Global Journal of Computer Science and Technology: C
Software & Data Engineering
Volume 16 Issue 4 Version 1.0 Year 2016
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

MAGED: Metaheuristic Approach on Gene Expression Data: Predicting the Coronary Artery Disease and the Scope of Unstable Angina and Myocardial Infarction

By E.Neelima & M.S.Prasad Babu

GITAM University

Abstract - The Genetic risk prediction strategies found in practice for coronary artery disease are not significant to estimate the scope of adverse cardiovascular events such as unstable angina and myocardial infarction. Hence in regard to this objective, this manuscript contributed a metaheuristic approach to predict coro- nary artery disease and the scope of unstable angina and myocardial infarction. The proposed metaheuristic is built from the gene expression data of blood samples collected from patients with coronary artery disease diagnosed, unstable angina and Myocardial Infarction. The data also includes gene expression data collected from the blood samples taken from the people clinically proven as salubrious (healthy). The relation between genes and gene expressions are considered as the state of input to devise the metaheuristic.

Keywords: micro array, coronary artery disease, unstable angina, myocardial infarction, gene expression data, gene expression profiling, metaheuristics, machine learning.

GJCST-C Classification : J.3, H.2.1

Strictly as per the compliance and regulations of:

MAGED: Metaheuristic Approach on Gene Expression Data: Predicting the Coronary Artery Disease and the Scope of Unstable Angina and Myocardial Infarction

E.Neelima [α] & M.S.Prasad Babu [σ]

Abstract - The Genetic risk prediction strategies found in practice for coronary artery disease are not significant to estimate the scope of adverse cardiovascular events such as unstable angina and myocardial infarction. Hence in regard to this objective, this manuscript contributed a metaheuristic approach to predict coro- nary artery disease and the scope of unstable angina and myocardial infarction. The proposed metaheuristic is built from the gene expression data of blood samples collected from patients with coronary artery disease diagnosed, unstable angina and Myocardial Infarction. The data also includes gene expression data collected from the blood samples taken from the people clinically proven as salubrious (healthy). The relation between genes and gene expressions are considered as the state of input to devise the metaheuristic. In order to find the confidence of the relation between gene and gene expression a bipartite graph is built between them. The experimental study evincing that the prediction performance of the proposed model is substantial that compared to other benchmarking models.

Keywords: micro array, coronary artery disease, unstable angina, myocardial infarction, gene expression data, gene expression profiling, metaheuristics, machine learning.

I. Introduction

Cardiovascular diseases are the critical reason of human deaths happening worldwide. The statistics indicating that this disease causes annually around 17.3 million deaths [1].The inadequate blood supply to the heart causes necrosis of myocardial tissue, which is clinically referred as Myocardial Infarction (MI).

The MI was claimed 7.6 million deaths among 58 million deaths worldwide in 2005 [2]. The advancements in clinical practices to diagnose and prevent MI are evinced to be not significant, since the count of human deaths due to MI is high that compared to the deaths caused by any other disease [1] [2].

The current diagnosis of MI is based on clinical symptoms including chest pain and difficulty to breath, ECG pattern variants, and potential drop and raise of blood floating in cardiac muscles (cardiac troponins also referred as cTns) [3]. Though the phenomenal advances in clinical diagnosis strategies found, still the substantial constraints are evinced in current clinical diagnosis strategies. The advances in hs-cTni assays [4] have evinced high detection of cardio vascular disease cases (Increased true positive rate) but significant normal cases have been labeled as cardio vascular prone (decreased true negative rate), which is a potential constraint. Another advanced approach of diagnose the cardio vascular disease diagnostic measure is the cardiac miRNAs as biomarkers [5]. The prediction outcomes of this model are trivial due to limited size and tissue specific expression. Hence it is obvious to have more significant and automated detection strategies, which are using the cardiac miRNAs as primary input [6]. The serum inflammatory markers such as BNP, CRP are also considered as cardiovascular biomarkers but the detection accuracy observed with slight improvement [7][8][9].

The acts such as clinical pathology and biology are the crucial to define cardiac biomarkers, which are expensive and less accurate. In contrast to this, the gene expression profiling quantifies the gene expressions formed by the large quantity of genes in order to identify biomarkers, which is analogous and concurrent across the multiple pathways. Hence the gene expression profiling is potential and feasible to quantify the biomarkers to diagnose cardio vascular diseases [10]. The biomarkers defined by Gene expression profiling are potential and those are not evinced by the pathology and biology based clinical processes.

The rest of manuscript describes the related work in section 2, the Metaheuristic Approach on Gene expression Data (MAGED) that followed by section 4, which elaborates the experimental study of the proposal. Finally the section 5 concludes the contribution of the manuscript.

II. Related Work

Gene expression analysis is a potential approach to discover profound biomarkers of cardio vascular diseases. The contemporary literature contains signifycant contributions in defining biomarkers through gene

Author α: Assistant Professor, Department of CSE GITAM University, Visakhapatnam, AP, INDIA. e-mail: eadha.neelima@gmail.com
Author σ: Professor, Department of CS & SE Andhra University, Visakhapatnam, AP, INDIA. e-mail: profmspbabu@gmail.com

expression analysis. Randi et al., [11] devised a gene expression analysis that conceded 482 genes associated to the composition of plaques found in arteries. Many of these genes were not considered for atherosclerosis in earlier diagnosis strategies. Archacki et al.,[12] proposed a gene expression profiling strategy that resulted 56 different genes for atherosclerosisprone and salubrious human coronary arteries. Among these 56, the 49 genes were not associated to coronary artery disease earlier. The model devised in [13] discovered set of genes those enables classification according to age and sex, which are having strong association with obstreperous CAD in the patients, who are not diagnosed as diabetic. The contributions in [14] and [15] profiled variant gene expressions to differentiate the cardio myopathies with influence of ischemic and non-ischemic conditions. Min KD et al., [16] contributed profiling and analysis of gene expressions to notice the divergent genes associated to congestive heart failure. Suresh R et al., [17] studied the salubrious and MI patients that discovered biomarkers and imbalanced pathways those significant evince the reappearance MI in patients effected once with MI.

Liew et al.,[18] defined sequence tags from gene expressions using microarray analysis that compares mRNA molecules found in cellular components of the blood with mRNA molecules found in9divergent human tissues comprising heart. The correlation observed from this comparison concluded that 84% of mRNA molecules were overlapped with mRNA molecules of heart and 80% were overlapped with mRNA molecules of other tissues. mRNA molecules of cellular components of the blood are costing minimal and feasible to access in order to substitute gene expression in other tissues.

The contributions found in contemporary literature are specific to discover the influential genes of Myocardial Infarction. None of these are capable to identify the given gene expression is prone to CAD under MI and UA or the expression is salubrious. This evinces the need of novel contributions to discover the state of a given gene expression is prone to CAD under MI and UA or salubrious. This helps to deploy the case based reasoning to treat the patients prone to CAD under MI and UA differently. In this regard this manuscript attempted to define metaheuristic approach on gene expression data (MAGED) to discover the state of a given gene expression is prone to CAD under MI and UA or salubrious. The MAGED is machine learning strategy that learns from the labeled gene expression data of Cardia Vascular Diseased, Unstable Angina, Myocardial Infarction and Salubrious cases.

III. Metaheuristic Approach on Gene Expression Data

The objective of the MAGED is to define a metaheuristic scale by the knowledge gained from the given gene expression data. In order to this the given gene expressions are partitioned into their respective categories of coronary artery disease (CAD), unstable angina (UA), Myocardial Infarction (MI) and salubrious (blood samples taken were diagnosed as healthy). The data also includes gene expression data collected from the blood samples taken from the people clinically proven as normal.

The genes involved in each gene expression are considered as features of the respective category. Since the gene expression contains dense number of genes and majority of them may be insignificant to respective category of the disease. Henceforth, the feature optimization process (see sec 3.1) will be carried out to eliminate these insignificant features. The gene range will be discretized further to compare two genes through equality by approximation (see sec 3.2). Afterwards the confidence of each feature towards all categories of gene expression data will be assessed (see sec 3.3) that follows the assessment of each gene expression confidence against the features of all categories (see sec 3.4). Further the confidence obtained for each feature and gene expression of respective category will be used as input to define the metaheuristic scales to estimate the scope of coronary artery disease, the unstable angina and myocardial infarction.

a) Feature Optimization

For each disease context considered, the gene expression dataset $D_i = \{e(i)_1, e(i)_2, ...e(i)_{|D_i|}\}$ of size $|D_i|$ will be considered for training towards defining metaheuristic scale. Each gene expression is represent-ted by sequence of genes for the set of features selected of respective diseases context. This description binds to all datasets ofgene expressions representing coronary artery diseases, Unstable Angina, Myocardial Infarction.

Let $D_n = \{e(n)_1, e(n)_2, ..., e(n)_{|D_n|}\}$ be the set of gene expressions collected from the blood samples of salubrious cases. The sets $F_i = \{f(i)_1, f(i)_2, ..., f(i)_{|F_i|}\}$ and $F_n = \{f(n)_1, f(n)_2, ..., f(n)_{|F_n|}\}$ are feature sets of gene expressions represented by D_i and D_n respectively. The attribute set $G(i)_j = \{g(ij)_1, g(ij)_2, .\ g(ij)_{|G(i)_j|}\}$ be the set of genes as values observed for feature $f(i)_j$ of gene expressions represented by D_i. Similarly the attribute set $G(n)_j = \{g(nj)_1, g(nj)_2, .\ g(nj)_{|G(n)_j|}\}$ be the set of genes as values observed for feature $f(n)_j$ of gene expressions represented by D_n.

Since the gene expression is the combination of numerous count of genes, the size of feature set can lead to process complexity. In order to overcome the

process complexity, the insignificant features should be identified and discarded. The feature $f(i)_j$ of F_i is said to be insignificant feature, if genes $G(i)_j$ of $f(i)_j$ are almost similar to the genes $G(n)_j$ of feature $f(n)_j$ of F_n .Hence to identify the insignificant features, we adopt hamming distance that applied on genes of each feature as vectors from each disease and normal cases. The hamming distance with 0 or less than the given threshold indicates that the respective feature is insignificant. The process of hamming distance is explored below:

i. *Hamming Distance*

The value of Hamming Distance obtained here is to denote the difference between genes assigned to same feature from gene expression data of diseased and normal cases. This is one of the significant strategy to assess the difference between to elements in coding theory.

The hamming distance between given vectors $CX = \{cx_1, cx_2, ..., cx_n\} \& CY = \{cy_1, cy_2, ..., cy_m\}$ of size n and m respectively will be measured as follows:

Let $CZ \leftarrow \phi$ // is a vector of size 0

$foreach\ \{i \exists i = 1, 2, 3, ... \max(n.m)\}$ Begin

$if\ (\{cx_i \exists cx_i \in CX\} - \{cy_i \exists cy_i \in CY\}) \equiv 0\ then$

$CZ \leftarrow \{cx_i \exists cx_i \in CX\} - \{cy_i \exists cy_i \in CY\}$

Else

$$CZ \leftarrow 1$$

End

$$hd_{CX \leftrightarrow CY} = \sum_{j=1}^{|CZ|} CZ\{i\}$$

// $hd_{CX \leftrightarrow CY}$ is the hamming distance between CX and CY , $CZ\{i\}$ is the i^{th} element of the vector CZ and $|CZ|$ is the size of the vector CZ

b) Gene and Gene Expression Confidence Assessment

Then these genes found for each optimal feature of respective gene expression data set and the gene expressions of that data set will be used further to assess the gene and gene expression confidence.

In order to this, initially the gene pairs will be defined such that each pair contains two genes and each gene representing different feature of the same dataset. Then we assess the associativity support of each gene pair. The associativity support can be described as the ratio of gene expressions contains that pair against the total number of gene expressions in respective dataset. The process of assessing associativity support of each gene pair is described in following section (see sec 3.2.1).

ii. *Assessing gene pair correlation*

Let P_i be the set and contains all possible unique gene pairs from respective dataset D_i. The possible unique gene pairs will found as follows:

For each gene expression $e(i)_j$ of respective dataset D_i, find all possible unique pairs of genes and add to P_i. Then correlation of each pair $\{p_j \exists p_j \in P_i\}$ as follows.

Let $\{g_k \exists g_k \in p_j\}$ and $\{g_l \exists g_l \in p_j\}$ be the two genes paired as $\{p_j \exists p_j \in P_i\}$, then the correlation $s(p_j)$ of the pair p_j is

$$s(p_i) = \frac{\sum_{v=1}^{|D_i|} \{1 \exists \{g_k, g_l\} \subseteq e(i)_v\}}{|D_i|}$$

//The ratio of number of gene expressions contain both genesagainst total number of genes

The correlation of each pair of genes found in gene expressions of each respective gene expression data set of coronary artery disease, unstable angina, myocardial infarction and normal cases should be estimated using the process explored in sec 3.2.1.

iii. *Assessing Gene and Gene Expression Confidence*

In order to assess the confidence of genes and gene expressions of respective gene expression dataset D_i, a mutual relation graph will be formed between gene expressions and genes of respective D_i. There will be an edge between a gene and gene expression if and only if the selected gene exists in that gene expression. Then each edge between gene and gene expression is weighted as follows.

$\forall_{j=1}^{|G(i)|} \{g_j \exists g_j \in G(i)\}$ Begin

$\forall_{k=1}^{|D_i|} \{e(l)_k \exists g_j \in e(l)_k\}$ Begin

$$w_{g_j} = 0$$

$\forall_{m=1}^{|e(l)_k|} \{g_m \exists g_m \in e(l)_k \wedge g_j \neq g_m\}$ Begin

$p_m = \{g_j, g_m\}$

$$w(g_j) += s(p_m)$$

End

$$w_{g_j \leftrightarrow e(l)_k} = \frac{w(g_j)}{|e(l)_k| - 1}$$

End

End

The weights obtained for edges between genes and gene expressions in mutual graph are further used to assess the gene and gene expression confidence towards respective CAD (coronary artery disease), UA (unstable angina), MI (myocardial infarction) and Normal datasets.

Further we measure the each feature confidence towards gene expression dataset D_i as follow

$\forall_{j=1}^{|G(i)|} \{g_j \exists G(i) \ni g_j\}$ Begin

$$c_{g_j \Rightarrow D_i} = \sum_{k=1}^{|D_i|} \{w(g_j) \exists e(i)_k \ni g_j \wedge D_i \ni e(i)_k\}$$

//aggregating the weight of gene g_j towards each gene expression $e(i)_k$ of respective dataset D_i and the same is considered as the respective gene confidence towards dataset D_i

End

Similarly each respective gene expression confidence towards gene expression dataset D_i is measured as follows

$$\overset{|D_i|}{\underset{j=1}{\forall}}\{e(i)_j \exists D_i \ni e(i)_j\}$$ Begin

$$c_{e(i)_j \Rightarrow D_i} = \sum_{k=1}^{|G(i)|}\{w(g_k) \otimes c_{g_k \Rightarrow D_i} \exists e(i)_j \ni g_k \wedge D_i \ni e(i)_j\}$$

//The sum of product of each gene weight and the respective gene confidence, such that the gene exists in selective gene expression is the confidence of that gene expression

End

The confidence of genes and gene expressions of each respective gene expression data set of CAD, UA, MI and salubrious cases should be estimated using the process explored in sec 3.2.2.

c) Defining metaheuristics to CAD, UA, MI and Salubrious scope

Further the confidence of gene expressions of gene expression datasets D_{CAD}, D_{UA}, D_{MI} and

$$m_{CAD} = \frac{\sum_{i=1}^{|D_{CAD}|}\{c_{e(CAD)_i \Rightarrow D_{CAD}} \exists D_{CAD} \ni e(CAD)_i\}}{|D_{CAD}|}$$ //Aggregate

mean of the respective gene expressions confidence of coronary artery disease gene expression dataset D_{CAD}

In order to identify the lower and upper bounds of m_{CAD}, the mean absolute distance of D_{CAD} is assessed as follows

$$m_{CAD}^{ad} = \frac{\sum_{i=1}^{|D_{CAD}|}\sqrt{\left(m_{CAD} - c_{e(CAD)_i \Rightarrow D_{CAD}}\right)^2}}{|D_{CAD}|}$$

Then the lower and upper bounds of m_{CAD} is assessed as

$ml_{CAD} = m_{CAD} - m_{CAD}^{ad}$ // lower bound of m_{CAD}

$mu_{CAD} = m_{CAD} + m_{CAD}^{ad}$ // upper bound of m_{CAD}

Similarly metaheuristics for UA (unstable angina), MI (Myocardial infarction) and salubrious (healthy) scope

$$m_{UA} = \frac{\sum_{i=1}^{|D_{UA}|}\{c_{e(UA)_i \Rightarrow D_{UA}} \exists D_{UA} \ni e(UA)_i\}}{|D_{UA}|}$$ //Aggregate

mean of the respective gene expressions confidence of Unstable Angina gene expression dataset D_{UA}

The mean absolute distance of D_{UA} is

$$m_{UA}^{ad} = \frac{\sum_{i=1}^{|D_{UA}|}\sqrt{\left(m_{UA} - c_{e(UA)_i \Rightarrow D_{UA}}\right)^2}}{|D_{UA}|}$$

Then the lower and upper bounds of m_{UA} is assessed as

$ml_{UA} = m_{UA} - m_{UA}^{ad}$ // lower bound of m_{UA}

$mu_{UA} = m_{UA} + m_{UA}^{ad}$ // upper bound of m_{UA}

$$m_{MI} = \frac{\sum_{i=1}^{|D_{MI}|}\{c_{e(MI)_i \Rightarrow D_{MI}} \exists D_{MI} \ni e(MI)_i\}}{|D_{MI}|}$$ //Aggregate

mean of the respective gene expressions confidence of myocardial Infarction gene expression dataset D_{MI}

The mean absolute distance of D_{MI} is

$$m_{MI}^{ad} = \frac{\sum_{i=1}^{|D_{MI}|}\sqrt{\left(m_{MI} - c_{e(MI)_i \Rightarrow D_{MI}}\right)^2}}{|D_{MI}|}$$

Then the lower and upper bounds of m_{MI} is assessed as

$ml_{MI} = m_{MI} - m_{MI}^{ad}$ // lower bound of m_{MI}

$mu_{MI} = m_{MI} + m_{MI}^{ad}$ // upper bound of m_{MI}

$$m_N = \frac{\sum_{i=1}^{|D_N|}\{c_{e(N)_i \Rightarrow D_N} \exists D_N \ni e(N)_i\}}{D_N}$$ //Aggregate mean of the respective gene expressions confidence of salubrious gene expression dataset D_N

The mean absolute distance of D_N is

$$m_N^{ad} = \frac{\sum_{i=1}^{|D_N|}\sqrt{\left(m_N - c_{e(N)_i \Rightarrow D_N}\right)^2}}{|D_N|}$$

Then the lower and upper bounds of m_N is assessed as

$ml_N = m_N - m_N^{ad}$ // lower bound of m_N

$mu_N = m_N + m_N^{ad}$ // upper bound of m_N

d) Predicting the state of gene expression

The metaheuristics devised (see section 3.3) will be used further to assess the CAD, UA and MI scope of a given gene expression e. The confidence of given gene expression

$$c_{e \Rightarrow CAD} = \frac{\sum_{i=1}^{|G(D_{CAD})|}\left\{c_{g_i \Rightarrow CAD} \otimes w(g_i) \exists g_i \in G(D_{CAD}) \wedge e \ni g_i\right\}}{\sum_{j=1}^{|G(D_{CAD})|}\left\{c_{g_j \Rightarrow CAD} \otimes w(g_j) \exists g_j \in G(D_{CAD})\right\}}$$

// the aggregate of product of each gene confidence and weight of that exists in $G(D_{CAD})$ and e, which divides by the aggregate of confidence of all genes exists in $G\ D(\ _{CAD}\ )$

Further the confidence of e towards D_{UA}, D_{MI} and D_N assessed as :

$$c_{e\Rightarrow UA} = \frac{\sum_{i=1}^{|G(UA)|}\left\{c_{g_i\Rightarrow UA}\otimes w(g_i)\exists g_i \in G(UA)\wedge e \ni g_i\right\}}{\sum_{j=1}^{|G(UA)|}\left\{c_{g_j\Rightarrow UA}\otimes w(g_j)\exists g_j \in G(UA)\right\}}$$

// the aggregate of product of each gene confidence and weight of that exists in $G(D_{UA})$ and e, which divides by the aggregate of confidence of all genes exists in $G(D_{UA})$.

$$c_{e\Rightarrow MI} = \frac{\sum_{i=1}^{|G(MI)|}\left\{c_{g_i\Rightarrow MI}\otimes w(g_i)\exists g_i \in G(MI)\wedge e \ni g_i\right\}}{\sum_{j=1}^{|G(MI)|}\left\{c_{g_j\Rightarrow MI}\otimes w(g_j)\exists g_j \in G(MI)\right\}}$$

// the aggregate of product of each gene confidence and weight of that exists in $G(D_{UA})$ and e, which divides by the aggregate of confidence of all genes exists in $G(D_{UA})$.

$$c_{e\Rightarrow N} = \frac{\sum_{i=1}^{|G(N)|}\left\{c_{g_i\Rightarrow N}\otimes w(g_i)\exists g_i \in G(N)\wedge e \ni g_i\right\}}{\sum_{j=1}^{|G(N)|}\left\{c_{g_j\Rightarrow N}\otimes w(g_j)\exists g_j \in G(N)\right\}}$$

// the aggregate of product of each gene confidence and weight of that exists in $G(D_{UA})$ and e, which divides by the aggregate of confidence of all genes exists in $G(D_{UA})$.

Then these confidence values of gene expression e with respect to CAD, UA, MI and N will be used to estimate the given expression state is salubrious, prone to coronary artery disease, Unstable Angina or Myocardial Infarction according to the following conditions.

$$(c_{e\Rightarrow CAD} \geq mu_{CAD}) \vee (c_{e\Rightarrow UA} \geq mu_{UA}) \vee (c_{e\Rightarrow MI} \geq mu_{MI})$$

Coronary Artery Disease Confirmed (highly prone to either of three disease conditions)

$$(c_{e\Rightarrow CAD} \geq m_{CAD}) \wedge (c_{e\Rightarrow UA} \geq ml_{UA}) \wedge (c_{e\Rightarrow MI} \geq ml_{MI})$$

Coronary Artery Disease Confirmed (prone to CAD and either or both of the UA and MI)

$$\text{If} \begin{pmatrix} (c_{e\Rightarrow CAD} \geq ml_{CAD}) \wedge \\ (c_{e\Rightarrow UA} \geq ml_{UA}) \wedge \\ (c_{e\Rightarrow MI} \geq ml_{MI}) \wedge \\ (c_{e\Rightarrow N} < m_{MI}) \end{pmatrix}$$

Then Prone to Coronary Artery Disease

$$\text{if} \begin{pmatrix} (c_{e\Rightarrow CAD} < ml_{CAD}) \wedge \\ (c_{e\Rightarrow UA} < ml_{UA}) \wedge \\ (c_{e\Rightarrow MI} < ml_{MI}) \wedge \\ (c_{e\Rightarrow N} > m_{MI}) \end{pmatrix}$$

Then Salubrious state Confirmed

$$\text{if} \begin{pmatrix} (c_{e\Rightarrow CAD} < m_{CAD}) \wedge \\ (c_{e\Rightarrow UA} < m_{UA}) \wedge \\ (c_{e\Rightarrow MI} < m_{MI}) \wedge \\ (c_{e\Rightarrow N} \geq mu_{MI}) \end{pmatrix}$$

Then Prone to Salubrious state

IV. Experimental Study

The experimental study was carried out on a set of gene expressions taken from multiple benchmark datasets [19]. The number of gene expressions used are 1114, which are the combination of coronary artery Disease (286 expressions), Unstable Angina (275 expressions), Myocardial Infarction (277 expressions) and salubrious condition (276 expressions). The gene expressions of respective category are considered as separate datasets labeled as D_{CAD}, D_{UA}, D_{MI} and D_N. Each dataset D_{CAD}, D_{UA}, D_{MI} and D_N partitioned into test and training sets. The 75% of gene expressions of each dataset are considered as training set and rest 25% of gene expressions considered as test set.

The metaheuristics obtained from the given training set were explored in table 1.

Table 1: The metaheuristics obtained from training data

Training Set	834 (CAD:214, UA:206,MI:207, N:207)
m_{CAD}	0.582474187
m^{ad}_{CAD}	0.095593654
ml_{CAD}	0.486880533
mu_{CAD}	0.678067841
m_{UA}	0.615957277
m^{ad}_{UA}	0.103864099
ml_{UA}	0.512093178
mu_{UA}	0.719821376
m_{MI}	0.646638853

m_{MI}^{ad}	0.099722167
ml_{MI}	0.546916686
mu_{MI}	0.74636102
m_N	0.631593026
m_{MI}^{ad}	0.068999373
ml_N	0.562593653
mu_N	0.700592398

Table 2 : The prediction statistics of the SDS

Test Set	280 (CAD:72, UA:69,MI:70, N:69)
True Positives	197
True Negatives	54
False Positives	15
False Negative	14
CAD, UA and MI gene expression Prediction Value (positive prediction value, PPV)	0.929245283
Salubrious gene expression Prediction Value (Negative Prediction value, NPV)	0.794117647
Detection Accuracy	0.896428571
AD, UA and MI gene expression prediction Rate (True Positive Rate)	0.933649289
Salubrious gene expression Prediction rate (True Negative Rate)	0.782608696

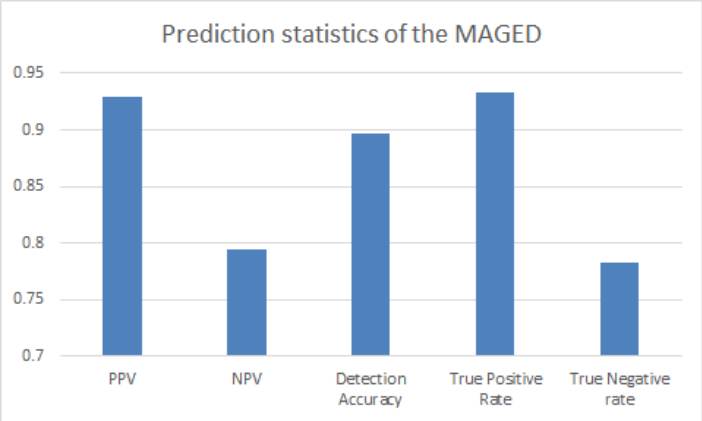

Figure 1: The prediction statistics observed for MAGED

The 280 (CAD: 72, UA: 69, MI: 70, N: 69) gene expressions were used to assess the prediction accuracy of the proposed MAGED. The MAGED assessed the given input gene expressions such that 197gene expressions are true positives (the detection of CAD, UA and MI gene expressions are true), 15gene expressions are false positive (falsely detected as CAD, UA or MI), 54gene expressions are true negatives (detecting gene expressions as salubrious is true) and 14gene expressions are false negative (detecting gene expressions as salubrious is false). Hence the CAD, UA or MI gene expression prediction value (also known as precision or positive prediction value) is 0.93, Salubrious Gene Expression prediction value is 0.79, the CAD, UA and MIgene expression detection rate (also known as sensitivity) is 0.93, the salubrious gene expression detection rate (also known as specificity) is 0.782 and the overall success rate (also known as accuracy, which is the ratio between true prediction of all types of gene expressions and all given number of gene expressions) is 0.90. These statistics indicating that the MAGED is find to significant to identify the CAD, UA and MI prone gene expressions with success percentage of 93% (since sensitivity is 0.93), but the detection of salubrious cases, the success rate is 78% (since specificity is 0.782). The computer aided medical diagnosis should

be more robust to deliver high sensitivity at the cost of specificity. Hence the Model MAGED is scalable and robust to predict the CAD, UA and MI prone gene expressions.The prediction statistics observed from the experimental study of the MAGED are visualized in fig1.

V. Conclusion

This paper introduced a learning model that device heuristics to scale the given patient record is disease prone or normal. The proposed learning model delivers two heuristics called Scale to Diseased health Scope and Scale to Normal Health Scope. In contrast to the existing benchmarking models, these heuristics are further used as scales to assess the given patient record is disease prone or normal. The medical records labeled as diseased and normal are used to device the heuristics $sdhs$ and $snhs$ respectively. In order to this all unique values of all the attributes are considered as features, and then the influence weight of these features towards their respective datasets. The influence weights further will be used to assess the influence weight of the each record in dataset. From these influence weights of the records of respective dataset will be used to assess the proposed heuristics. The experimental results are optimistic and concluding the prediction accuracy and robustness. This work can be extended to identify the impact of feature correlation towards minimizing the process and computational complexity of the learning process.

References Références Referencias

1. Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, et al. Heart disease and stroke statistics—2015 update: a report from the American Heart Association. Circ. 2015; 131:29–32.
2. Mendis, S., Thygesen, K., Kuulasmaa, K., Giampaoli, S., Mähönen, M., Blackett, K. N., & Lisheng, L. (2011). World Health Organization definition of myocardial infarction: 2008–09 revision. International journal of epidemiology,40(1), 139-146.
3. Thygesen K, Alpert JS, White HD. Universal definition of myocardial infarction. Europ Heart J. 2007; 28:2525–2538.
4. Eggers KM, Lind L, Venge P, Lindahl B. Will the universal definition of myocardial infarction criteria result in an over diagnosis of myocardial infarction?. The Amer J of Card. 2009; 103:588–591.
5. Wang Z, Luo X, Lu Y, Yang B. miRNAs at the heart of the matter. J of Mol Med. 2008; 86:771–783.
6. de Planell-Saguer M, Rodicio MC. Detection methods for microRNAs in clinic practice. Clin Biochem. 2013; 46:869–878. doi: 10.1016/j.clinbiochem.2013.02.017 PMID: 23499588.
7. Melander O, Newton-Cheh C, Almgren P, Hedblad B, Berglund G, Engström G, et al. Novel and conventional biomarkers for prediction of incident cardiovascular events in the community. The J of the Amer Med Assoc. 2009; 302:49–57.
8. Shah T, Casas JP, Cooper JA, Tzoulaki I, Sofat R, McCormack V, et al. Critical appraisal of CRP measurement for the prediction of coronary heart disease events: new data and systematic review of 31 prospective cohorts. Inter J of Epid. 2009; 38: 217–231.
9. Wilson PWF, Pencina M, Jacques P, Selhub J, D'Agostino R, O'Donnell CJ. C-reactive protein and reclassification of cardiovascular risk in the Framingham Heart Study. Circ: Card Qual and Outc. 2008; 2: 92–97.
10. Pedrotty DM, Morley MP, Cappola TP. Transcriptomic biomarkers of cardiovascular disease. Prog in Card Dis. 2012; 55: 64–69.
11. Randi AM, Biguzzi E, Falciani F, Merlini P, Blakemore S, Bramucci E, et al. Identification of differentially expressed genes in coronary atherosclerotic plaques from patients with stable or unstable angina by cDNA array analysis. J of Throm and Haem. 2003; 1: 829–835.
12. Archacki S, Angheloiu G, Tian XL, Tan FL, DiPaola N, Shen GQ, et al. Identification of new genes differentially expressed in coronary artery disease by expression profiling. Phys Genom. 2003; 15: 65–74.
13. Elashoff MR, Wingrove JA, Beineke P, Daniels SE, Tingley WG, Rosenberg S, et al. Development of a blood-based gene expression algorithm for assessment of obstructive coronary artery disease in nondiabetic patients. BMC Med Genom. 2011; 4: 4–26.
14. Kittleson MM, Ye SQ, Irizarry RA, Minhas KM, Edness G, Conte JV, et al. Identification of a gene expression profile that differentiates between ischemic and nonischemic cardiomyopathy. Circ. 2004; 110: 3444–3451.
15. Kittleson MM, Minhas KM, Irizarry RA, Ye SQ, Edness G, Breton E, et al. Gene expression analysis of ischemic and nonischemic cardiomyopathy: shared and distinct genes in the development of heart failure. Phys Genom. 2005; 21: 299–307.
16. Min KD, Asakura M, Liao Y, Nakamaru K, Okazaki H, Takahashi T, et al. Identification of genes related to heart failure using global gene expression profiling of human failing myocardium. Bioch and Biophy Res Comm. 2010; 393: 55–60.
17. Suresh R, Li X, Chiriac A, Goel K, Terzic A, Perez-Terzic C, et al. Transcriptome from circulating cells suggests dysregulated pathways associated with long-term recurrent events following first-time myocardial infarction. J of Mol and Cell Card. 2014; 74: 13–21.
18. Liew CC, Ma J, Tang HC, Zheng R, Dempsey AA. The peripheral blood transcriptome dynamically reflects system wide biology: a potential diagnostic tool. The J. of Lab and Clin Med. 2006; 147: 126–[132].
19. https://www.ebi.ac.uk/ega/datasets