HMFC: Hybrid MODLEM – Fuzzy Classifier for Liver Diseases Diagnose

Diaa Salama AbdElminaam[1],Walaa H. El-Ashmawi[2], Shaimaa Abdallah Ibraheem[3]
[1]Information Systems department, Faculty of Computers and Informatics, Benha University, Egypt
[2]Computer Science department, Faculty of Computers and Informatics, Suez Canal University, Egypt
[3]Computer Science department, Higher Technological Institute, 10th of Ramadan City, Egypt

Abstract, *Diagnose of Common diseases like liver is very complicated because its symptoms can be vague and easily confused with other health problems. In some cases, a person may have no symptoms at all but the liver may already have suffered significant damage. Many mathematical approach can deal with vagueness and uncertainty like fuzzy set theory (FS) although FS is good for classification but it requires more knowledge and experience in setting rules of inference based.in this paper an intelligent diagnosing model called hybrid modlem2-fuzzy classifier (HFMC) is proposed to enhance risk classification accuracy of liver diseases. Rough set approaches (RST) used to generate automated perfect rules to improve accuracy of classification of FS. The proposed model implemented in two phases, in first phase rules generated by laplace-Modlem2 and use othor RST algorithmes (LEM2, Laplace-Modlem2) as comparaison way .in second phase build fuzzy inference system based on the rules generated in first phase. The proposed model gives result of 99.14 % classification accuracy of rule generation.*

Keywords: *Rough Set theory, Fuzzy Set Theory, Uncertainty, Rules Generation, liver diseases, classification accuracy.*

Received August 29, 2018; Accepted November 25, 2018

1. Introduction

Many countries start a lot of initiatives to develop their health care systems with investments in health information technology (IT). The goal of these initiatives is to use technology to improve the health care system by reducing costs, increasing patient safety and improving quality of care. Improving health care is a common goal for these countries [1, 3, 5] .

One of important field in health and medical care field is diseases diagnose process. Health-care professionals use symptoms, tests and signs as references that used to determine the most likely diagnosis when illness is exist. Symptoms are also used to create a list of the possible differential diagnoses. These differential diagnosis is the basic point to safe patient's life by choosing initial treatments. Liver disease is common and dangerous disease in recent years because pollution of water, air and food. Most of patients with liver disease have multiple symptom's forms and values, but these symptoms not appear at earlier stage. It also may cause vague and uncertain symptoms, as loss of energy and weakness [3]. So our goal to provide early liver diseases diagnose system with high accuracy no way for errors because it doesn't depend on physician's expert or knowledge. There are many researcher proposed heart and liver disease diagnose model. A Fuzzy Expert System for heart disease diagnosis is designed with accuracy of 94% while rules not extracted automatically but it based on knowledge of physicians [7]. A Decision support system has been proposed for diagnosis of Congenital Heart Disease based on fuzzy set theory with 90% classification accuracy [18]. Computational intelligence techniques for Liver Patient Classification are presented which evaluates the selected classification algorithms (J-48, Multi-Layer Perceptron, Support Vector Machine, Random Forest and Bayesian Network) for the classification of liver patient datasets [6]. A fuzzy expert system for the diagnosis of Cirrhosis is designed. Numbers of rules used are 85 rules in this system and this large number which lake of accuracy and it lead to time consuming [9]. An intelligent diagnosis system for liver disease is proposed based on artificial neural network, rough and fuzzy set. But Using LEM algorithm lead to 6 rules are generated with accuracy of 96% and this not a correct classification [2].a short summary on the use of rough sets in the medical informatics domain presented, focusing on applications of medical image segmentation, pattern classification and computer assisted medical decision making presented in [13].The previous show importance of Knowledge discovery techniques in business and medical development in recent days. It is becoming very effective one in extract useful and meaningful information. There are many researchers are interested in proposing models based on discovery knowledge approaches. pawlak present a lot of Rough sets benefits in [19]. RST algorithms are

included in a lot of proposed hybrid techniques in knowledge discovery area. RST is an effective technique when compared with other approaches [15]. RST discovered as very useful and effective technique for rule generations and attribute selection. Classification techniques and reduction algorithms presented with RST development in [19].intelligent models are proposed for diagnose diseases based on RST and its hybridization [21].

The aim of this study is to build a hybrid modlem2-fuzzy classifier (HFMC) for liver diseases by combining RS approaches which known as Modified Lem2 (MODLEM2) under Laplace measure and fuzzy set. The model is proposed to analyzes the liver data and generate rules automatically and employ it in inference based of fuzzy set .this model can aid in prediction of the Pathological risk stages of the patient.

The rest of the paper is described as follows: In Section 2 the basic concept of rough set theory is presented; Section 3 presents LEM2 algorithm; section 4 presents Modified Lem2 (MODLEM2); section 5 presents fuzzy set theory concept. Section 6 presents the proposed model HFMC the liver diseases diagnoses section 7 presents the experimental results of proposed model. Finally Section 8 shows the conclusion of this paper

2. Rough Set Theory Concepts

Rough set is defined by topological operations called approximations that has pair of basic concepts, named the lower and the upper approximations of sets [8] as shown in figure 1.

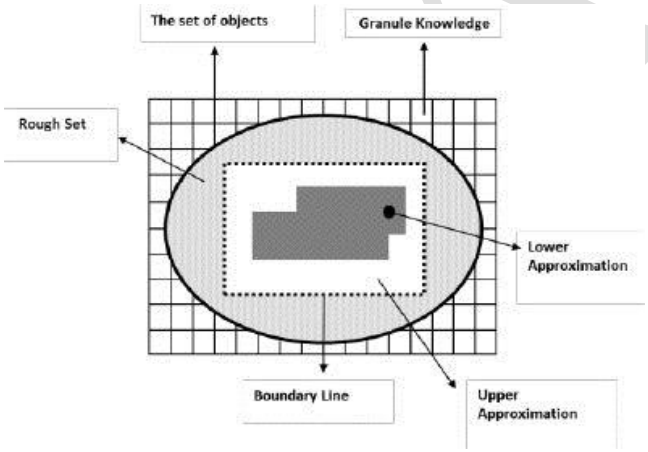

Figure 1. Rough set theory.

RST depend on equivalence relation R or indiscernibility relation Iq to define the knowledge from data table which is 4-tuple information system IS = (U,Q,V,f)

For each set of attributes $q \subset C$ an indiscernibility relationI_q is defined in the following way: two objects, x_i and x_j are indiscernible by the set of attributes q in C if $f(x_i,q) = f(x_j,q)$ for every$q \subset C$. I_q has equivalence class which called elementary set or knowledge granules in q because it shows the smallest discernible groups of objects. For any element x_i ofCl, the equivalence class of x_i in relation I_q is showed as $[x_i]_{I(q)}$.An indiscernibility relation formed as follows:

$$I_q = \{(x_i,x_j) \in U \times U: f(x_i,q) = (x_j,q), \text{for every } q \in Q \quad (1)$$

2.1. Lower and Upper Approximations

Lower approximation of Cl denoted by $\underline{P}(Cl)$) and the P-upper approximation of Cl denoted by$\overline{P}(Cl)$. As shown in figure 2 the Lower and upper approximation can be defined, respectively, as:

$$\underline{P}(Cl) = \{x \in U: I_q(x) \subseteq Cl\} \quad (2)$$

$$\overline{P}(Cl) = \{x \in U: I_q(x) \cap Cl = \emptyset\} \quad (3)$$

The elements of $\underline{P}(Cl)$ are all and only objects $x \in U$ which belong to the equivalence classes generated by the indiscernibility relation I_q contained in Cl.The elements of $\overline{P}(Cl)$ are all and only objects $x \in U$ that belong to those equivalence classes generated by the indiscernibility relation $I_p$containing at least one object x belonging to Cl.

The P-boundary of Cl, denoted by $B_{nP}(Cl)$, and is defined as:

$$B_{nP}(Cl) = \overline{P}(Cl) - \underline{P}(Cl) \quad (4)$$

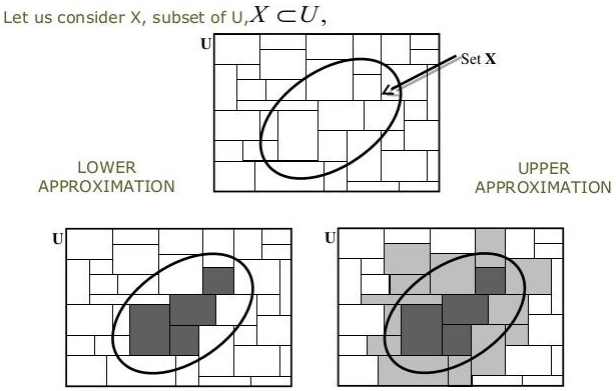

Figure 2. Upper , lower and boundery approximation.

The quality of approximation can be determined by lower and upper approximations [8,19]. The accuracy values range's is between (0, 1) and this show how the information derived from the original data.

$$\text{The approximation accuracy} = \frac{\text{no of object tbelongs to lower approximation}}{\text{no of object tbelongs to upper approximation}} \quad (5)$$

The quality of classification of decision class determined as:

$$\text{The classification accuracy} = \frac{\text{no of object that correctly calssified}}{\text{no of object in U}} \quad (6)$$

The approximation accuracy determined based on attributes set whether the set is a rough (accuracy ≠1) or a crisp [20].

2.2. Core and Reduct of Attributes

There are a subset of attributes which fully describes the knowledge in the dataset; those attribute set is called a reduct [19]. Reduct is a subset of attributes $RED \subseteq P$.The $core$ is the intersection of all reducts. Attributes from $core$ cannot be removed from the data sample without deteriorating the knowledge to be discovered.

2.3. Decision Rules generation by Rough Approximations

The rules are generated in form of "if... then...", when the condition part defines values defined by one or more condition attributes (C-elementary sets) and the decision part defines an assignment to one or more decision classes (D-elementary sets).

The generated rules extracted by different RST algorithm have three types Minimum set of rules (LEM2 algorithm) [4], Exhaustive rules (modlem2 algorithm) and satisfactory rules (satisfactory algorithm) [19].This discretization process is very important process to converting numerical attribute to nominal attribute [16]. Many of RST algorithms developed for generate decision rules

3. LEM2 Algorithm

LEM2 (Learning from Examples Module version-2) is used as modules in the algorithm LERS for learning from examples based on rough set.LEM2 computes a local covering followed by converting them to rule sets. The set of all cases labeled with the same decision value is called a concept. For attribute value pair $(q, v) = t$, denoted [t], is the set of all examples that for attribute q have value v. Let B be a concept and let T be a set of attribute-value pairs. Concept B depends on a set T if and only if

$$\emptyset \neq [T] = \bigcap_{t \in T}[t] \subseteq B \quad (7)$$

Set T is said to be a minimal complex of B if and only if B depends on T and T is minimal. Let τ be a non-empty collection of non-empty sets of attribute value pairs. We say that τ is a local covering of B if and only if each member T of τ is a minimal complex of B, $\bigcup_{T \in \tau}$ [T] = B, and τ is minimal, i.e. τ has the smallest possible number of elements [10, 22] .

For each concept B the LEM2 induced rules by computing local covering τ. Any set T, a minimal complex which is a member of τ,is computed from attribute-value pairs selected from the set T(𝔾) of attribute-value pairs relevant with a current goal 𝔾. i.e., pairs whose blocks have nonempty intersection with 𝔾. The initial goal 𝔾 is equal to the concept and then it is iteratively updated by subtracting from 𝔾 the set of examples described by the set of minimal complexes computed so far. Attribute-value pairs from T which are selected as the most relevant, i.e., on the basis of maximum of the cardinality of $[t] \cap \mathbb{G}$, if a tie occur, on the basis of the smallest cardinality of $[t]$. The last condition is equivalent to the maximal conditional probability of goal𝔾, given attribute-value pair t. LEM2 algorithms described as follow:

```
Procedure LEM2
Input: a set B;
Output: a single local coverting T of set B;
Begin
G := B ;
T := ø;
While G ≠ ø  do
        T := ø;
        T(G) := {t|[t] ∩ G ≠ ø};
        While T = ø or not ([T] ⊆ B) do
            select a pair t ∈T(G) with the highest attribute priorty, if a tie occur, select a pair t ∈T(G) such that |[t] ∩ G| is maximum; if another tie occurs, select a pair t ∈T(G) with the smallest cardinality of [t]; if a further tie occurs, select first pair;
            T:= T ∪ {t}
            G :=  [t] ∩ G;
            T(G) := {t|[t] ∩G≠ø}
            T(G) := T(G) – T;
        end{while};
        For each t in T do
            If [T-{t}] ⊆ then T := T –{t};
        T := T ∪ {T}
        G := B - Ur∈T [T];
end{while};
for each T ∈ T do
  if Us∈T-[T] [s] = B then T:= T –{T};
end{ procedure}
```

4. MODLEM2 Algorithm

MODLEM (Modified Learning from Examples Module) is an extension of Lem2 algorithm. MODLEM appeared despite it can generate rules from numerical and missing values of attribute but it requires preliminary preprocessing. In latest years MODEM2(Modified Learning from Examples Module,version2) not require preliminary preprocessing because it was combined with local approximations that use attribute-value pairs which known as MODLEM-Entropy and MODLEM-Laplace. MODLEM2 can treat with attribute's continuous values without need of preliminary discretization and extract extended minimal rules. MODLEM2 is able to generate rule in form of at most and at least instead the simple form of equality like LEM2 as follow:

Attribute ≤ value; attribute ≥ value or lower bound ≤ attribute ≤ upper bound.

Entropy measure generate rules covering only small number of examples [10, 11]. Theses rule's conditions constructed on cutting points that reduce the value of Entropy measure. The generated conditions have the syntax value ≤ attribute or attribute ≤ value.

Laplace measure which preferring higher values of Laplace accuracy to generate rules than entropy. In Laplace rule's conditions constructed on cutting points that reduce the value of Laplace measure. The generated conditions have the syntax value ≤ attribute or attribute ≤ value.

MODLEM2- entropy measure is better in generating rule strength than Laplace measure while Laplace method is better than entropy method in inducing the number of rules [11, 12]. MODLEM2 algorithms described as follow:

Procedure MODLEM

Input: a set B; a set C of attributes
Output: a single local coverting T of set B;
begin
G := B ;
T := ø;
While G ≠ ø do
T := ø; {candidate of condition part of the rule}
S:= U; { set of objects currently covered by T}
While T = ø or not ([T] ⊆ B) do
t := ø; {candidate for elementary condition}
t_eval := ∞; { evaluation measure for t}
for each attribute a ∈ C do
Find_best_condition (a, S, new_t, eval_new_t); (look for best condition for a)
If eval_new_t < t_eval **then** (evaluate if new condition is better than previous)
t := new_t;
t_eval := new_eval;
end;
end {for};
T := T ∪{t};
S := S ∩ {t};
end{while not ([T] ⊆ B };
For each elementary condition t ∈ T **do**
If [T-{t}] ⊆ B then T := T –{t};
T := T ∪ {T}
G := B - $U_{r\in T}$ [T];
end{while C≠ϕ};
for each T ∈ T **do**
if $U_{s\in T-[T]}$ [s] = B then T:= T –{T};
end{ procedure}

Procedure Find_best_condition
input attribute a; set S of objects;
output current best condition best_t
Evaluation measure eval_best_t of the best condition);
begin
best_t := ϕ;
eval_best_t : = ∞;
sort H; (create a list of values for attribute a and objects from S sorted according to increasing values; H(i) is the ith value in the list)
for i:= 1 to length (H) -1 **do**
v := (H(i) +H(i+1))/2; {consider only mid-points between values
characterizing examples
belonging to
different classes}
S_1:= { x∈ S | a(x) < v}; { a(x) is the value of attribute a for object x}
S_2 : = {x ∈S | a(x) ≥ v};
eval_t := ($|S_1|/|S_1 \cup S_2|$) *Ent(S_1) + ($|S_2|/|S_1 \cup S_2|$) * Ent(S_2);
(Ent entropy calculated for S_j)
If eval_t < eval_best **then**
If $|C \cap S_2| \geq |C \cap S_1|$ **then**
Best_t := (c ≥ v) **else** best_t := (c < v);
Eval_best_t := eval_t;
end if
end
end{ procedure}

5. Fuzzy Set Theory Concept

The theory of fuzzy has advanced in different ways and in many applications [17]. It introduced by Zadehand Dieter Klaua that supports many concepts. Fs provide an important contribution in data mining to provide fuzzy data mining [14]. Fs are sets whose elements have membership degree. It is one on most important approach that can handle uncertainty, vagueness of any dataset. Fuzzy set theory supports concept of membership function for each element in set; the value of membership function interval [0, 1].
Fuzzy inference system is one of most approaches deal with uncertainty and vagueness which uses concept of FSs for diagrammatic representation and mapping from a given input to a driven output. FIS have two types of FIS Mamdani and Sugeno. Architecture of FIS is showed in figure 3. FS theory consists of:

- (Fuzzifcation process) input/output membership functions which refer to how to define the fuzzy set.
- Inference Base rules, in which generate rules Based on fuzzy logic/FSs operators.
- Defuzzifcation and output.

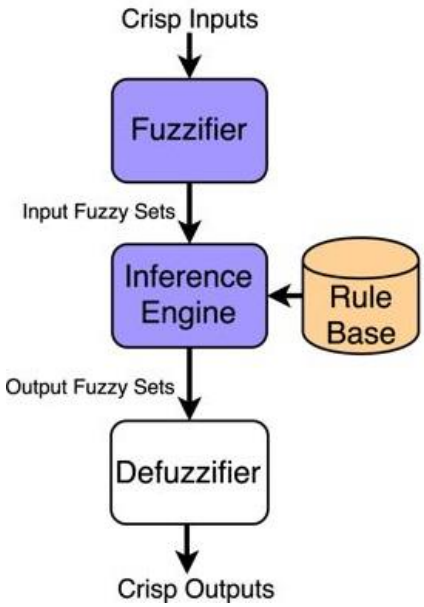

Figure 3. Fuzzy Set Theory Architecture.

6. Hybrid Modlem2-Fuzzy Classifier for Liver Diseases Diagnoses Model (HMFC)

The aim of this study diagnose liver diseases based on hybrid Modlem -Fuzzy classifier (HFMC) as shown in figure 4 .this proposed model consist of two phases . First phase is rules generation using Lapalce-Modelem2 algorithm then compare result with different RST approaches (LEM2 and Entropy-Modlem2) .in the second phase FIS used as classifier to diagnose liver diseases.

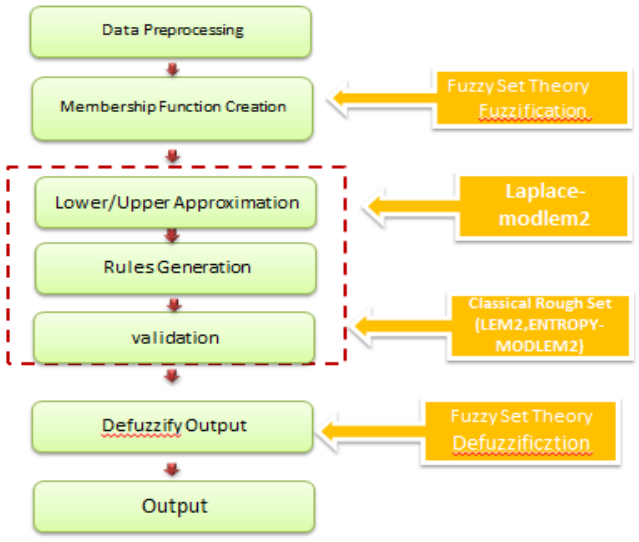

Figure 4. Hybrid Modlem2-Fuzzy Classifier model.

6.1. A Case Study of Liver Diseases Diagnoses

The dataset was collected from medical data warehouse of liver disease. Dataset consist of 466 cases, it consists of five measured variables as follow:

- Total Bilirubin: in the total bilirubin there are three values refers to low, medium and high.
- Serum Albumin: the Serum albumin is one of most vital factor that effect on liver which has three values that refer to high, medium and low.
- Prothrombin Time (PT): is a blood test that measures the time it takes for the blood plasma to coagulate. It has three values which refer to high, medium and low.
- Ascites: It usually occurs when the liver stops working correctly; this lead to pain and swelling in the abdomen, And Queasiness. It has three values hat refer to high, Medium and low.
- Hepatic Encephalopathy: it decrease function of brain that come as a result of highly liver disease, it has three values refer to high, medium and low.

Result is the decision attribute, it refer to levels of liver risk that has values (A, B, C) refer to low risk, moderate risk, high risk

6.1.1. Data Preprocessing

Data are processed before used in our model based on The Child-Pugh score which consists of five clinical features that used to assess the prognosis of chronic liver disease and cirrhosis. Table.1 presents the Child-Pugh score for liver diseases.

Table 1. Child-Pugh score for liver diseases.

Factor	1 point	2 point	3 point
Total bilirubin (µmol/L)	<34	34-50	>50
Serum albumin (g/L)	>35	28-35	<28
PT INR	<1.7	1.71-2.30	>2.30
Ascites	None	Mild	Moderate to Severe
Hepatic encephalopathy	None	Grade I-II (or suppressed with medication)	Grade III-IV (or refractory)

The result is the output which has A, B, C classes based on Child-Pugh score, class A refers to total point (5-6), class B refers to total point (7-9) and finally class C refers to total point (10-15). Figure.5 shows liver diseases attributes with its nominal values after preprocessing and scoring step. decision attribute is result, risk values(A, B, C) refer to different liver diseases risk levels which mean low risk, moderate risk, high risk.

```
**ATTRIBUTES

TotalBiliurbin: [1, 2, 3]
SerumAlbmin: [1, 2, 3]
ProthrombinTime: [1, 2, 3]
Ascites: [1, 2, 3]
HepaticEncehalopathy: [1, 2, 3]
Result: [A, B, C]

decision: Result
```

Figure 5. Liver diseases attributes with its nominal values.

6.1.2. First Phase

In this phase we generate automated rules using different rough set algorithms. the rules generated by MODLEM2 algorithm with its two measures(Entropy& Laplace) then use LEM2 algorithm as a comparison method to ensure efficiency of generated rules. But to generate perfect minimal rules we must ensure accuracy of approximation.

Accuracy of Approximation. The quality of approximation for (A, B, C) decision classes is presented in table.2 as below.

Table 2. Accuracy of approximation for the three classes.

class	No. object	Lower approximation	Upper approximation	accuracy
A	12	10	14	0.71
B	76	74	74	0.94
C	378	378	378	1
Total Quality Of Approximation :			**99.1%**	

The approximation results indicate perfect accuracy for the three classes. Class "A" is the "low risk" class. There are 12 objects belonging to that class while upper approximation has 10 object and lower approximation include 14 object, so the approximation accuracy for "A" is 0.71. Class "B" means the "moderate risk" class, which accuracy reaches to 0.94. The accuracy of "c" class is one that refers to "high risk".

Rule Generation. We generates rules using MODLEM2. Based on laplace and entropy measures to extract the "extended minimum cover rules" then use LEM2 algorithm to generate minimal covering rules for comparing accuracy result.

6.1.2.1. Rules Generated by MODLEM2 Entropy measure.

The total number of rules generated by entropy measures is 23 rules. Sample of these rules are shown in table.3

Table 3. Rules generated by Entropy-MODLEM2

Id	Condition	Decision
1	(SerumAlbmin in {1,2}& ProthrombinTime = 1 & Ascites = 2 & (HepaticEncehalopathy in{1,2}	Result= A
2	SerumAlbmin = 1 & ProthrombinTime = 1& HepaticEncehalopathy = 2	Result= A
3	(TotalBiliurbin = 2) & (SerumAlbmin = 1) & (HepaticEncehalopathy = 1)	Result= A
4	TotalBiliurbin = 1) & (SerumAlbmin = 2) & (Ascites = 1) & (HepaticEncehalopathy = 1)	Result= A
5	(TotalBiliurbin = 1) & (SerumAlbmin = 1) & (ProthrombinTime = 2) & (Ascites = 1)	Result= A
6	)TotalBiliurbin = 1) & (ProthrombinTime = 3) & (Ascites = 1(	Result= B
7	)TotalBiliurbin = 1) & (ProthrombinTime = 3) & (HepaticEncehalopathy = 1(	Result= B
8	(TotalBiliurbin = 1) & (SerumAlbmin = 3) & Ascites = 1	Result= B
9	(TotalBiliurbin = 2) & (SerumAlbmin in {2,3} & Ascites in {1,2} & HepaticEncehalopathy = 1	Result= B

The accuracy of classification of entropy-modlem2 is validated with k-fold cross validation technique in 22 second with classification accuracy 98.27% as shown in figure 6.

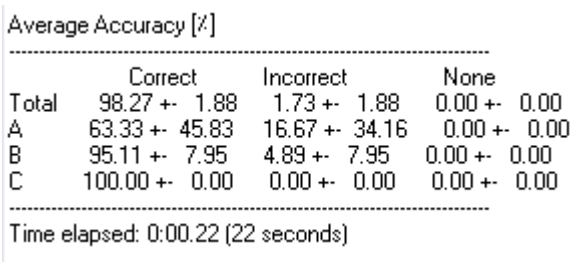

Figure 6. Entropy-DOMLEM classification accurcy.

6.1.2.2. Rules Generated by MODLEM2 Laplace measure.

The total number of rules generated by Laplace measures is 28 rules .sample of these rules shown in table. 4.

Table 4. Rules generated by Laplace -MODLEM2.

id	Condition	decision
1	)TotalBiliurbin = 2) & (SerumAlbmin = 1) & (HepaticEncehalopathy = 1(	Result= A
2	(TotalBiliurbin = 1) & (SerumAlbmin = 2) & (Ascites = 1) & (HepaticEncehalopathy = 1(	Result= A
3	(TotalBiliurbin = 1) & (ProthrombinTime = 1) & (HepaticEncehalopathy = 2)	Result= A
4	(TotalBiliurbin = 1) & (SerumAlbmin = 1) & (ProthrombinTime = 1) & (Ascites = 2)	Result= A
5	(TotalBiliurbin = 1) & (SerumAlbmin = 1) & (ProthrombinTime = 2) & (Ascites = 1)	Result= A
6	(TotalBiliurbin = 1) & (SerumAlbmin = 1) & (ProthrombinTime=2 & (HepaticEncehalopathy = 2)	Result= B
7	)TotalBiliurbin = 1) & (ProthrombinTime = 1) & (Ascites = 3(	Result= B
8	)ProthrombinTime = 1) & (Ascites = 1) & (HepaticEncehalopathy = 3(	Result= B
9	)TotalBiliurbin = 3) & (SerumAlbmin = 1) & (Ascites = 1(	Result= B
10	)TotalBiliurbin = 1) & (SerumAlbmin = 2) & (HepaticEncehalopathy = 2(	Result= B

The accuracy of classification of modlem2-laplace is validated with k-fold cross validation technique with 99.1% and elapsed time 21 second as shown in figure 7.

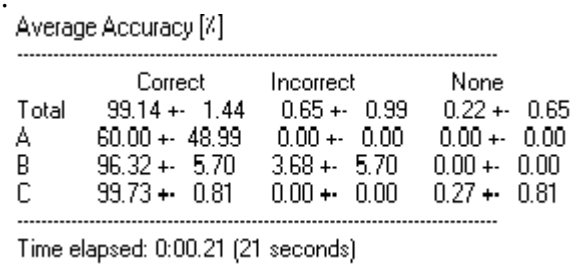

Figure 7. Laplace -DOMLEM clssification accuracy.

6.1.2.3. Rules Generated by LEM2

Rules generated with this algorithm are minimal rules. This means the set of rules does not have any redundant rules, and these rules are certain, such that there are a total of 31 rules generated from the data. The following table.5 shows sample of the minimum cover rules obtained (about 10 rules).

Table 5. Rules generated by LEM2.

Id	Condition	Decision
1	(TotalBiliurbin = 1) & (ProthrombinTime = 1) & (HepaticEncehalopathy = 2)	Result= A
2	(TotalBiliurbin = 1) & (SerumAlbmin = 2) & (ProthrombinTime = 1) & (HepaticEncehalopathy = 1)	Result= A
3	(TotalBiliurbin = 1) & (ProthrombinTime = 2) & (Ascites = 1) & (HepaticEncehalopathy = 1)	Result= A
4	(TotalBiliurbin = 1) & (ProthrombinTime = 1) & (Ascites = 2) & (HepaticEncehalopathy = 1)	Result= A
5	(TotalBiliurbin = 2) & (SerumAlbmin = 1) & (HepaticEncehalopathy = 1)	Result= A
6	(TotalBiliurbin = 2) & (ProthrombinTime = 2) & (Ascites = 1) & (HepaticEncehalopathy = 1(	Result= B
7	(TotalBiliurbin = 3) & (SerumAlbmin = 1) & (Ascites = 1(	Result= B
8	(TotalBiliurbin = 1) & (ProthrombinTime = 3) & (Ascites = 1)	Result= B
9	(TotalBiliurbin = 1) & (ProthrombinTime = 1) & (HepaticEncehalopathy = 3)	Result= B
10	(TotalBiliurbin = 2) & (Ascites = 2) & (HepaticEncehalopathy = 1)	Result= B

The classification accuracy of rules generated using LEM2 algorithm is 98.7%. Figure 8 show the result of k-fold cross validation technique for rules generated that validates result in 16 second.

Average Accuracy [%]

	Correct	Incorrect	None
Total	98.71 +- 1.72	1.29 +- 1.72	0.00 +- 0.00
A	60.00 +- 48.99	10.00 +- 30.00	0.00 +- 0.00
B	98.06 +- 3.94	1.94 +- 3.94	0.00 +- 0.00
C	99.46 +- 1.62	0.54 +- 1.62	0.00 +- 0.00

Time elapsed: 0:00.16 (16 seconds)

Figure 8. LEM2 result of cross-fold validation technique.

6.1.2.4. Comparison Between Proposed Algorithms

Finally as comparison between the three algorithms LEM2, MODLEM2 with ENTROPY and MODLEM2 with LAPLACE, we find that the second algorithm (MODLEM2) is the better algorithm in number of rule generated and accuracy of classification that needed for diagnose of liver diseases.we find that entropy–modlem2 is the best in number of rules while the lplace-modlem is the best in classification accuracy .so we will use laplace –modlem2 in our proposed HMFC based on its perfect accuracy.

6.1.3. Second Phase

In this phase we develop FIS for diagnose or classify Pathological stages of liver diseases through using the rules generated in first phase from Laplace -modlem2 algorithm in FIS as inference rule based without need to external expert of physicians.We develop the fuzzy expert system by defining Membership functions, fuzzy rule base, fuzzifcation and defuzzifaction.

We firstly, construct the membership functions for five input variables and one output variable. Then rules generated from the first phase are employed as inference rule base. In the last step calssification result are generated in form of defuzzification

6.1.3.1. Input Variables

We will present five input variable with their fuzzy set ranges and their membership function for all of these sets in FIS. First input variable is Total Bilirubin which its Fuzzy set ranges are shown in Table 6.While Membership functions are trapezoidal and triangular for these fuzzy sets as shown in figure 9.

Table 6. Total bilirubin's Fuzzy sets.

Input variable	range	Fuzzy set
Total Bilirubin	< 34(<2)	1
	34-50 (2-3)	2
	>50 (>3)	3

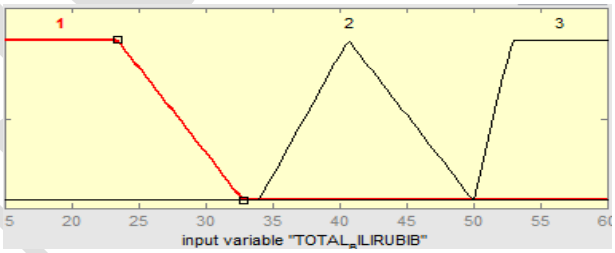

Figure 9. Mmbership function of Total Bilirubin.

Second input variable is Serum Albumin which its Fuzzy set ranges are shown in table 7.While the membership functions are trapezoidal and triangular as shown in figure 10.

Table 7. Serum Albumin Fuzzy sets.

Input variable	range	Fuzzy set
Serum Albumin	>3.5	1
	2.8-3.5	2
	<2.8	3

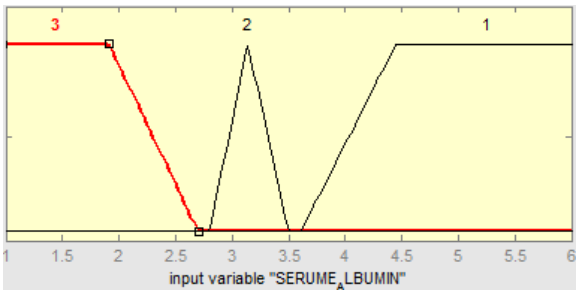

Figure 10. Membership function of Serum Albumin.

Third input variable is Prothrombin Time which its Fuzzy set ranges are shown in table 8. Membership functions are trapezoidal and triangular for fuzzy sets as shown in figure 11.

Table 8. Prothrombin Time Fuzzy sets.

Input variable	range	Fuzzy set
Prothrombin Time	<4.0	1
	4.0-6.0	2
	> 6	3

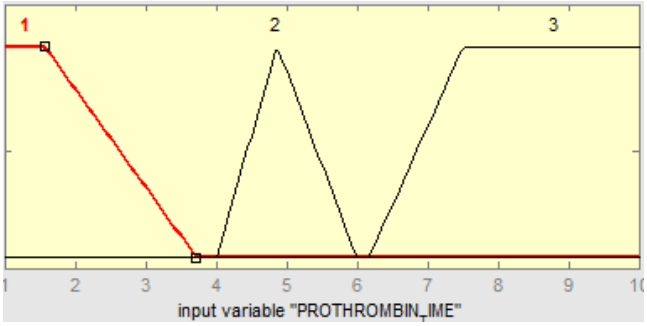

Figure 11. Membership function of Prothrombin Time.

Fourth input variable is Ascites which its Fuzzy set ranges are shown in table 9.while Membership functions for fuzzy sets are trapezoidal and triangular as shown in figure 12.

Table 9. Ascites Fuzzy sets.

Input variable	range	Fuzzy set
Ascites	none (<2)	1
	mild (2.0-3.0)	2

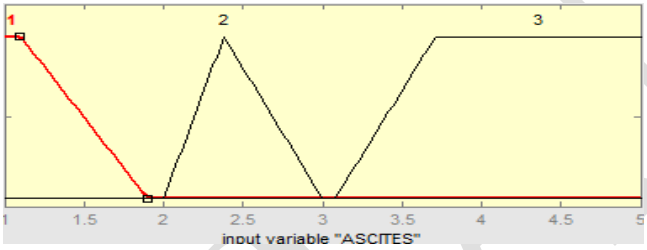

Figure 12. Membership function of Ascites.

Fifth input variable is Hepatic Encephalopathy which its Fuzzy set ranges are shown in table 10. Membership functions are trapezoidal and triangular for fuzzy sets as shown in figure 13.

Table 10. Hepatic Encephalopathy Fuzzy sets.

Input variable	range	Fuzzy set
Hepatic Encephalopathy	None	1
	Grade I-II (2-3)	2
	Grade III IV(>3)	3

6.1.3.2. Output Variable

Our model proposed to identify pathologies risk stages of liver. The fuzzy set ranges of output in range [0, 15] as shown in table 11.while the Membership functions of these fuzzy sets are triangular and are shown in figure 14.Output's Fuzzy set A, B, C refers to Low risk, moderate risk and High risk.

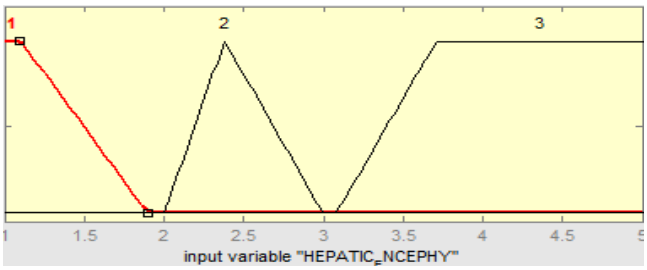

Figure 13. Membership function of Hepatic Encephalopathy.

Table 11. Result Fuzzy sets.

Output variable	range	Fuzzy set
Result	5-6	A
	(7-9)	B
	(10-15)	C

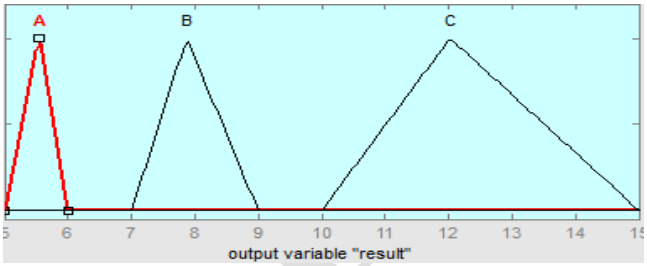

Figure 14. Membership function of Result.

6.1.3.3. Fuzzy Rule Base

The rules which generated before through first phase with RST using laplace-modlem2 algorithm are entered to FIS system as fuzzy rule base. The rule base consists of 28 well defined extended minimal rules that determine the risk stages. Sample of these rules in FIS is shown in Figure 15.

1. If (TOTAL_BILIRUBIB is 1) and (PROTHROMBIN_TIME is 1) and (HEPATIC_ENCEPHY is 1) then (result is A) (1)
2. If (TOTAL_BILIRUBIB is 1) and (SERUME_ALBUMIN is 2) and (PROTHROMBIN_TIME is 1) and (HEPATIC_ENCEPHY is 1) then (result is A) (1)
3. If (TOTAL_BILIRUBIB is 1) and (PROTHROMBIN_TIME is 2) and (ASCITES is 1) and (HEPATIC_ENCEPHY is 1) then (result is A) (1)
4. If (TOTAL_BILIRUBIB is 1) and (PROTHROMBIN_TIME is 1) and (ASCITES is 2) and (HEPATIC_ENCEPHY is 1) then (result is A) (1)
5. If (TOTAL_BILIRUBIB is 2) and (SERUME_ALBUMIN is 1) and (HEPATIC_ENCEPHY is 1) then (result is A) (1)
6. If (TOTAL_BILIRUBIB is 2) and (PROTHROMBIN_TIME is 2) and (ASCITES is 1) and (HEPATIC_ENCEPHY is 1) then (result is B) (1)
7. If (TOTAL_BILIRUBIB is 3) and (SERUME_ALBUMIN is 1) and (ASCITES is 1) then (result is B) (1)
8. If (TOTAL_BILIRUBIB is 1) and (PROTHROMBIN_TIME is 3) and (ASCITES is 1) then (result is B) (1)
9. If (TOTAL_BILIRUBIB is 1) and (PROTHROMBIN_TIME is 1) and (HEPATIC_ENCEPHY is 3) then (result is B) (1)
10. If (TOTAL_BILIRUBIB is 2) and (ASCITES is 2) and (HEPATIC_ENCEPHY is 1) then (result is B) (1)

Figure 15. :fuzzy rules that generated by entropy- modlem2.

6.1.3.4. Fuzzifcation and Defuzzifacation

Inference mechanism of our system relies on MAMDANI model. The FIS result of rules and surface viewer is shown in figure 16,17 and 18.figure.16 represents that the liver risk with its values while total bilirubin is 40 , Serum Albumin is 3.5, Prothrombin Time is 7.95 , ascites is 2.57 and Hepatic Encephalopathy is 2.51 as tested data the result will be 12.3 .figure 17 and 18 represent the relation between different input variable and result output and how these symptoms effect on result of liver risks.

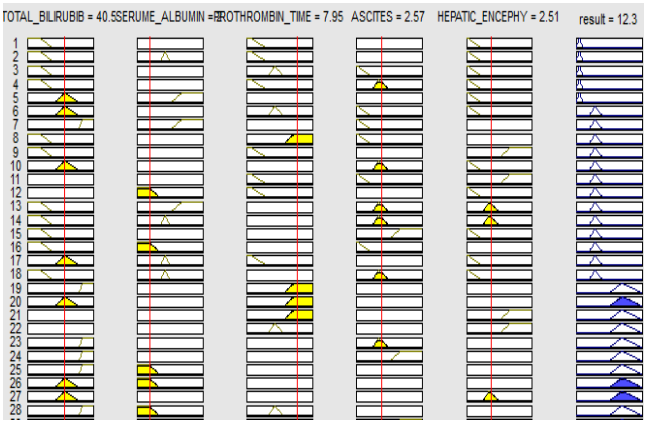

Figure 16. Value of rules viewer.

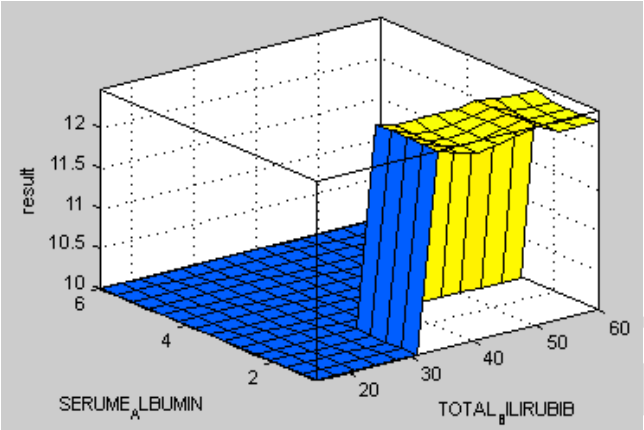

Figure 17. Relation between Serum Album, total bilirubin and result.

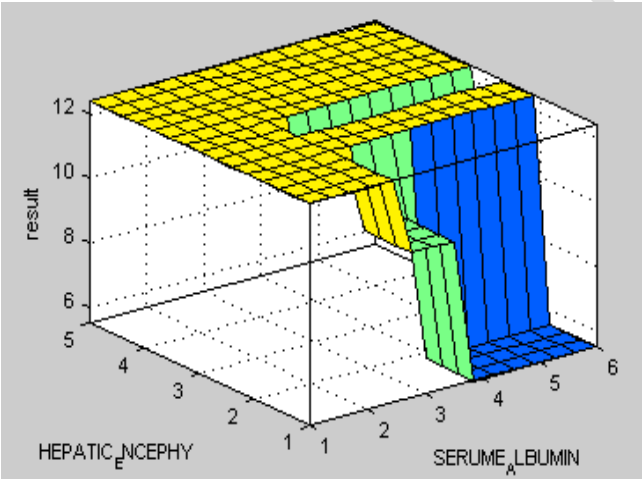

Figure 18. relation between Hepatic Encephalopathy, Serum Albumin and result.

7. Conclusion

This study builds robust expert system for diagnoses diseases of liver based on hybrid model Modlem2-fuzzy set HRFC. Since using RST or FS alone cannot provide accurate result and require more knowledge and experience. The proposed model improves fuzzy set while determining the content of the rules is usually achieved by knowledge acquisition with an expert which, for a large Application can be very time consuming, as well as not automated. HMFC has two stages, in first stage the laplace-modlem2 used mainly to automatic generation of extended minimal cover rules. In the second stage, construct fuzzy inference system for liver diseases through using rules generated by RST in first stage for diagnose or classify Pathological stages (risks) of the patient in liver diseases diagnose system .The experimental results presented in this study shows that the proposed HMFC model is better accuracy than using rough sets or fuzzy systems separately. When rules that differentiate between pathology stages of liver diseases are generated with 99.14% as classification accuracy this mean that combining rough set and fuzzy system approaches enhance classification accuracy.

References

[1] Agache L ., "Endotypes of allergic diseases and asthma: An important step in building blocks for the future of precision medicine " , allergology international journal, vol. 65, no. 3, pp. 243–252 , 2016.

[2] Bashir S., Qamar U. , Khan F.H., " IntelliHealth: a medical decision support application using a novel weighted multi-layer classifier ensemble framework" . *Journal of biomedical informatics*, Vol. *59*, pp.185-200. ,Feb 2016

[3] Castro D., "Meeting national and international goals for Improving health care: The role of information technology in Medical research ", in Science and Innovation Policy, 2009 Atlanta Conference on Science, Technology and Innovation Policy, pp. 1-9, Atlanta , GA , 2-3 October , 2009.

[4] Greco S, Benedetto M, Roman S. "Decision rule approach". In Multiple criteria decision analysis, pp. 497-552. Springer, New York, NY, 2016.

[5] Gulia A., Vohra R., P. Rani, " Liver patient classi_cation using intelligent techniques", International Journal of Computer Science and Information Technologies, vol. 5, no. 4, pp. 5110- 5115, 2014.

[6] Gulia N., Vohra R., Rani P., "Liver patient classification using intelligent techniques", International Journal of Computer Science and Information Technologies, vol. 5, no. 4, pp. 5110-5115, 2014.

[7] Hettiarachchi C, Do H, Choi B., "Risk-based test case prioritization using a fuzzy expert system". Information and Software Technology. Vol. 69, PP.1-5, Jan 2016.

[8] Hu JH, Shang JL, Zhou KP, Chen YK, Ning YL, Liu L, Aliyu MM. "Hazard degree identification of goafs based on scale effect of structure by RS-TOPSIS method" . Journal of Central South University , Vol. 22 , No. 2 , PP. 684-92. 2015 .

[9] Ibraheem S.A.A., Elkader H.M.A., Selim I. , Hussein R, . " Hybrid Rough-Fuzzy Classifier for Liver Disease Diagnosis" . International Journal of Electronics and Information Engineering, Vol. 4 , No. 2, pp.60-70. Jan 2016

[10] Jerzy W. , Grzymala-busse, "a local version of the MLEM2 algorithm for rule induction" , Fundamenta Informaticae journal , Vol 100 Issue 1-4, Pages 99-116 , January 2010.

[11] Jerzy W. , Grzymala-Busse, " A comparison of three strategies to rule induction from data with numerical attributes" , the International Workshop on Rough Sets in Knowledge Discovery, associated with the European Joint Conferences on Theory and Practice of Software , pp. 132–140 , April 5–13, 2003 Warsaw,Poland,

[12] Jerzy W., Grzymala-Busse, Stefanowski J. , "Three discretization methods for rule induction". International Journal of Intelligence Systems , Vol. 16 , PP. 29–38., 2001.

[13] Kumar S.S., Inbarani H.H , . "Cardiac arrhythmia classification using multi-granulation rough set approaches". International Journal of Machine Learning and Cybernetics, Vol. 9, No. 4, pp.651-666. 2018

[14] Marsala C., Bouchon-Meunier B., "Fuzzy data mining and Management of interpretable and subjective information" , Fuzzy Sets and Systems, vol. 281, pp. 252-259, 2015

[15] Mohamad M., Selamat A., Krejcar O., Kuca K., "A recent Study on the rough set theory in multi-criteria decision analysis Problems" , 7th International Conference(ICCCI 2015), PP. 21–23 , Madrid, Spain, September, 2015.

[16] Mohamad M., Selamat A., "Recent Study on the Application Of Hybrid Rough Set and Soft Set Theories in Decision Analysis Process", the 29th International Conference on Industrial Engineering and Other Applications Of Applied Intelligent Systems, IEA/AIE, pp 713-724, Morioka, Japan, August 2-4, 2016.

[17] Popli L. , Singh B. P., "Fuzzy sets and artifcial ntelligence: A survey" , International Journal of Advanced Research in Computer Science and Software Engineering, vol. 4, no. 10, pp. 308-313, 2014.

[18] Samuel O.W., Asogbon G.M., Sangaiah A.K., Fang P. , Li G. "An integrated decision support system based on ANN and Fuzzy_AHP for heart failure risk prediction". Expert Systems with Applications, Vol. 68, pp.163-172 ,February 2017.

[19] Sudha M., Kumaravel A., "Quality of classification with LERS system in the data size context".Applied Computing And Informatics, Volume 14, Issue 1, pp 1-6 January 2018.

[20] Sudha M., Kumaravel A., "Comparative Analysis between Rough set theory and Data mining algorithms on their Prediction" , Global Journal of Pure and Applied Mathematics , Vol. 13, No. 7, pp.3249–3260, 2017.

[21] Wiktorowicz K., "Design of state feedback adaptive fuzzy controllers for second-order systems using a frequency stability criterion". IEEE Transactions on Fuzzy Systems, Vol. 25, No. 3, pp.499-510, 2017.

[22] Zhang Q., Xie Q., Wang G. , "A survey on rough set theory And its applications", CAAI Transactions on Intelligence Technology , Vol. 1, Issue. 4, PP 323-333 , October 2016.

Diaa Salama Abdul-Minaam was born on November 23, 1982, in KafrSakr, Sharkia, Egypt. He received the B.S from Faculty of Computers &Informatics, Zagazig University, Egypt in 2004 with grade very good with honor, and obtains the master degree in information system from the faculty of computers and information, menufia university, Egypt in 2009 specializing in Cryptography and network security. He obtained his Ph.D. degree in information system from the faculty of computers and information, menufia university, Egypt in 2015. He is currently a Assistance Professor in Information systems department, Faculty of Computers and Information, Benha University, Egypt since 2011.

Walaa H. El-Ashmawi is an Assistant Professor at Suez Canal University, Faculty of Computers and Informatics, Computer Science Department. Egypt. She received her B.Sc. in 2001 and M.Sc. in 2008 from Egypt. She got her PhD. Degree in 2013 from Computer Science Department, College of Information Science and Engineering, Hunan University, China.

Shimaa Abd Allah Ibraheem was born on November 12, 1981 in benha, kaluobya, Egypt. I received the B.S from Faculty of Computers and Information, Zagazig University, Egypt in 2003 with grade very good, and master degree from faculty of computer and Information, Monofya University, in 2016. She submitted for PHD degree in 2018.she was working in higher technology institute Egypt as teaching assistance at computer science department. She specializing in uncertainty and knowledge discovery approaches. Shaimaa has published papers under artificial intelligence area in international journals, international conferences, local journals and local conferences.