# *Using Decision Tree for Selecting Volunteers for Biomedical Study Based on Laboratory Report*

Ibrahim Dweib[a], Munther Melhim[b]
Department of Computer science
[b]Sultan Qaboos University, Muscat - Al Khoudh, PC 123, Oman
dweib@squ.edu.om
[b]Jordan Center for Pharmaceutical Research
P.O. Box 950435 Amman 11195 Jordan
m.melhim@jcpr-jo.com

## ABSTRACT

Bioequivalence studies are used to assess whether the test drug is equivalent to the reference drug. In those studies, they compare between both the reference and test drugs by investigating the effect of both on a sample of qualified human being. In this work, a Knowledge-Based System was established based on the human experts in order to distinguish between qualified and nonqualified volunteers whom are able to participate in the study. Therefore, we considered many methods for representing expertise in a knowledge-based system to simulate the human behavior. Among these, the Rule Base Method is the most appropriate one.

## KEYWORDS

Decision tree, Rule based, Bioequivalence studies, Knowledge-Based.

## 1. INTRODUCTION

Bioequivalence studies are used to assess whether the test drug is equivalent to the reference drug. The reference drug is the drug produced by the originator company, where full chemical, physical and biological investigation was established. Test drug is the drug produced by manufacturer other than the originator. This manufacturer shall establish physical and chemical investigation by meeting pharmacopoeia requirements. However the biological equivalence between the test and the reference medication is only established by performing bioequivalence studies.

Jordan Center for Pharmaceutical Research (JCPR) is specialized research center in conducting bioequivalence studies. Usually not less than 24 volunteer are used to perform such biostudies where in the first period 50 % of those volunteers designed to take test medication and the other 50% will take reference medication. During in the second period, volunteers whom took test will take reference medication and who took reference will take test medication.

Blood samples will be taken from volunteers on time schedule previously specified depending on the pharmacokinetic data of the study medication.

Volunteers participated in the study shall pass certain criteria (inclusion criteria) and shall not meet any of the exclusion criteria as described in the Study Case Report Form (CRF). One of the initial and primary inclusion criteria is that the volunteer shall be healthy as judged by the clinical investigation and laboratory investigation.

CRF is designed to describe every detail regarding volunteer participation in the study. It contains 19 parameters under the heading Laboratory Investigation covering Chemical and Hematology aspects. Each parameter is constrained by a minimum and a maximum limit. If the result is within the range, the clinical relevance is 1, which is normal and acceptable. Otherwise it is 3 which is un-acceptable. Some of these parameters are related to each other, that's mean, some of the abnormal parameters within ± tolerance

values are acceptable depending on some other normal parameters, which needs for doctor's expert decision.

Artificial Intelligence (AI) is a research and study for finding the relationship between cognitive science and computation theory. In order to represent it as either data structures or problem solving methodology, search technique or knowledge representation forms [1].

Knowledge and know-how of humans that have been accumulated for many years comprise both general and technical knowledge. It includes group of knowledge that can be re-arranged and utilized according to algorithmic-procedures, and a type of knowledge that can not be formulated as algorithms. In decision-making the treatment of knowledge that can hardly be formulated as algorithms become quite important. AI has become popular in late 1960's. It deals with the above mentioned type of knowledge and its application was extended to many fields [2].

Knowledge engineers and AI professionals try to build expert systems in different areas to overcome the shortages in human experts in some fields and reduce the cost of consultants of experts in other fields.

Early studies in intelligent system such as Mycin, Casnet, PIP and Internist-I have shown to out perform manual practice of diagnosis in several diseases domain. Neural Network for classification of fuzzy pattern of HIV and Aids is using unsupervised learning. It simulates the function of human brain to perform tasks as human does [w.1].

In our study, we have four choices to solve the problem using AI techniques. In section 2, the four techniques, Neural Networking, Fuzzy Logic, Clustering and Rule-Based will be discussed. In section 3 the methodology, rule based, will be discussed. Implementation for the system is shown in section 4 and the conclusions and recommendations of the study will be discussed in the last section, references will be after that.

### 1.1 Mission:

Building a Knowledge-Based System that is able to distinguish between qualified and non qualified volunteers to participate in the study. This system will be established based on the human knowledge in this field. Therefore, we considered many methods for representing this expertise in a knowledge-base system to simulate the human behavior. Also, this system will ensure that only qualified volunteer whom passing certain inclusion criteria and laboratory tests will be accepted to participate in the study.

### 1.2 Objectives:

a. Reduce the cost and expenses of studies.
b. Reduce the time and effort consumed due to huge paper work needed for each study.
c. Standardize the work in order to decrease human interference to the lowest acceptable limit.
d. Save the physician time to be used for more complicated situations.

## 2. PROBLEM SOLVING TECHNIQUES

The main features in medical diagnoses and prediction using AI technique will make the consultation to be more interactive. As clinical decision making inherently requires reasoning under uncertainty, expert systems and fuzzy logic will be suitable for dealing with partial evidence and with uncertainty regarding the effects of proposed interventions [w.1].

### 2.1 Neural Networking:

A neural network is a powerful data-modeling tool that is able to capture and represent complex input / output relationships. The motivation for the development of neural network technology stemmed from the desire to develop an artificial system that could perform "intelligent" tasks similar to those performed by the human brain. Neural networks resemble the human brain in the following two ways:

a. A neural network acquires knowledge through learning.
b. A neural network's knowledge is stored within inter-neuron connection strengths known as synaptic weights.

The true power and advantage of neural networks lies in their ability to represent both linear and non-linear relationships and in their ability to learn these relationships directly from the data being modeled. Traditional linear models are simply inadequate when it comes to modeling data that contains non-linear characteristics.[w.2]

The appropriate choice of the type of neural network (supervised, unsupervised, or reinforced) depends on availability of data. Supervised learning requires pairs of data consisting of input patterns and the correct outputs, which are sometimes difficult to obtain. Unsupervised training classifies input patterns internally and does not exceed expected results. The data requirements for unsupervised training are thus much easier and less costly to meet, but the capability of the network is significantly less than for supervised learning. A compromise between supervised and unsupervised training is reinforcement learning, which requires input and only a grade or reward signal as the desired output [3].

**2.2 Fuzzy Logic:**

Fuzzy set theory implements classes or grouping of data with boundaries that are not sharply defined. Fuzzy logic is ideal for controlling nonlinear systems and for modeling complex systems where an inexact model exists or where ambiguity or vagueness is common. A typical fuzzy system consists of a rule base, membership functions and an inference procedure.

The construction of a fuzzy model includes both the structure and the parameter identifications. In the structure identification, not only the system variables need to be chosen but also proper partitions of both input and output spaces are required. After identifying the structure of a fuzzy model, the parameter identification follows in order to adjust the membership functions and the parameters used in the fuzzy rules such that a more satisfactory performance can be obtained [4].

**2.2.1 Fuzzy classification**: It is one of the important applications of fuzzy logic. In a fuzzy classification system, a case can properly be classified by applying a set of fuzzy rules based on the linguistic terms of its attributes. Fuzzy classification systems are capable of handling perceptual uncertainties such as the vagueness and ambiguity involved in classification problems. The most important task to accomplish a fuzzy system is to find a set of fuzzy rules suitable for a specific classification problem [5].

**2.2.2 Fuzzy modeling and identification:** Fuzzy logic models can be developing for expert knowledge or from process (input-output) data. In the first case, fuzzy models can be extracted from expert knowledge of the process. The expert knowledge can be expressed in terms of linguistics, which is sometimes faulty and requires the model to be tuned. Where the second method is used if there is no knowledge about the process, when the rules and membership functions can be extracted directly from the data by clustering the input / output space.

**2.3 Clustering:**

**2.3.1 Clustering technique:** The presented clustering technique gives us information about the influence of particular variables or attributes of the data set on special clusters. This knowledge can be used in classification tasks to determine or detect class defining attributes. Without ignoring one data attribute for the whole classification it is possible to reduce the influence of that attribute on only some clusters. In that way, attribute weights could help to partition the whole data set into smaller data parts depending on the same attributes. Analyzing the smaller parts with a reduced number of attributes would reduce computation effort [6].

**2.3.2 Cluster analysis** can be defined as "a wide variety of procedures that can be used to create a classification, these procedures empirically form "clusters" or group of highly similar entities". One can say that the objective of cluster analysis "is to sort the observations into groups such that the degree of "natural association" is high among numbers of the same group and low between numbers of different groups". The complexity of such task is easily recognized due to the number of possible arrangements. Even a small number of elements (25) to be clustered in a small number of groups (5) raise a very large number of possibilities (2,436,684,974,110,751).[w.3]

**2.3.3 Fuzzy Clustering:** Clustering algorithms are mainly concerned with partitioning the data into a number of subsets, within each subset; the elements are similar to each other. On the other hand, elements from different subsets are as different as possible. There are different fuzzy clustering techniques based on unsupervised learning such as relation criterion functions, object criterion functions and C-mean, etc. Most of the clustering is being applied for diagnosis.

### 2.4 Rule-Based System:

The goal of production system is to model human performance in problem solving. In a rule-base system, the condition action pairs are represented as if…. Then…. Rules, where the if portion corresponding to the condition and then portion corresponding to the conclusion or action [10]. Shouman used it for Diagnosing Heart Disease Patients [11].

#### 2.4.1 Techniques for Problem Solving Based on Rule-Based are:

a) **Goal-drive problem solving**: In a goal-driven expert system, the goal expression is initially placed in the working memory. The system matches rule conclusions with the goal, selecting one rule and placing its premises in the working memory. The problem's goal decomposes into simple sub-goals. The system works back from the original goal until all the sub-goals in the working memory are known to be true, indicating that the hypothesis has been verified.

b) **Data-driven problem solving:** Breadth-first search is even more common in data-driven reasoning, the contents of working memory is compared with the conditions of each rule in the rule base by ordering of the rule base. If the data in the working memory supports a rule's firing, the result is placed in the working memory and then control moves on to the next rule. Once all rules have been considered, search starts again at the beginning of the rule set.

**2.4.2 Best-first search:** It is a general algorithm for heuristically searching, heuristic is the study of the methods and rules of discovery and invention which is used when a problem may not have an exact solution because of the inherent ambiguities in the problem statement or variable data or when a problem may have an exact solution but the computational cost of finding it may be prohibitive, any state space graph. It is equally applicable to data and goal driven searches and supports a variety of heuristic evaluation functions. It will continue to provide a basis for examining the behavior of heuristic search. Because of its generality, best-first search can be used with a variety of heuristics, ranging from subjective estimates of states to sophisticated measures based on the probability of a state leading to a goal. The goal of the best-first search is to find the goal state by looking as few state as possible; the more informed the heuristic, the fewer states are processed to find out the goal.

## 3. THE METHODOLOGY USED

It was found that the rule-based method is more adequate to represent the knowledge of the concerned domain.

### 3.1 Domain Identification:

The domain can be defined by JCPR blood laboratory investigation as described in the CRF relevant to the study going to be conducted. It is worthy to mention that the nature, number and comprehensiveness of

these laboratory investigations are dependent on the study medication as well as the study protocol; another thing also to be mentioned is that these parameters as described in this work represent some but not limited to all parameters as used by JCPR.

This domain includes 19 parameters for their volunteers. These parameters will be used to ensure that the participated volunteers are healthy. Each parameter is constrained by a minimum and a maximum limit, as listed in Table (1).

Clinical Investigators whom are physician working in this field are consulted to provide the necessary expert and information regarding:

1) If some results of the volunteer parameters are out of normal range, whether those data results can be acceptable or not.
2) How far and the extent that can be acceptable for the out of range data.
3) How to correlate different blood parameters together.

### 3.2 Identification of Parameters Relation:

Through the discussion with the expert, the following relations had been constructed among blood test parameters: [7]

a) Parameters 1, 2 and 4.
b) Parameters 1, 2, 3, 5 and 6.
c) Parameters 5 and 6.
d) Parameters 8, 9, 10, 11 and 12.
e) Parameters 13 and 14.
f) Parameters 7, 8, 9, 10, 11, 15, 16, 18 and 19 are critical.
g) Parameter 17 is flexible within plus minus 5 around the range.

Table (1): Blood Laboratory Investigation

| No. | Parameter | Unit | Reference Range* | Clinical Relevance** |
|---|---|---|---|---|
| 1 | Sodium | Mmol/l | 133 – 151 | (1) (2) (3) |
| 2 | Potassium | Mmol/l | 3.5 – 5.5 | (1) (2) (3) |
| 3 | Total protein | Mg/dl | 6.4 – 8.3 | (1) (2) (3) |
| 4 | Glucose | Mg/dl | 70 – 110 | (1) (2) (3) |
| 5 | Creatinine | Mg/dl | 0.6 – 1.1 | (1) (2) (3) |
| 6 | Urea | Mg/dl | 10 – 50 | (1) (2) (3) |
| 7 | Uric acid | Mg/dl | 3.6 – 8.2 | (1) (2) (3) |
| 8 | Bilirubin, Total | Mg/dl | Up to 1.0 | (1) (2) (3) |
| 9 | AST | U/l | Up to 37.0 | (1) (2) (3) |
| 10 | ALT | U/l | Up to 42.0 | (1) (2) (3) |
| 11 | GGT | U/l | 10.0-47.0 | (1) (2) (3) |
| 12 | Alkal. Phosphatase | U/l | 80 – 306 | (1) (2) (3) |
| 13 | Heamoglobin | Gm/dl | 14.0 – 18.0 | (1) (2) (3) |
| 14 | Heamatocrit | % | 42.0 – 52.0 | (1) (2) (3) |
| 15 | Erythrocytes | / ul | 4.7 – 6.10 $X10^6$ | (1) (2) (3) |
| 16 | Leukocytes | / ul | 4.5 – 10.0 $X10^3$ | (1) (2) (3) |
| 17 | Platelet Count | / ul | 150 – 450 $X10^3$ | (1) (2) (3) |
| 18 | HbsAg | N.A | Negative | N.A |
| 19 | HIV-Ab | N.A | Negative | N.A |

***** Relevance Range varies according to material used in testing.*
****** N.A: not applicable. (1) Normal. (2) Outside normal range, no clinical relevance. (3) Outside normal range, clinically relevant.*
*The physician responsible for drug safety as per the Combined Delegation and Signature List determines clinical relevance.*
***a -** Values outside normal range with clinical relevance at entry examination are an exclusion criterion.*
***b -** A control investigation should be performed if any value out of range with clinical relevance is found at final examination. All such values should be regarded as adverse events.*

### 3.3 Identification of Parameters Tolerance:

From the previous section, relations from *a* to *e*. Just one parameter in each relation has a chance to be abnormal within lower tolerance (LT) and upper tolerance (UT) values from its acceptable normal range where the other parameters in this relation shall be within its normal range (i.e. within its acceptable range). It has been well established in the medical field that the limits of the laboratory tests are not fixed since the human exhibits many variations between volunteers and even among the same volunteer. Table (2) shows for each parameter how far out of range data which can be accepted within the normal range. [8]

For example, parameter 1 (sodium) is acceptable if its value lies between the normal range, not less than 133 and not more than 151. However, according to the expert physician, even values +9 of the upper limit and values –9 below the lower limit, (124 - 160) is consider as an acceptable value if parameter 2 (Potassium) and parameter 4 (Glucose) are within their acceptable ranges (i.e. parameter 2 has a value within 3.5 to 5.5 and parameter 4 has a value within 6.4 to 8.3).

### 3.4 Evaluation of Parameters:

From the previous discussion, we conclude that any volunteer has clinically relevance 3 for any parameter is not fit for the study and must be excluded. For the other volunteers who have clinically relevance 1 or 2 for some parameters, those are considered healthy and accepted for the study. Volunteer, who has clinically relevance 1 for all parameters, is considered 100% healthy from the physician viewpoint. Others can have evaluation less than 100% according to number of parameters having clinical relevance 2. From the 19 parameters, just 6 parameters could be judged to clinical relevant 2. Evaluation to the healthy volunteer can be considered according to these parameters, each parameter has clinical relevant 2 reduce the evaluation by 0.1 (10%). Since all these parameters have the same weight from physician viewpoint. However, this evaluation will help to prioritize the volunteers according to healthy condition. [9]

Table (2): Parameters Tolerance Limits

| No. | Parameter | LT | UT |
|---|---|---|---|
| 1 | Sodium | 9 | 9 |
| 2 | Potassium | 0.2 | 0.2 |
| 3 | Total protein | 0.2 | 0.2 |
| 4 | Glucose | 10 | 10 |
| 5 | Creatinine | 0.1 | 0.1 |
| 6 | Urea | 5 | 5 |
| 7 | Uric acid | - | - |
| 8 | Bilirubin, Total | - | - |
| 9 | AST | - | - |
| 10 | ALT | - | - |
| 11 | GGT | - | - |
| 12 | Alkal. Phosphatase | 0 | 4 |
| 13 | Heamoglobin | 1 | 1 |
| 14 | Heamatocrit | 2 | 2 |
| 15 | Erythrocytes | - | - |
| 16 | Leukocytes | - | - |
| 17 | Platelet Count | 5 | 5 |
| 18 | HbsAg | - | - |
| 19 | HIV-Ab | - | - |

### 3.5 Rules Construction:

The state diagram in Figure (1) represents the flow of information as usually followed by the human expert to take the decision for accepting or rejecting the volunteer.

Where

$State_i$ : represents the clinical relevance for parameter (i) according to Table (1)

$T_i$ : Represents the tolerance for parameter (i).

$\rightarrow$ : Represents the output of previous state to the next state.

Exclusion: Means that the volunteer shall be excluded from the study.

Inclusion: Means that the volunteer can participate in the study.

## 4. IMPLEMENTATION

### 4.1 Database Design:

The database was designed taking into account the future direction of the JCPR to build an integrated system for its whole work.

**4.1.1 Entity Relationship Model (ERM):** Figure (2) is represented the ERM for the

system which shows the system entities and their attributes and the relationships gathering these entities.

**4.1.2 Tables Creation:** The system is consisting of six tables and which can be defined as follows:

**a.** Study table: which contains information about each study to be done by the center for a defined sponsor, study attributes are:
1- St_no: study identification.
2- St_name: study name.
3- St-date: starting date of study.
4- Sp_no: sponsor identification for who study will be done.
5- St_activeIngredient: study active ingredient.

**b.** Volunteer table: which contains information about the volunteers who could participate in a defined study, volunteer attributes are:
1- V_no: volunteer identification.
2- V_name: volunteer name.
3- V-initial: volunteer initial used as a special code for a volunteer by the center.
4- V_age: volunteer age.
5- V_address: volunteer address.
6- V_telephone: volunteer telephone

**c.** Sponsor table: which contains information about the sponsor who could ask for study to be done, sponsor attributes are:
1- Sp_no: sponsor identification.
2- Sp_name: sponsor name.
3- Sp-contact: sponsor contact person.
4- S_telephone: sponsor telephone.

**d.** Blood tests table: It contains information about the result of laboratory blood test for every volunteer in each specified study according to the parameters in table (1) in section 3.2.

**e.** Clinical relevance blood test: which contains derived information depending on the blood test table information and the constraints governing the tested parameters, these information's are v_no, st_no, clinical relevance, 1, 2 or 3 for each tested parameters in the blood table in addition to the final decision and volunteer evaluation.

**f.** Reference tables: contains information about the tested parameters, their minimum, maximum limits, upper and lower tolerances.

### 4.2 Rules Implementation:

The rules constructed in section 3.5 were implemented using the ***IF condition Then action*** and the logical operators "AND" and "OR" to get the desired conclusions for selecting and evaluating the volunteers as follow:

```
Procedure selection (Var decision:
integer; var TE: real);
Begin
TE:=1.0;    { volunteer desired
evaluation= 100%}
  { decision=1 or 2 accepted, 3 not
accepted}
{ exclusion rules}
If ((State19=3) or (State18=3) or
   (State16=3) or (State15=3) or
   (State17=3) or (State18=3) or
   (State9=3) or (State10=3) or
   (State11=3)) Then    decision:=0;
   TE=0; return;
If ((State12=3) and Not ((SV12 >=
   S12_Min - LT12) and (SV12 <=
   S12_Max + UT12))) Then
   decision:=0; TE=0; return;
If ((State13=3) and Not ((SV13 >=
   S13_Min - LT13) and (SV13 <=
   S13_Max + UT13))) Then
   decision:=0; TE=0; return;
If ((State14=3) and Not ((SV14 >=
   S14_Min - LT14) and (SV14 <=
   S14_Max + UT14))) Then
   decision:=0; TE=0; return;
If ((State17=3) and Not ((SV17 >=
   S17_Min - LT17) and (SV17 <=
   S17_Max + UT17))) Then
   decision:=0; TE=0; return;
If ((State5=3) and Not ((SV5 >=
   S5_Min - LT5) and (SV5 <=  S5_Max
   + UT5))) Then  decision:=0; TE=0;
   return;
If ((State6=3) and Not ((SV6 >=
   S6_Min - LT6) and (SV6 <=  S6_Max
   + UT6))) Then  decision:=0; TE=0;
   return;
If ((State3=3) and Not ((SV3 >=
   S3_Min - LT3) and (SV3 <=  S3_Max
   + UT3))) Then  decision:=0; TE=0;
   return;
```

```
If ((State1=3) and Not ((SV1 >=
   S1_Min – LT1) and (SV1 <=  S1_Max
   + UT1))) Then  decision:=0; TE=0;
   return;
If ((State2=3) and Not ((SV2 >=
   S2_Min – LT2) and (SV2 <=  S2_Max
   + UT2))) Then  decision:=0; TE=0;
   return;
If ((State4=3) and Not ((SV4 >=
   S4_Min – LT4) and (SV4 <=  S4_Max
   + UT4))) Then  decision:=0; TE=0;
   return;
{Evaluation rules}
If ((State12=3) and ((SV12 >= S12_Min
   – LT12) and (SV12 <=  S12_Max +
   UT12))) Then begin  State12=2; TE=
   TE – 0.1; end;
If ((State13=3) and ((SV13 >= S13_Min
   – LT13) and (SV13 <=  S13_Max +
   UT13)) and (SV14=1)) Then begin
   State13=2; TE= TE – 0.1; end;
If ((State14=3) and ((SV14 >= S14_Min
   – LT14) and (SV14 <=  S14_Max +
   UT14)) and (SV13=1)) Then begin
   State14=2; TE= TE – 0.1; end;
If ((State17=3) and ((SV17 >= S17_Min
– LT17) and (SV17 <=  S17_Max +
UT17))) Then begin State17=2; TE= TE
– 0.1; end;
If ((State5=3) and ((SV5 >= S5_Min –
LT5) and (SV5 <=  S5_Max + UT5)) and
(SV6=1)) Then begin State5=2; TE= TE
– 0.1; end;
If ((State6=3) and ((SV6 >= S6_Min –
LT6) and (SV6 <=  S6_Max + UT6)) and
(SV5=1)) Then begin  State6=2; TE= TE
– 0.1; end;
If ((State3=3) and ((SV3 >= S3_Min –
   LT3) and (SV3 <=  S3_Max + UT3))
   and (SV1=1) and (SV2=1) and
   (SV5=1) and (SV6=1))
   Then begin  State3=2; TE= TE –
   0.1; end;
If ((State1=3) and ((SV1 >= S1_Min –
   LT1) and (SV1 <=  S1_Max + UT1))
   and (SV2=1) and (SV4=1)) Then
   begin State1=2; TE= TE – 0.1; end;
If ((State2=3) and ((SV2 >= S2_Min –
LT2) and (SV2 <=  S2_Max + UT2)) and
(SV1=1) and (SV4=1)) Then begin
State2=2; TE= TE – 0.1; end;
If ((State4=3) and ((SV4 >= S4_Min –
   LT4) and (SV4 <=  S4_Max + UT4))
   and (SV1=1) and (SV2=1)) Then
   begin State4=2; TE= TE – 0.1;end;
{ decision for evaluation rules}
If (TE=1) then decision :=1 else
decision :=2;
Return;     {end of procedure
selection}
```

## 5. CONCLUSION

The task of knowledge acquisition was hard and it was difficult to accumulate the knowledge and expertise of the physicians' experts. Furthermore, the confirmation of this knowledge by another expert made the acquisition even more difficult.

Implementation of this work will save the time and the money and it will reduce the human errors to the lowest possible extent.

Such system will be considered as the first step in building an integrated system for the JCPR.

JCPR by adapting this system will go a further step toward establishing a solid quality system, I mean to say establishing a stable system which is almost independent on the persons whom performing such activities, consequently insuring reproductively of the volunteer evaluation process.

The system is easy to be implemented, flexible enough to cope with the amendments which are usually occurred on the blood test parameters and their constraints.

The system provides a tool for preservation of volunteers information as a soft copy as required by the regulation with effect from the clinical study. It is more significant to mention that such information about the volunteers shall keep for more than ten years.

## REFERENCES:


[1] Owidi, H. Hussein lectures note 2003.

[2] Ali, D.H., design and development of knowledge-based system in plastics selection and evaluation, master thesis, march 1997.

[3] Nagendra, S., Practical Aspects of using Neural Networks: Necessary Preliminary Specifications. GE Research &Development Center, 97CRD173, January 1998.

[4] Huang, Y.P. and Chen, Y.R., Preprocessing data patterns to simplify the identification of fuzzy models. International Journal System Science, 1999, volume 30, number 5, pages 479-489.

[5] Chen, S.M. and et al., A new Method for Generating Fuzzy Rules from Numerical Data for Handling Classification Problems. Applied Artificial Intelligent, 15: 645-664, 2001.

[6] Keller, A. and Klawonn, F., Fuzzy Clustering with Weighting of Data variables. Interntional Journal of

Uncertainty, Fuzziness and Knowledge-Based Systems Vol. 8, No. 6 (2000) 735-746.

[7] Al-Hour, Jehad, Ibn Al-Haytham Hospital, Amman – Jordan, interview 18/7/2014.

[8] Al-Hour, Jehad and Jaber, Wael, Ibn Al-Haytham Hospital, Amman – Jordan, interview 25/7/2014.

[9] Qaswal, Jameel, Isalmic Hospital, Amman – Jordan, interview 21/8/2014.

[10] Yo-Ping Huang and Vu Thi Thanh Hoa. 2009. General criteria on building decision trees for data classification. In Proceedings of the 2nd International Conference on Interaction Sciences: Information Technology, Culture and Human (ICIS '09). ACM, New York, NY, USA, 649-654. DOI=10.1145/1655925.1656042

[11] Mai Shouman, Tim Turner, and Rob Stocker. 2011. Using decision tree for diagnosing heart disease patients. In Proceedings of the Ninth Australasian Data Mining Conference - Volume 121 (AusDM '11), Peter Vamplew, Andrew Stranieri, Kok-Leong Ong, Peter Christen, and Paul J. Kennedy (Eds.), Vol. 121. Australian Computer Society, Inc., Darlinghurst, Australia, Australia, 23-30.

**Web sites:**

[W.1] ]Wan Ishak, W.H. and Siray, F., *Artificial Intelligence in Medical Application: an Exploration*. www.hi-europe.co.uk/files/2002/9980.htm . 5/6/2014.

[W.2] What is neural network?, http://www.nd.com/welcome/whatisnn.htm, 30/6/2014.

[W.3] Painho, M. and Bacao, F., *Using Genetic Algorithm in Clustering problems*. GeoComputation, 2000, updated 16 April 2002. http://www.geocomputation.org/2000/GC015/Gc015.htm. 15/6/2014.

Figure (1): State Diagram represents the flow of information

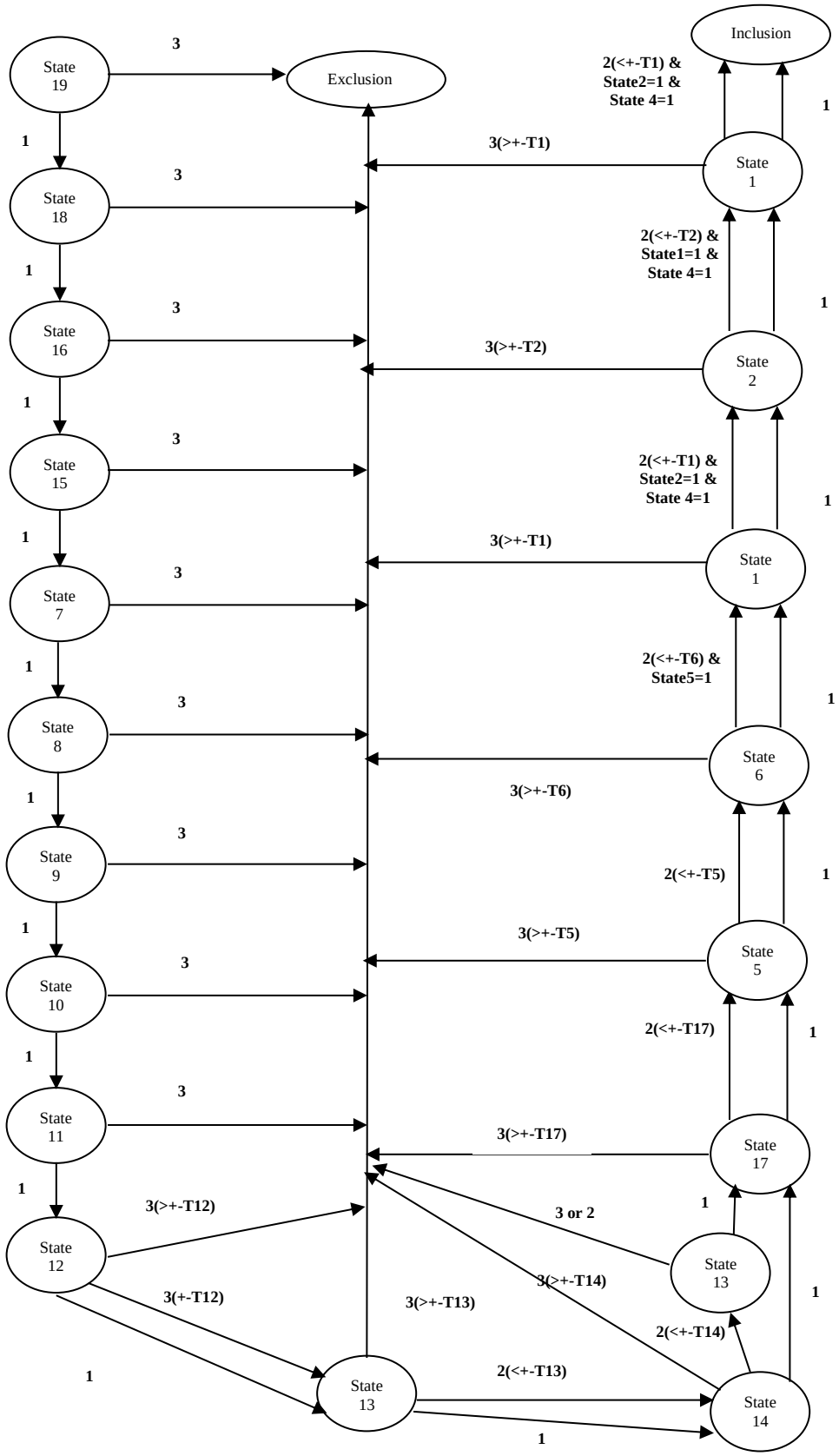
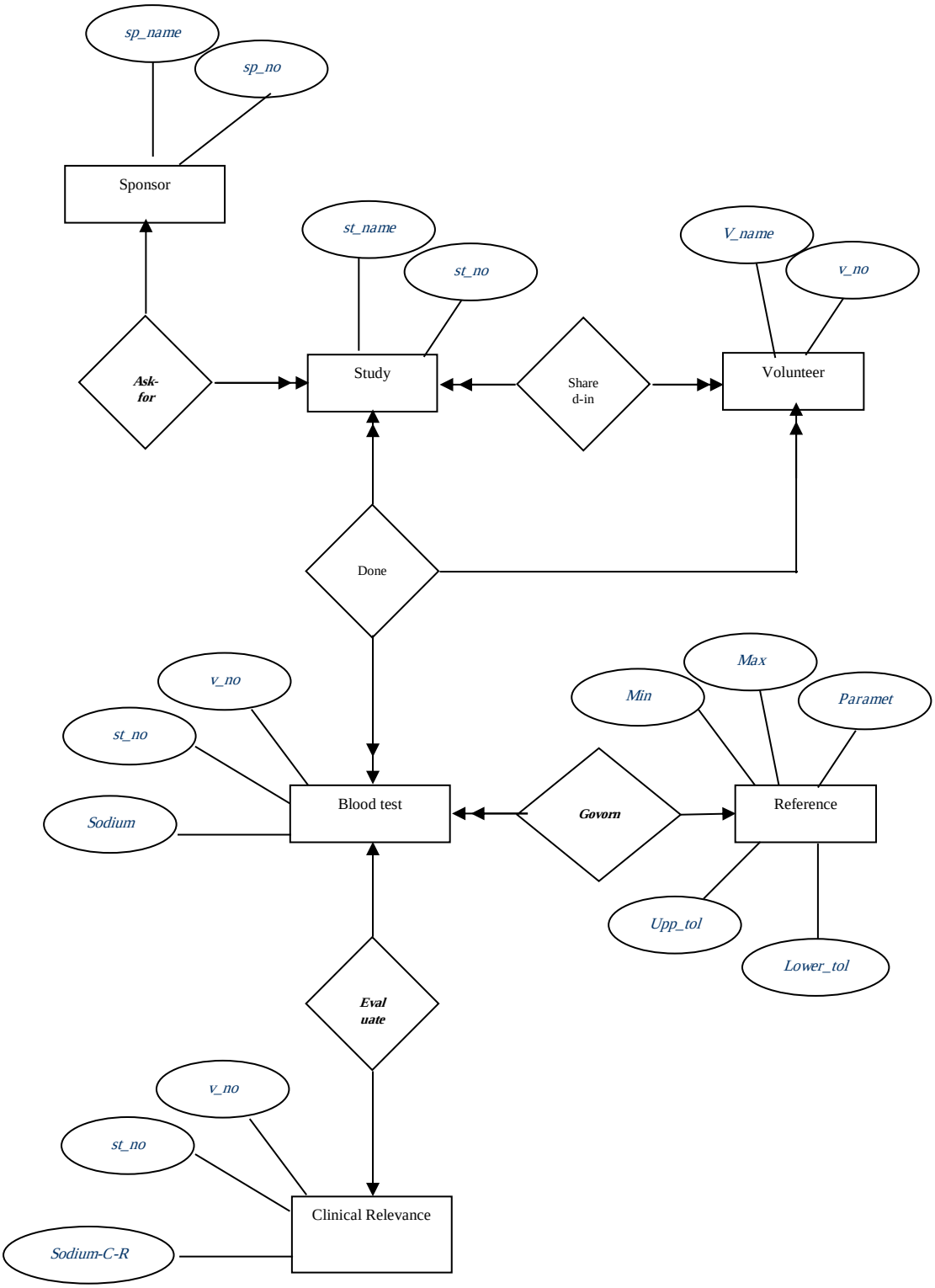


Figure (2) Entity Relationship Model (ERM)




** The attributes connected to each entity are not all its attributes, the complete list of attributes are listed in the subsection 4.1.2.