Interval Type II Fuzzy Rough Set Rule Based Expert System to Diagnose Chronic Kidney Disease

Mona Abdolkarimzadeh[1], M. H. Fazel Zarandi[1(✉)], and O. Castillo[2]

[1] Department of Industrial Engineering,
Amirkabir University of Technology, Tehran, Iran
{mabdolkarimzadeh, zarandi}@aut.ac.ir
[2] Tijuana Institute Technology, Tijuana, Mexico
ocastillo@tectijuana.mx

Abstract. Chronic kidney disease is a worldwide public health problem with an increasing incidence and prevalence, poor outcomes, and high cost. Diagnosis of Chronic Kidney Disease has always been a challenge for physicians. This paper presents an effective method for diagnosis of Chronic Kidney Disease based on interval Type-II fuzzy. This proposed system includes three steps: pre-processing (feature selection), Type-II fuzzy classification, and system evaluation. Fuzzy Rough QuickReduct algorithm feature selection is used as the preprocessing step in order to exclude irrelevant features and to improve classification performance and efficiency in generating the classification model. Rough set theory is a very useful tool for describing and modeling vagueness in ill-defined environments. In the type-II fuzzy classification step, an "indirect approach" is used for II fuzzy system modeling by implementing the Sugeno index for determining the number of rules in the fuzzy clustering approach. In the proposed system, the process of diagnosis faces vagueness and uncertainty in the final decision. The results that were obtained show that interval Type-II fuzzy has the ability to diagnose Chronic Kidney Disease with an average accuracy of 90%.

Keywords: Chronic kidney disease · Interval type-II fuzzy · Rough set
Diagnosis · Feature selection

1 Introduction

1.1 Chronic Kidney Disease

Chronic kidney disease includes conditions that damage your kidneys and decrease their ability to keep you healthy by doing the jobs listed. If kidney disease gets worse, wastes can build to high levels in your blood and make you feel sick. Also, kidney disease increases your risk of having heart and blood vessel disease. These problems may happen slowly over a long period of time. When kidney disease progresses, it may eventually lead to kidney failure, which requires dialysis or a kidney transplant to maintain life. The number of persons with kidney failure who are treated with dialysis

G. A. Barreto and R. Coelho (Eds.): NAFIPS 2018, CCIS 831, pp. 559–568, 2018.
https://doi.org/10.1007/978-3-319-95312-0_49

and transplantation is projected to increase from 340 000 in 1999 to 651 000 [1]. Unfortunately, chronic kidney disease is underdiagnosed and undertreated, resulting in lost opportunities for prevention [2, 3] in part because of a lack of agreement on a definition and classification of stages in the progression of chronic kidney disease [4] and a lack of uniform application of simple tests for detection and evaluation. Chronic kidney disease affects approximately 11% of the U.S. adult population (20 million people from 1988 to 1994). The prevalence of earlier stages of disease (10.8%) is more than 100 times greater than the prevalence of kidney failure (0.1%). Adverse outcomes of chronic kidney disease, including loss of kidney function and development of kidney failure can often be prevented or delayed through early detection and treatment.

1.2 Fuzzy Logic System

The theory of Fuzzy logic was introduced by Prof. Zadeh. In this theory an element belongs to a set according to the membership function values. Theory of FSs is an expansion of the traditional sets theory in which an element either is or is not a set member [5]. The fuzzy logic systems (FLSs) are well known for their ability to model linguistics and system uncertainties. Due to this ability, FLSs have been successfully used for many real world applications, including modeling and controlling [6–8].

1.3 Interval Type-II Fuzzy

Type II fuzzy sets have grades of membership that are themselves fuzzy. A type II membership grade can be any subset in [0, 1]. When the secondary memberships are either zero or one, we call them interval type II sets [9]. As Type II fuzzy logic is better suited for modeling linguistic terms [10] in this study, we use the Type II FLS and introduce a type II fuzzy system for diagnosing Chronic Kidney disease. A type II fuzzy set denoted as $\widetilde{A}$, is characterized by a type-II membership function $\mu_{\widetilde{A}}(x,u)$:$U \times I \rightarrow I$ where $x \in U$ and $u \in J_x \subseteq [0,1]$ i.e.

$$\widetilde{A} = \{((x,u), \mu_{\widetilde{A}}(x,u)) | \forall x \in X, \forall u \in J_x \subseteq [0,1]\} \tag{1}$$

Where $0 \leq \mu_{\widetilde{A}}(x,u) \leq 1$. $\widetilde{A}$ can also be expressed as:

$$\widetilde{A} = \int_{x \in X} \int_{u \in J_x} \mu_{\widetilde{A}}(x,u) / (x,u), J_x \subseteq [0,1] \tag{2}$$

The upper membership function (UMF) and lower membership function (LMF) of $\widetilde{A}$ are two type 1 membership function that bound the FOU. The UMF of $\widetilde{A}$ is the upper bound of the FOU($\widetilde{A}$) and denoted $\overline{\mu_{\widetilde{A}}}(x)$ $\forall x \in X$, and the LMF is the lower bound of the FOU($\widetilde{A}$) and denoted $\underline{\mu_{\widetilde{A}}}(x)$ $\forall x \in X$.

$$\overline{\mu_{\tilde{A}}}(x) = \overline{FOU}(\tilde{A}), \forall x \in X \tag{3}$$

$$\underline{\mu_{\tilde{A}}}(x) = \underline{FOU}(\tilde{A}), \forall x \in X \tag{4}$$

Figure 1 shows the bounds of type-II membership function for Gaussian MF. A structure of a type-II fuzzy logic system shows in Fig. 2.

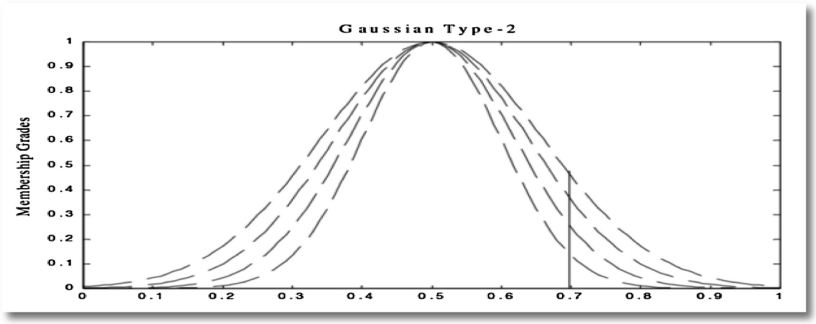

Fig. 1. The type-II membership function [10]

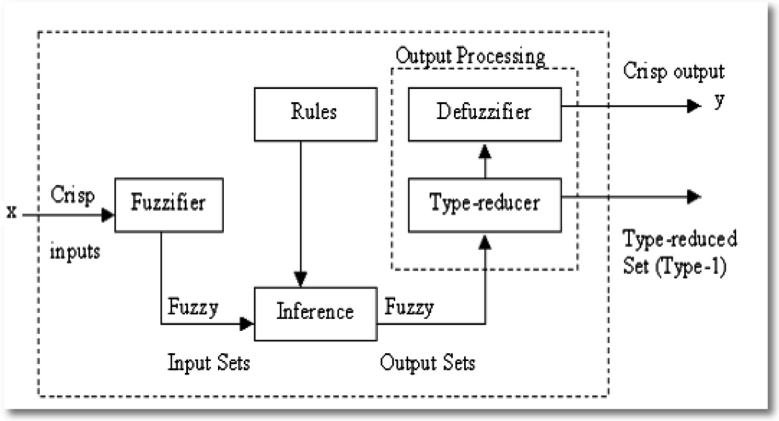

Fig. 2. A structure of a type-II fuzzy logic system [10]

Figure 2 shows the structure of an IT2 FLS. IT2 FLS contain the four mentioned major components (rules, fuzzifier, inference engine, and output processor) but the only difference between T1 and T2 structures is in the output processing part. In type-I

FLSs, output processing consists of a defuzzifier which transforms the fuzzy output of the system into a crisp value. But, output processing component in an IT2 FLS has two parts: Type reducer and defuzzifier. So before defuzzifying the output, it should be transformed from type-II to type-1. After type reduction, the output becomes a type-I FS and then we can implement various dufuzzification methods to obtain the crisp output [10]. Due to this ability, I2FLSs have been successfully used for many real world applications, including modeling and controlling [11–13].

1.4 Rough Set Theory

These two denominators (fuzzy and rough) have been successfully used in various uncertainty information processing systems. The RST, attributed by prof. Pawlak, is based on the research in the logical properties of information systems, and the uncertainty in information systems which are expressed by a boundary region [14]. RST has been generalized in many ways to tackle various problems. In particular, in 1990, Dubois and Prade [15] combined concepts of vagueness expressed by membership degrees in fuzzy sets [16] and indiscernibility in RST to obtain fuzzy rough set theory (FRST). FRST has been used e.g., for feature selection, instance selection, classification, and regression. There are many application areas that have been addressed by FRST, see e.g. [17–21].

For the sake of simplicity we assume that R is an equivalence relation. Let X is a subset of U. R-lower approximation of X ($R_*(x)$) and R-upper approximation of X ($R^*(x)$) and R-boundary region of X ($RN_R(X)$) are as follows:

$$R_*(x) = \bigcup_{x \in U} \{R(x) \subseteq X\} \tag{5}$$

$$R^*(x) = \bigcup_{x \in U} \{R(x) : R(x) \cap X \neq \emptyset\} \tag{6}$$

$$RN_R(X) = R^*(x) - R_*(x) \tag{7}$$

The paper is organized as follows: in Sect. 2, the used database is explained. In Sect. 3, the proposed feature selection is explained. In Sect. 4, the proposed type fuzzy system modeling is presented. Finally, in Sect. 5, the discussion and conclusion are presented.

2 Chronic Kidney Disease (CKD) Dataset

In this study, the Chronic Kidney database gathered from the Chamran Hospital in Tehran, Iran [22]. This data set contains 600 samples, 2 classes and fifteen features for each sample. These classes are assigned to the values that named as patient and healthy. The attributes of Chronic Kidney dataset are given in Table 1.

Table 1. The attributes of chronic kidney disease dataset

The number of attribute	The name of attribute	The values of attribute
1	Sex	Male – Female
2	Age*	2–100
3	Blood pressure max*	6–22
4	Blood pressure min	5–13
5	FBS	41–600
6	Bacteria*	Yes, No
7	Blood urea*	5–138
8	Serum creatinine*	0.5–12
9	Na - sodium*	120–150
10	K - potassium	2–8
11	Hemoglobin*	4–21
12	Rbc - red blood cells	2–7
13	Wbc - white blood cells*	0.4–40
14	Diabetes	Yes, No
15	Anemia	Yes, No

3 Feature Selection

The number of features in the raw dataset can be enormously large. This enormity may cause serious problems to many data mining systems. Feature selection is one of the oldest existing methods that deal with these problems. A method is used to compute reducts for fuzzy rough sets, where only the minimal elements in the discernibility matrix are considered. First, relative discernibility relations of conditional attribute are defined and relative discernibility relations are used to characterize minimal elements in the discernibility matrix. Then, an algorithm to compute the minimal elements is developed. Finally, novel algorithms to find proper reducts with the minimal elements are designed [23]. In general, there are two methods for choosing a feature by Rough sets: Measure the dependencies between features and Detection Matrix Method. In the first method, the degree of dependence between the features is calculated by the Eq. 8.

$$\gamma(c,d) = \frac{|POS_c(d)|}{U} \tag{8}$$

$$POS_c(d) = U_{X \in U/IND(d)} \underline{C}(X) \tag{9}$$

Which in Eq. 9, C is a set of conditional properties, and $POS_c(d)$ denotes a set of samples that are obtained in the positive region resulting from the division of samples into equivalence classes and finally a set the features that have the most dependency are introduced as optional features.

This method was used and the most important variables between the possible candidates were selected. Based on the results of this feature selection method, the

number of features was reduced to 8, which show by star in Table 1, and we used these features in our proposed system.

4 Type - II Fuzzy System Modeling

4.1 Determining the Number of Rules

In a fuzzy clustering algorithm, we should use a cluster validity index to determine the most suitable number of clusters. In this study, we used the validity index proposed by Fukuyama and Sugeno [24]. This validity index can find the number of clusters as the minimum of its function with respect to c. This index is defined as:

$$FS(c) = \sum_{i=1}^{c} \sum_{j=1}^{n} \mu_{ij}^{m} ||x_j - a_i||^2 - \sum_{i=1}^{c} \sum_{j=1}^{n} \mu_{ij}^{m} ||a_i - \overline{a}||^2 = J_m(\mu, a) + K_m(\mu, a) \quad (10)$$

Where $\overline{a} = \sum_{i=1}^{c} a_i / c$. $J_m(\mu, a)$ is the FCM objective function which measures the compactness and $K_m(\mu, a)$ measures the separation. This cluster validity index is implemented to determine the most suitable number of clusters or rules. The best number of clusters based on this cluster validity index is obtained in five clusters. So, the system contains five rules.

4.2 The Proposed Type - II Fuzzy Model

In the, we obtain fuzzy model with five rules, eight inputs and one output. The inputs are age, blood pressure (max), bacteria, urea, creatinine, Na, hemoglobin and wbc. The output of our rule-base is an interval type II fuzzy set that must be type reducted and then defuzzify. We used centroid type reduction and defuzzifier. The proposed system used the mamdani fuzzy inference method. Figures 4, 5 and 6 show the memberships functions of samples of features. In the proposed model, Gaussian membership function was used. The numbers of rules consist five.these rules are as follow:

Rule 1: IF (Age isr in1cluster1) AND (blood pressure (max) isr in2cluster1) AND (bacteria isr in3cluster1) AND (urea isr in4cluster1) AND (creatinine isr in5cluster1) AND (Na isr in6cluster1) AND (hemoglobin isr in7cluster1) AND (wbc isr in8cluster1) THEN (out isr cluster1).

Rule 2: IF (Age isr in1cluster2) AND (blood pressure (max) isr in2cluster2) AND (bacteria isr in3cluster2) AND (urea isr in4cluster2) AND (creatinine isr in5cluster2) AND (Na isr in6cluster2) AND (hemoglobin isr in7cluster2) AND (wbc isr in8cluster2) THEN (out isr cluster2).

Rule 3: IF (Age isr in1cluster3) AND (blood pressure (max) isr in2cluster3) AND (bacteria isr in3cluster3) AND (urea isr in4cluster3) AND (creatinine isr in5cluster3) AND (Na isr in6cluster3) AND (hemoglobin isr in7cluster3) AND (wbc isr in8cluster3) THEN (out isr cluster3).

Rule 4: IF (Age isr in1cluster4) AND (blood pressure (max) isr in2cluster4) AND (bacteria isr in3cluster4) AND (urea isr in4cluster4) AND (creatinine isr in5cluster4) AND (Na isr in6cluster4) AND (hemoglobin isr in7cluster4) AND (wbc isr in8cluster4) THEN (out isr cluster4).

Rule 5: IF (Age isr in1cluster5) AND (blood pressure (max) isr in2cluster5) AND (bacteria isr in3cluster5) AND (urea isr in4cluster5) AND (creatinine isr in5cluster5) AND (Na isr in6cluster5) AND (hemoglobin isr in7cluster5) AND (wbc isr in8cluster5) THEN (out isr cluster5).

Figure 3 represents the type-II fuzzy rules of the proposed system.

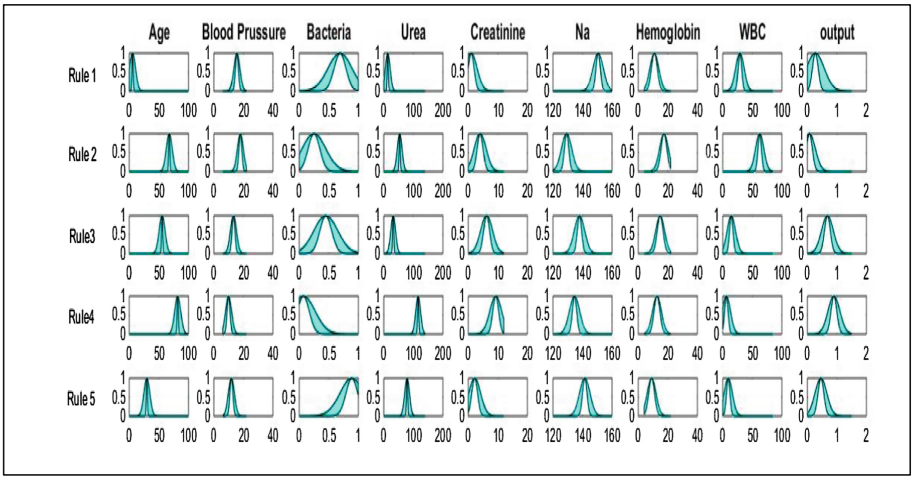

Fig. 3. Type-II-fuzzy rule-based

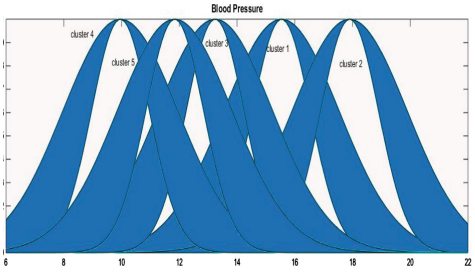

Fig. 4. Membership function of blood pressure

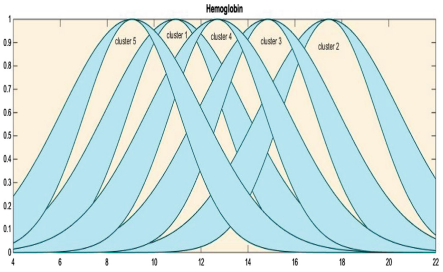

Fig. 5. Membership function of hemoglobin

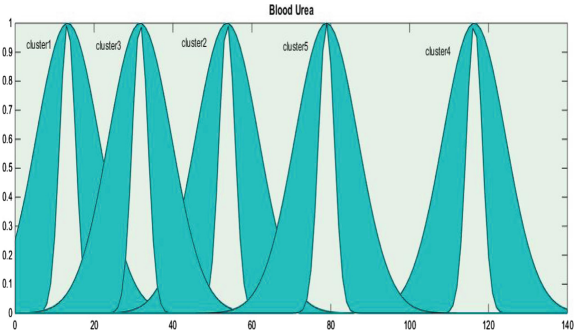

Fig. 6. Membership function of blood urea

4.3 Performance Evaluation

In this study, we used classification accuracy as criteria for evaluating the performance of the proposed system. For this purpose, we divided the CKD data set to training data and testing data. Training data consists of 480 sample data for modeling and developing the system and 120 sample data as testing data for evaluating the proposed system. By using confusion matrix method, the classification accuracy of the proposed system for diagnosis of chronic kidney disease was obtained about 90% (Eq. 11). Table 2 represents the test results of 120 testing data. As you can see in Table 3, the accuracy of the proposed method is greater than the method used in the previous article with the same data.

Table 2. The result of confusion matrix

Testing data	Class - healthy	Class - disease
Class - healthy	47	6
Class - disease	5	62

$$accuracy = \frac{47 + 62}{120} = 0.90 \tag{11}$$

Table 3. Comparison methods

Methods	Accuracy
Type - I fuzzy [6]	80%
Proposed method	90%

5 Conclusion

This paper represents an Interval type-II fuzzy rule-based expert system as an assistance system for diagnosing chronic kidneys function disease. This system uses the results of the prescribed measurement of chronic kidney as input data and by entering

the input data, the output of the system will be a crisp value. In this study, we focused on identifying the rules and the parameters of the type-II fuzzy system. We used an Interval type-II fuzzy classification based on Sugeno index and FCM algorithm for determining the number of clusters and values of parameters. The classification accuracy of the proposed system for diagnosis of chronic kidney disease was obtained about 90%.

References

1. United States Renal Data System: Excerpts from the 2000 U.S. renal data system annual data report: atlas of end stage renal disease in the United States. Am. J. Kidney Dis. **36**, S1–S279 (2000)
2. McClellan, W.M., Knight, D.F., Karp, H., Brown, W.W.: Early detection and treatment of renal disease in hospitalized diabetic and hypertensive patients: important differences between practice and published guidelines. Am. J. Kidney Dis. **29**, 368–375 (1997). PMID: 9041212
3. Coresh, J., Wei, G.L., McQuillan, G., Brancati, F.L., Levey, A.S., Jones, C., et al.: Prevalence of high blood pressure and elevated serum creatinine level in the United States: findings from the third National Health and Nutrition Examination Survey (1988–1994). Arch. Intern. Med. **161**, 1207–1216 (2001). PMID: 11343443
4. Hsu, C.Y., Chertow, G.M.: Chronic renal confusion: insufficiency, failure, dysfunction, or disease. Am. J. Kidney Dis. **36**, 415–418 (2000). PMID: 10922323
5. Zadeh, L.A.: Fuzzy logic = computing with words. IEEE Trans. Fuzzy Syst. **4**, 103–111 (1996)
6. Fazel Zarandi, M.H., Abdolkarimzadeh, M.: Fuzzy rule based expert system to diagnose chronic kidney disease. In: Melin, P., Castillo, O., Kacprzyk, J., Reformat, M., Melek, W. (eds.) NAFIPS 2017. AISC, vol. 648, pp. 323–328. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-67137-6_37
7. Abdolkarimzadeh, L., Azadpour, M., Fazel Zarandi, M.H.: Two hybrid expert system for diagnosis air quality index (AQI). In: Melin, P., Castillo, O., Kacprzyk, J., Reformat, M., Melek, W. (eds.) NAFIPS 2017. AISC, vol. 648, pp. 315–322. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-67137-6_36
8. Fazel Zarandi, M.H., Seifi, A., Ershadi, M.M., Esmaeeli, H.: An expert system based on fuzzy bayesian network for heart disease diagnosis. In: Melin, P., Castillo, O., Kacprzyk, J., Reformat, M., Melek, W. (eds.) NAFIPS 2017. AISC, vol. 648, pp. 191–201. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-67137-6_21
9. Liang, Q., Mendel, J.M.: Interval type-2 fuzzy logic systems: theory and design. IEEE Trans. Fuzzy Syst. **8**(5), 535–550 (2000)
10. Mendel, J.M., John, R.I.: Type-2 fuzzy sets made simple. IEEE Trans. Fuzzy Syst. **10**(April), 117–127 (2002)
11. Husseini, Z.M., Fazel Zarandi, M.H.: Type-2 fuzzy approach in multi attribute group decision making problem. In: Melin, P., Castillo, O., Kacprzyk, J., Reformat, M., Melek, W. (eds.) NAFIPS 2017. AISC, vol. 648, pp. 73–81. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-67137-6_8
12. Fazel Zarandi, M.H., Seifi, A., Esmaeeli, H., Sotudian, Sh.: A type-2 fuzzy hybrid expert system for commercial burglary. In: Melin, P., Castillo, O., Kacprzyk, J., Reformat, M., Melek, W. (eds.) NAFIPS 2017. AISC, vol. 648, pp. 41–51. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-67137-6_5

13. Sadat Asl, A.A., Fazel Zarandi, M.H.: A type-2 fuzzy expert system for diagnosis of leukemia. In: Melin, P., Castillo, O., Kacprzyk, J., Reformat, M., Melek, W. (eds.) NAFIPS 2017. AISC, vol. 648, pp. 52–60. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-67137-6_6
14. Pawlak, Z.: Rough sets. Int. J. Comp. Sci. **11**, 341–356 (1982)
15. Dubois, D., Prade, H.: Rough fuzzy sets and fuzzy rough sets. Int. J. Gen Syst **17**, 91–209 (1990)
16. Zadeh, L.A.: Fuzzy sets. Inf. Control **8**, 338–353 (1965)
17. Huang, B., Zhuang, Y., Li, H., Wei, D.: A dominance intuitionistic fuzzy-rough set approach and its applications. Appl. Math. Model. **37**, 7128–7141 (2013)
18. Yu, X.D.: A new patterns recognition method based on fuzzy rough sets. Appl. Mech. Mater. **380–384**, 3795–3798 (2013)
19. Bhatt, R.B., Gopal, M.: FRCT: fuzzy-rough classification trees. Pattern Anal. Appl. **11**, 73–88 (2008)
20. Leung, Y., Fischer, M.M., Wu, W.-Z., Mi, J.-S.: A rough set approach for the discovery of classification rules in interval-valued information systems. Int. J. Approx. Reason. **47**, 233–246 (2008)
21. Zarandi, F., Hossein, M., Kazemi, A.: Application of rough set theory in data mining for decision support systems (DSSs). J. Optim. Ind. Eng. 25–34 (2010)
22. Chamran hospital in iran. http://www.chamranhospital.ir
23. Hu, Q., Yu, D., Guo, M.: Fuzzy preference based rough sets. Inf. Sci. **180**(10), 2003–2022 (2010)
24. Fukuyama, Y., Sugeno, M.: A new method of choosing the number of clusters for the fuzzy c-means method. In: Proceeding of Fifth Fuzzy Systems Symposium, pp. 247–250 (1989)