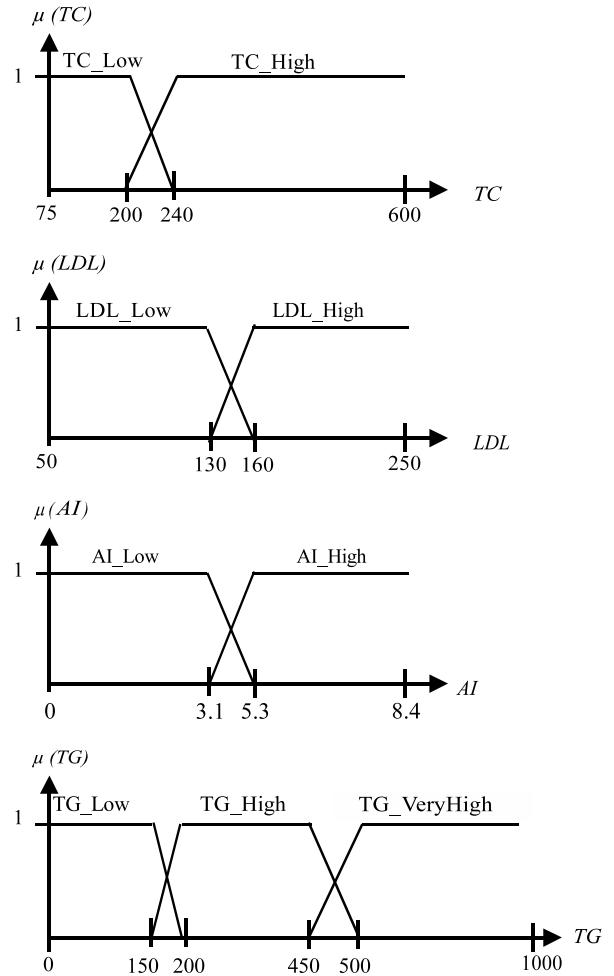


Fig. 2. Membership Functions for the fuzzification of the variables Total Cholesterol TC , Low-Density Lipids LDL , Atherogenic Index AI and Triglycerides TG .

definition for Total Cholesterol TC fuzzification as an example of what has been done with the other variables. It can be noticed that the intervals of those fuzzy sets agree with the limits followed by medical convention in Table 1.

$$\mu_{TC_{Best}}(TC) = \begin{cases} 1 & \text{if } TC < 200 \\ \frac{240 - TC}{240 - 200} & \text{if } 200 \leq TC \leq 240 \\ 0 & \text{if } TC > 240 \end{cases} \quad (1)$$

$$\mu_{TC_{High}}(TC) = \begin{cases} 0 & \text{if } TC < 200 \\ \frac{TC - 200}{240 - 200} & \text{if } 200 \leq TC \leq 240 \\ 1 & \text{if } TC > 240 \end{cases}$$

For $75 \leq TC \leq 600$

For this FIS, it was determined that all 4 variables should count for any possible situation. The total possible number of combinations resulting from all the membership functions of these fuzzified variables is equal to the maximum

number of inference rules that can be obtained, i.e. $2 \times 2 \times 2 \times 3 = 24$.

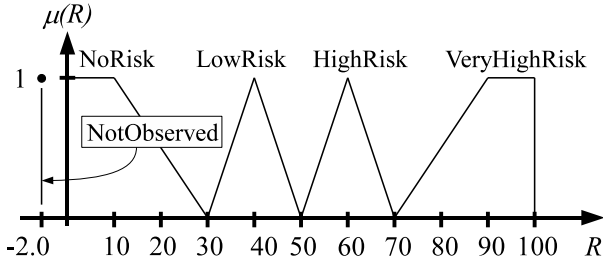


Fig. 3. Membership Functions for Risk R fuzzification.

This is true because we are using the AND operator ($\wedge$) to perform the inference method, which it is expressed in a general form in (2).

$$\begin{aligned} \mu_m(CT, LDL, AI, TG; R) = \\ = \min [(\mu_i(CT) \wedge \mu_j(LDL) \wedge \mu_k(AI) \wedge \mu_h(TG)) ; \\ \mu_p(R)] \end{aligned} \quad (2)$$

But elaborating the set of rules merely by such a purely combinatorial way would result in an imprecise design for this FIS assisting medical diagnosis. The reason is that some combinations are impossible to achieve in reality, due to the metabolism. It is never watched that a situation where simultaneously a very high level of TC appears with very low LDL levels together in a human individual. Fuzzy rules must be composed with close medical assistance. Every rule is the result of discussions among medical and mathematical specialists for this system.

This system's output is Risk R , a qualification that spans from 0 to 100. Fig. 3 depicts membership function shapes designed for this variable. It is possible to realize that four of this membership functions relate directly with what can be established as a risk condition to a patient being diagnosed. A fourth membership function relates with a singleton denoted as "not-observed", defined for those combinations not occurring in a set of lipid levels for a particular individual, as was explained above. The $R = -2$ value indicates a non-operative result for this variable.

Different fuzzy values μ_m are the result to apply each of the different inference rules to the CT , LDL , AI and TG values affecting the consequent R . Fig. 4 shows an example of aggregation of two shot rules and the resultant Risk area, following a Mamdani-type operation, where rule shooting is governed by *minimum* operator.

Once the shootings are performed, all of them are combined through (3)

$$\mu_S(R) = \max\{\mu_1(R), \mu_2(R), \dots, \mu_n(R)\} \quad (3)$$

where n is the total number of rules and μ_S is the fuzzy state for Risk R .

To translate this aggregated set into a crisp number that represents the diagnosis through a qualification by R , we propose a defuzzification method that we identify as the "mean point of absolute maximum", mathematically expressed by (4)

$$R_o = \frac{\min\{u \in N^+(A_i)\} + \max\{u \in N^+(A_i)\}}{2} \quad (4)$$

where the top values reached for every single membership function is obtained by (5)

$$N(A_i) = \{x \in X | \mu(x) = \max\{\mu_{A_i}(x)\}\} \quad (5)$$

being N^+ the larger value, defined for instance as in (6)

$$N^+(A_i) = \max\{\mu(N(A_i))\} \quad (6)$$

The "mean point of absolute maximum" defuzzification method proposed here is graphically depicted in Fig. 5.

It is not the centroid method that is commonly used in many similar systems, because it has the disadvantage of not reaching the extremes of the R variable in the way we have designed it. Besides, this proposed method discriminates portions of resultant areas of the aggregation, stressing the highest with the intention to enforce the membership function with such a value and making clearer the diagnosis obtained.

4. RESULTS

The FIS presented here can be interpreted as an $\mathbb{R}^4 \rightarrow \mathbb{R}$ mapping or more precisely as a function of the form $R = R(CT, LDL, AI, TG)$. As an example of its direct application, table 2 contains results when applied to ten individuals. Each row includes values for Total Cholesterol TC , Low-Density Lipids LDL , Atherogenic Index AI and Triglycerids TG that result from the blood plasma analysis made to patients. The last column shows Risk R evaluated for the inference system.

According to those membership functions shown in Fig. 3, one person can then be qualified definitely as "NoRisk", other two patients as "Low Risk" and "High Risk" respectively, and other two of them as "Very High Risk". However, the rest are rather a transition between two adjoint sets, which can be useful in regarding the possible transient that a clinical history of such cases could mean.

Table 2. TC , LDL , AI , TG values for ten different individuals subject to blood test. The last column shows the Risk R level calculated by the FIS designed

TC [75, 600] (mg/dl)	LDL [50, 250] (mg/dl)	AI [0, 8.4]	TG [0, 1000] (μg/ml)	R [1, 100]
200	102.3	4.5	164	21.3
247	139.5	5.3	384	91.2
183	109.4	4.3	158	16.8
212	130.1	4.0	165	33.9
242	120.4	5.1	169	70.0
212	127.3	4.8	300	64.8
167	93.1	4.7	600	83.5
171	84.4	2.6	400	39.8
320	118.0	8.1	600	94.9
143	56.5	2.3	120	5.14

Graphical interpretations spanning for Risk R can be achieved if two of the antecedent variables are used in order to elaborate an $\mathbb{R}^2 \rightarrow \mathbb{R}$ mapping. For example, $R = R(TC, TG)$ have the graphical representation observed in Fig. 6, i.e. it is the resulting surface when TC and TG

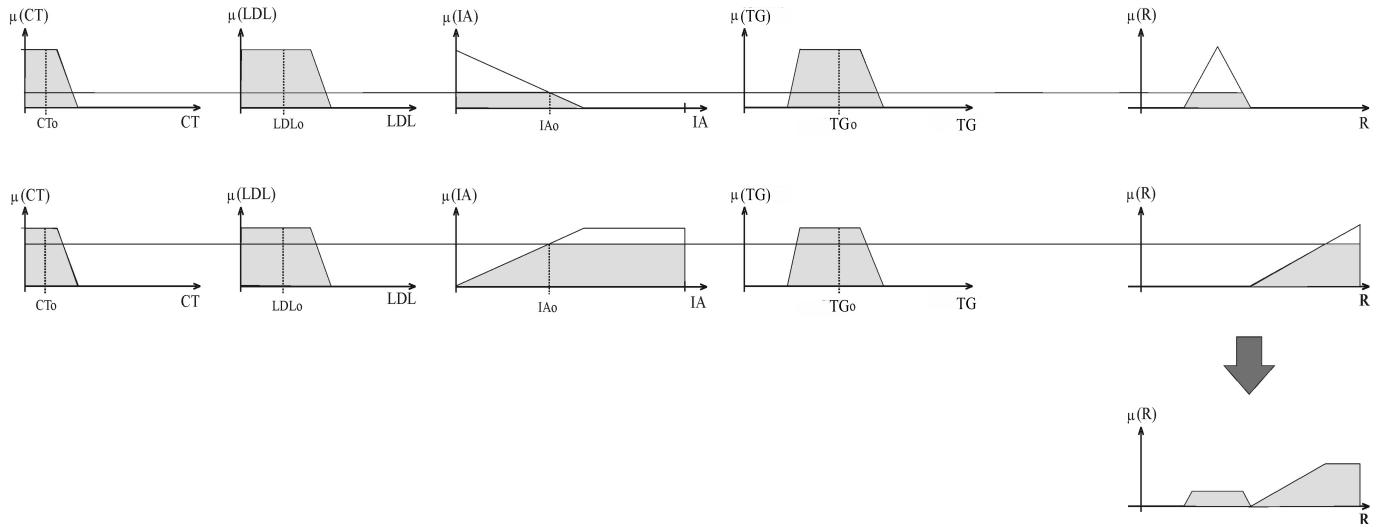


Fig. 4. Graphical representation of rule shooting and Mamdani-type aggregation.

are left as independent variables and LDL and AI are set fixed at $LDL = 128$ mg/dl and $AI = 4.5$.

Deepest regions, in darkest color, correspond to those trends to values of $R = -2$, where related values have no congruency in real situations, as mentioned in Section 3.

Other two-independent-variable combinations are possible and their respective visualizations. However, even though Fig. 6 is useful, to obtain a complete diagnosis it is necessary to take into account the whole set of four variables, establishing relations among all of them simultaneously.

This result is expected. Classification performed is a consequence of the medical criteria that has been translated to the membership functions, rules and operations carried out. However, it is not only a classification what can be reached. As table 2 shows, the objective of emulate a medical diagnosis based purely in four strongly related lipid variables is possible.

5. CONCLUSION

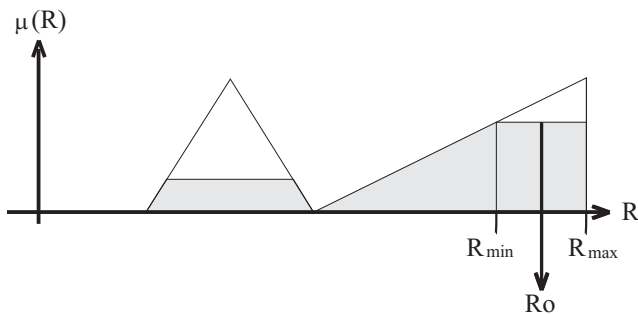


Fig. 5. Mean point of absolute maximum defuzzification method.

Fuzzy inference systems (FIS) have been used since many years ago due to they are relatively easy to design and to program. The application fields in which it is possible to find them is very broad, and the medical practice has been involved with this tool. The very delicate and laborious part in the design of almost all those possible FIS is the selection of variables that constitute the antecedent, where

input information is entered and processed, if the objective is to approximate the outcome to the most real values possible.

In this case, variables selected are those defined by medical convention, with accepted values following the most strict consensus. In a similar way, rules have been composed after discussion sessions among those of us that performed the programming of the FIS and those whose expertise is in the medical field. By these two strong characteristics it is possible to say that expert knowledge has been introduced in the FIS presented here.

Results obtained back up those statements, due to the values obtained for the output value Risk R is in accordance with those results that could be achieved by a physician, if only takes into account that set of four lipid variables. Even though a system like that presented here has not the intention to substitute any real expert, it does constitute a valuable aid to perform a previous and partial diagnosis for any patient.

For the obtention of the output value Risk R , it was necessary to defined a defuzzification method named the

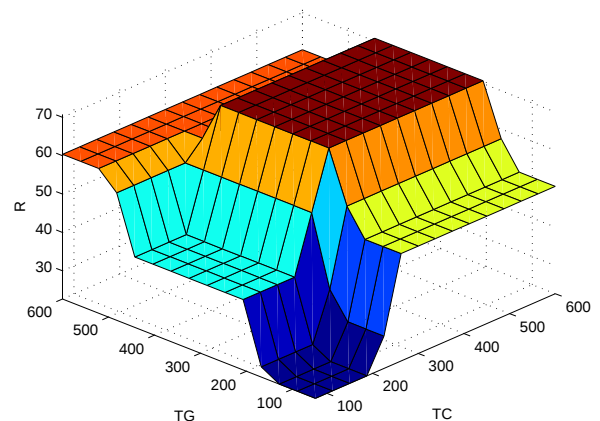


Fig. 6. Graphical representation of $R = R(TC, TG)$.

“mean point of absolute maximum”. It has as a main advantage that the whole range of values for that output variable are covered, unlike other methods like centroid, for example.

Another feature that was designed into this method is that it identifies portions of resultant surfaces of the aggregation, stressing the highest and making clearer the diagnostic obtained.

REFERENCES

- Choi, Y. and Krishnapuram, R. (1995). A fuzzy-rule-based image enhancement method for medical applications. In *Proceedings of the 8th IEEE Symposium on Computer-Based Medical Systems*, 75–80.
- Cohen J.D., Cziraky, M.J., Cai, Q., Wallace, A., Wasser, T., Crouse, J.R., Jacobson, T.A. (2010). 30-Year trends in serum lipids among United States adults: results from the National Health and Nutrition Examination Surveys II, III, and 19992006. In *Am. J. Cardiol.*; 106:969–975.
- Dobiášová, M. (2004). Atherogenic Index of Plasma [Log(Triglycerides/HDL-Cholesterol)]: Theoretical and Practical Implications. *Clinical Chemistry*, 50(7), 1113–1115.
- Gorzalczany, M.B. and McLeish, M. (1992). Combination of neural networks and fuzzy sets as a basis for medical expert systems. In *Fifth Annual IEEE Symposium on Computer-Based Medical Systems*, 412–420.
- Hudson, D.L. and Cohen, M.E. (1994). Fuzzy logic in medical expert systems. *IEEE Engineering in Medicine and Biology*, 13(5), 693–698.
- Lagrost, L. (1994). Regulation of cholesteryl ester transfer protein (CETP) activity: review of in vitro and in vivo studies. *Biochimica et Biophysica Acta*, 1215.
- Miller, M., Stone, N.J., Ballantyne, C., Bittner, V., Criqui, M.H., Ginsberg, H.N., Goldberg, A.C., Howard, W.J., Jacobson, M.S., Kris-Etherton, P.M., Lennie, T.A., Levi, M., Mazzone, T. and Pennathur, S. (2011). Triglycerides and Cardiovascular Disease: A Scientific Statement From the American Heart Association. *Circulation*, 123:2292–2333.
- Nguyen, H.P., Nguyen, B.T., Ding, L., and Hirota, K. (2001). Case based reasoning using fuzzy set theory and the importance of features in medicine. In *IFSA World Congress and 20th NAFIPS International Conference*, volume 2, 872–876. 9th joint.
- NIH Consensus Development Panel on Triglyceride, High-Density Lipoprotein, and Coronary Heart Disease (1993). NIH Consensus Conference: triglyceride, high-density lipoprotein, and coronary heart disease. *JAMA*, 269:505–510.
- Ritsch, A., Scharnagl, H., Eller, P., Tancevski, I., Duwensee, K., Demetz, E., Sandhofer, A., Boehm, B.O., Winkelmann, B.R., and März, J.R.P.W. (2010). Cholesteryl ester transfer protein and mortality in patients undergoing coronary angiography: The Ludwigshafen risk and cardiovascular health study. *Circulation*, 121(3), 366–374.
- Serhatlioglu, S., Bozgeyik, Z., Ozkan, Y., Hardalac, F., and Güler, I. (2003). Neurofuzzy classification of the effect of diabetes mellitus on carotid artery. *Journal of Medical Systems*, 27(5), 457–464.
- Tall, A.R. (1995). Plasma lipid transfer proteins. *Annual Review of Biochemistry*, 64, 235–257.
- Udupa, J.K., Wei, L., Samarasekera, S., Miki, Y., van Buchem, M.A., and Grossman, R.I. (1997). Multiple sclerosis lesion quantification using fuzzy-connectedness principles. *IEEE Transactions on Medical Imaging*, 16(5), 598–609.
- Wong, K.W., Gedeon, T., and Koczy, L. (2004). Construction of fuzzy signature from data: an example of SARS pre-clinical diagnosis system. In *Fuzzy Systems, 2004. Proceedings. 2004 IEEE International Conference on*, volume 3, 1649–1654.