

# A Fuzzy-Mining Approach for Solving Rule Based Expert System Unwieldiness in Medical Domain

Oladipupo O. Olufunke<sup>1</sup>, Uwadia. O. Charles<sup>2</sup> and Ayo K. Charles<sup>2</sup>, Ajith Abraham<sup>3,4</sup> and Vaclav Snasel<sup>4</sup>

<sup>1</sup>Department of Computer Science and Information Sciences, Covenant University, Ota, Nigeria.

<sup>2</sup>Department of Computer Science, University of Lagos, Lagos, Nigeria

<sup>3</sup>Machine Intelligence Research Labs (MIR Labs), Scientific Network for Innovation and Research Excellence, WA, USA

<sup>4</sup>IT4Innovations, VSB - Technical University of Ostrava, Ostrava, Czech Republic

frajooje@yahoo.com, couwadia@yahoo.com, ckayome@yahoo.com, ajith.abraham@ieee.org, vaclav.snasel@vsb.cz

## Abstract

Over the years, one of the challenges of a rule based expert system is the possibility of evolving a compact and consistent knowledge-base with a fewer numbers of rules that are relevant to the application domain, in order to enhance the comprehensibility of the expert system. In this paper, the hybrid of fuzzy rule mining interestingness measures and fuzzy expert system is exploited as a means of solving the problem of unwieldiness and maintenance complication in the rule based expert system. This negatively increases the knowledge-base space complexity and reduces rule access rate which impedes system response time. To validate this concept, the Coronary Heart Disease risk ratio determination is used as the case study. Results of fuzzy expert system with a fewer numbers of rules and fuzzy expert system with a large numbers of rules are presented for comparison. Moreover, the effect of fuzzy linguistic variable risk ratio is investigated. This makes the expert system recommendation close to human perception.

Keywords: Expert system (ES), Fuzzy expert system (FES), Coronary Heart Disease (CHD), Rule Support and Confidence

## 1. Introduction

Expert system (ES) is an intelligent computer program that uses knowledge and inference procedures to solve problems that are difficult enough to require significant human expertise for their solutions [1],[2][3]. There are different approaches into modeling expert systems: the rule based approach, black-board system approach, frame-based approach, open-based expert architecture and object-oriented approach [4]. However, all these approaches have their limitations.

The rule based expert systems collect the small fragments of human knowledge into a knowledge-base in the form of *if-then* rules that are used to reason through a problem, using appropriate knowledge [5]. An important advantage here is that within the domain of the knowledge-base, a different problem can be solved using the same program without re-programming efforts [5]. Rule based expert systems are easy to formulate; they emulate human cognitive process and decision-making ability; and finally, they represent knowledge in a structured homogeneous and modular way. However, there are several limitations of these systems. These limitations include large numbers of rules in the knowledge base that causes the system to become unwieldy because of the presence of rules that might not be relevant to the application domain. This inevitably complicates maintenance especially in the case of subtle updates [6]. Medical domain diseases are attached to different ranges of data, for instance, a disease like coronary heart disease might be common to a certain range of blood pressure measurement, and including a rule outside this range might amount to a redundant rule. The presence of such a rule in an expert system's knowledge base causes the system to become unwieldy and reduces the compactness of the expert system [11] which ultimately increases the expert system memory space, complicates the system maintenance, and reduces rule access rate which impedes system response time. For this reason, this paper adopts a hybrid approach that combines, fuzzy rule mining expert driven approach and fuzzy expert system. This isolates interesting rules based on the expert knowledge in the application domain. Ultimately this leads to reduction in the number of rules in the knowledge-base in order to reduce the system space complexity, accelerates the system response time, and simplifies the systems maintenance.

A fuzzy expert system (FES) is an expert system that consists of fuzzification, inference, knowledge-base, and defuzzification subsystems. It uses a collection of fuzzy membership functions and rules instead of Boolean logic to reason about data in the inference mechanism [7]. Fuzzy *if-then* rules and fuzzy reasoning are the backbone of fuzzy

expert systems, which are the most important modelling tools based on fuzzy set theory. In literature, different approaches have been used to generate fuzzy rule base [8,9,10,11,12,13,14]. For instance, to formulate classical rule-base, the input space is divided into multidimensional portions and then actions are assigned to each of the portions. The standard structure of a fuzzy system is such that given M dimensions where each dimension is partitioned into N subspaces and where there are up to  $N^M$  rules in the fuzzy system [11]. The larger the N, the larger the number of rules and, according to Phayung [15], if all the possible rules are used, then the system would not be compact because of the redundant rules. This has three negative effects on expert systems: 1) it increases the knowledge-base memory usage, since extra space is needed to store the redundant rules; 2) the existent of large number of rules reduces the rule access rate which ultimately slows down the response time of the ES; 3) it make the knowledge-base unwieldy which complicates the system maintenance. Hence, the motivation for this paper is the need to reduce the number of rules in the rule-base in other to enhance the compactness of a rule based expert system.

The remaining part of this paper is organized as follows: Section 2 reviews related research works in the area of fuzzy experts systems, section 3 presents our proposed methodology and elucidates on the methodology for carrying out the objectives, and section 4 reports the experimental design and analysis of the results obtained. The paper is concluded in section 5 with an indication of our future research direction.

## 2. Related Research

### 2.1 Rule Based Expert System

Recently, the introduction of expert system in the field of medicine in areas such as, diagnosis, treatment of illnesses, and determination of risk has highly increased in spite of very high complexity and uncertainty in this field. In [17], classification based data mining technique was used on a medical data set of diabetic patients for diagnosis. In their approach, the rule-base was generated from the classification process which is basically determined by the data and not reflecting the expert knowledge. In [18], a hybrid fuzzy-neural based medical diagnosis system was proposed. The system was not concerned with how to calculate the best membership function for each fuzzy data that also has effect on the accuracy of the system. In addition to that, many medical diagnosis systems and risk deterministic systems have been implemented with classical fuzzy rule formulation. In [19], a fuzzy expert system for diagnosis of prostate cancer was designed and in [20] a fuzzy expert system for determination of Coronary Heart Disease (CHD) risk ratio was designed to determine the patient risk ratio over ten years. In their work, four determinant factors were considered to determine the risk ratio: age, cholesterol, high density of lipoprotein cholesterol and blood pressure. Based on the determinant factors and the fuzzy sets (subspaces), 108 rules were evolved for the rule-base, and according to Phayung, if all possible rules are considered, then the system tends to become less compact, unwieldy and complicates the system maintenance especially in the case of subtle updates [15].

Our approach seeks to solve the problem of unwieldiness in a rule based expert system by adopting a hybrid methodology that combines the fuzzy rule mining interestingness measures and fuzzy expert system. The objectives of this approach is to eliminate redundant rules in the expert system in order to reduce the ES space complexity, increase rule access rate, accelerate the system response time and simplify system maintenance. To validate this approach, the determination of the risk ratio for Coronary Heart Disease (CHD) is used as the case study.

### 2.2 Fuzzy Database

A fuzzy database  $D = \{t_1, t_2, \dots, t_n\}$  where  $t_i$  represents the  $i^{\text{th}}$  record in D. Given  $I = \{i_1, i_2, \dots, i_m\}$  where  $i_j$  represents the  $j^{\text{th}}$  attributes in D. I, is called an itemset since it contains a set of items. Then,  $\forall i_k \exists$  several fuzzy sets  $F_{i_k}$  associated with  $i_k$  and  $f_{i_k}^j$  represent the  $j^{\text{th}}$  fuzzy set in  $F_{i_k}$ . For example, if attribute Age has three fuzzy sets; Young, Middle and Old, we will have

$$F_{Age} = \{\text{YoungAge}, \text{MiddleAge}, \text{OldAge}\}.$$

The fuzzy set and the corresponding membership functions are to be provided by the domain expert [20]. Table 1a is a database instance from Coronary Heart Disease (CHD) patient quantitative record: Age; years and Cholesterol; mg/dL. Table 1b is a database instance from Coronary heart disease patient fuzzified record.

Given a database D with attribute I, and those fuzzy sets  $F_{i_k}$  associated with attributes in I, we can evolve fuzzy *if-then* rule such as

$$\text{if } X \text{ is } A \text{ then } Y \text{ is } B$$

In the above rule X and Y are ordered subsets of I and are disjoint, which means that they share no common attribute. A is a fuzzy set in X and B is a fuzzy set in Y. “X is A” is the rule antecedent and “Y is B” is the rule consequent. An itemset  $\langle X, A \rangle$  is said to be frequent if its support value satisfied the minimum support threshold. The frequent itemset obtained  $\langle Z, C \rangle$  is used to generate all possible rules.

Table 1a

An instance of CHD patient quantitative record

	Age	Cholesterol	Risk
$t_1$	30	180	0
$t_2$	35	190	0
$t_3$	48	260	8
$t_4$	57	300	8
$t_5$	65	250	18

(a)

Table 1b

An instance of CHD patient fuzzified record

	Young Age	Middle Age	Old Age	Low Cholesterol	Normal Cholesterol	High Cholesterol	Very Low Risk	Low Risk	Normal Risk	High Risk	Very High Risk
$t_1$	1.00	0.00	0.00	0.50	0.50	0.00	1.00	0.00	0.00	0.00	0.00
$t_2$	0.50	0.50	0.00	0.25	0.75	0.00	1.00	0.00	0.00	0.00	0.00
$t_3$	0.00	1.00	0.00	0.00	0.50	0.50	0.00	0.70	0.30	0.00	0.00
$t_4$	0.00	0.65	0.35	0.00	0.00	1.00	0.00	0.70	0.30	0.00	0.00
$t_5$	0.00	0.25	0.75	0.00	0.75	0.25	0.00	0.00	0.70	0.30	0.00

(b)

If the union of antecedent  $\langle X, A \rangle$  and consequent  $\langle Y, B \rangle$  has sufficient support and the rule has high confidence, then, the rule is said to be interesting. According to Gyenesi [21], the term rule support and rule confidence are defined thus:

- i) The fuzzy support value is calculated by the summation of all votes of each record with respect to the specified itemset. Then, it is divided by the total number of records. Each record contributes a vote which falls in [0,1]. The fuzzy support of itemset  $\langle X, A \rangle$  in transaction set D is

$$Support = \frac{Sum\ of\ votes\ satisfying\ \langle X, A \rangle}{Number\ of\ records\ in\ D} \quad (1)$$

$$FS_{\langle X, A \rangle} = \frac{\sum_{t_i \in D} \prod_{x_j \in X} d_{x_j}(a_j, t_i, x_j)}{|D|} \quad (2)$$

In calculating the vote of  $t_i$  other operator other than  $\prod$  (*mul*) can be used e.g. *min*, *max*. The support value does not only reflect the number of records supporting the itemset, but also indicates the degree of relevance in the application domain.

- ii) The fuzzy confidence value of a rule is given as follows:

$$Support = \frac{Sum\ of\ votes\ satisfying\ \langle Z, C \rangle}{Sum\ of\ votes\ satisfying\ \langle X, A \rangle} \quad (3)$$

$$FS_{\langle\langle X,A \rangle, \langle Z,C \rangle\rangle} = \frac{\sum_{t_i \in D} \prod_{z_j \in Z} d_{z_j}(c_j, t_i, z_j)}{|D|} \quad (4)$$

$$= \frac{\sum_{t_i \in D} \prod_{x_j \in X} d_{x_j}(a_j, t_i, x_j)}{|D|} \quad (5)$$

where  $Z = X \cup Y, C = A \cup B$ .

**Example 1:** This example illustrate the calculation of the fuzzy support value and the fuzzy confidence value for the fuzzified database shown in table 1 (b).

Let  $X = \{\text{Age, Cholesterol}\}, A = \{\text{YoungAge, LowCholesterol}\}$  and

$Y = \{\text{CHDRisk}\}, B = \{\text{VeryLowRisk}\}.$

The support of itemset  $\langle X, A \rangle$  is

$$FS_{\langle X, A \rangle} = \frac{0.5 + 0.125 + 0 + 0 + 0}{5} = 0.625$$

The fuzzy confidence value of the rule, “If Age is young and Cholesterol is Low then CHDRisk is Low” is as follows.

$$FC_{\langle\langle X,A \rangle, \langle Y,B \rangle\rangle} = \frac{0.5 + 0.125 + 0 + 0 + 0}{0.5 + 0.125 + 0 + 0 + 0} = 1$$

The result, indicate that to a degree of 0.625, the rule is

relevant within this domain and the confidence is 1.

### 3. Proposed Methodology

The proposed methodology consists of two phases: the interesting rule mining phase and fuzzy expert system design phase. The first phase consists of 5 steps. In step 1, the existing data is collected and stored as the fact-base. The CHD risk determinant factors are identified to compose a fact-base. In step 2, membership expressions for the input variables and the output variable are modeled, while fuzzy rules are generated based on the input variables dimension and their subspaces. In step 3 and step 4, the quantitative attributes values in the fact-base are transformed into a fuzzy database using the membership expressions modelled in step 2. Finally, in step 5, relevant fuzzy rules are extracted from rules generated in step 3 in the form of *if-then* rules based on the interestingness measures of support and confidence. These interesting fuzzy rules are later used to compose the rule-base for the fuzzy expert system design phase. Notice that step 2 and 3 can be executed independently in parallel. Fig. 1 shows an activity workflow of the proposed methodology.

#### 3.1 Data Collection & Identification of Determinant Factors

Data was collected from an on-line database and from literature for non-smoking men and women based on their CHD determinant factors and their risk ratio to compose the fact-base. Determination of CHD risk ratio over a period of years depends on various items such as weight, smoking, sex, disease history of patients, family genetic factors, blood sedimentation, etc. [20]. In this study, we assume that age, year, cholesterol level, mg/dL, high density lipoprotein cholesterol, mg/dL (HDL-C) and systolic blood pressure level, and mm/Hg essentially determine the CHD risk ratio for the next 10 years [20].

#### 3.2 Fuzzy Interval Partitioning

As it is well known, real world medical knowledge is often characterized by inaccuracy. Medical terms do not usually have a clear-cut interpretation. Fuzzy logic makes it possible to define inexact medical entities via fuzzy sets [23]. In the last decades, a number of hybrid techniques based on fuzzy sets and rules have appeared and been applied to medical systems [24, 25]. One of the reasons is that fuzzy logic provides capabilities for approximate reasoning, *i.e.*, reasoning with inaccurate (or fuzzy) values expressed as linguistic terms, and allows for overlapping of element within the neighboring linguistic terms, which in turn prevents over estimation of boundary values [5]. In the application domain of CHD, the input linguistic variables are: age, cholesterol, HDL-C, blood pressure, and output linguistic variable in

CHD risk ratio. The linguistic variables are partitioned according to expert doctors' analysis. Table 2 shows the linguistic variables and their fuzzy sets.

The trapezoidal membership function is used for the fuzzy sets membership calibration because of its support for the fuzzy sets data ranges. For instance, age is partitioned into YoungAge, MiddleAge and OldAge as the linguistic values such that  $\forall x \in Age$ , the fuzzy membership models and the graph are shown below in equation (6) and figure 2 respectively:

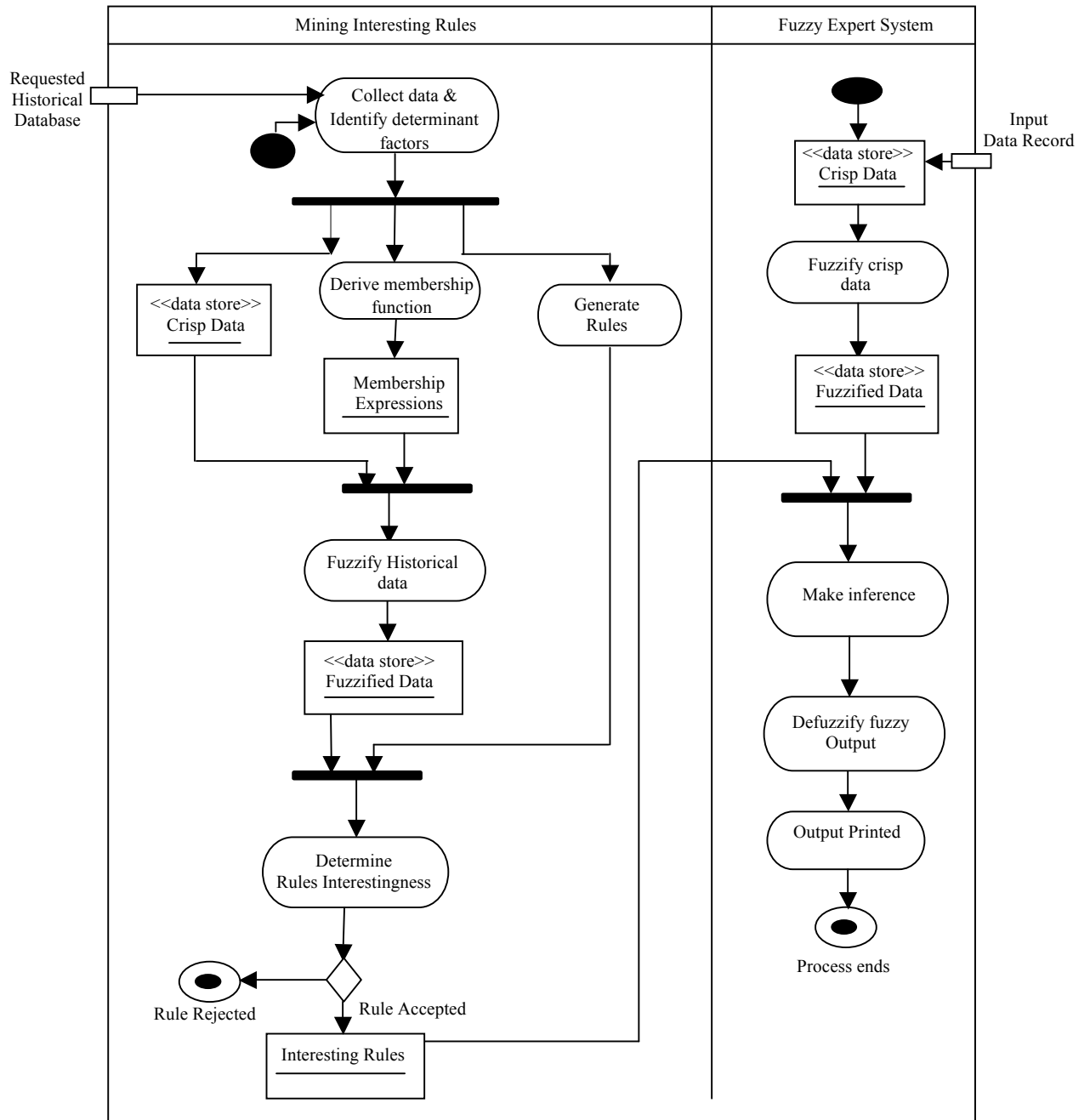


Fig. 1. Partitioned Activity Workflow of the Proposed Methodology

The trapezoidal membership function is used for the fuzzy sets membership calibration because of its support for the fuzzy sets data ranges. For instance, age is partitioned into YoungAge, MiddleAge and OldAge as the linguistic values such that  $\forall x \in \text{Age}$ , the fuzzy membership models and the graph are shown below in equation (6) and figure 2 respectively:

Table 2  
Linguistic terms for membership function in fuzzy model

Linguistic variable	Fuzzy set	Domain
Age	YoungAge, MiddleAge, OldAge	Input
Cholesterol	Low, Normal, High	Input
HDL-C	Low, Middle, High	Input
Blood Pressure	Low, Middle, High, Very High	Input
CHD risk ration	Very Low, Low, Middle, High, Very high	Output

$$\mu_{\text{youngAge}}(x) = \begin{cases} 1 & x \leq 20 \\ \frac{40-x}{20} & 20 \leq x \leq 40 \end{cases} \quad \mu_{\text{middleAge}}(x) = \begin{cases} \frac{x-20}{20} & 20 \leq x \leq 40 \\ 1 & 40 \leq x \leq 50 \\ \frac{(60-x)}{10} & 50 \leq x \leq 60 \end{cases} \quad \mu_{\text{oldAge}}(x) = \begin{cases} \frac{(x-50)}{10} & 50 < x \leq 60 \\ 1 & 60 \leq x \leq 70 \end{cases} \quad \dots (6)$$

For other determinant factors the linguistic expressions are similarly determined.

### 3.3 Construction of the Fuzzy Rules.

Fuzzy expert systems make decision and generate output values based on knowledge provided by the designer in the form of IF {condition} THEN {action} rules. The rule-base specifies qualitatively how the output of the system; ‘‘CHD Risk Ratio’’ is determined for various instances of the input variables of Age, Cholesterol, HDL-C and Blood pressure. If the standard structure of constructing a rule-base is to be followed for this experiment based on the 4 dimensions input variables and their subspace, we would have 108 rules for the fuzzy expert system rule-base as done in [20]. A sample of the rule-base structure is shown in table 3.

### 3.4 Fuzzification Process and Rule Interestingness

In *fuzzification process*, the crisp input values from the fact-base are converted to membership degrees (fuzzy values). The fuzzied fact-base is used to determine the support and confidence for each rule in order to determine the interestingness of each rule. A rule is interesting if it has sufficient support and a high confidence value. Rule support and confidence indicate the relevance of the rule to the application domain. The rule support and confidence is calculated according to a method proposed by Gyenesei [22]. From example 1:

If (Age is YoungAge) and (Cholesterol is Low) and (HDL-C is High) and (Blood\_Pressure is High) then (CHD\_Risk is VeryLow)

it gives 0.01 *support* and 1.00 *confidence* with *product* operator,  
0.02 *support* and 1.00 *confidence* with *min* operator, and  
0.66 *support* and 1.04 *confidence* with *max* operator

This rule is said to be interesting if the minimum support is 0.01 and minimum confidence is 0.5 i.e. The rule is relevant in modelling a fuzzy expert system to determine CHD risk ratio in medical domain.

### 3.5 Development of a Fuzzy Expert System with Interesting Rules

Reasoning in a fuzzy expert system includes three stages: fuzzification, inference, defuzzification. In fuzzification, the crisp input values (from the fact-base) are converted to membership degrees (fuzzy values). In the inference stage, the MIN method operator is used for the combination of a rule's conditions, to produce the membership value of the conclusion, and the MAX method operator is used to combine the conclusions of the rules. In defuzzification, the centroid method is used to convert a fuzzy output to a crisp value [5].

## 4. Experimentation Design

### Phase 1: Mining of Interesting Rule

A program was written in C# programming language to determine the interestingness of the rules. A pilot implementation of the proposed methodology was undertaken using SQL server 2005 as the database management system for data storage. To determine each rule interestingness, the minimum support and minimum confidence values were varied to observe their effect on the rule base, which in turn determines the compactness and accuracy of the proposed methodology. Fig. 3 shows the snapshot of the interface for mining interesting rules as implemented by C# programming language implementation for mining interesting rules. Table 4 presents the interesting rules with 0.01 minimum support value and 0.5 minimum confidence value.

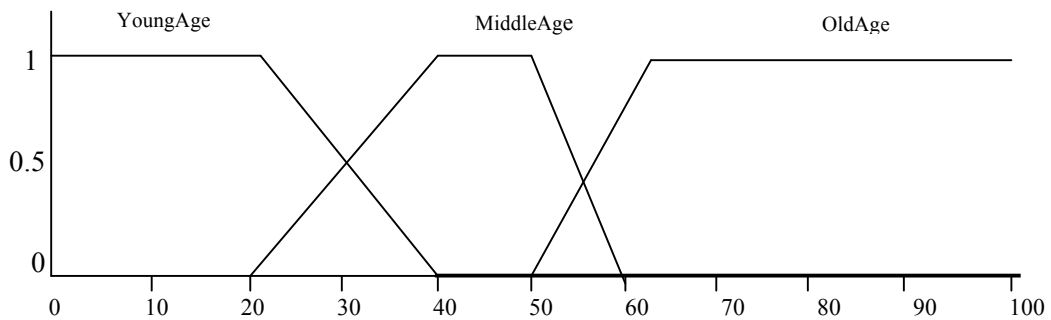


Fig. 2. Membership function for input variable "Age" years

Table 3  
Rule-base structure

Rule No	Rule
1	OldAge,HighCholesterol,HighHDLc,MiddleBloodPressure->MiddleRisk
2	OldAge,HighCholesterol,HighHDLc,HighBloodPressure->MiddleRisk
6	YoungAge,LowCholesterol,LowHDLc,HighBloodPressure->VeryLowRisk
10	YoungAge,NormalCholesterol,LowHDLc,HighBloodPressure->VeryLowRisk
18	MiddleAge,LowCholesterol,LowHDLc,HighBloodPressure->LowRisk
21	MiddleAge,NormalCholesterol,LowHDLc,MiddleBloodPressure->LowRisk
24	MiddleAge,HighCholesterol,LowHDLc,LowBloodPressure->MiddleRisk
25	MiddleAge,HighCholesterol,LowHDLc,MiddleBloodPressure->MiddleRisk
27	MiddleAge,HighCholesterol,LowHDLc,VeryHighBloodPressure->VeryHighRisk
30	OldAge,LowCholesterol,LowHDLc,HighBloodPressure->VeryHighRisk
--	--
--	--
--	--
96	MiddleAge,HighCholesterol,HighHDLc,LowBloodPressure->LowRisk
106	OldAge,NormalCholesterol,HighHDLc,HighBloodPressure->MiddleRisk
108	OldAge,HighCholesterol,HighHDLc,LowBloodPressure->LowRisk

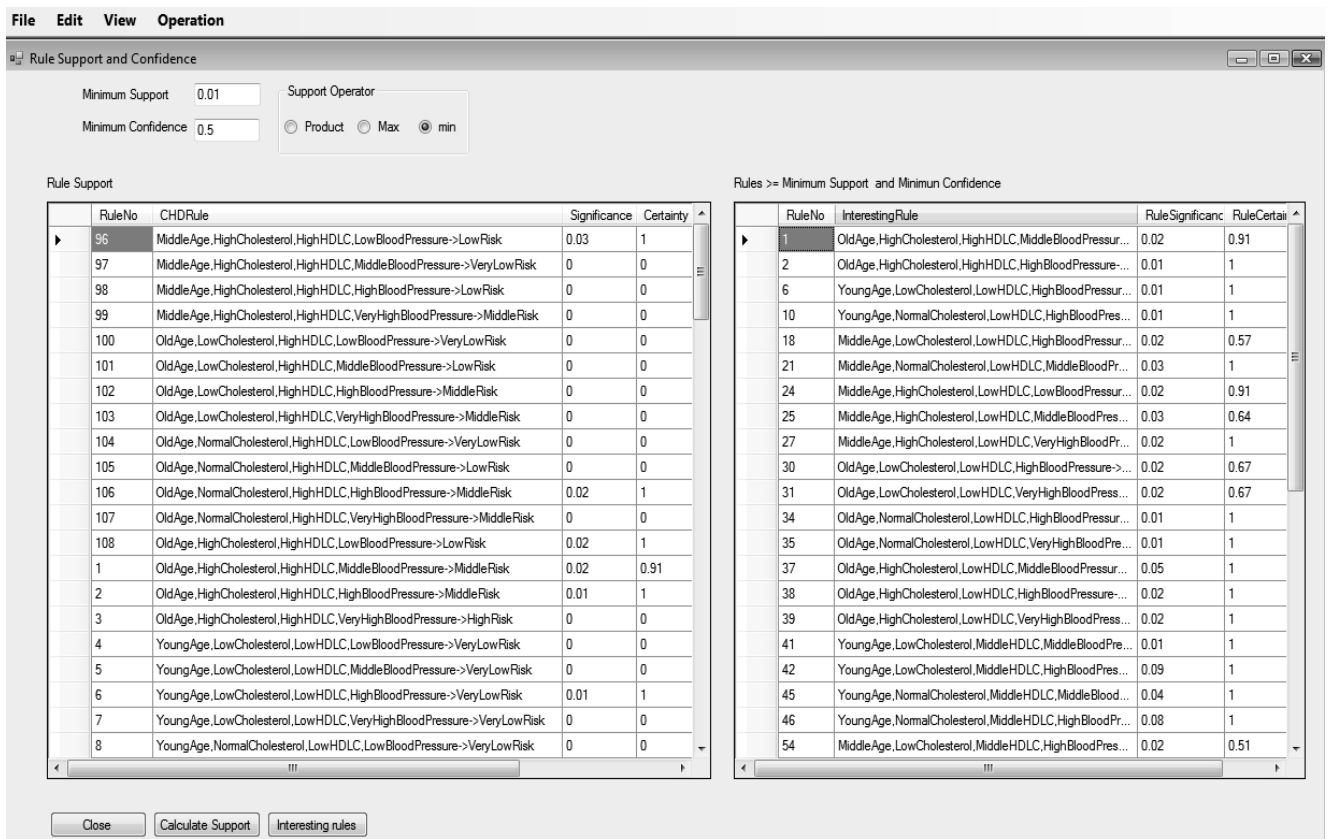


Fig. 3. Snapshot of C# implementation to determine rule interestingness



Table 4

Interesting rules at 0.01 min support and 0.5 min confidence.

S/n	Rule No	Rule	Support	Confidence
1	1	OldAge,HighCholesterol,HighHDL,LowBloodPressure->MiddleRisk	0.02	0.91
2	2	OldAge,HighCholesterol,HighHDL,HighBloodPressure->MiddleRisk	0.01	1
3	6	YoungAge,LowCholesterol,LowHDL,HighBloodPressure->VeryLowRisk	0.01	1
4	10	YoungAge,NormalCholesterol,LowHDL,HighBloodPressure->VeryLowRisk	0.01	1
5	18	MiddleAge,LowCholesterol,LowHDL,HighBloodPressure->LowRisk	0.02	0.57
6	21	MiddleAge,NormalCholesterol,LowHDL,MiddleBloodPressure->LowRisk	0.03	1
7	24	MiddleAge,HighCholesterol,LowHDL,LowBloodPressure->MiddleRisk	0.02	0.91
8	25	MiddleAge,HighCholesterol,LowHDL,MiddleBloodPressure->MiddleRisk	0.03	0.64
9	27	MiddleAge,HighCholesterol,LowHDL,VeryHighBloodPressure->VeryHighRisk	0.02	1
10	30	OldAge,LowCholesterol,LowHDL,HighBloodPressure->VeryHighRisk	0.02	0.67
-	--	--	--	--
-	--	--	--	--
-	--	--	--	--
37	96	MiddleAge,HighCholesterol,HighHDL,LowBloodPressure->LowRisk	0.03	1
38	106	OldAge,NormalCholesterol,HighHDL,HighBloodPressure->MiddleRisk	0.02	1
39	108	OldAge,HighCholesterol,HighHDL,LowBloodPressure->LowRisk	0.02	1

## Phase 2: Fuzzy Expert System Design

The Matlab™ fuzzy logic toolbox is used to simulate the fuzzy expert system using the Mandani fuzzy inference. The Trapezoidal membership function (trapmf) is chosen to model membership function for the input variables (Age, Cholesterol, HDL-C and Blood pressure ) fuzzy set; and output variable; CHD risk ratio. The rule view for the fuzzy expert system is shown in fig. 4.

## 5. System Evaluation

To evaluate the proposed methodology, cases of 20 non-smoking men, whose CHD risk ratio has been successfully determined by Adult Treatment panel (ATP III), were used [20]. Two fuzzy expert systems were simulated: one with all possible number of rules (108 rules), and the other with interesting rules (39 rules) where support value is 0.01 and confidence value is 0.5. The two fuzzy expert systems were based on the same membership function, fuzzy inference engine and defuzzification method. The outputs from the systems were reported in crisp value and fuzzy linguistic value. The result showing the patient lab test record, ATP III risk ratio, FES with 108 rules and FES with 39 rules risk ratio and the linguistic values is presented in Table 5.

## 5.1 Result Analysis on Compactness and Accuracy

According to Phayung [15] fuzzy expert system compactness relates to three aspects: a small number of linguistic terms in each dimension, a small number of fuzzy rules in the rule base, and a small number of conditions in the rule premise or antecedent part. From the result obtained (see Table 5), the FES with interesting rule is more compact compared to the FES with all possible rules. Although they have the same number of linguistic term, and number of conditions in their premise, the later has 108 rules while the former has 39 rules, which makes the expert system more compact. The outputs (CHD risk ratio) from both systems are not significantly different. Moreover, considering the fuzzy linguistic value for the CHD risk ratio, the similarity in the output is up to 90%. Comparing the two results with ATP III, the former has 80% accuracy while the later has 70% accuracy. With these observations, it is evident that a fuzzy expert system could perform more accurately even when few interesting rules are considered. This prevents unwieldiness of the rule-base, and maintenance complication which amplified memory usage and reduces rules accessing rate of the FES. A graphical overview of the analysis is shown with the charts in Figs. 5 and 6.

In Fig. 6, the linguistic values for CHD % risk: VeryLow, Low, Middle, High and VeryHigh, are represented with 1,2,3,4,5, respectively. The graph shows that in many instances the risk ratios fall under the same linguistic value.

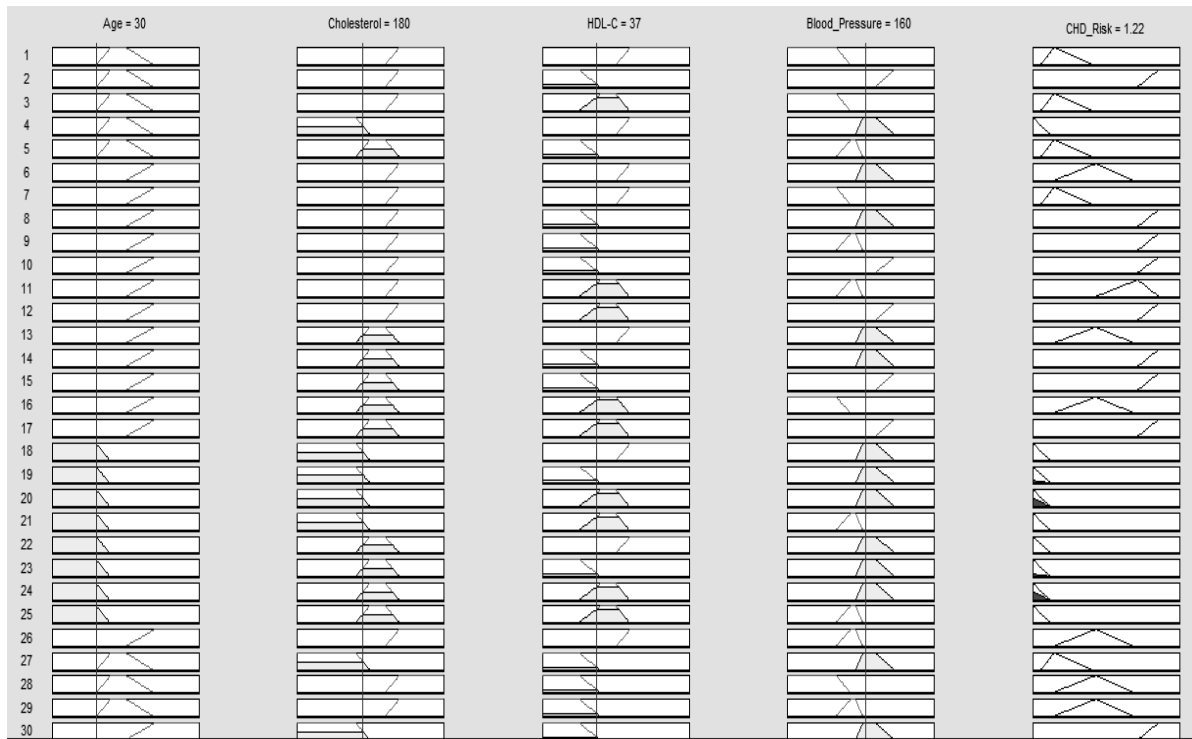


Fig. 4. FES CHD risk indices for a patient record.

Table 5  
Comparison of CHD indices for non-smoking men

Patient no.	Age	Cholesterol	HDL	Blood Pressure	ATP III	FES With 108 Rules	FES With 39 Rules	ATP III CHD risk Linguistic	FES CHD risk Linguistic value	FES CHD risk Linguistic value
1	30	180	37	160	0	1.22	1.22	VeryLow	VeryLow	VeryLow
2	35	190	45	145	0	5.68	5.20	VeryLow	Low	Low
3	48	260	33	120	8	11.2	12.4	Low	Low	Low
4	57	300	67	110	8	9.68	10.6	Low	Low	Low
5	65	250	54	170	18	18.9	18.6	Middle	Middle	Middle
6	75	290	25	135	30	31.8	31.2	VeryHigh	VeryHigh	VeryHigh
7	30	160	49	160	0	1.22	1.22	VeryLow	VeryLow	VeryLow
8	40	310	33	140	8	14.7	14.7	VeryLow	Middle	Middle
9	55	300	26	200	30	25.9	28.9	VeryHigh	High	High
10	60	230	39	110	11	11.3	12.2	Low	Low	Low
11	70	210	45	130	16	14.7	14.7	Middle	Middle	Middle
12	30	240	50	150	0	1.22	1.22	VeryLow	VeryLow	VeryLow
13	35	180	65	160	0	5.52	1.22	VeryLow	Low	*VeryLow
14	45	300	47	155	9	14.7	14.7	Low	Middle	Middle
15	55	300	49	160	16	17.3	18.2	Middle	Middle	Middle
16	65	250	41	140	18	15.4	16.1	Middle	Middle	Middle
17	70	260	38	190	30	26.7	31.2	VeryHigh	High	*VeryHigh
18	44	210	37	180	5	7	7	Low	Low	Low
19	55	150	30	200	11	17.3	16.6	Middle	Middle	Middle
20	66	150	26	200	28	24.5	26.4	High	High	High

## 5.0 Conclusion

In this paper, we proposed a fuzzy-mining expert system that introduces interestingness measure of support and confidence into a rule-base fuzzy expert system to enhance the compactness of a fuzzy expert system. The viability of the proposed methodology is confirmed by designing a fuzzy expert system for determining CHD risk ratio for non-smoking men. The result from the experiment reveals that the new approach has the potential to reduce the rule base by 63% and still produces a reasonable result of 90% accuracy. With these observations, it is evident that a fuzzy expert system could perform more accurately even when few interesting rules are considered. This will solve the problem of unwieldiness and maintenance complication of a rule-base FES, which negatively amplifies memory usage requirement of the system, reduces the rule access rate and maintenance complication. In the future, the effectiveness of other soft computing techniques like genetic algorithm and rough sets for rule optimization in FES would be investigated.

## Acknowledgement

This work has been partly supported by the European Regional Development Fund in the IT4Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070) and by the Bio-Inspired Methods: research, development and knowledge transfer project, reg. no. CZ.1.07/2.3.00/20.0073 funded by Operational Programme Education for Competitiveness, co-financed by ESF and state budget of the Czech Republic.

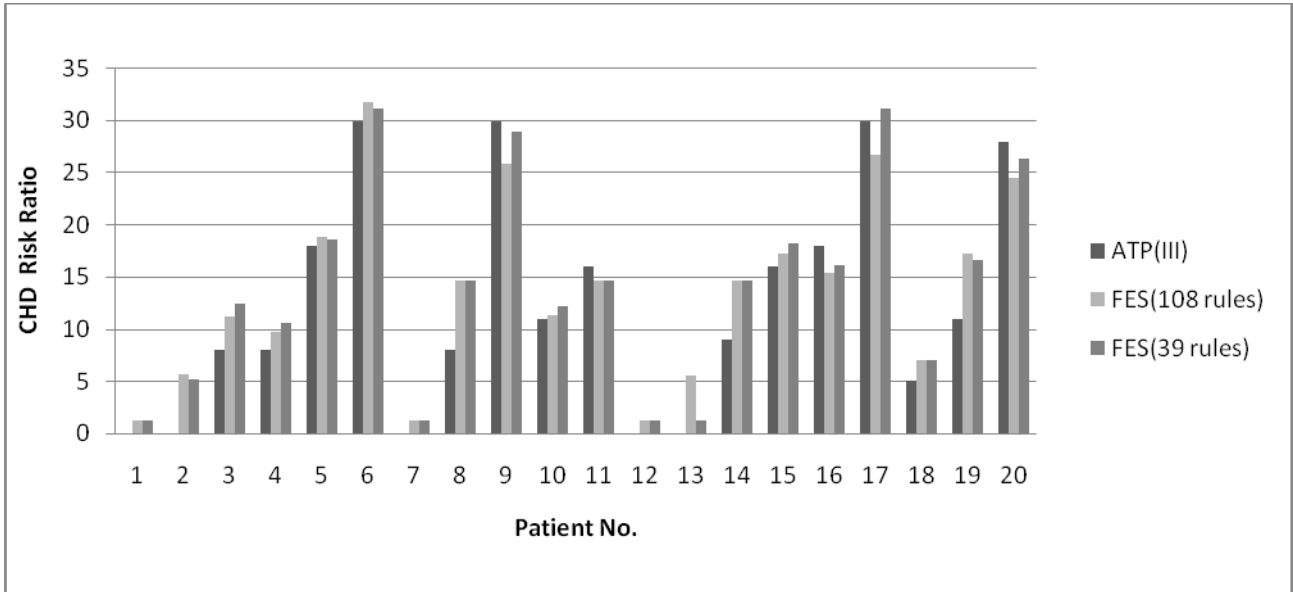


Fig. 5. ATP III, FES (108 rules) and FES (39 rules) risk ratio diagramatic representation

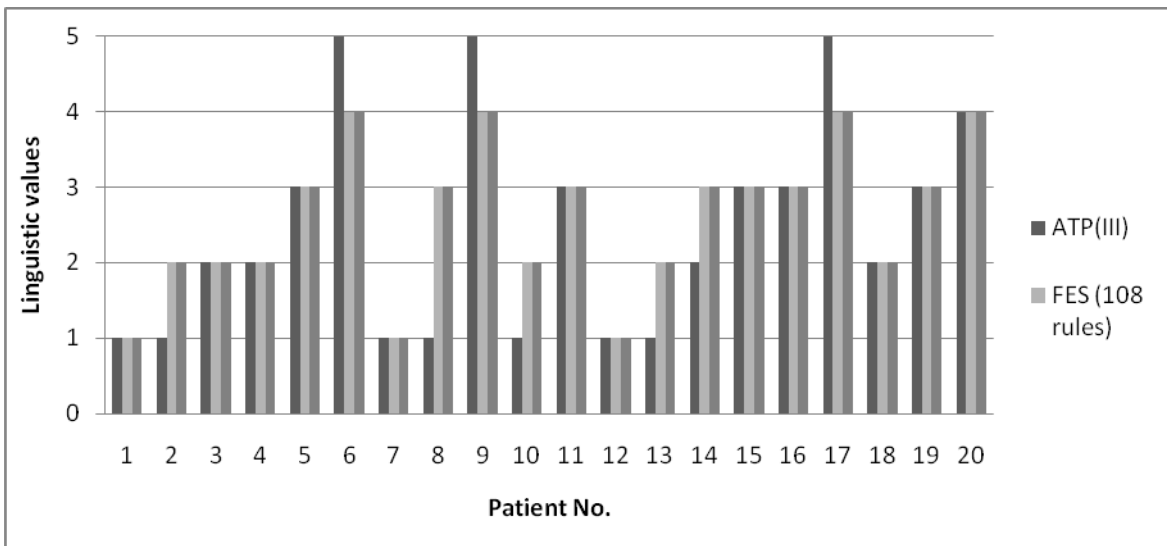


Fig. 6. The CHD risk ratio linguistic value diagramatic representation

## References

- [1] E.A. Feigenbaum, Knowledge Engineering in the 1980s. Stanford: Department of Computer Science, Stanford University, Stanford CA. (1982).
- [2] V.L. Patel, E.H. Shortliffe, M. Stefanelli, P. Szolovits, M.R. Berthold, R. Bellazzi, A. Abu-Hanna. The coming of age of artificial intelligence in medicine (2009) Artificial Intelligence in Medicine, 46 (1), pp. 5-17.
- [3] S.-H. Liao, Expert system methodologies and applications-a decade review from 1995 to 2004. (2005) Expert Systems with Applications, 28 (1), pp. 93-103

- [4] S. Aly and I. Vrana, "Toward efficient modeling of fuzzy expert systems: a survey" *AGRIC. ECON.- CZECH*, vol. 52(10). (2006). Pp. 456-460.
- [5] A. Ajith Abraham, "Rule-based Expert Systems: Handbook of Measuring System Design", edited by Peter H. Sydenham and Richard Thorn. John Wiley & Sons, Ltd. ISBN: 0-470-02143-8. (2005)
- [6] R. Kaula and L.C. Lander, "A module-base conceptual framework for large-scale expert system". *Industrial Management & data System*, (1995) 95(2) pp. 15-23
- [7] M. Schneider, G. Lagholz, A. kandel and G. Chew, "Fuzzy Expert System Tools" John Wiley & Sons, USA ISBN-10: 0471958670. (1996).
- [8] S. Satya, R. Rajkumar and T. Ashutosh, "Development of fuzzy Expert System for Customer and Service advisor categorization within contact center environment" 10th World online Conference on Soft Computing in Industrial Applications, Applications of soft computing: Recent Trends Series: Advances in Soft Computing, (2005).Vol. 36
- [9] S. H. Christoph, "A Hybrid Fuzzy-Neural Expert System for Diagnosis" *Proceedings of the international Joint Conference on Artificial Intelligence (IJCAI)*, Montreal, Canada ISBN ~ ISSN:1045-0823, 978-1-558-60363-9, (1995). Pages 494-500.
- [10] B. I. Norbik and S. Bharanidharan, "Novel Attack Detection Using Fuzzy Logic and Data Mining" "Proceedings of the 2006 International Conference on Security & Management", SAM 2006, Las Vegas, Nevada, USA, June 26-31, (2006).CSREA Press 2006, ISBN 1-60132-001-9
- [11] J.E. Dickerson and J.A. Dickerson, "Fuzzy Network Profiling for Intrusion Detection" *Proceedings of NAFIPS 19th International Conference of the North American Fuzzy Information Processing Society*, Atlanta, July, (2000). 301-306.
- [12] Y. Bai, D. Wang, *Fundamentals of Fuzzy Logic Control-Fuzzy Sets, Fuzzy Rules and Defuzzifications* (2006) *Advanced Fuzzy Logic Technologies in Industrial Applications*, pp. 17-36.
- [13] V. Snasel, P. Krömer, J. Platoš, A. Abraham. *The evolution of fuzzy classifier for data mining with applications* (2010) *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6457 LNCS, pp. 349-358. ISBN: 3642172970; 978-364217297-7
- [14] P. Krömer, S. Owais, J. Platoš, V. Snášel. *Towards new directions of data mining by evolutionary fuzzy rules and symbolic regression. Computers and Mathematics with Applications* 66 (2), pp. 190-200. (2013)
- [15] M. Phayung, "Quantitative measures of a Fuzzy Expert System" *Final Report IEEE Neural Network Council Student summer Research*. (2001).
- [16] A. Barr and E. A. Feigenbaum, "The handbook of artificial intelligence" (Vols. I and II). Los Altos, CA: Kaufmann. (1982).
- [17] K. Harleen and K.W. Siri, "Emperical Study on Applications of Data mining Techniques in Healthcare." *Journal of Computer Science* 2(2), (2006). ISSN 1549-3636, pp. 194-200.
- [18] S. Moein, S.A. Monayjemi and P. Moallem, "A Novel Fuzzy-Neural Based Medical Diagnosis System" *Proceedings of World Academy of Science, Engineering and Technology*. (2008). Vol. 27. ISSN 1307-6884.
- [19] I. Saritas, N. Allahverdi and U. Sert, "A Fuzzy Expert System Design for Diagnosis of Prostate Cancer", in *Proc. Intern. Conference on Computer Systems and Technologies CompSysTech'2003- CompSysTech'2003*, Sofia, Bulgaria, 18-20 June (2003)
- [20] A. Navruz and T. Serhat, "Design a Fuzzy Expert System for Determining of Coronary Heart Disease Risk" *ACM International Conference Proceeding Series; Vol. 285*, ISBN:978-954-9641-50-9 *Proceedings of the 2007 international conference on Computer systems and technologies*, (2007). Article No. 36.
- [21] A. Gyenesei, "A fuzzy approach for mining quantitative association rules," *Acta Cybern.*, vol. 15, no. 2, (2001). pp. 305–320.
- [22] J. Bayliss, "Framingham risk score to predict 10 year absolute risk of CHD event west Hertfordshire cardiology" from Wilson PWF, et al *Prediction of coronary heart disease using risk factor categories. Circulation* 1998;97: (2001) 1837-47.
- [23] M. F. Abbod, D. G. von Keyserlingk, D. A. Linkens and M. Mahfouf, *Survey of Utilization of Fuzzy Technology in Medicine and Healthcare*, *Fuzzy Sets and Systems*, (2001).120, 331–349.
- [24] H. P. Nguyen and V. Kreinovich, "Fuzzy Logic and Its Applications in Medicine", *International Journal of Medical Informatics*, (2001) 62, 165–173.
- [25] C. Koutsojannis and I. Hatzilygeroudis, "Fuzzy-Evolutionary Synergism in an Intelligent Medical Diagnosis System" *Lecture Notes in Computer Science, Knowledge-Based Intelligent Information and Engineering Systems Springer Berlin / Heidelberg*, 0302-9743 (Print) 1611-3349 (Online) Volume 4252/(2006). pp. 1313-1322
- [26] S. Gottwald, *Universes of fuzzy sets and axiomatizations of fuzzy set theory. Part I: Model-based and axiomatic approaches. Studia Logica* (2006). 82: 211-144.
- [27] ATP III Online Calculator, <http://cvdrisk.nhlbi.nih.gov/calculator.asp> (last access May 2013)

[28] L. Zadeh, "The role of soft computing and fuzzy logic in conception, design and deployment of intelligent systems". Proceedings International Workshop on Soft Computing in Industry, Muroran, Japan, April (1996), pp 136-137.