World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:10, No:7, 2016

From Type-I to Type-II Fuzzy System Modeling for Diagnosis of Hepatitis

Shahabeddin Sotudian, M. H. Fazel Zarandi, I. B. Turksen

***Abstract*—**Hepatitis is one of the most common and dangerous diseases that affects humankind, and exposes millions of people to serious health risks every year. Diagnosis of Hepatitis has always been a challenge for physicians. This paper presents an effective method for diagnosis of hepatitis based on interval Type-II fuzzy. This proposed system includes three steps: pre-processing (feature selection), Type-I and Type-II fuzzy classification, and system evaluation. KNN-FD 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. In the fuzzy classification step, an "indirect approach" is used for fuzzy system modeling by implementing the exponential compactness and separation index for determining the number of rules in the fuzzy clustering approach. Therefore, we first proposed a Type-I fuzzy system that had an accuracy of approximately 90.9%. In the proposed system, the process of diagnosis faces vagueness and uncertainty in the final decision. Thus, the imprecise knowledge was managed by using interval Type-II fuzzy logic. The results that were obtained show that interval Type-II fuzzy has the ability to diagnose hepatitis with an average accuracy of 93.94%. The classification accuracy obtained is the highest one reached thus far. The aforementioned rate of accuracy demonstrates that the Type-II fuzzy system has a better performance in comparison to Type-I and indicates a higher capability of Type-II fuzzy system for modeling uncertainty.

***Keywords*—**Hepatitis disease, medical diagnosis, type-I fuzzy logic, type-II fuzzy logic, feature selection.

I. Introduction

A. Hepatitis Diseases

HEPATITIS is a viral infection that was also transmitted by blood or blood products in the past when there was no test available screening this infection. Hepatitis occurs due to one of these three viruses [1]; hepatitis A, hepatitis B, and hepatitis C. Moreover, the Epstein Barr Virus can transform into hepatitis which leads to inflammation of the liver. In addition, there are some viruses and bacteria that produce hepatitis D and E, varicella (chickenpox), and cytomegalovirus (CMV). The most important types of hepatitis, which are hepatitis A, hepatitis B, and hepatitis C, can be explained as given below [2]:

Sh. Sotudian is with the Department of Industrial Engineering Amirkabir University of Technology, Tehran, Iran (phone: +989157918122; e-mail: shahab7290@aut.ac.ir).

M. H. Fazel Zarandi is with the Department of Industrial Engineering Amirkabir University of Technology, Tehran, Iran (e-mail: zarandi@aut.ac.ir).

I. B. Turksen is with the Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, Canada (e-mail: urksen@mie.utoronto.ca).

Hepatitis A is the most common form of hepatitis in children. It is known as "infectious hepatitis" and caused by the hepatitis A virus (HAV). This virus lives in the stools (feces or poop) of infected individuals.

Hepatitis B is known as "serum hepatitis". It arises because of the hepatitis B virus (HBV). This virus is diffused from infected body fluids, such as blood, saliva, semen, vaginal fluids, tears, and urine, contaminated blood transfusion, shared contaminated needles or syringes for injecting drugs, sexual activity with an HBV-infected person, and transmission from HBV-infected mothers to their newborn babies.

Hepatitis C occurs because of the hepatitis C virus (HCV). It spreads by direct contact with an infected person's blood. Hepatitis C causes chronic liver disease and liver transplantation, and is becoming an increasing cause of concern in the world. The symptoms of this hepatitis type are similar to those of hepatitis A and B. The hepatitis C virus is diffused by sharing drug needles, getting a tattoo or body piercing with unsterilized tools, blood transfusions (especially ones that occurred before 1992; since then the US blood supply has been routinely screened for the disease), transmission from mother to newborn, and sexual intercourse.

The signs and symptoms of hepatitis are malaise (a general ill feeling), fever, muscle aches, loss of appetite, nausea, vomiting, diarrhea, and jaundice (the yellowing of the skin and whites of the eyes). All A, B and C viral hepatitis conditions can be diagnosed and followed through the use of readily available blood tests [3]. A physician commonly takes decisions by evaluating the current test results of a patient, or compares the patient with other patients under the same condition by referring to the previous results and decisions. Therefore, it is very difficult for a physician to diagnose hepatitis.

B. Fuzzy Logic System

Fuzzy set theory was first introduced by Zadeh in 1965 [4]. 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. Type-1 FLSs (T1 FLSs) are the best known and widely used types of FLS.

The concept of a Type-II fuzzy set was first introduced by Zadeh as an extension of Type-I fuzzy set [5]. A Type-I fuzzy set is characterized by the fuzzy membership function, i.e., the membership grade for each element is a fuzzy set in an interval [0], [1].Such sets can be used in situations where there are uncertainties about the membership values. As more complex models, T2 FSs are considered to be potentially

suitable for modeling uncertainty. The additional complexity arises from the inclusion of a footprint of uncertainty (FOU) and a third dimension, offering extra degrees of freedom to T2 FSs in comparison to T1 FSs [6], [7]. The most important application of fuzzy sets theory is fuzzy rule-based systems. These kinds of systems constitute an extension of classical rule-based systems because they deal with fuzzy rules instead of classical logic rules.

A rule-based fuzzy logic system is comprised of four elements: rules, fuzzifier, inference engine and output processor that are inter-connected, a T1 FLS is depicted in Fig. 1.

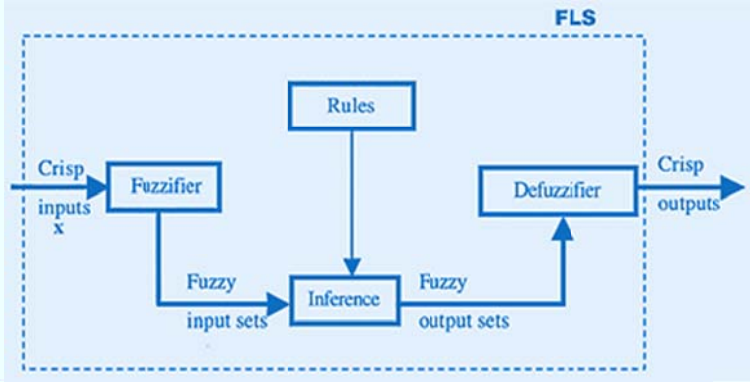

Fig. 1 Type-I fuzzy logic system [9]

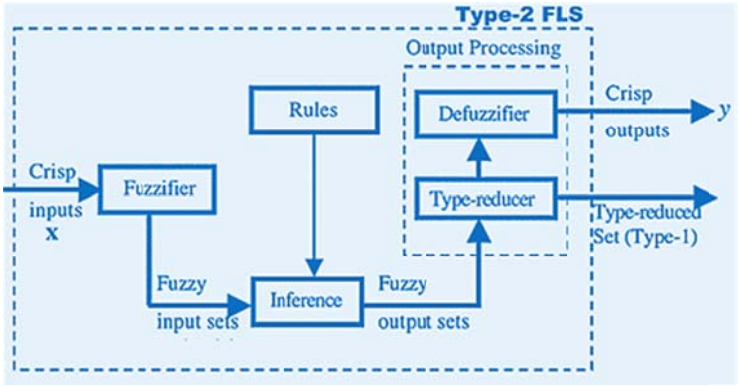

Fig. 2 Type-II fuzzy logic system [9]

The difference between T1 FLS and T2 FLS is in the output processing, so that, there is a type-reducer in output processing besides other modules. Fig. 2 represents the structure of a T2 FLS. Once the rules have been established, an FLS can be viewed as a mapping from inputs to outputs. Rules are the heart of an FLS. They may be provided by experts or extracted from numerical data. In either case, the rules can be expressed as a collection of IF–THEN statements. The IF-part of a rule is its antecedent, and the THEN-part of a rule is its consequent [8].

C. Interval Type-I Fuzzy

There are two types of Type-II fuzziness: Interval-valued Type-II and generalized Type-II fuzzy. Interval-valued Type-II fuzzy is a special Type-II fuzzy, where the upper and lower bounds of membership are crisp and the spread of membership distribution is ignored considering the assumption that membership values between upper and lower values are uniformly distributed. For Generalized Type-II fuzzy the upper and lower membership values are defined, as well as the spread of membership values between these bounds (either probabilistically or fuzzily). As mentioned above, Type-II membership function provides additional degrees of freedom in fuzzy logic systems, which can be very useful in many uncertain situations [9]. A Type-II fuzzy set $\tilde{A}$ may be represented as:

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

where $\mu_{\tilde{A}}(x,u)$ is the Type-II fuzzy membership function in which $0 \leq \mu_{\tilde{A}}(x,u) \leq 1$. $\tilde{A}$ can also be defined as:

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

where $\int \int$ denotes union over all admissible x and u. J_x is called primary membership of x. Additionally, there is a secondary membership value corresponding to each primary membership value that defines the possibility for primary memberships. Whereas the secondary membership functions can take values in the interval of [0], [1] in generalized Type-II fuzzy logic systems, they are uniform functions that only take on values of 1 in interval Type-I fuzzy logic systems.

Since the general Type-II fuzzy logic systems are computationally very demanding, the use of interval Type-II fuzzy logic systems is more commonly seen in the literature, due to the fact that the computations are more manageable [10].

When all $\mu_{\tilde{A}}(x,u)$ are equal to 1, then $\tilde{A}$ is an interval Type-II fuzzy logic systems. The special case of "(2)" might be defined for the interval Type-II fuzzy logic systems:

$$\tilde{A} = \int_{x\in X} \int_{u\in J_x} 1/(x,u) \qquad J_x \subseteq [0,1] \tag{3}$$

The upper membership function (UMF) and lower membership function (LMF) of $\tilde{A}$ are two T1 MFs that bound the FOU. The UMF is associated with the upper bound of FOU ($\tilde{A}$) and is denoted $\bar{\mu}_{\tilde{A}}(x)\ \forall x \in X$, and the LMF is associated with the lower bound of FOU($\tilde{A}$) and is denoted $\underline{\mu}_{\tilde{A}}(x)\ \forall x \in X$ [10]:

$$\begin{aligned} \bar{\mu}_{\tilde{A}}(x) &\equiv \overline{\mathrm{FOU}(\tilde{A})} \qquad \forall x\in X \\ \underline{\mu}_{\tilde{A}}(x) &\equiv \underline{\mathrm{FOU}(\tilde{A})} \qquad \forall x\in X \end{aligned} \tag{4}$$

It should be noted that Type-II fuzzy sets can model and minimize the effects of uncertainties in rule-based fuzzy logic systems. The effects of uncertainties can be minimized by optimizing the parameters of the Type-II fuzzy sets during a training process. The purpose of this study is to demonstrate the higher ability of Type-II fuzzy systems to modeling uncertainty. The paper is organized as follows: In Section II, previous works for diagnosis of hepatitis diseases in literature is presented. In Section III, the used hepatitis database is explained. In Section IV, the feature number of hepatitis disease dataset is reduced from 19 to 10. In Section V, the proposed Type-I fuzzy system modeling is presented. In section VI, some reasons for using Type-II fuzzy system, instead of the Type-I fuzzy system, are presented. In section VII, the proposed Type-II fuzzy system modeling is explained. Finally, in Section VIII, the discussion and conclusion are presented.

II. Literature Review

Up to now, many studies have been performed in the diagnosis of hepatitis literature. In some cases, articles attempted to increase the classification accuracy. Although classification accuracy is an important feature of a system, this study focuses on generated fuzzy-rules and the values of membership function's parameters. Table I presents the classification accuracy of previous hepatitis diagnosis methods [11].

III. Hepatitis Disease Dataset

In this study, the hepatitis database obtained from the UCI repository of machine learning databases is used [12]. This hepatitis disease dataset requires determination of whether the patients having hepatitis will either live or die. The purpose of the dataset is to predict the presence or absence of the hepatitis disease given the results of various medical tests carried out on a patient. This database contains 19 attributes, which have been extracted from a larger set of 155. The hepatitis dataset contains 155 samples belonging to two different classes (32 "die" cases, 123 "live" cases). There are 19 attributes, 13 binary attributes and six with 6–8 discrete values. The attributes of Hepatitis dataset are given in Table II.

IV. Feature Selection

The number of features (attributes) and instances 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 deals with these problems. Its objective is to select a minimal subset of features according to some reasonable criteria so that the original task can be equally achieved well, if it was not better. By choosing a minimal subset of features, irrelevant and redundant features are removed according to the criterion. Simpler data can lead to more concise results and their better comprehensibility [1].

TABLE I
Classification Accuracies Obtained by Other Methods in Literature [11]

Author	Method	Accuracy (%)
"Grudzinski et al [19]"	Weighted 9NN	92.9
"Grudzinski et al [20]"	18NN, stand. Manhattan	90.2
"Grudzinski et al [20]"	15NN, stand. Euclidean	89.0
"Adamczak et al [21]"	FSM with rotations	89.7
"Adamczak et al [21]"	FSM without rotations	88.5
"Adamczak et al [21]"	RBF (Tooldiag)	79.0
"Adamczak et al [22]"	MLP+BP (Tooldiag)	77.4
"Šter and Dobnikar [23]"	LDA	86.4
"Šter and Dobnikar [23]"	Naive Bayes and semi-NB	86.3
"Šter and Dobnikar [23]"	QDA	85.8
"Šter and Dobnikar [23]"	1NN	85.3
"Šter and Dobnikar [23]"	ASR	85.0
"Šter and Dobnikar [23]"	FDA	84.5
"Šter and Dobnikar [23]"	LVQ	83.2
"Šter and Dobnikar [23]"	CART (decision tree)	82.7
"Šter and Dobnikar [23]"	MLP with BP	82.1
"Šter and Dobnikar [23]"	ASI	82.0
"Šter and Dobnikar [23]"	LFC	81.9
"Norbert Jankowski [24]"	Inc Net	86.0
"Özyıldırım et al [25]"	MLP	74.37
"Özyıldırım et al [26]"	RBF	83.75
"Özyıldırım et al [26]"	GRNN	80.0
"Polat and Gunes [1]"	FS-AIRS with fuzzy res.	92.59

Present research used the k-nearest neighborhood functional dependency (KNN-FD) approach proposed by [13]. This feature selection algorithm combines feature wrapper and feature filter approaches in order to identify the significant input variables in systems with continuous domains. This method utilizes functional dependency concept, correlation coefficients and K-nearest neighborhood (KNN) method to implement the feature filter and feature wrappers. Four feature selection methods independently select the significant input variables and the input variable combination, which yields the

best result respective to their corresponding evaluation function and selected as the winner [13]. This method was used and the most important variables between the possible candidates was selected.

TABLE II
THE ATTRIBUTES OF HEPATITIS DISEASE DATABASE

The number of attribute	The name of attribute	The values of attribute
1	Age	10, 20, 30, 40, 50, 60, 70, 80
2	Sex	Male, female
3	Steroid	Yes, No
4	Antivirals	Yes, No
5	Fatigue	Yes, No
6	Malaise	Yes, No
7	Anorexia	Yes, No
8	Liver big	Yes, No
9	Liver firm	Yes, No
10	Spleen palpable	Yes, No
11	Spiders	Yes, No
12	Ascites	Yes, No
13	Varices	Yes, No
14	Bilirubin	0.39, 0.8, 1.2, 2.0, 3.0, 4.0
15	Alk phosphate	33, 80, 120, 160, 200, 250
16	SGOT	13, 100, 200, 300, 400, 500
17	ALBUMIN	2.1, 3.0, 3.8, 4.5, 5.0, 6.0
18	PROTIME	10, 20, 30, 40, 50, 60, 70, 80, 90
19	HISTOLOGY	Yes, No

Based on the results of this feature selection method, the number of features was reduced to 10 by removing age, sex, antivirals, anorexia, liver big, spleen palpable, bilirubin, protime, histology values and we used the other features in our proposed system.

V. TYPE-I FUZZY SYSTEM MODELING

A. 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

Fazel Zarandi et al. [14]. This validity index V_{ECAS} (an Exponential compactness and separation index) can find the number of clusters as the maximum of its function with respect to c. This index is defined as [14]:

$$V_{ECAS} = ECAS(c) = \frac{EC_{comp}(c)}{\max_c(EC_{comp}(c))} - \frac{ES_{sep}(c)}{\max_c\left(ES_{sep}(c)\right)} \quad (5)$$

where $EC_{comp}(c)$ and$ES_{sep}(c)$ are Exponential compactness and Exponential separation measures respectively, and are defined as:

$$EC_{comp}(c) = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^{m} \; exp\left(-\left(\frac{\|x_i - v_j\|^2}{\beta_{comp}} + \frac{1}{c+1}\right)\right) \quad (6)$$

$$ES_{sep}(c) = \sum_{i=1}^{c} exp\left(-\min_{i \neq k}\left\{\frac{(c-1)\|v_i - v_k\|^2}{\beta_{sep}}\right\}\right) \quad (7)$$

in which,

$$\beta_{comp} = (\sum_{k=1}^{n} \|x_i - \bar{v}\|^2/n(i)$$

and

$$\beta_{sep} = (\sum_{l=1}^{c} \|v_l - \bar{v}\|^2/c) \text{ with } \bar{v} = (\sum_{j=1}^{n} x_j/n).$$

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 three clusters. So, the Type-I system contains three rules.

B. The Proposed Type-I Fuzzy Model

The determination of fuzzy rules from data is an important issue for solving tasks like building fuzzy controllers, fuzzy classifiers, or supporting decision-making processes.

For many application problems, classifiers can be used to support a decision-making process. In some areas like medical, it is not preferable to use black box approaches. The user should be able to understand the classifier and to evaluate its results. Fuzzy rule-based classifiers are especially suitable because they consist of simple linguistically interpretable rules and do not have some drawbacks of symbolic or crisp rule-based classifiers. Classifiers must often be created from data by a learning process because there is not enough expert knowledge to determine their parameters completely [15].

In the Type-1 fuzzy model, we obtain its model with three rules, ten inputs, and one output. The inputs are steroid, fatigue, malaise, liver firm, spiders, ascites, varices, Alk-phosphate, SGOT and albumin. We use Mamdani-style inference, min–max operators and centroid defuzzification methods. In the proposed model, Gaussian membership function was used for fuzzy sets description. The rule-based of the proposed system consists of three general rules. The rules of the proposed system are as follows:

1. If (STEROID is in1cluster1) and (FATIGUE is in2cluster1) and (MALAISE is in3cluster1) and (LIVER_FIRM is in4cluster1) and (SPIDERS is in5cluster1) and (ASCITES is in6cluster1) and (VARICES is in7cluster1) and (ALK_PHOSPHATE is in8cluster1) and (SGOT is in9cluster1) and (ALBUMIN is in10cluster1) then (output is out1cluster1)
2. If (STEROID is in1cluster2) and (FATIGUE is in2cluster2) and (MALAISE is in3cluster2) and (LIVER_FIRM is in4cluster2) and (SPIDERS is in5cluster2) and (ASCITES is in6cluster2) and (VARICES is in7cluster2) and (ALK_PHOSPHATE is in8cluster2) and (SGOT is in9cluster2) and (ALBUMIN is in10cluster2) then (output is out1cluster2)
3. If (STEROID is in1cluster3) and (FATIGUE is in2cluster3) and (MALAISE is in3cluster3) and (LIVER_FIRM is in4cluster3) and (SPIDERS is in5cluster3) and (ASCITES is in6cluster3) and (VARICES is in7cluster3) and (ALK_PHOSPHATE is in8cluster3) and (SGOT is in9cluster3) and (ALBUMIN is in10cluster3) then (output is out1cluster3)

For a better view of the rule-based, Fig. 3 represents the fuzzy rules of the proposed system.

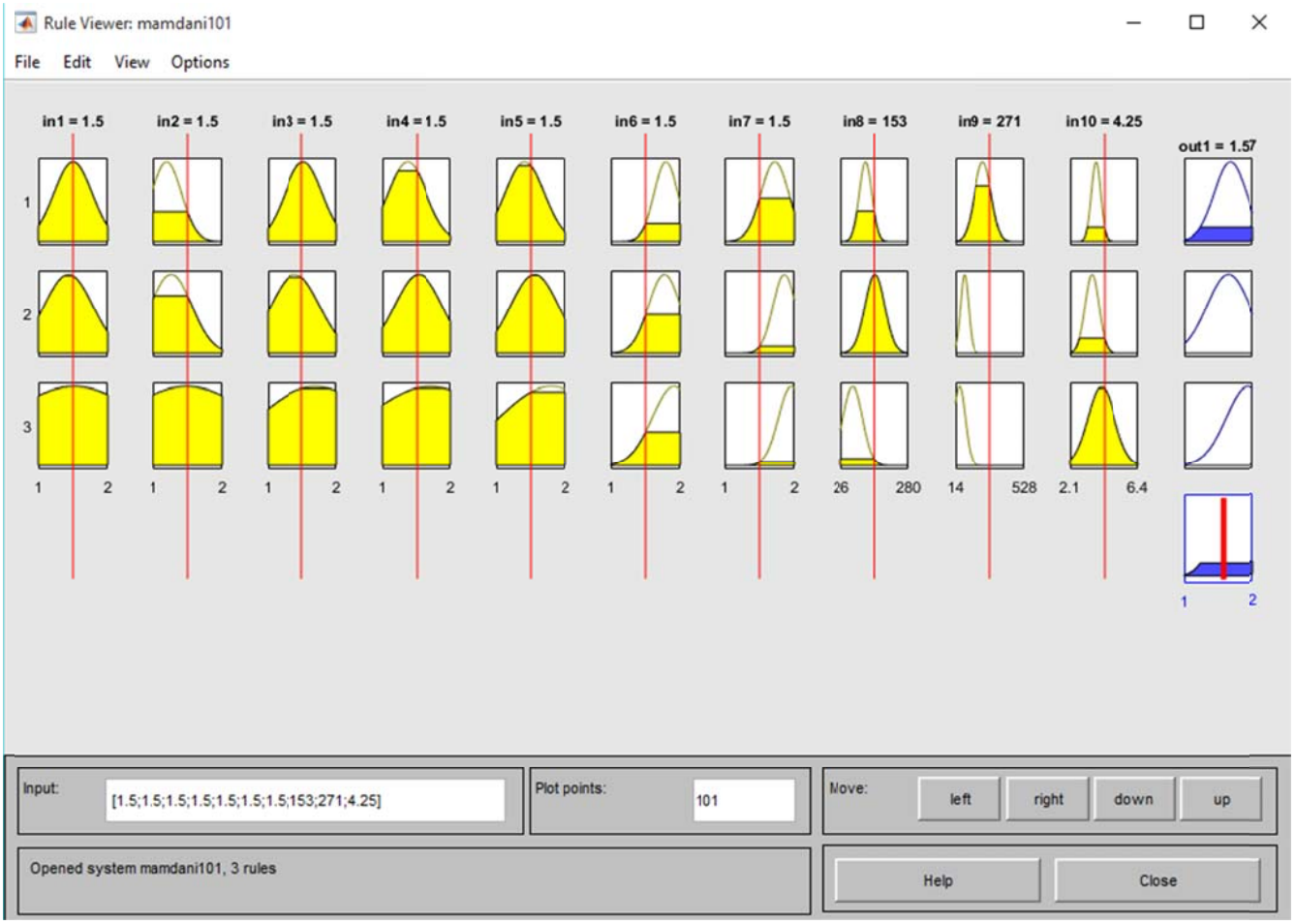

Fig. 3 Type-I Fuzzy Rule-Based

C. Performance Evaluation

We used classification accuracy for evaluating the performance of the proposed system. For this purpose, the entire dataset is divided into two sets. One is used for training the system and the other to test them. The training set consists of 122 samples. These 122 samples consist of both the classes (live or die). The test set contains 33 samples. These samples are used to check the performance of the proposed system.

Table III gives the details of the training and test samples. The classification accuracy of the Type-I system for diagnosis of Hepatitis diseases has obtained about 90.9 %.

TABLE III
DETAILS OF THE TRAINING AND TEST SAMPLES

Dataset	Total Samples	Class	
		Live	*Die*
Training data	122	98	24
Testing data	33	25	8

VI. From Type-I to Type-II Fuzzy System Modeling

Researches done in recent years have undergone a significant increase toward more complex forms of fuzzy logic such as Interval Type-II fuzzy logic systems (IT2 FLSs) and more recently, general Type-II FLSs (T2 FLSs) [16], [17].

This transition was motivated by this concept that Type-I fuzzy sets (T1 FSs) can only handle a limited level of uncertainty, whereas real-world applications are often faced with multiple sources and high levels of uncertainty [6]. In the last section, we presented a Type-I fuzzy system for diagnosis of Hepatitis, but we were convinced to use Type-II fuzzy system due to following reasons:

1. The diagnosis of Hepatitis is a complicated process dealing with uncertainty.
2. The higher capability of the Type-II fuzzy system for modeling uncertainty.
3. The structure and semantic of this issue are closer to Type-II fuzzy system.

It should be noted that Type-II fuzzy sets can model and minimize the effects of uncertainties in rule-based fuzzy logic systems. In the next section, we presented a Type-II fuzzy system for diagnosis of Hepatitis.

VII. Type-II Fuzzy System Modeling

A. Determining the Number of Rules

In Section V.A, we introduce exponential compactness and separation index proposed by Fazel Zarandi et al. [14]. This cluster validity index is implemented to indicate a proper number of clusters or rules. The best number of clusters based

on the current cluster validity index is five clusters. So, the Type-II system contains five rules.

B. The Proposed Type-II Fuzzy Model

In the Type-II fuzzy model, we obtain a fuzzy model with five rules, ten inputs, and one output. The inputs are the steroid, fatigue, malaise, liver firm, spiders, ascites, varices, Alk- phosphate, SGOT and albumin. The output of our rule-base is an interval Type-II fuzzy set that must be type- reduced and then defuzzify. We used centroid type reduction and defuzzifier. The proposed system uses the Mamdani fuzzy inference method, in which the output membership function is a fuzzy set so it is difficult to understand the output. Due to this fact, the centroid method is used for defuzzification which takes the fuzzy set as input and crisp value as an output.

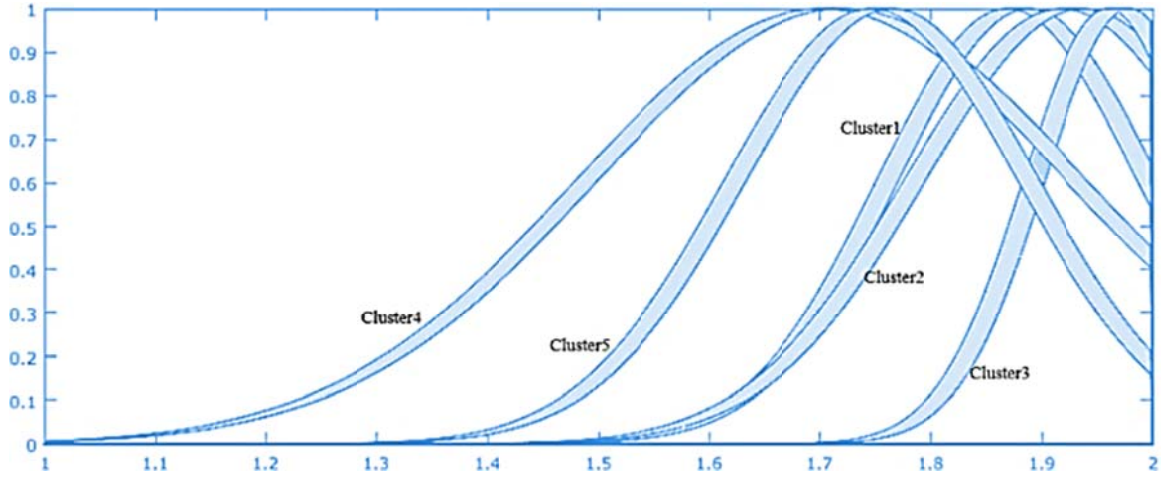

Fig. 4 Membership functions of feature 7

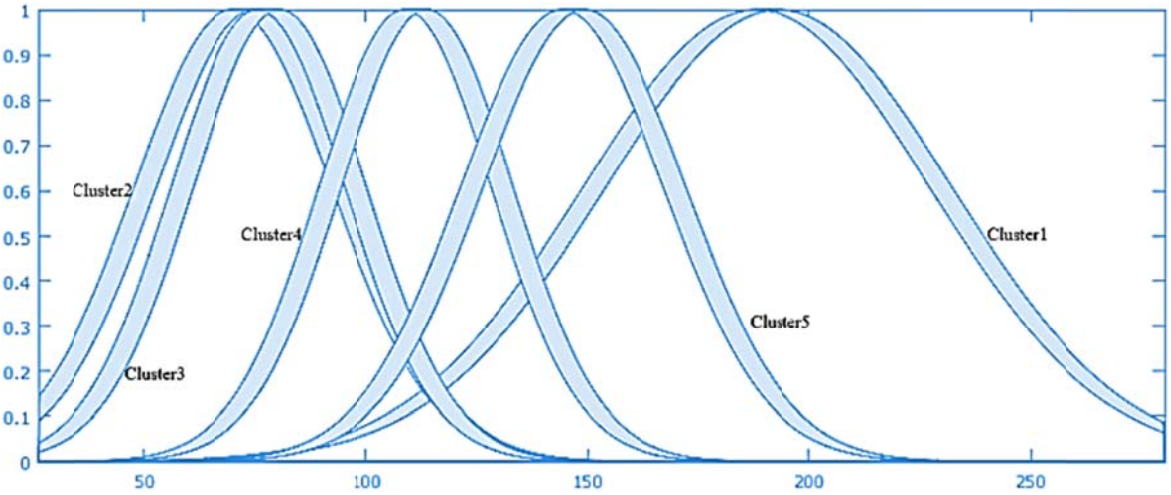

Fig. 5 Membership functions of feature 8

By applying the proposed Type-II fuzzy model, the parameters of the proposed diagnosis system were determined. These parameters for Features 7, 8 and 9 are represented below:

1. Feature 7: Varices

Varices are dilated blood vessels usually that exist in the esophagus or stomach. They cause no symptoms unless they rupture and bleed. Bleeding from varices is a life-threatening complication of portal hypertension. Portal hypertension is an increase in the pressure within the portal vein (the vein that carries blood from the digestive organs to the liver) due to blockage of blood flow throughout the liver. The most common cause of portal hypertension is cirrhosis of the liver. Cirrhosis is scarring which accompanies the healing of injured liver caused by hepatitis, alcohol, or other less common causes of liver damage [18]. Fig. 4 demonstrates the membership function of feature 7.

2. Feature 8: Alk Phosphate

Alk Phosphate (ALP) is an enzyme found in all body tissues. There are many different forms of ALP called isoenzymes. The structure of the enzyme depends on where it is produced in the body. Tissues with higher amounts of ALP include the liver, bile ducts, and bone. The alkaline phosphatase is the most frequently used test to detect

obstruction in the biliary system. Elevation of this enzyme may be found in a large number of disorders as common as gallstone disease, alcohol abuse, and drug-induced hepatitis [18]. Fig. 5 demonstrates the membership function of this feature.

3. Feature 9: SGOT

Serum glutamic oxaloacetic transaminase (SGOT) is an enzyme that normally present in liver and heart cells. SGOT is released into the blood when the liver or heart is damaged. Thus the blood SGOT levels are elevated with liver damage (for example viral hepatitis) or with an insult to the heart (for example a heart attack) [18]. Fig. 6 demonstrates the membership function of this feature.

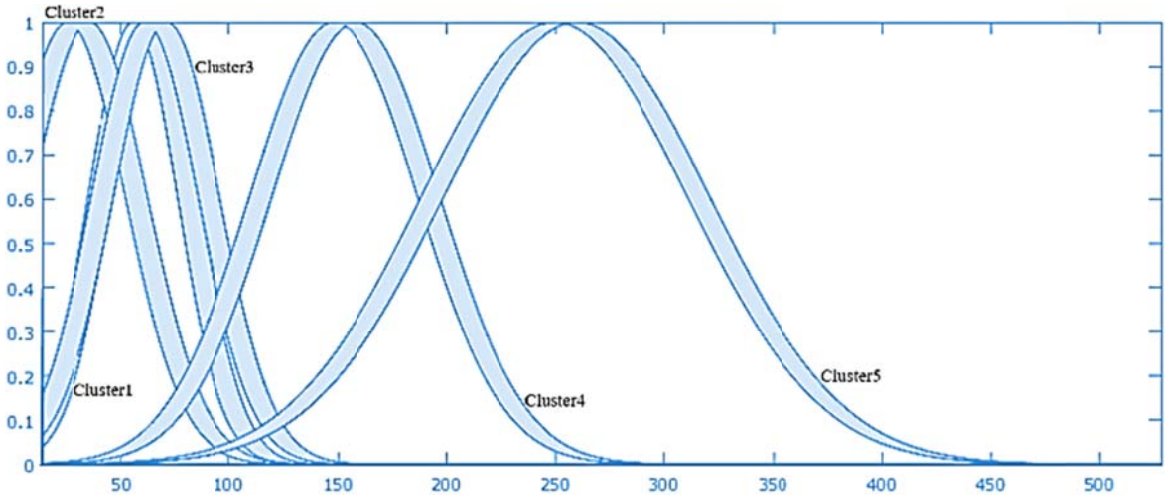

Fig. 6 Membership functions of feature 9

In the proposed interval Type-II classifier, Gaussian membership function was used for fuzzy sets description. The rule-based of the proposed system consists of five general rules. These rules are as follows:

1. If (STEROID is in1cluster1) and (FATIGUE is in2cluster1) and (MALAISE is in3cluster1) and (LIVER_FIRM is in4cluster1) and (SPIDERS is in5cluster1) and (ASCITES is in6cluster1) and (VARICES is in7cluster1) and (ALK_PHOSPHATE is in8cluster1) and (SGOT is in9cluster1) and (in10 is in10cluster1) then (out1 is out1cluster1)
2. If (STEROID is in1cluster2) and (FATIGUE is in2cluster2) and (MALAISE is in3cluster2) and (LIVER_FIRM is in4cluster2) and (SPIDERS is in5cluster2) and (ASCITES is in6cluster2) and (VARICES is in7cluster2) and (ALK_PHOSPHATE is in8cluster2) and (SGOT is in9cluster2) and (in10 is in10cluster2) then (out1 is out1cluster2)
3. If (STEROID is in1cluster3) and (FATIGUE is in2cluster3) and (MALAISE is in3cluster3) and (LIVER_FIRM is in4cluster3) and (SPIDERS is in5cluster3) and (ASCITES is in6cluster3) and (VARICES is in7cluster3) and (ALK_PHOSPHATE is in8cluster3) and (SGOT is in9cluster3) and (in10 is in10cluster3) then (out1 is out1cluster3)
4. If (STEROID is in1cluster4) and (FATIGUE is in2cluster4) and (MALAISE is in3cluster4) and (LIVER_FIRM is in4cluster4) and (SPIDERS is in5cluster4) and (ASCITES is in6cluster4) and (VARICES is in7cluster4) and (ALK_PHOSPHATE is in8cluster4) and (SGOT is in9cluster4) and (in10 is in10cluster4) then (out1 is out1cluster4)
5. If (STEROID is in1cluster5) and (FATIGUE is in2cluster5) and (MALAISE is in3cluster5) and (LIVER_FIRM is in4cluster5) and (SPIDERS is in5cluster5) and (ASCITES is in6cluster5) and (VARICES is in7cluster5) and (ALK_PHOSPHATE is in8cluster5) and (SGOT is in9cluster5) and (in10 is in10cluster5) then (out1 is out1cluster5)

For a better view of the rule-based system, Fig. 7 represents the Type-II fuzzy rules of the proposed system.

C. Performance Evaluation

Like the Type-I fuzzy system, we used classification accuracy for evaluating the performance of the proposed Type-II fuzzy system. As mentioned above, the test set contains 33 samples. These samples are used to check the performance of the proposed system. The classification accuracy of the Type-II fuzzy system for diagnosis of the Hepatitis diseases was obtained at about 93.94 %.

According to the results, interval Type-II fuzzy has the ability to diagnose hepatitis with the average accuracy of 93.94 %, which is a better performance compared with the Type-I fuzzy system.

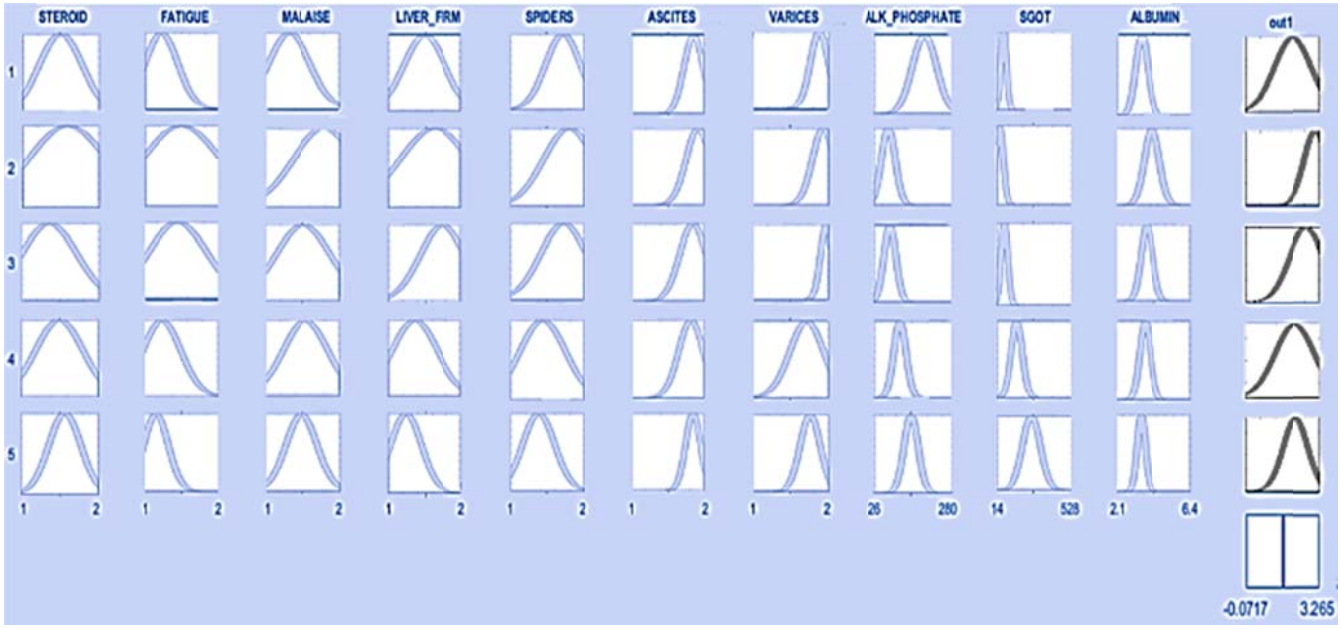

Fig. 7 Type-I I Fuzzy Rule-Based

VIII. Discussion and Conclusion

In this study, an effective method for the diagnosis of the Hepatitis disease based on interval Type-II fuzzy was proposed. Firstly, we proposed a Type-I fuzzy system for diagnosis of Hepatitis diseases. The obtained classification accuracy rate of this system was about 90.9 %. Due to the structure and semantics of hepatitis diagnosis and higher capability of the Type-II fuzzy system for modeling uncertainty, we used the Type-II fuzzy system. The extra parameters of Type-II fuzzy provide additional degrees of freedom, making it possible to minimize the effects of vagueness. According to the obtained results, the Type-II fuzzy has the ability to diagnose Hepatitis with the average accuracy of about 93.94 %. The obtained classification accuracy is the highest one reached so far. As expected, the accuracy of the Type-I I fuzzy system was higher than the Type-I fuzzy system; highlighting the ability of the Type-II fuzzy system for modeling the uncertainty.

References

[1] K. Polat and S. Güneş, "Hepatitis disease diagnosis using a new hybrid system based on feature selection (FS) and artificial immune recognition system with fuzzy resource allocation," *Digital Signal Processing,* vol. 16, no. 6, 2006, pp. 889-901.

[2] "The Nemours Foundation/KidsHealth ", (Online). Available: http://kidshealth.org/parent/infections/bacterial_viral/hepatitis.html. (Accessed 31 Dec 2015).

[3] E. Dogantekin, A. Dogantekin and D. Avci, "Automatic hepatitis diagnosis system based on Linear Discriminant Analysis and Adaptive Network based on Fuzzy Inference System," *Expert Systems with Applications,* vol. 36, no. 8, 2009, pp. 11282-11286.

[4] L. A. Zadeh, "Fuzzy sets," *Information and Control,* vol. 8, 1965, pp. 338-353.

[5] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning," *Information Sciences,* vol. 8, 1975, pp. 199-249.

[6] J.M. Mendel, "Uncertain Rule-Based Fuzzy Logic Systems:Introduction and New Directions," *Prentice Hall,* 2001.

[7] D. Wu and . J. M. Mendel, "Uncertainty measures for interval type-2 fuzzy sets," *Information Sciences,* vol. 177, 2007, pp. 5378-5393.

[8] J.M. Mendel, J. Robert and L. Feilong , "Interval type-2 fuzzy logic systems made simple," *Fuzzy Systems, IEEE Transactions,* vol. 14, no. 6, 2006, pp. 808-821.

[9] M.H Fazel Zarandi, M. Zarinbal and M. Izadi, "Systematic image processing for diagnosing brain tumors: A Type-I I fuzzy expert system approach," *Applied soft computing,* vol. 11, no. 1, 2011, pp. 285-294.

[10] J.M. Mendel and Robert I. Bob John. "Type-2 fuzzy sets made simple." *Fuzzy Systems, IEEE Transactions on*, vol. 10, no. 2, 2002, pp. 117-127.

[11] D. Çalişir and E. Dogantekin, "A new intelligent hepatitis diagnosis system: PCA-LSSVM," *Expert Systems with Applications,* vol. 38, no. 8, 2011, pp. 10705-10708.

[12] C. L. Blake and C. J. Merz, (1996). UCI repository of machine learning databases (Online) Available: http://www.ics.uci.edu./~mlearn/MLReporsitory.html.

[13] Ö. Uncu and I. B. Türkşen, "A novel feature selection approach: combining feature wrappers and filters," *Information Sciences,* vol. 177, no. 2, 2007, pp. 449-466.

[14] M.H Fazel Zarandi,M.R Faraji and M. Karbasian. "An exponential cluster validity index for fuzzy clustering with crisp and fuzzy data." *Sci. Iran. Trans. E: Ind. Eng*, vol. 17, no. 2, 2010, pp. 95-110.

[15] D. Nauck and K. Rudolf , "Obtaining interpretable fuzzy classification rules from medical data," *Artificial intelligence in medicine,* vol. 16, no. 2, 1999, pp. 149-169.

[16] L. A. Zadeh, "Toward extended fuzzy logic—A first step," *Fuzzy Sets and Systems,* vol. 160, 2009, pp. 3175-3181.

[17] L. Qilian and J. M. Mendel, "Interval type-2 fuzzy logic systems:theory and design," *IEEE Transactions on Fuzzy Systems,* vol. 8,2000, pp. 535-550.

[18] "Canadian Liver Foundation (CLF)," (Online). Available: http://www.liver.ca/liver-disease. (Accessed 31 Dec 2015).

[19] W. Duch, K. Grudzinski, and G. Diercksen. "Neural minimal distance methods." In Proc. 3-rd Conf. on Neural Networks and Their Applications, Kule, Poland, 1997, pp. 183-188.

[20] W. Duch and Karol Grudziński. "Ensembles of similarity-based models." *In Intelligent Information Systems*, 2001, pp. 75-85.

[21] W. Duch and R. Adamczak. "Statistical methods for construction of neural networks." *In ICONIP*, 1998, pp. 639-642.

[22] W. Duch, R. Adamczak, and Diercksen, "Neural Networks from Similarity Based Perspective." *New Frontiers in Computational Intelligence and its Applications*. Ed. M. Mohammadian, IOS Press, Amsterdam, 2000, pp.93-108.

[23] B. Ster, and Andrej Dobnikar. "Neural networks in medical diagnosis: Comparison with other methods." *Proceedings of the International Conference EANN*. vol. 96. 1996, pp. 427-430.

[24] N. Jankowski. "Approximation and classification in medicine with incnet neural networks." *Machine Learning and Applications*. Workshop on Machine Learning in Medical Applications, 1999, pp. 53-58.
[25] L. Ozyilmaz and Y. Tulay "Artificial neural networks for diagnosis of hepatitis disease." In *Neural Networks, 2003. Proceedings of the International Joint Conference on*, vol. 1, 2003, pp. 586-589.
[26] R.A. Vural, L. Özyılmaz, and T. Yıldırım. "A comparative study on computerised diagnostic performance of hepatitis disease using ANNs." In *Computational Intelligence*, 2006, pp. 1177-1182.