*2012 4th International Conference on Bioinformatics and Biomedical Technology*
*IPCBEE vol.29 (2012) © (2012) IACSIT Press, Singapore*

# Rule Induction as a Technique of Detecting Severity of Myocardial Infarction

Raquib Ridwan [1], Md Ibnea Sina Bony [2] and Kamrul Hasan [3]

[1] Habib Bin Muzaffar

[2] Raiyan Rahman

**Abstract.** Myocardial Infarction commonly known as heart disease is a very acute disease and requires an intelligent system to diagnose the intensity level of the disease and take immediate and prompt decisions. The main barrier lies in scouring through a large data set and attribute value pairs. This paper represents a novel technique of building a diagnostic model of detecting and describing the acuteness of myocardial infarction using Rule induction method. This paper also proposes a unique scientific model that can be utilized in hospitals for administering the right drug to patients depending on the decision of our intelligent system. This paper follows a definite method to explain the outcome of our model. Finally the result and our proposed model are sequentially explained and in the conclusion we have explained the limitation of our proposal and given ideas for further study.

**Keywords:** Concept, discrimination, local covering, attribute value pair, minimal complex, cardinality.

## 1. Introduction

In the realm of knowledge based systems, genomic signal processing is a sophisticated disciplinary field that enables human experts to extract knowledge about diseases from clinical case study [1]. With the interrelation engineering and medical science this Knowledge acquisition has changed the direction of active research towards biomedical engineering/technology for years. Many rules have been developed and also implemented so far in the clinical study following some constructive ways and methods which include observation of symptoms; classifying the attributes and finally generating a rule set with minimal attribute computation in order to detect and check the acuteness of the disease as well. Decision making is normally based on available evidence. For example, a piece of evidence may be the result of a test or an observation. A decision making process consists of a sequence of steps. In each step, one makes a decision of acceptance or rejection if the available evidence is sufficiently strong. Otherwise we move to the next step in which a rule of deferment in last step is refined into a set of rules of acceptance, rejection or deferment by adding new evidence. The process will be continued till a certain condition is met [2]. In this paper, a novel computational scientific method has been proposed as an interactive model generation process for drug delivery system for myocardial infarction. A very popular rule induction technique focusing on LERS (Learning from Examples using Rough Sets) is applied in order to point out the acuteness of myocardial infarction by reducing the large number of attribute value pairs associated with a certain heart condition to a definite rule set. Machine learning techniques are applied to the defined data set to develop a concept representation. This concept representation will be modeled later on in this paper.

## 2. Methods, Materials and Diagnostic Model

Utilizing the rule induction method technique, a diagnostic model for myocardial infarction detection has been proposed using LEM2 algorithm (Learning from Examples module, Version 2).This algorithm is a very important component of LERS which corresponds to characterization, discrimination and observation [3]. Besides that the proven capability of this algorithm to give faster and better results compared to other methods such as LEM1 and AQ method has prompted us to use the LEM2 algorithm for designing our diagnostic model. Hence LEM2 serves as the best rule induction method/technique to develop a diagnostic model for detection of the degree of severity of myocardial infarction.

## 2.1. Clinical study

To determine the severity of myocardial infarction, attributes should be considered. After a thorough study on the factors that give rise to myocardial infarction, we have been able to identify the four most significant attributes or markers for the cause of myocardial infarction. Firstly we studied some patients and analyzed data from expert professionals in this field. We finally arranged our Data in a tabular form in which, cases are labeled as rows and variables are labeled as attributes followed by a Decision in Table-2. The attribute values were finally converted to its nominal form according to Table-1 so that we can use simple and efficient algorithm of LEM2 for extracting the most definite rule set.

| Attributes | Lower limit | Upper limit | Decision |
|---|---|---|---|
| Troponin-T | 0 ng/ml | 0.5 ng/ml[9] | Normal |
| | 0.5 ng/ml | 2.0 ng/ml[9] | High |
| | 2.0 ng/ml | >2.0 ng/ml[9] | Very High |
| CK-MB | 0 ng/ml | 5.0 ng/ml[4] | Normal |
| | 5.0 ng/ml | >5.0 ng/ml[4] | High |
| GPI-BB | 0 ng/ml | 7.9 ng/ml[5] | Normal |
| | 7.9 ng/ml | >7.9 ng/ml[5] | High |
| LDH | 180units/L | 200units/L [8] | Normal |
| | 200units/L | >200units/L [8] | High |

Table-1: The boundaries of the attributes

| Case | Troponin-T level | CK MB Level Is high | Glycogenphosphorylase Isoenzyme BB (GPI-BB) Level is high | LDH (Lactate Dehydrogenase) Level is high | Decision (severity of heart) |
|---|---|---|---|---|---|
| 1 | Very High | Yes | Yes | Yes | Red |
| 2 | Very High | No | No | Yes | Yellow |
| 3 | Normal | Yes | No | Yes | Green |
| 4 | Normal | No | Yes | No | Green |
| 5 | High | No | Yes | Yes | Yellow |
| 6 | Normal | Yes | Yes | Yes | Yellow |
| 7 | High | No | No | No | Green |
| 8 | High | Yes | Yes | No | Red |
| 9 | Very High | No | Yes | Yes | Red |
| 10 | High | No | Yes | No | Yellow |
| 11 | High | Yes | No | Yes | Green |
| 12 | Normal | No | No | Yes | Green |
| 13 | Very high | Yes | No | No | Red |
| 14 | Normal | Yes | No | No | Green |
| 15 | Very high | Yes | Yes | No | Red |
| 16 | Very high | No | No | No | Yellow |

Table-2: part of the obtained database of Myocardial Infarction Severity

# 3. Rule Induction

## 3.1. Definition

Induction is the process of trying to perform generalization from a specific set of examples. It is one of the most versatile techniques of machine learning. The assumption is that some underlying regularity always exists in the specific set of examples, the nature of which is unknown. The task is to discriminate between the classes from a given a set of examples that have been pre classified. The generalized expression is used as a form similar to,

*if,*
*(attribute – 1, value − 1) and (attribute – 2, value − 2) and (attribute-3, value-3)….and (attribute – n, value − n)*
*then*
*(decision, value).* [6]

Some rule induction systems include more complex rules, in which values of attributes may be expressed by proposition of some values or by a value subset of the attribute domain. All the attribute values are taken in its nominal form for the convenience of easy yet effective calculation.

## 3.2. LEM2

LEM2 (Learning from Examples Module version-2) is one of the derivatives of LERS (Learns from Examples Rough Sets) [7]. LEM2 computes a local covering followed by converting them to rule sets. We use the notation and terminology of [6]. The set of all cases labeled by the same decision value is called a *concept*. For an attribute- value pair $(a,v)=t$, a block of *t*, denoted by [*t*] is the set of all cases for which attribute *a* has value *v*. Let *B* be a concept. We say that the set *B* *depends* on a set *T* of attribute- value pairs $(a,v)=t$ if and only if $\emptyset \neq [T] = \bigcap_{t\in T}[t] \subseteq B$. A set $T$ is said to be a *minimal complex* of *B* if and only if *B* depends on $T$ and there is no proper subset $T'$ of $T$ such that $B$ depends on $T'$. Let $\tau$ be a non-empty collection of non-empty sets of attribute value pairs. We say that $\tau$ is a *local covering* of $B$ if and only if each member $T$ of $\tau$ is a minimal complex of $B$, $\bigcup_{T\in\tau}[T]=B$, and $\tau$ is minimal, i.e. $\tau$ has the smallest possible number of elements.

## 3.3. Logical representation of theLEM2 algorithm

(Input: a set *B* which is a concept,
Output: a single local covering $\tau$ of set *B*);
begin
$G := B$;
$\tau := \emptyset$
while $G \neq \emptyset$;
begin
$T := \emptyset$;
$T(G) := \{t \mid [t]\cap G \neq \emptyset\}$;
while $T = \emptyset$ or $[T] \not\subseteq B$
begin
select a pair $t\in T(G)$ such that $|[t]\cap G|$ is
maximum; if a tie occurs then select a pair $t\in T(G)$ with the smallest cardinality of [*t*];
if another tie occurs, select first pair;
$T := T \cup \{t\}$;
$G := [t]\cap G$;
$T(G) := \{t \mid [t]\cap G \neq \emptyset\}$;
$T(G) := T(G) - T$;
end {while}
for each $t\in T$ do
if $[T-\{t\}] \subseteq B$ then $T := T - \{t\}$;
$\tau := \tau \cup \{T\}$;
$G := B - \bigcup_{T\in\tau}[T]$;
end {while};
for each $T\in\tau$ do
if $\bigcup_{S\in\tau-\{T\}}[S]=B$ then $\tau := \tau - \{T\}$;
end {procedure}.[6]

## 3.4. Flow chart representation of LEM2 algorithm

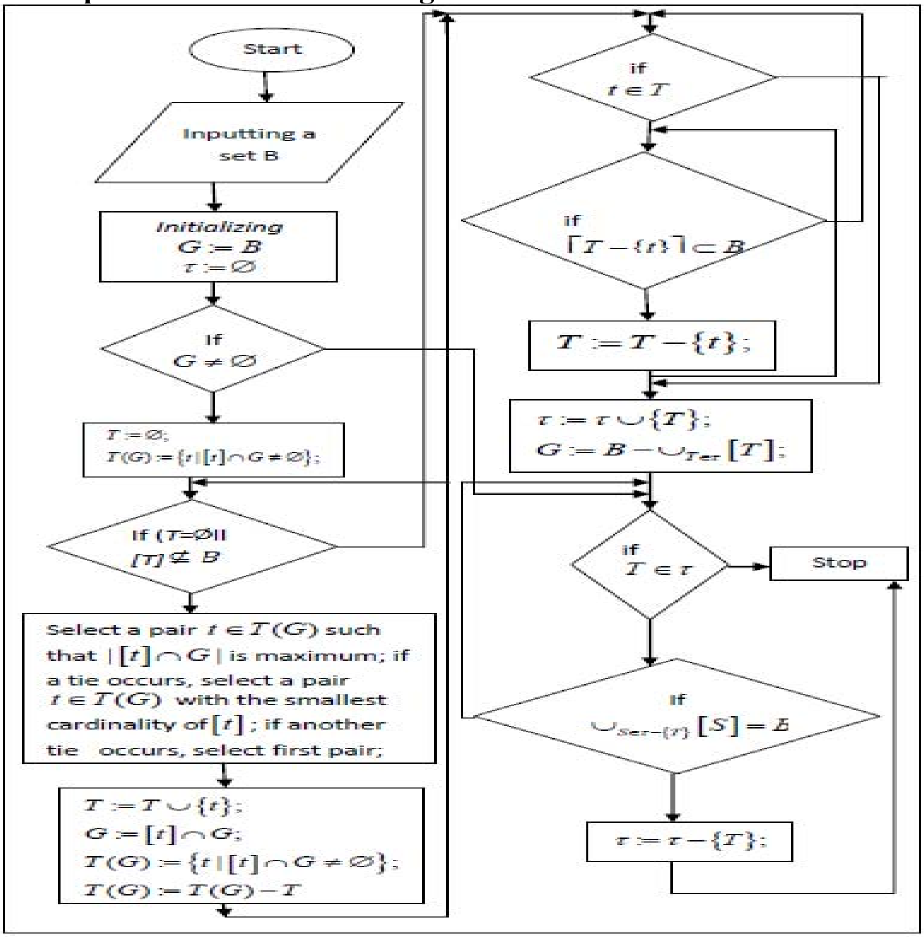


Fig. 1: Flow chart representation of LEM2

# 4. Data Analysis

From the analysis of the data set we have found out that three stages or degrees explaining the severity of myocardial infarction can be obtained. We have symbolized them with color notation by denoting red as the most severe heart condition, yellow as the moderate condition and green as the mild condition. Finally based on the decision of either Red or Yellow or Green, we have used the LEM2 algorithm to find out a minimal rule set for each of these conditions.

## 4.1. In the case of Red

The concept corresponding to the decision "red" is the set {1, 8, 9, 13, 15}. This concept is taken as the set B in LEM2. Thus, the starting value of *G* is *B*. We see that *T* (*G*) is the following set as: {(Troponin-T level, Very high), (Troponin-T level, High), (CK MB level high, Yes), (CK MB level high, No), (GPI-BB level high, Yes), (GPI-BB level high, No), (LDH level high, Yes), (LDH level high, No)}. Moving to the next part of the algorithm, we get *t* = (Troponin-T level, Very high), *T* = {(Troponin-T level, Very high)}. Our new *G* is {1, 2, 9, 13, 15, 16} ∩ {1, 8, 9, 13, 15} = {1, 9, 13, 15}. The final *T* (*G*) of this loop is {(CK MB level high, Yes), (CK MB level high, No), (GPI-BB level high, Yes), (GPI-BB level high, No), (LDH level high, Yes), (LDH level high, No)}. We have [*T*] = {1, 2, 9, 13, 15, 16} which is not a subset of *B*. Thus, we have to repeat the internal loop. This time we get *t* = (CK MB level high, Yes), *T* = {(Troponin-T level, Very high), (CK MB level high, Yes)}, *G* = {1, 13, 15}. This time [*T*] = {1, 13, 15} so that $[T] \subseteq B$ . Hence, the internal loop has terminated and the latest *T* is a minimal complex. Continuing, we get $\tau$ = {{(Troponin-T level, Very high), (CKMB level high, Yes)}} and the new *G* = {8, 9} which is non-empty. So, the external loop has to be repeated.

The new *T* (*G*) for *G* = {8, 9} is the set {(Troponin-T level, Very high), (Troponin-T level, High), (CK MB level high, Yes), (CK MB level high, No), (GPI-BB level high, Yes), (LDH level high, Yes),( LDH level high, No)}. This time, the internal loop terminates when we have the minimal complex $T$ = {(Troponin-T level, high), (CKMB level high, Yes), (GPI-BB level high, Yes)} and we get $G$ = {9}.

Applying the external loop with $G$ = {9}, we see that the internal loop terminates when $T$ = {(Troponin-T level, very high), (CK MB level high, No), (GPI-BB level high, Yes)}.

This time, running the following part of the LEM2 algorithm:

for each $t \in T$ do

if $[T - \{t\}] \subseteq B$ then $T := T - \{t\}$

We see that $t$ = (CKMB level high, No) has to be subtracted giving the minimal complex $T$ = {(Troponin-T level, very high), (GPI-BB level high, Yes)}. We get $G$ = Ø and the algorithm terminates with the following local covering $\tau$ :

{{(Troponin-T level, Very high), (CK MB level high, Yes)}, {(Troponin-T level, high), (CKMB level high, Yes), (GPI-BB level high, Yes)}, {(Troponin-T level, very high), (GPI-BB level high, Yes)}}.

From the above local covering, we have the following rules:

Rule Set

RULE 1: (*Troponin-T level, Very high*) and (*GPI-BB level high,Yes*) → (*heart condition, red*).
RULE 2: (*Troponin-T level, Very high*) and (*CK MB level high,Yes*) → (*heart condition, red*).
RULE 3: (*Troponin-T level, high*) and (*CK MB level high, Yes*) and (*GPI-BB level high, Yes*) → (*heart condition, red*).

### 4.2. In the case of yellow

Following the same algorithm as we did for Red, we found that the rule set for identifying the moderate condition of myocardial infarction is

Rule Set

RULE 4: (*Troponin- T level, high)* and (*CK MB level high, No*) and (*GPI-BB level high, Yes*) → (*heart condition, yellow*).
RULE 5: (*Troponin-T level, Very High*) and (*CK MB level high, No*) and (*GPI-BB level high, No*) → (*heart condition*, *yellow*).
RULE 6: (*Troponin-T level, Normal)* and (*CK MB level high, Yes*) and (*GPI-BB level high, Yes*) → (*heart condition, yellow*).

### 4.3. For the case of green

RULE 7: (*Troponin-T level, Normal*) and (*GPI-BB level high, No*) → (*heart condition, green*).
RULE 8: (*Troponin-T level, high*) and (*GPI-BB level high, No*) → (*heart condition, green*).
RULE 9: (*Troponin-T level, Normal*) and (*CK MB level high, No*) → (*heart condition, green*).

## 5. Results and Discussion

In the context of classification based techniques all the decisions made by Rule Induction method, LEM2 may not be 100% accurate in all cases. In our case of Myocardial Infarction, the Rule Induction technique using LEM2 algorithm was able to give us a perfect result which hence can be used to develop an interactive scientific approach to diagnose Myocardial Infarction degree of severity. Modeling of the problem can be represented.

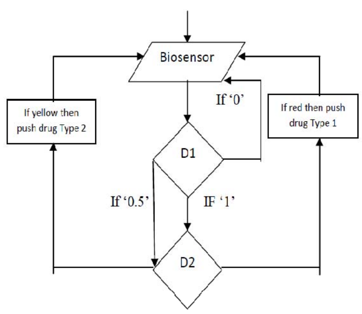


Fig. 2: Interactive scientific model to diagnose Myocardial Infarction Using Rule Induction approach

In short, the biosensor will be used to monitor the condition of the heart, and D1 is the systems-on-chip microcontroller which is embedded with the source code of our algorithm. The output will generate three possible voltage levels namely 0.5 to indicate the case for yellow, and 1 to indicate the case for red and finally 0 for the green case. Finally the other microprocessor D2 will ensure the selection of the right drug to be administered to the patient.

## 6. Conclusions

In this paper Rule induction method using LEM2 algorithm was used to generate unique and accurate rule sets for detecting the severity of Myocardial Infarction. Finally based on this algorithm we have provided a proposed diagnostic model that can be implemented in the hospitals for quick administration of proper drug to the patients. However the table that we generated for performing the LEM2 algorithm was based on nominal values, i.e. Yes, Very High, High, normal, No only. But in practical cases numerical values of the attributes are extracted and besides that we have skipped ranges that can exist between very high and high. Our future goal will be to use the elaborate numerical values of the attributes and finally implement a fuzzy logic technique to get multiple existing states apart from very high and high only. This will enable us to properly identify the severity of myocardial infarction in a more detailed and complete fashion.

## 7. References


[1] On Rule Induction Method Based Rough Sets in Diagnostic Expert System, LI Ai-Ping, Jia Yan, WU Quan-Yuan R. Dewri, and N. Chakraborti. Simulating recrystallization through cellular automata and genetic algorithms. *Modelling Simul. Mater. Sci. Eng*. 2005, **13** (3): 173-183.

[2] A Model of Machine Learning Based on User Preference of Attributes, Yiyu Yao, Yan Zhao, Jue Wang and Suqing Han

[3] LEM2-Based Rule Induction via Clustering Decision Classes, Masahiro Inuiguchi, Masayo Tsurumi, Daisuke Fukuda, Kazuki Yamanaka

[4] CARDIAC TROPONIN-T AND CK-MB (MASS) LEVELS IN CARDIAC AND NON CARDIAC DISEASE, G. Chinnapu Reddy, G. Kusumanjali, A.H.R. Sharada and Pragna Rao.Indian Journal of Clinical Biochemistry, 2004, 19 (2) 91-94

[5] Biomarkers of Myocardial Infarction in Patients Undergoing Gastrointestinal Cancer Surgery, Giuseppe Lippi, Gian Luca Salvagno, Anna Maria Minicozzi, Martina Montagnana, Claudio Cordiano, Giovanni De Manzoni, Gian Cesare Guidi.

[6] Chapter 1, RULE INDUCTION, Jerzy W. Grzymala-Busse, University of Kansas

[7] LEM2 Theory and Implementation of the Learnable Evolution Model, Guido Cervone, George Mason University

[8] Pagana, Kathleen Deska. Mosby's Manual of Diagnostic and Laboratory Tests. St. Louis: Mosby, Inc., 1998.

[9] Wu, Alan, editor. Cardiac Markers. Washington, DC:American Association of Clinical Chemistry (AACC) Press,1998