An application for medical decision making with the fuzzy soft sets

Murat Kirişci[a*]

[a]*Department of Mathematical Education, Hasan Ali Yücel Education Faculty, Istanbul University-Cerrahpaşa, Istanbul, Turkey*
[*]*Corresponding author E-mail: mkirisci@hotmail.com*

Article Info

Keywords: *Cleveland dataset, decision making, defuzzification, fuzzy soft set, triangular fuzzy number, trapezoidal fuzzy number*

2010 AMS: *Primary 03E75; Secondary 03E72, 68T37, 94D05*

Received: *X XXXXXX XXXX*

Accepted: *X XXXXXX XXXX*

Available online: *X XXXXXX XXXX*

Abstract

In the present study, for the medical decision making problem, the proposed technique related to the fuzzy soft set by Celik-Yamak through Sanchez's method was used. The real dataset which is called Cleveland heart disease dataset applied in this problem.

1. Introduction

Many terms that we use randomly in everyday life usually have a fuzzy structure. Verbal or numerical expressions we use, while describing something, explaining an event, commanding and in many other cases include fuzziness. People use terms that do not express certainty when explaining an event and deciding on a situation. According to the age of the person, old, middle, young, very old and very young are called. Depending on the slope and ramp condition of the road, the car's gas or brake pedal is press slightly slower or slightly faster. All these are examples of how the human brain behaves in uncertain and vagueness situations, and how it evaluates, identifies, and commands events.

After the fuzzy set theory, which uses fuzzy logic rules was developed by Lotfi A. Zadeh and published in its original 1965 paper [9], the examination of uncertainty systems has gained a new dimension. Fuzzy sets are characterized by membership functions. In fact, these membership functions are nothing more than fuzzy numbers(FNs). Such a set can be characterized by a membership function that assigns membership values to each of the elements from 0 to 1. Members that are not included in the set are assigned membership values of 0, and those who are included in the set are assigned membership values of 1. The elements that are not included in the set are assigned values between 0 and 1 according to the uncertainty situation.

The concept of soft set has been defined by Molodtsov. Molodtsov proposed the concept of the soft set(SS), a completely new approach to modeling uncertainty. The soft set theory(SST) has a rich application potential in many ways. Some of these applications have been shown by Molodtsov in his pioneering work. This theory was applied to many areas of uncertainty such as information systems, decision making problems, optimization theory, algebraic structures and mathematical analysis. Maji and et al.[4] investigated the SST for decision making problems. In [5], the authors defined operations "AND", "OR", union-intersection of two SSs and also gave some examples. The same authors established a hybrid model known as fuzzy soft set(FSS), which is a combination of SS and FS [3]. Actually, the concept of FSS is an extension of crisp SS.

In this study, we will consider the approach of Çelik-Yamak [1], obtained from the Sanchez's model [8] and we will use real dataset, which is called Cleveland heart disease dataset [10], for the problem of medical decision-making.

2. Decision Making for Imprecision

The problems we face in our lives are often not clear and precise. So we use various decision-making mechanisms to solve our problems. With these mechanisms we use, we try to make the most right decision by reducing uncertainties. Therefore, improved mathematical tools for uncertainty and imprecision are needed. FS theory has been used quite extensively to deal with such imprecisions. In this section, we will give some definitions of the theories of SSs and FSSs.

Email addresses: **mkirisci@hotmail.com (Murat Kirişci)**

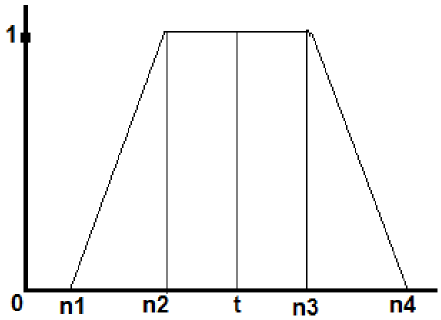

Figure 2.1: A trapezoidal FN

The FS has emerged as a generalization of the classical set concept. If we choose a non-empty set X, then a function $f_A(x) : X \to [0,1]$ is called FS on X and represented by

$$A = \{(x_i, f_A(x_i)) : f_A(x_i) \in [0,1]; \forall x_i \in X\}.$$

FS A on X can be expressed by, set of ordered pair as follows:

$$A = \{(x, f_A(x)) : x \in X\}.$$

The fuzzy subset of X is a non-empty subset $\{(x, f(x)) : x \in X\}$ for some function $f : X \to [0,1]$ [2]. Consider the function $f : \mathbb{R} \to [0,1]$ as a subset of a non-empty base space $\mathbb{R}$. Suppose that the function f satisfies the following properties:

i. f is a normal, that is, there exist $x_i \in \mathbb{R}$ such that $f(x_i) = 1$,

ii. f is a fuzzy convex, that is, $f[kx + (1-k)y] \geq \min\{f(x), f(y)\}$, where $x, y \in \mathbb{R}$ and $k \in [0,1]$,
iii. f is upper semi-continuous,
iv. The closure of the set C is compact, where the set $C = \{x \in \mathbb{R} : f(x) \geq 0\}$.

Therefore, the function f is called a *fuzzy number* [2].

A triangular FN $\widehat{m}$ is represented by a triplet (m_1, m_2, m_3) and its the membership function is

$$f(x) = \begin{cases} 0 & , \quad x < m_1 \\ \frac{x-m_1}{m_2-m_1} & , \quad m_1 \leq x \leq m_2 \\ \frac{x-m_3}{n_2-n_3} & , \quad n_2 \leq x \leq n_3 \\ 0 & , \quad x > n_3 \end{cases}$$

A trapezoidal FN $\widehat{n}$ is a piecewise function, parameterized by a quadruplet (n_1, n_2, n_3, n_4) (Figure 2.1). The membership function of a trapezoidal FN is given by

$$f(x) = \begin{cases} 0 & , \quad x < n_1 \\ \frac{x-n_1}{n_2-n_1} & , \quad n_1 \leq x \leq n_2 \\ 1 & , \quad n_2 \leq x \leq n_3 \\ \frac{x-n_4}{n_3-n_4} & , \quad n_3 \leq x \leq n_4 \\ 0 & , \quad x > n_4 \end{cases}$$

SS theory developed by Molodtsov [6] is a suitable tool for solving uncertainties in non parametric situations and is a natural generalization of FS theory. Since SS Theory is a natural generalization of FS theory, it has been applied in a wide range of fields ranging up to from mathematics to engineering from economics to optimization. To deal with a collection of approximate description of objects, a generalized parametric gizmo is used known as SS.

In approximate description, there are two value sets which are called predicate and approximate. Initially, the object description has an approximate by nature and so there is no require to present the concept of exact solution. The SS theory is very convenient and simply effective in performance due to the nonentity of any limitations on the approximate descriptions. With the aid of words and sentences, real sentences, real number, function, mapping and so on; any parameter can be operate that we desire.

Now, we will give definitions of Soft Set and Soft Subset.

Definition 2.1. *Consider X, E as initial universe and parameters sets. Take $\mathscr{P}(X)$ as a power set of X. Let $R \subset E$. Give the mapping $\mathscr{F} : R \to \mathscr{P}(X)$. Then, the pair $(\mathscr{F}, R)$ is called a soft set(SS) on X.*

This definition can be also expressed as:

A SS on X is a parameterized family of subsets of the universe X. $\mathscr{F}(\alpha)$ may be considered as the set of $\alpha-$aprroximate elements of the SS $(\mathscr{F},R)$, for $\alpha \in R$.

Definition 2.2. *Let $(\mathscr{F},R)$ and $(\mathscr{G},S)$ be two SSs on X. If the following conditions are hold, then it is said that $(\mathscr{F},R)$ is a soft subset(SSS) of $(\mathscr{G},S)$:*

i. $R \subset S$,
ii. *$\mathscr{F}(\alpha)$ and $\mathscr{G}(\alpha)$ are identical approximations, for all $\alpha \in R$.*

We will denote the SSS by $(\mathscr{F},R)\hat{\subset}(\mathscr{G},S)$.

The AND and OR operator of SSs are defined as:

Definition 2.3. *Let $(\mathscr{F},R)$ and $(\mathscr{G},S)$ be two SSs. Define $\mathscr{H}(\lambda,\mu)=\mathscr{F}(\lambda)\cap\mathscr{G}(\mu)$, $((\lambda,\mu)\in R\times S)$. Then, $(\mathscr{F},R)\wedge(\mathscr{G},S)=(\mathscr{H},R\times S)$ is called $(\mathscr{F},R)$ AND $(\mathscr{G},S)$.*

Define $\mathscr{K}(\lambda,\mu)=\mathscr{F}(\lambda)\cup\mathscr{G}(\mu)$, $((\lambda,\mu)\in R\times S)$. Then, $(\mathscr{F},R)\vee(\mathscr{G},S)=(\mathscr{K},R\times S)$ is called $(\mathscr{F},R)$ OR $(\mathscr{G},S)$.

Choose set of k objects and set of parameters as $X=\{a_1,a_2,\dots a_n\}$ and $\{R_1,R_2,\dots,R_i\}$, respectively. Let $E\supseteq\{R_1\cup R_2\cup\cdots,R_i\}$ and each parameter set R_i represent the ith class of parameters and the elements of R_i represents a specific property set. Assumed that the property sets can be shown as FSs.

Definition 2.4. *Let $\mathscr{P}(X)$ denotes the set of all FSs of X and $R_i \subset E$. Give $\mathscr{F}_i : R_i \to \mathscr{P}(X)$ be a mapping. Then, $(\mathscr{F}_i,R_i)$ is called a fuzzy soft set(FSS) over X.*

Definition 2.5. *Let $(\mathscr{F},R)$ and $(\mathscr{G},S)$ be two FSSs on X. If the following conditions are hold, then it is said that $(\mathscr{F},R)$ is a fuzzy soft subset(FSSS) of $(\mathscr{G},S)$:*

i. $R \subset S$,
ii. *$\mathscr{F}(\alpha)$ is a FSS of $\mathscr{G}(\alpha)$, for all $\alpha \in R$.*

We write $(\mathscr{F},R)\hat{\subset}(\mathscr{G},S)$.

3. Method

3.1. FSS Method

In this study, we will use a technique developed with FSS theory [1]. We build a patient-attributes$(m\times n)$ matrix P_A using the FSS. The entries of this matrix are FNs $\hat{s}$ parameterized by a triplet $(s-1,s,s+1)$. Later on, the attributes-disease degrees$(n\times k)$ matrix A_D which is also called weighted matrix is built. In this matrix, each entry represents the weight of the attributes for a disease degree. Generally, these matrices are given as below:

$$P_A=\begin{bmatrix}\widehat{u_{11}} & \widehat{u_{12}} & \widehat{u_{13}} & \cdots & \widehat{u_{1n}}\\ \widehat{u_{21}} & \widehat{u_{22}} & \widehat{u_{23}} & \cdots & \widehat{u_{2n}}\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ \widehat{u_{m1}} & \widehat{u_{m2}} & \widehat{u_{m3}} & \cdots & \widehat{u_{mn}}\end{bmatrix} \quad \text{and} \quad A_D=\begin{bmatrix}\widehat{v_{11}} & \widehat{v_{12}} & \widehat{v_{13}} & \cdots & \widehat{v_{1k}}\\ \widehat{v_{21}} & \widehat{v_{22}} & \widehat{v_{23}} & \cdots & \widehat{v_{2k}}\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ \widehat{v_{n1}} & \widehat{v_{n2}} & \widehat{v_{n3}} & \cdots & \widehat{v_{nk}}.\end{bmatrix}$$

The patient-disease degrees matrix$(m\times k)$ P_D can be obtained from transformation operation $P_A\otimes A_D$. The general form of the matrix P_D as follows:

$$P_D=\begin{bmatrix}\widehat{w_{11}} & \widehat{w_{12}} & \widehat{w_{13}} & \cdots & \widehat{w_{1k}}\\ \widehat{w_{21}} & \widehat{w_{22}} & \widehat{w_{23}} & \cdots & \widehat{w_{2k}}\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ \widehat{w_{m1}} & \widehat{w_{m2}} & \widehat{w_{m3}} & \cdots & \widehat{w_{mk}}\end{bmatrix}$$

where $w_{ij}=(A,B,C)$ such that $A=\sum_{k=1}^{n}(a_{ik}-1)(b_{kj}-1)$, $B=\sum_{k=1}^{n}a_{ik}b_{kj}$ and $C=\sum_{k=1}^{n}(a_{ik}+1)(b_{kj}+1)$.
We can calculate the defuzzification of t of triangular FN (m_1,m_2,m_3) as

$$t=\frac{m_1+m_2+m_2+m_3}{4} \tag{3.1}$$

[1]. Therefore, each fuzzifying value of the above matrix will be defuzzify. Thus, we will obtain crisp diagnosis matrix D_F. The matrix D_F as below:

Table 1: Attributes of Cleveland dataset

Attributes	Fullname
a_1	age in years
a_2	chest pain type
a_3	resting blood pressure (in mm Hg)
a_4	serum cholestoral in mg/dl
a_5	fasting blood sugar ¿120 mg/dl
a_6	resting electrocardiographic results
a_7	maximum heart rate achieved
a_8	ST depression induced by exercise relative to rest
a_9	the slope of the peak exercise ST segment
a_{10}	number of major vessels (0-3) colored by flourosopy
a_{11}	3:normal; 6:fixed defect; 7:reversable defect

$$D_F = \begin{bmatrix} z_{11} & z_{12} & z_{13} & \cdots & z_{1k} \\ z_{21} & z_{22} & z_{23} & \cdots & z_{2k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ z_{m1} & z_{m2} & z_{m3} & \cdots & z_{mk} \end{bmatrix}$$

It will be decided according to the crisp values of the matrix D_F. If $\max z_{ij} = z_{ik}$ $(0 \leq j \leq k)$, then the disease degree of patient p_i is k $k = 1,2,3,4$ or 0(absence).

3.2. Dataset

Input variables are taken from Cleveland dataset [10]. This data set contains 303 patients, 11 attributes and 5 outcomes.

This database contains 76 attributes. However, it is understood that 14 of these attributes can be used because of the analysis. The outcomes are given as degree of disease. It is integer valued from 0 (no presence) to 4. tests with the Cleveland database have intensively on simply attempting to distinguish presence (values 1,2,3,4) from absence (value 0).

Let's choose patient sets $U = \{p_1, p_2, p_{24}, p_{25}, p_{75}, p_{303}\}$ from the Cleveland dateset.

The attributes of Cleveland dataset is given in Table 1. The outcomes are disease degrees as 1,2,3,4 and 0(absence).

According to the membership function, the age attribute is taken into account as follows: $0-20(0.0-0.2)$, $21-40(0.3-0.5)$, $41-60(0.6-0.8)$; $61+(0.9-1.0)$. In this data set, the attributes trestbps, chol, thalach, oldpeak are measured as lowest 94.0 highest 200.0; lowest 126.0 highest 564.0; lowest 71 highest 195; lowest 0.0 highest 5.6, respectively. We will give values between 0.0 and 1.0 to these measurements.

3.3. Algorithm

i. Input the SS $(\mathscr{F}, R)$ to get the patient-attribute matrix,
ii. Input the SS $(\mathscr{G}, S)$ to get the attribute-disease degree matrix,
iii. Compute the transformation operation $P_A \otimes A_D$ to make the patient-disease degree matrix,
iv. Defuzzify all the elements of the patient-disease degree matrix by Equation 3.1 and to obtain the defuzzify matrix,
v. Find k for which $z_{ik} = \max z_{ij}$.

4. Application

In this section, using the Celik-Yamak approach obtained by Sanchez's method, we give an application of FSS theory[7] for medical decision making. For this application, it will be used the Cleveland Dataset. This data set contains 303 patients.

Now, we construct the matrices for medical decision making by algorithm in previous section:

The patient-attributes matrix P_A is given by

$$P_A = \begin{bmatrix} \widehat{9} & \widehat{2} & \widehat{5} & \widehat{3} & \widehat{7} & \widehat{8} & \widehat{6} & \widehat{2} & \widehat{9} & \widehat{1} & \widehat{5} \\ \widehat{9} & \widehat{7} & \widehat{5} & \widehat{3} & \widehat{1} & \widehat{8} & \widehat{4} & \widehat{3} & \widehat{5} & \widehat{8} & \widehat{1} \\ \widehat{7} & \widehat{6} & \widehat{4} & \widehat{3} & \widehat{1} & \widehat{9} & \widehat{5} & \widehat{6} & \widehat{1} & \widehat{6} & \widehat{8} \\ \widehat{8} & \widehat{7} & \widehat{4} & \widehat{4} & \widehat{2} & \widehat{9} & \widehat{4} & \widehat{5} & \widehat{5} & \widehat{6} & \widehat{9} \\ \widehat{5} & \widehat{8} & \widehat{3} & \widehat{4} & \widehat{2} & \widehat{5} & \widehat{8} & \widehat{1} & \widehat{7} & \widehat{3} & \widehat{2} \\ \widehat{4} & \widehat{6} & \widehat{4} & \widehat{3} & \widehat{1} & \widehat{2} & \widehat{7} & \widehat{1} & \widehat{2} & \widehat{5} & \widehat{3} \end{bmatrix}.$$

In this matrix, rows are shown in patients $\{p_1, p_2, p_{24}, p_{25}, p_{75}, p_{303}\}$ and columns are shown attributes $\{a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9, a_{10}, a_{11}\}$. For example, $\mathscr{F}(p_1) = \{a_1/\widehat{9}, a_2/\widehat{2}, a_3/\widehat{5}, a_4/\widehat{3}, a_5/\widehat{7}, a_6/\widehat{8}, a_7/\widehat{6}, a_8/\widehat{2}, a_9/\widehat{9}, a_{10}/\widehat{1}, a_{11}/\widehat{5}\}$.

The attributes-disease degrees matrix A_D consist of the rows and columns which are attributes $\{a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9, a_{10}, a_{11}\}$ and disease degrees $1 = d_1, 2 = d_2, 3 = d_3, 4 = d_4, 0 = d_0$, respectively. For example, $\mathscr{G}(a_1) = \{d_1/\widehat{4}, d_2/\widehat{3}, d_3/\widehat{1}, d_4/\widehat{2}, d_5/\widehat{9}\}$.

$$A_D = \begin{bmatrix} \widehat{4} & \widehat{3} & \widehat{1} & \widehat{2} & \widehat{9} \\ \widehat{8} & \widehat{9} & \widehat{7} & \widehat{5} & \widehat{1} \\ \widehat{7} & \widehat{8} & \widehat{9} & \widehat{8} & \widehat{1} \\ \widehat{3} & \widehat{5} & \widehat{8} & \widehat{9} & \widehat{2} \\ \widehat{9} & \widehat{8} & \widehat{4} & \widehat{2} & \widehat{5} \\ \widehat{3} & \widehat{2} & \widehat{2} & \widehat{1} & \widehat{9} \\ \widehat{3} & \widehat{1} & \widehat{4} & \widehat{9} & \widehat{2} \\ \widehat{9} & \widehat{7} & \widehat{8} & \widehat{1} & \widehat{5} \\ \widehat{8} & \widehat{9} & \widehat{7} & \widehat{1} & \widehat{8} \\ \widehat{5} & \widehat{8} & \widehat{3} & \widehat{2} & \widehat{9} \\ \widehat{8} & \widehat{4} & \widehat{9} & \widehat{5} & \widehat{2} \end{bmatrix}$$

The patients-disease degrees matrix P_D is obtained from the matrices P_A and A_D. Therefore, the rows and columns of the matrix P_D are shown in patients and disease degrees, respectively.

$$P_D = \begin{bmatrix} \widehat{336} & \widehat{301} & \widehat{287} & \widehat{217} & \widehat{314} \\ \widehat{296} & \widehat{307} & \widehat{255} & \widehat{231} & \widehat{313} \\ \widehat{320} & \widehat{284} & \widehat{296} & \widehat{218} & \widehat{283} \\ \widehat{372} & \widehat{341} & \widehat{341} & \widehat{235} & \widehat{327} \\ \widehat{270} & \widehat{267} & \widehat{254} & \widehat{215} & \widehat{227} \\ \widehat{211} & \widehat{209} & \widehat{206} & \widehat{192} & \widehat{161} \end{bmatrix}$$

where

$$\widehat{336} = (223, 336, 454), \quad \widehat{301} = (191, 301, 433), \quad \widehat{287} = (179, 287, 417), \quad \widehat{217} = (122, 217, 322), \quad \widehat{314} = (215, 314, 432),$$

$$\widehat{296} = (186, 296, 428), \quad \widehat{307} = (200, 307, 436), \quad \widehat{255} = (150, 255, 382), \quad \widehat{231} = (107, 231, 305), \quad \widehat{313} = (267, 313, 431),$$

$$\widehat{320} = (208, 320, 454), \quad \widehat{284} = (175, 284, 415), \quad \widehat{296} = (196, 296, 425), \quad \widehat{218} = (128, 218, 330), \quad \widehat{283} = (185, 283, 403),$$

$$\widehat{372} = (253, 372, 513), \quad \widehat{341} = (225, 341, 479), \quad \widehat{341} = (228, 341, 477), \quad \widehat{235} = (138, 235, 354), \quad \widehat{327} = (222, 327, 454),$$

$$\widehat{270} = (166, 270, 396), \quad \widehat{267} = (232, 267, 390), \quad \widehat{254} = (164, 254, 375), \quad \widehat{215} = (133, 215, 319), \quad \widehat{227} = (137, 227, 339),$$

$$\widehat{211} = (124, 211, 318), \quad \widehat{209} = (121, 209, 317), \quad \widehat{206} = (125, 206, 307), \quad \widehat{192} = (124, 192, 280), \quad \widehat{161} = (82, 161, 256).$$

The defuzzifying matrix D_F constructed from equation (1):

$$D_F = \begin{bmatrix} 674,5 & 306,5 & 292,5 & 219,5 & 318,75 \\ 301,5 & 312,5 & 260,5 & 218,5 & 331 \\ 325,5 & 289,5 & 323,25 & 223,5 & 288,5 \\ 377,5 & 346,5 & 316,75 & 220,5 & 332,5 \\ 275,5 & 289 & 261,75 & 240,5 & 232,5 \\ 216 & 214 & 211 & 197 & 165 \end{bmatrix}$$

Decision: It seen that in the matrix D_F, heart disease degree of the patient p_1 is 1. That is, the maximum score is measured in the first person. Additionally, when the available data are calculated with the FSS, the highest patient with 2 disease degrees is p_2, the highest patient with 3 disease degrees is p_{24}, the highest patient with 4 disease degrees is p_{75} and the highest measured person with no heart disease is p_{25}.

5. Conclusion

In this paper, an application of FSS theory developed by Sanchez and implemented by Celik-Yaman by fuzzy arithmetic operations has been studied. In this application, the real dataset which called Cleveland dataset is used.

The relation matrix and weighted matrix is formed according to the given algorithm and the patient diagnostic matrix was obtained by transformation operation. The fuzz,fying values of the patient diagnostic matrix were defuzzifying and obtained the crisp diagnosis matrix. This study shows the contribution of FSSs in decision making problems and especially in medical diagnosis.

References

[1] Y. Çelik, S. Yamak, *Fuzzy soft set theory applied to medical diagnosis using fuzzy arithmetic operations*, Journal of Inequalities and Applications 2013, **2013**:82. DOI:10.1186/1029-242X-2013-82

[2] M. Kirişci, *Integrated and Differentiated Spaces of Triangular Fuzzy Numbers*, Fas. Math. 2017; **59**:75–89. DOI:10.1515/fascmath-2017-0018

[3] P.K. Maji, R. Bismas, A.R. Roy, *Fuzzy soft sets*, J.Fuzzy Math. 2001; **9**(3):677–692.

[4] P.K. Maji, A.R. Roy, R. Biswas *An application of soft sets in a decision making problem*, Comput. Math. Appl. 2002; **44**: 1077–1083

[5] P.K. Maji, R. Bismas, A.R. Roy, *Soft set theory*, Comput. Math. Appl. 2003; **45**:555–562.

[6] D. Molodtsov, *Soft Set Theory-First Results*, Comput. Math. Appl. 1999; **37**:19–31.

[7] A.R. Roy, P.K. Maji, *A fuzzy soft set theoretic approach to decision making problems*, Journal of Computational and Applied Mathematics. 2007; **203**:412–418.

[8] E. Sanchez, *Inverse of fuzzy relations, application to possibility distributions and medical diagnosis*, Fuzzy Sets Syst. 1979; **2**(1):75–86.

[9] L. A. Zadeh, *Fuzzy sets*, Inf. Comp. **8**, (1965), 338–353.

[10] UC Irvine Machine Learning Repository, *Cleveland heart disease data details*[Online], (2010), https://archive.ics.uci.edu/ml/datasets/heart+Disease [Accessed 9th September 2018].