Profiling using Fuzzy Set QCA for Medical Diagnosis

The Case of Anemia

Stavroula Barbounaki
Merchant Marine Academy of Aspropyrgos
Athens, Greece
e-mail: sbarbounaki@yahoo.gr

Nikos G. Dimitrioglou
National Technical University of Athens
Athens, Greece
e-mail: nikosdimitri@mail.ntua

Dimitris K. Kardaras
Athens University of Economics and Business
Athens, Greece
e-mail: Kardaras@aueb.gr; dkkardaras@yahoo.co.uk

Ilias Petrounias
The University of Manchester
Manchester, UK
e-mail: ilias.petrounias@manchester.ac.uk

***Abstract*— Despite the fact that anemia is a common disease, its diagnosis can be elusive. The signs and symptoms of anemia are generally unreliable in predicting the degree of anemia. Its diagnosis is mainly based on information of patient history and results of diagnostic tests that measuring indicators correlated to the disease. This paper suggests an approach to anemia diagnosis for adults by utilizing the fuzzy set Qualitative Comparative Analysis (FsQCA), which is not been used previously in medical diagnosis. The FsQCA, as an extension of QCA, is using fuzzy sets and entails the analysis of necessary and sufficient conditions to produce the some outcome, such as morbidity and severity. This paper aims to produce a set of causal configurations that can be used to assess and diagnose a medical case. The data set is collected from medical data sources. The factors to be considered are physiological indicators, such as hemoglobin, ferritin, mean corpuscular volume and hemoglobin, as well as age, gender and comorbidity. The anemia diagnosis is the outcome set used in this study. The proposed approach is tested for its accuracy and validity.**

Keywords-medical diagnosis; intelligent systems; fuzzy logic; QCA; healthcare.

I. INTRODUCTION

Fuzzy sets were introduced by Zadeh [1] as a formalism that can represent and manipulate problems that are ill-structured due to the uncertainty that characterizes them. It was specifically designed to mathematically represent uncertainty and vagueness and to provide formalized tools for dealing with the imprecision intrinsic to many problems.

Fuzzy logic attempts, to 'simulate' the human mind in order to improve the cognitive modeling of a problem [2]. The notion of an infinite-valued logic was introduced in [1] seminal work "Fuzzy Sets" where he described the mathematics of fuzzy set theory. This theory proposes that membership functions [3] represent linguistic variables, i.e., fuzzy sets, by assigning a number over the range of real numbers [0,1]. If an element does not belong to the fuzzy set with certainty, it is then assigned with 0, which represents the false and if an element certainly belongs to the fuzzy set, it is then assigned to 1, which represents the true. Every element, which is a member of a fuzzy set, has some grade of membership. The function that associates a number to each element (x) of fuzzy set (A) of the universe (X)is called membership function $\mu(x)\_A$, $x \in X$. Fuzzy logic expands to a great number of applications from power utilities and glass processing to washing machines and videos to medical, financial, supply chain management, decision support systems in various business areas [2].

In the field of medicine, computerized diagnostic systems have been proposed to assist physicians. Due to the probabilistic nature of choosing a diagnosis, it is possible that a physician will select the most likely diagnosis instead of the correct one [4]. To overcome uncertainty, many computerized diagnostic systems have been developed for miscellaneous diseases, such as diabetes [5], cancer [6], and cardiovascular diseases [7]. The accuracy of some systems matches actually the diagnostic abilities of physicians [8]. In particular, fuzzy logic and hybrid systems can be used to assist physicians in this decision process [9]. This approach uses fuzzy logic, to represent uncertainty better in order to improve diagnosis accuracy in a clinical context. For instance, a neuro-fuzzy network has been proposed to determine anemia level of a child [10]. The results indicate that the predicted anemia level values are very close to the measured values. Furthermore, an adaptive neuro-fuzzy inference system was developed to diagnose the iron deficiency anemia [11].

This paper suggests an approach to anemia diagnosis for adults by utilizing the fuzzy set Qualitative Comparative Analysis (FsQCA), which is not been used previously in medical diagnosis. Anemia is a common blood disorder in which Red Blood Cell (RBC) mass is reduced or the concentration of hemoglobin in blood is very low, resulting in a decrease in the oxygen-carrying capacity of it. The definition is population-based and varies by gender and race [12]. Despite the fact that anemia is a common disease, its diagnosis can be elusive. The signs and symptoms of anemia are generally unreliable in predicting the degree of anemia. Its diagnosis is mainly based on information of patient history and results of diagnostic tests that measuring indicators correlated to the disease. The most common diagnostic tests are Complete Blood Count (CBC) and Peripheral Blood Smear (PBS) assessment. However, the exact range of normal values for a given anemia factor is not well-defined among the medical community. Hence, values found that are near the upper or lower normal limits can be misinterpreted, possibly leading to false diagnosis. FsQCA, as an extension of QCA, is using fuzzy sets and entails the analysis of necessary and sufficient conditions to produce the some outcome, such as morbidity and severity.

II. METHODOLOGY

This paper proposes an approach to analyze patient medical data and diagnose a case of anemia as well as to determine the type of anemia. The proposed approach has been used and tested with a sample medical data set of 20 patient health records. This paper utilizes the FsQCA in order to produce causal combinations that best lead to reliable diagnosis. The FsQCA is particularly important for investigating intertwined relationships between multiple factors that affect a dependent variable. The dependent variable in this paper, i.e., the outcome set is the diagnosis regarding the existence of anemia [13]. The FsQCA analyses the sets of relationships among causes. In FsQCA variables are modelled as sets. FsQCA may produce alternative multiple paths, i.e., alternative causal combinations that can produce high consistency and coverage outcomes [14][15]. This analysis is performed separately for men and women due to different ranges of normal values for each of the variables used in diagnosis. Pregnants and vegans are excluded for our study, as their specific conditions have great impact on the level of indicators, such as hemoglobin.

The steps of the proposed approach follow:

- Step 1: Fuzzification of Medical Data.
- Step 2: Apply Qualitative Comparative Analysis.
  - Produce the truth table of all possible permutations of the terms considered.
  - Calculate membership degrees for each combination.
- Step 3: Calculate Consistency and Coverage of the solutions.
- Step 4: Suggest final diagnosis.

The following sections exemplify the steps of the proposed approach.

A. Fuzzification of Medical Data

The vector Exam-Data (XD) models every set of medical data that is available and represents a person whose case for anemia is considered. The Exam-Data vector is an (1×n) vector, which considers the same variables with the same order as in the first row of the Anemia-Matrix.

The values in the vector are the membership function values that result from the fuzzification process, which is performed for each one of the variables of the medical examinations before storing it in the vector. The membership functions of the Triangular Fuzzy Numbers (TFN) used in this research are the following in the case of hemoglobin:

$$\mu_{LOW} = \begin{cases} 1, & 0 \le x \le 30 \\ \frac{(34-x)}{4}, & 30 \le x \le 34 \\ 0, & 34 < x \end{cases} \quad (1)$$

$$\mu_{HIGH} = \begin{cases} 0, & 0 \le x \le 38 \\ \frac{(x-38)}{4}, & 38 \le x \le 42 \\ 1, & 42 < x \end{cases} \quad (2)$$

$$\mu_{NORMAL} = \begin{cases} \frac{(x-28)}{5}, & 28 \le x \le 33 \\ 1, & 33 \le x \le 40 \\ \frac{(-x-44)}{4}, & 40 < x < 44 \\ 0, & 44 < x \end{cases} \quad (3)$$

Similarly, membership functions are defined for all the variables.

B. Apply Qualitative Comparative Analysis.

Produce the truth table of all possible permutations of the terms considered. Each permutation is a possible causal combination. Calculate membership degrees for each combination. Its calculation is performed drawing on the fuzzy sets operations theory. Assume two fuzzy sets $\tilde{A}$ and $\tilde{B}$ then:

The fuzzy union, is defined as

$$\mu_{(A \cup B)} = \mathrm{MAX}(\mu_A, \mu_B) \quad (4).$$

The fuzzy intersection is defined as

$$\mu_{(A \cap B)} = \mathrm{MIN}(\mu_A, \mu_B) \quad (5)$$

and the fuzzy complement is calculated as

$$\mu_{-A} 1 - \mu_A \quad (6).$$

All variable expect hemoglobin, are fuzzified using the TFN formula (7).

$$f_a(x) \begin{cases} \frac{(x-a)}{(m-a)}, & a \le x \le m, m \ne a \\ \frac{(x-b)}{(m-b)}, & m \le x \le b, m \ne b \\ 0, & othrwise \end{cases} \quad (7)$$

where a, m, b are real numbers. During fuzzification calculations, the appropriate linguistic scale is identified and then the corresponding mean of the fuzzy number is considered to represent the level of each variable.

TABLE I. LINGUISTIC SCALES AND CORRESPONDING TFNS FOR AGE, MCV, MCH, FERRITIN, COMORBIDITY TESTS USED IN THIS STUDY

Linguistic scale	Triangular fuzzy scale	Mean of fuzzy numbers
Very High	(0.75, 1.00, 1.00)	1.00
High	(0.50, 0.75, 1.00)	0.75
Medium	(0.25, 0.50, 0.75)	0.50
Low	(0.00, 0.25, 0.50)	0.25
Very Low	(0.00, 0.00, 0.25)	0.00

C. Calculate Consistency and Coverage of the solutions

Calculate the consistency and the coverage of the solutions using formulas (8) and (9), respectively:

$$Consistency\,(X \leq Y) = \frac{\sum \mathrm{MIN}(X,Y)}{\sum x} \tag{8}$$

$$Coverage = \frac{\sum \mathrm{MIN}(X,Y)}{\sum x} \tag{9}$$

where *(X)* is the membership degree of each causal combination and *(Y)* is the membership degree of the outcome set.

D. Simplify Causal Combinations and Suggest final diagnosis

Identify best combinations in terms of consistency and coverage. Thus, this study considers a consistently rate above a threshold that is set at 0.8 and the highest possible coverage. The produced causal combinations that satisfy consistency and coverage thresholds are then simplified into the final set of causal combinations, which can be used for anemia diagnosis.

III. DATA ANALYSIS AND RESULTS

This paper analyses the data set shown in Table 2. This data refers to the women data set, showing the values for each medical test as they are produced after fuzzification. For hemoglobin, fuzzification is per-formed by applying formulas (1)-(3). For the rest of the variables fuzzification id used by taking the mean value of the appropriate fuzzy set as shown in Table 1. Similar calculations are performed for the men data set.

TABLE II. THE WOMEN PATIENT DATA SET

Patient	Age	Hg	MCV	MCH	Fr	Como/ty	Outcome
P1	0.75	0.75	0.5	0.5	0.5	1	0
P2	0.5	0.5	0	0	0.25	0	1
P3	1	0.5	0.5	0.5	0.75	1	1
P4	0.5	0.25	0.25	0.25	0	1	1
P5	0.5	0.5	0.75	0.75	0.25	1	1
P6	0.5	0.25	0.5	0.25	0	1	1
P7	0.5	0.75	0.25	0	0.25	1	1
P8	0	0.75	0.25	0.25	0.25	0	1
P9	1	0.75	0.5	0.5	0.5	1	0
P10	0.75	0.75	0.5	0.5	0.5	1	0
P11	0.25	0.5	0.5	0.5	0.75	0	1
P12	0.25	0	0	0	0	1	1
P13	0.5	0.5	0.5	0.5	0.5	1	1
P14	0.75	0.75	1	1	0.25	1	1
P15	0.25	0.25	0.5	0.5	0	1	1
P16	0.75	0.25	0.5	0.5	1	1	1
P17	0.25	0.5	0.5	0.5	0.75	0	1
P18	0.25	0	0	0	0	1	1
P19	0.5	0.5	0.5	0.5	0.5	1	1

The outcome set indicates whether the patient is diagnosed with anemia or not, coded with 1 and 0 respectively. Table 3, shows part of the total of 2^5=64 possible truth table permutations.

TABLE III. THE TRUTH TABLE (PART OF) SHOW ALL POSSIBLE PERMUTATIONS OF THE TERMS

Causal Permutation	Age	Hg	MCV	MCH	Fr	Comorbidity
1	0	0	0	0	0	0
2	0	0	0	0	0	1
3	0	0	0	0	1	0
4	0	0	0	0	1	1
5	0	0	0	1	0	0
6	0	0	0	1	0	1
7	0	0	0	1	1	0
8	0	0	0	1	1	1
9	0	0	1	0	0	0
10	0	0	1	0	0	1
11	0	0	1	0	1	0
12	0	0	1	0	1	1
13	0	0	1	1	0	0

 ISBN: 978-1-61208-681-1

Causal Permutation	Age	Hg	MCV	MCH	Fr	Comorbidity
14	0	0	1	1	0	1
15	0	0	1	1	1	0

TABLE IV. CAUSAL COMBINATIONS' CONSISTENCY AND COVERAGE FOR SEVEN PATIENTS

Causal Permutation	P1	P2	P3	P4	P5	P7	P8
1	0.25	0	0	0.5	0.25	0.5	0.25
2	0.25	0	0	0.5	0.25	0.5	0.25
3	0.25	0	0	0.5	0.25	0.5	0.25
4	0.25	0	0	0.5	0.25	0.5	0.25
5	0.25	0	0	0.5	0.25	0.5	0.25
4	0.25	0	0	0.5	0.25	0.5	0.25
5	0.25	0	0	0.5	0.25	0.5	0.25
6	0.25	0	0	0.5	0.25	0.5	0.25
7	0.25	0	0	0.5	0.25	0.5	0.25
8	0.25	0	0	0.5	0.25	0.5	0.25
9	0.25	0	0	0.5	0.25	0.5	0.25
…							
33	0.25	0	0	0.5	0.25	0.5	0.25
34	0.25	0	0	0.5	0.25	0.5	0.25
35	0.25	0	0	0.5	0.25	0.5	0.25
36	0.25	0	0	0.5	0.25	0.5	0.25
37	0.25	0	0	0.5	0.25	0.5	0.25
38	0.25	0	0	0.5	0.25	0.5	0.25
39	0.25	0	0	0.5	0.25	0.5	0.25
40	0.25	0	0	0.5	0.25	0.5	0.25
41	0.25	0	0	0.5	0.25	0.5	0.25
42	0.25	0	0	0.5	0.25	0.5	0.25
43	0.25	0	0	0.5	0.25	0.5	0.25
44	0.25	0	0	0.5	0.25	0.5	0.25
45	0.25	0	0	0.5	0.25	0.5	0.25
46	0.25	0	0	0.5	0.25	0.5	0.25
47	0.25	0	0	0.5	0.25	0.5	0.25
48	0.25	0	0	0.5	0.25	0.5	0.25
49	0.25	0	0	0.5	0.25	0.5	0.25
50	0.25	0	0	0.5	0.25	0.5	0.25
51	0.25	0	0	0.5	0.25	0.5	0.25
52	0.25	0	0	0.5	0.25	0.5	0.25
53	0.25	0	0	0.5	0.25	0.5	0.25
…							
60	0.25	0	0	0.5	0.25	0.5	0.25
61	0.25	0	0	0.5	0.25	0.5	0.25
62	0.25	0	0	0.5	0.25	0.5	0.25
63	0.25	0	0	0.5	0.25	0.5	0.25
64	0.25	0	0	0.5	0.25	0.5	0.25

The cells in the truth table take the value (1) or (0) representing true or false. Thus, permutation number 7 is read (Age=false, AND Hemoglobin=false, AND MCV=false, AND MCH=true, AND Ferritin=true, AND Comorbidity=false). Next, the membership degrees for all combinations for each patient are calculated drawing on the fuzzy sets operations theory, i.e., formulas (4,5,6). Table 4 shows the membership degrees for only a selected set of combinations, for simplicity.

The membership degree of combination number 7, for patient 1, see cell (Causal permutation1, P1) in Table 4, is calculated as follows by using formulas (5) and (6): Consider combination number 7 membership degree = (Age=false Hemoglobin =false MCV=false MCH=true Ferritin=true Comorbidity =false) = (not (Age), not (Hemoglobin), not (MCV), MCH, Ferritin, not (Comorbidity)). Thus, drawing on data shown in Table 2 for patient P1, the (Age=false) = ((1- (Age)) = (1-0.75) =0.25. Similar calculations are performed for all terms. Thus, for P1, =min (0.7; 0.25; 0.5; 0.5; 1; 1)=0.25. After all membership degrees are calculated, for all causal combinations for all patients, the consistency and coverage degrees are determined. Table 5 shows the results for a selected set of causal combinations.

TABLE V. CAUSAL COMBINATIONS' CONSISTENCY AND COVERAGE

Causal Permutation	Consistency	Coverage
1	1.000	0.077
2	0.929	0.250
3	1.000	0.077
4	0.833	0.096
5	1.000	0.038
…	…	…
33	1.000	0.058
34	0.846	0.212
35	1.000	0.038
36	0.800	0.154
37	1.000	0.019
38	0.778	0.135

 ISBN: 978-1-61208-681-1

Causal Permutation	Consistency	Coverage
39	1.000	0.019
40	0.778	0.135
41	1.000	0.019
42	0.900	0.173
43	1.000	0.019
44	0.889	0.154
45	1.000	0.019
46	0.900	0.173
47	1.000	0.019
48	0.889	0.154
…		
60	0.778	0.135
61	1.000	0.019
62	0.846	0.212
63	1.000	0.019
64	0.778	0.135

The consistency for combination number 2 is calculated, by applying formula (8) as follows:

Consider the outcome column shown in Table 2. Also consider the variables' membership degrees of combination number 2, for all patients as shown in Table 4.

Then,

min{min(0.25;0)+min(0.25;1)+min(0.25;0)+…+min(0.25;1)= min(0.25+0.25+0.25+…+0.25)=3.25 (0.25+0+0+0.5+0.25+…+0)=3.5.

Therefore, the consistency for combination number 2= =0.928.

Regarding the coverage, by applying formula (9), 3.5 and 14 thus coverage=0.25. The FsQCA theory assumes that the best causal combinations exhibit as high as possible consistency and coverage. However, the higher the consistency is, the lower the coverage becomes. By selecting the causal combinations with a consistency higher than the threshold value of 0.8, the final set is defined by selecting those combinations that exhibit the higher possible coverage as well. The selected set of causal combinations (i.e., combination 2, 34, 42, 46 and 62) are highlighted in Table 5. Table 6 shows the selected combinations.

TABLE VI. THE TWO NECESSARY AND SUFFICIENT CAUSAL COMBINATIONS

Causal Permutation	Age	Hg	MCV	MCH	Fr	Comorbidity
2	0	0	0	0	0	1
34	1	0	0	0	0	1
42	1	0	1	0	0	1
46	1	0	1	1	0	1
62	1	1	1	1	0	1

The closer look of Table 6, may lead to the elimination of the Ferritin test from the causal combinations since it seems it does not affect the results. Thus, by restructuring the causal combinations, the results show that the following alternative paths lead to anemia diagnosis.

- ✓ Comorbidity OR
- ✓ Age AND Comorbidity OR
- ✓ Age AND MCV AND Comorbidity OR
- ✓ Age AND MCV AND MCH AND Comorbidity OR
- ✓ Age AND Hemoglobin AND MCV AND MCH AND Comorbidity.

Furthermore, Hemoglobin can also be omitted, since Hemoglobin appears in only one combination (number 62) and it is the only additional factor to combination number 46. Thus, the set of causal combinations is as follows:

- ✓ Comorbidity OR
- ✓ Age AND Comorbidity OR
- ✓ Age AND MCV AND Comorbidity OR
- ✓ Age AND MCV AND MCH AND Comorbidity OR

IV. CONCLUSION AND FUTURE RESEARCH

This paper suggests the use of FsQCA in medical diagnosis and supports its use with the development of the necessary models and a prototype that was put to practice with real world test data. Data is collected from a hospital in Athens, Greece. The data protection act is satisfied in the sense that no personal data is held in the database. It only contains examinations results and diagnoses. FsQCA are used as a decision support method in order to assist physicians with manipulating and interpreting borderline cases such as the handling of the upper or lower normal limits of medical data that may raise ambiguity in problem formulation thus leading to faulty diagnoses. For future research, this paper suggests the testing of the model with more data. Large data sets will assist in evaluating the approach. Further, the analysis of large data sets may increase the complexity of applying the FsQCA method, thus FsQCA may be applied at a second stage after pruning an initial set of causal combinations. In addition, the method can be applied in other areas of medical problems.

REFERENCES

[1] L. A. Zadeh, "Fuzzy sets", Information and control, vol. 8, no. 3, pp. 338-353, 1965.

[2] P. Craiger and M.D. Coovert, "Modeling dynamic social and psychological processes with fuzzy cognitive maps" In Fuzzy Systems, IEEE World Congress on

Computational Intelligence. Proceedings of the Third IEEE Conference, pp. 1873-1877, 1994.
[3] H.J. Zimmermann (1991). “Fuzzy Set Theory and its Applications”, Kluwer Academic Publishers.
[4] M. Graber, R. Gordon, and N. Franklin, Reducing diagnostic errors in medicine: what's the goal? Academic Medicine, vol 77, no. 10, pp. 981-992, 2002.
[5] K. Polat and S. Güneş, “An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease”, Digital Signal Processing, vol. 17, no. 4, pp. 702-710, 2007.
[6] A. Keleş, A. Keleş, and U. Yavuz, “Expert system based on neuro-fuzzy rules for diagnosis breast cancer”, Expert systems with applications, vol. 38, no. 5, 5719-5726, 2011.
[7] S. Muthukaruppan and M.J. Er., “A hybrid particle swarm optimization based fuzzy expert system for the diagnosis of coronary artery disease”, Expert Systems with Applications, vol. 39, no. 14, pp. 11657-11665, 2012.
[8] V. Pabbi, “Fuzzy Expert System for medical diagnosis”, Internatinal Journal of Scientific and Research Publications, vol. 5 no. I, pp. 1-3, 2015.
[9] N. Allahverdi, “Design of fuzzy expert systems and its applications in some medical areas” International Journal of Applied Mathematics, Electronics and Computers, vol. 2, no. 1, pp. 1-8, 2014
[10] N. Allahverdi, A. Tunali, H. Işik, H. Kahramanli, “A Takagi–Sugeno type neuro-fuzzy network for determining child anemia”, Expert Systems with Applications, vol. 38, no. 6, pp. 7415-7418, 2011.
[11] L. Azarkhish, M.R. Raoufy and S. Gharibzadeh, “Artificial intelligence models for predicting iron deficiency anemia and iron serum level based on accessible laboratory data”, Journal of medical systems, vol. 36, no. 3, pp. 2057-2061, 2012.
[12] M.J. Cascio and T.G. DeLoughery, “Anemia” Medical Clinics, vol. 101, no. 2, pp. 263-284, 2017.
[13] S. Chari, A. Tarkiainen, H. Salojärvi, “Alternative pathways to utilizing customer knowledge: A fuzzy-set qualitative comparative analysis”, J Bus Res, vol 69, pp. 5494–5499, 2016.
[14] B. Rihoux and C.C. Ragin, “Configurational comparative methods: Qualitative comparative analysis (QCA) and related techniques”, Sage Publications, 2008.
[15] D. Skarmeas, C.N. Leonidou, C. Saridakis “Examining the role of CSR skepticism using fuzzy-set qualitative comparative analysis”, J Bus Res, vol. 67, pp. 1796–180, 2014.