

# Intelligent Heart Diseases Prediction System Using A New Hybrid Metaheuristic Algorithm

Mrs S.M Uma

Research Scholar  
Computer Science and Engineering  
Kings College of Engineering

Dr.E.Kirubakaran

Outsourcing Unit  
BHEL  
Trichy

**Abstract**— Data mining is a crucial step in discovery of knowledge from large datasets. In recent years, Data mining has found its significant hold in every field including health care. Mining process is more than the data analysis which includes classification, clustering, and association rule discovery. Predicting the outcome of a disease is one of the most interesting and challenging tasks in which to develop data mining applications. our research focus on this aspect of medical diagnosis by learning pattern through the collected data of heart dataset and to develop intelligent medical decision support system to help physicians. Different classification methods such as Neural Networks and Decision Trees were applied to predict the presence of heart disease and to identify the most significant factor which contributes for the cause of the disease, while classification rule discovery was used to identify the effect of diet, lifestyle, and environment on the outcome of the disease. This paper presents a hybrid swarm intelligence optimization algorithm called GA-ACO algorithm that combines the evolution ideas of genetic algorithms and ant colony optimization based on compensation for solving healthcare problem. Thus the aco-ga model along with the extended definition for identifying heart disease provided a good classification accuracy based model.

**Keywords:** Active learning, domain expert, Gene selection, ant colony optimization, genetic algorithm

## I. INTRODUCTION

Swarm Intelligence is an important research topic based on the collective behaviour of decentralized and self organized. It consists of a population which stimulates the animal behaviour in the real world. Now there are many swarm intelligence optimization algorithms, such as genetic algorithms, particle swarm optimization, ant colony optimization, bee colony algorithm, differential evolution, fish-warm algorithm,.....etc. With the rapid development of the social economy, science and technology, the large-scale optimization problems are becoming too complicated to obtain satisfactory results by using the single swarm intelligence optimization algorithm. So it is necessary to utilize hybrid swarm intelligence optimization algorithm to solve all kinds of the complex large-scale optimization problems.

The hybrid swarm intelligence optimization algorithm is to utilize the single swarm intelligence optimization algorithm of complementary advantages and the

value-added information to overcome the insufficiencies of the single swarm intelligence optimization algorithm in order to enhance the efficiency of solving the complex large-scale optimization problems. Many researchers presented some personalization hybrid swarm intelligence optimization algorithms according to the practical problems on the applications of health care in the real world in various fields, such as artificial intelligence, biology, mathematics, physics, and operations research.

A major challenge facing healthcare organization is the provision of quality services at affordable cost. Quality service implies diagnosing patients correctly and administering Treatments that are effective. Poor clinical decision can lead to disastrous consequences which are therefore unacceptable. Hospitals must also minimize the cost of clinical tests. They can achieve these results by employing appropriate computer-based information and decision support system. However, ACO suffers from following three shortcomings. first, the ants are easy to be trapped in local extrema, second, the convergence speed is sometimes too slow[1]. Third, the result obtained are sensitive to initial values[2-3]. In other words, ACO performs well as a local search algorithm other than a global search algorithm[4] considering genetic algorithm (GA) is accomplished in robust global search. Therefore, we combine the GA and ACO, producing a GACO algorithm for Optimization problem.

A new hybrid heuristic approach named GACO algorithm is presented to solve the healthcare problem. In this paper, we present a attribute selection method using ACO with GA. First, the wilcoxon rank sum test (deng et al.2004) is used to preprocess The original attribute expression data, and then the proposed by hybrid GA/ACO is adopted to select the most important attribute subset using tenfold cross validation scheme. Finally, a classifier is trained based on the attribute subset obtained from ACO/GA and used to predict the testing samples.

The organization of this paper is as follows. The proposed GA/ACO method is presented in detail in sec2. In sec3 presented the experimental results. Conclusion of this paper are addressed in sec.4.

## II. PROPOSED METHOD

### 2.1 GENETIC ALGORITHM

Genetic algorithm (GA) is an adaptive optimization search algorithm simulating the evolutionary ideas of natural selection (Goldberg 1989). A genetic algorithm represent a non-traditional search and optimization method that is finding important applications in engineering optimization Further, each genetic algorithm is a computer simulation that operates on a constant-sized population of individuals, called the search space. The individuals in the initial population are stochastically selected, recombined, mutated, and either eliminated or retained according to their relative levels of fitness[5].

With some modification of the genetic operators, the real coded GAs perform better than the binary-coded GAs for problem. suppose that the parameter space is  $(O,M)$  To implement a GA we discretise  $(O,M)$  coding mechanism give rise to equally spaced

possible decision. We define the grid spacing  $\Delta$  of a code to be difference between two consecutive coded decision .For a binary code of fixed length  $L$ , grid spacing  $\Delta = m/(2^L-1)$  in using a binary coded we determine how many bits we should use in the binary string to give a grid spacing ,which is finite enough. For real coding the precision is given by the fixed decimal point representation of real numbers on a computing where for eg  $\Delta = 0.0000000001$  for a ten decimal point representation. The crossover operator of a real coded Gas is performed by the borrowing concept of convex combination. The random mutation operator is used to change the attribute with a random number in the problem's domain.

Genetic algorithm goes through many generation of explorative and exploitative search until it convergence to a near optimal solution. we briefly introduce the genetic algorithm for healthcare problem used in our hybrid algorithm.

#### Procedure genetic algorithm

```

p ← Generate Random Population()
repeat
  (p1,p2) ← Select parent pair(p)

  s ← crossover(p1,p2)
  s ← Mutation(p1,pm)
  p ← population Update(p,s)
Until termination condition met
end

```

Figure 1: Genetic Algorithm for Healthcare Problem

Fig1. & Fig2shows that, a population of solutions  $P$  is generated randomly. Then,within each evolutionary cycle, two candidate solutions designated as parents

are selected. The two parents undergo crossover to produce an offspring solution  $s$ . After that,  $s$  mutates with a probability of mutation rate  $Pm$ . Finally, the resulting  $s$ , together with the existing population  $P$ , constitutes a new population for the next evolutionary cycle. This process is repeated until the termination condition is met. The genetic operators involved are briefly described as follows.

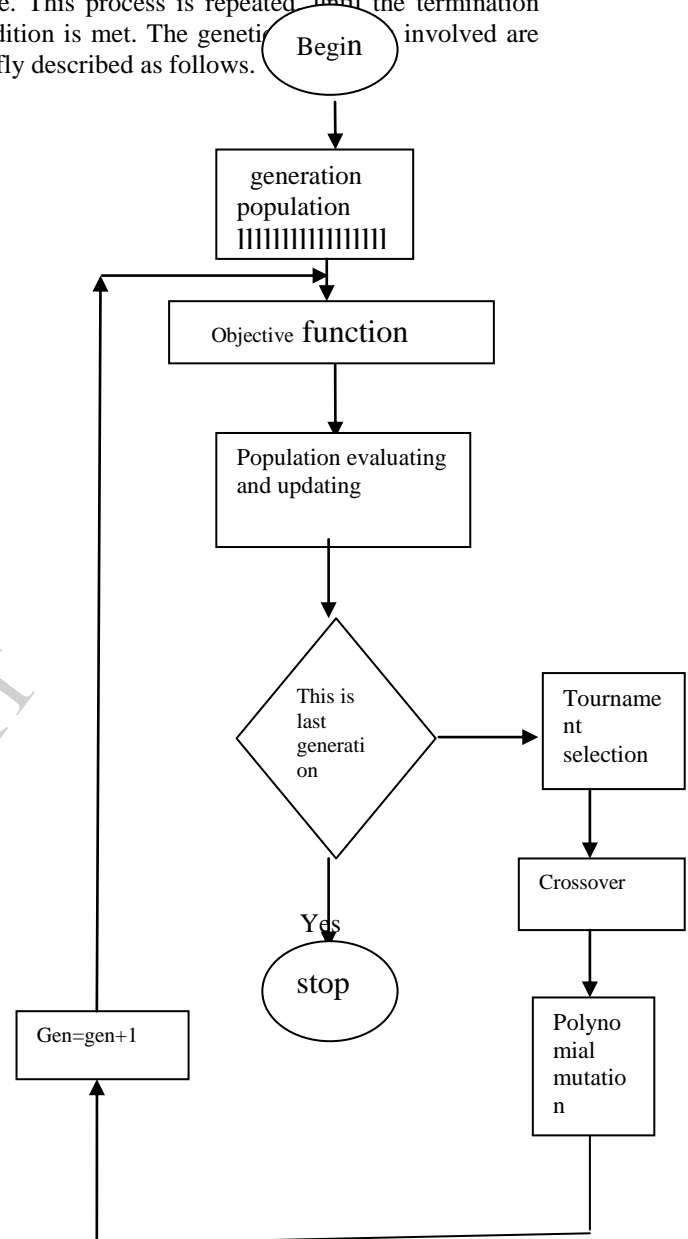


Fig 2.Flowchart of GA

Hence, in this work, we have used a real coded genetic algorithm. Real parameters are used directly,and optimization is easier than for binary coded GAs. However, in this case, the main problem that arises is the question of how to use a pair of real parameters to create a new pair of offspring and how to perturb a variable for the mutation. [6]have provided a good overview of many real parameters used for crossover and mutation operators.

In our genetic algorithm, tournament selection is used as the parents selection procedure. In this mechanism, two individuals are randomly selected from the reproduction pool. The better individual is designated as one parent. In this way, two parents are selected and the better parent is designated as the first parent for reproducing one offspring by crossover. crossover is used in GA to inherit constructive information from parents through the generation. mutation serves as the secondary search operator of GA for exploring new search regions by altering a certain number of genes of a chromosome randomly.

## 2.2 Ant colony Optimization

Ant colony optimization (ACO) is a class of constructive metaheuristic algorithms. It imitates the behaviors of real ants of a colony when foraging for food. Each artificial ant agent constructs a solution based on the constructive information, which is termed pheromone, provided by previous ant agents that have already built solutions. After having built new solutions, the artificial ants update the pheromone traces, taking into account the quality of the existing solutions. We briefly introduce the ACO algorithm for health care used in our hybrid algorithm. ACO algorithms can be used in any optimization problem, if the following aspects are provided [8]: *Figure2. show the general framework of the ant colony optimization for healthcare in our implementation*

### Procedure ant colony optimization

```

P ← Generate Random Population()
Ph ← pheromoneinitialization()
Repeat
    Ph ← Pheromone Update(p)
    S ← solution construction (ph)
    P ← Population Update(P,S)
Until termination condition met
End ant colony optimization
    
```

Figure2. AntColonyOptimization for HCP

### 2.2.1 Pheromone initialization

At the beginning of each iteration of the WHILE loop, the pheromone values on the edges between all terms are initialized with the same amount of pheromone. The initial pheromone is calculated according to equation (1)

$$\tau_{ij} = \frac{1}{\sum_{i=1}^a b_i} \tag{1}$$

Where a is the total number of attribute and bi is the number of possible values that can be taken on by attribute. since all the pheromone values are the same, the first ant has no historical information to guide its search.

### 2.2.2 Solution construction in ACO Metaheuristic :

At each step, each ant constructs a solution. Then each ant will share its solution feedback with the entire colony by updating a global data structure called the pheromone matrix. This data structure simulates the pheromone trails. Each entry in the pheromone matrix shows the desirability of each solution component. At the end of each iteration, the pheromone associated with each solution component is reinforced based on the quality of the solution that comprises the particular solution component. In subsequent iterations, [9] ants will use the pheromone intensities of available solution components to guide solution construction. As a result of repeated pheromone reinforcements, a subset of solution components will emerge to have pheromone intensities much higher than the others. An ant incrementally adds terms in the antecedent part of the rule that is constructing. the selection of the next term is subject to the restriction that the attribute  $A_i$  of that term should not be already present in the current partial rule. in other words, once a term has been included in the rule then no other terms containing that attribute can be considered. subject to this restriction, the probability of selection of a term for addition in the current partial rule is given by the equation(2)

$$\rho_{ij} = \frac{\tau_{ij} \eta_{ij}}{\sum_{i=1}^a x_i \sum_{j=1}^b \tau_{ij} \eta_{ij}} \tag{2}$$

$\eta$  is a problem-dependent heuristic value for term j  
T is the number of pheromone currently available (at time t) on the connection between attribute i and value j

The best subset contains the least number of features that most contribute to accuracy. The whole search space for optimization contains all possible subsets of features, meaning that its size is:  $2^n$ , where n is the dimensionality. Therefore, this search for an optimal set of attributes would be time-consuming and computationally expensive if n is large.

### 2.2.4 Heuristic function

The heuristic value is calculated using the entropy function. However, we believe that the ACO algorithm does not need accurate information in this heuristic value since the idea of the pheromone should

compensate the small potential errors in the heuristic values. [10] In other words, a simpler heuristic value may do the job as well as the complex one. As a result, we propose in this paper the following easily computable density estimation equation(3):

$$\eta_{ij} = \frac{\text{majority\_class}T_{ij}}{|T_{ij}|} \quad (3)$$

Where majority\_class  $\tau_{ij}$  is the majority class

partition in  $\tau_{ij}$ . The heuristic value when considering

the first term of the rule antecedent is calculated on the basis of the following Laplace-corrected confidence of a term specified in the equation(4)

$$\eta_{ij} = \frac{|term_j \text{ class}_k| + 1}{|term_j| + \text{total\_classes}} \quad (4)$$

Where Total\_classes is the number of classes present in the data set. This heuristic function has the advantages of penalizing the terms that would lead to very specific rules and thus helps to avoid overfitting. For example, if a term occurs in just one training sample and its class is the chosen class, then its confidence is 1 without the Laplace correction.

### 2.2.5 Rule Termination

An ant continues to add terms in the rule it is constructing and stops only when all the samples covered by the rule have homogenous class label [11]. The other possibility of termination is that there are no more terms to add. Due to these termination criteria, it might happen that a constructed rule may cover only one training sample.

### 2.2.6 Quality of a rule

When an ant has completed the construction of a rule its quality is measured. The quality, Q, of a rule is computed by using  $\text{confi}(t+1) = (1 - \text{dence and coverage of the rule and is given by:}$

$$Q = (\text{TP} / \text{Covered}) + (\text{TP} / \text{N}) \quad (5)$$

Where TP is the number of samples covered by the rule that have the same class label as that of the rules consequent, covered is the total number of samples covered by the rule, and N is the number of samples in the training set yet uncovered by any rule in the discovered rule set. The second portion is added to encourage the construction of rules with wider coverage. Equation (5) has been used in Ant Miner+, whereas Ant Miner uses sensitivity multiplied by specificity as the quality measure.

### 2.2.7 Pheromone Update

The pheromone values are updated so that the next ant can make use of this information in its search. The amount of pheromone on the links between consecutive Terms occurring in the rule is updated according to the equation (6)

$$\tau_{ij}(t+1) = (1 - \rho)\tau_{ij}(t) + (1 - 1/1 + Q)\tau_{ij}(t) \quad (6)$$

Where  $\tau_{ij}(t)$  is the pheromone value encountered by Ant  $i$  between term  $i$  and term  $j$ . The pheromone evaporation rate is represented by  $\rho$  and Q is the quality of the rule constructed by Ant  $i$ . Equation (6) is according to the Ant system and has been previously used in Antminer3, but with a different equation for determining Q. The equation updates pheromone by first evaporating a percentage of the previously occurring pheromone and then adding a percentage of the pheromone dependant on the quality of the recently discovered rule [12]. If the rule is good then the pheromone added is greater than the pheromone evaporated and the terms become more attractive for subsequent ants. The evaporation in the equation improves exploration, otherwise the presence of a static heuristic function tends to make the ants quickly converge to the term selected by the first few ants.

### 2.3 Hybrid ACO/GA

The main idea of hybrid ACO/GA algorithm is to integrate the GA operators into the ACO algorithm. Fig 1 shows the flowchart of the hybrid ACO/GA and details are presented as below:

Step1: Generate initial population. Randomly generates  $M \times N$  initial population with binary system. M is the number of pheromone in the swarm and N stands for the length of an individual.

Step2: Compute fitness. All the individuals are evaluated by a fitness function.

Step3: Perform the parameters of ACO, generate the distribution of pheromone according to the optimization problem, calculate the transition probability of each ant and move Ant according to probability

Step4: Judge termination. If an update pheromone with new fitness cannot satisfy termination condition, go to

step5, perform GA process

Step6: Compute fitness. This step is the same as step2.

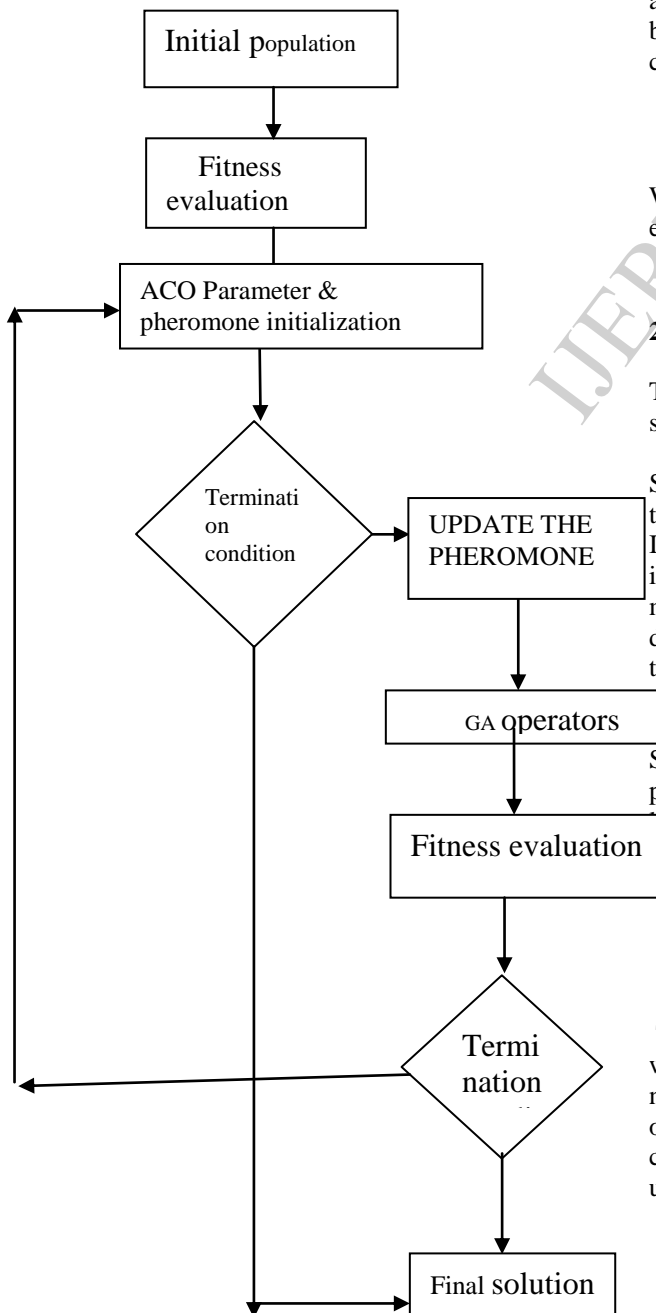
Step7: judge termination. Once the termination condition is met, output the final solution, otherwise go to step 3. The maximum number of iteration is considered as the termination criterion.

**Procedure GA –ACO**

```

p ← Generate Random Population
ph ← Phomoneinitialization
pg ← 0.5
repeat
    if(bapplyGA(Pg) = TRUE
        (P1,P2) ← Select parent pair (P)
        S ← Solution Construction (ph)
        S ← Localsearch(s)
        P ← Population update (p,s)
        Pg ← update pg ()
    until termination condition met
End GA-ACO
    
```

**Fig3.Flowchart of GACO**



In the fine search stage,GACO algorithm determines the exact position of the optimal point .The flowchart of our GACO is depicted in Fig3.

ACO keeps constructing new solutions,which replaces the worst solution in the population and improves the average quality of the parent population members.The improvement enhances the GA convergences and increase the chance of further producing better solution for GA.

**2.3.1Adaptive control parameter for P**

To determine GA and ACO weight in order to achieve an appropriate balance in explorative and exploitative behaviors. In our work, we adopted an adaptive control over the parameter P.

$$P_g = F(T) = \int f(x)dx$$

Where f(x) is the probability density function of exponential distribution.

$$f(x) = \lambda e^{-\lambda x}$$

**2.4 ATTRIBUTE SELECTION**

The basic procedure of hybrid PSO/GA based attribute selection method.Details are described as follows

Step1. The gene expression data is preprocessed by the Kolomogorov-Smirnov Test.

In this stage ,all the training sample are firstly divided into training sample and testing sample using tenfold method.The ten fold procedure is(1) the n sample are divided randomly into 10 subsets of equal size.(2) 9 of the 10 subsets are used for attribute selection and to in a classifier using attribute selected.(3) the trained subset is used to test the performance. Secondly, the Kolomogorov-Smirnov Testis performed on the training sample. In this process, a large number of redundant, noisy data are filtered by Kolomogorov-Smirnov Test.The statistic formula is as follows

$$s(f) = 1/nh \sum_{i=1}^n K(x - X_i/h)$$

The estimate of f at point x is obtained using a weighted function of observations in the h- neighbourhood of x. The weight given to each of the observation in the neighbourhood depends on the choice of kernel function. Some kernel functions are uniform kernel function like

$$K(u) = 1/2I[|u| \leq 1]$$

Gives a better estimate of the underlined density.

Increasing the bandwidth is equivalent to increasing the amount of smoothing in the estimate. Very large  $h \rightarrow \infty$  would give a flat estimate and  $h \rightarrow 0$  will lead to a needlepoint estimate giving a noisy representation of the data. For instance, with Gaussian kernel, the optimal (MISE) bandwidth is

Dataset	Total no of attributes	R/I/N	INSTANCES	CLASSES
Cleveland	5	5/0/0	297	5
statlog	5	1/4/0	270	2
Hepatitis	9	2/8/0	80	2
Demagology	20	0/20/0	358	6

$h_{opt} = 1.06\sigma n^{-1/5}$  each attribute is evaluated and ranked according to equation.

Step2: The final attribute selection is performed using hybrid ACO/GA on the training sample.

## 2.5 Experimental results

The proposed method is evaluated on four public dataset Four data sets from UCI data repository [12], Cleveland Heart Disease, Statlog-Heart, Hepatitis and Dermatology were used for this study (Table I). The Cleveland dataset consists of 13 attributes and 297 instances from the sample.statlog-heart dataset consists of 13 attributes and 270 instances from the sample.dermatology contains 34 attributes and 358 instances from the sample.The experiments are implemented using tenfold CV as discussed in sec 2.4.As the training set and testing set changing under the tenfold strategy. the data sets were reduced using CV criterion and the new reduced data sets are shown in Table II. For second comparison, single PSO and single GA methods are also used for gene selection and compared to the proposed hybrid ACO/GA method. The corresponding parameter settings of the two algorithms are the same as the hybrid ACO/GA method.

Table 3 shows the average classification accuracies (%) obtained by the three methods in 10 runs. The first number shows the average classification accuracy and the number in parenthesis is the average number of attribute selected. From Table3 ,we can see that the best classification results are 95.1%on satlog data using the hybrid ACO/GA methods, whereas, only 94.6% classification accuracy is obtained by single ACO and GA method.we can draw a conclusion that

ACO algorithm With GA operators integrated in has better performance for attribute selection compared to single ACO and GA.

TABLE I: Data set description

Dataset	Total no of attributes	R/I/N	INSTANCES	CLASSES
Cleveland	13	13/0/0	297	5
statlog	13	1/12/0	270	2
Hepatitis	19	2/17/0	80	2
Demagology	34	0/34/0	358	6

TABLE II: Reduced Dataset

Table 3: Testing accuracies (%) obtained by the three methods on four dataset

Method	Cleveland	statlog	Hepatitis	Demagology
Single ACO	87.1(19.8)	94.6(22.3)	91.8(29.4)	
Single GA	87.1(17.5)	94.6(23.1)	91.6(28.9)	
Hybrid ACO/GA	88.7(16.3)	95.1(21.0)	93.4(25.9)	

## III. CONCLUSIONS

In this paper a hybrid ACO/GA method is proposed for attribute selection and tested on four public datasets. Firstly all the training samples are divided into training samples and testing sample using tenfold method. Secondly Kolomogorov-Smirnov Test is adopted to find a attribute subset on the training sample.Thirdly,the proposed hybrid ACO/GA hybrid algorithm for healthcare application.We further introduced an adaptive Parameter control mechanism into the algorithms to balance the explorative and exploitative behaviors of the two heuristic at different stages of the algorithm .The simulation results were compared with another recently reported hybrid algorithm and we showed that our ga-aco hybrid algorithm .

## REFERENCES

- [1] E.Bonabeau,M.Dorigo and G.Theraulaz,"Inspiration for optimization from social insect behavior, " Nature,vol.406,pp.39-42,july 2000

- [2] M.Dorigo,G.D.Caro and L.M.Gambardella,"Ant algorithm for discrete optimization,"*Artif.Life*,vol.5,no.2,pp.137-172,1999
- [3] J.I Deneubourg,S.Arun,S.Goss,and J.M Pasteels,"The self-organizing exploratory pattern of the argentine ant,"*J.Insect Behav.*,vol. 3,pp. 159-168,1990.
- [4]M.Dorigo and T.Stutzle,"The ant colony optimization metaheuristic:Algorithms,applications and advances,"in *Handbook of Metaheuristics* F,Glover and G.Kochenberger,Eds. Norwell,MA:Kluwer.
- [5]Albayrak m.allahverdi N. (2011) Development a new mutation operator to solve the travelling salesman problem by aid of Genetic Algorithms.*Expert Syst Appl* 38(3):1313-1320
- [6]Weinberg CR,Darden TA et al (2001b)Gene selection for sample classification based on gene expression data:study of sensitivity of choice of parameters of the GA/KNN method.*Bioinformatics* 17:1131-1142
- [7]Matthew Settles,Terence Soule "Breeding Swarms:A GA/PSO Hybrid"
- [8] A GA-ACO-Local Search Hybrid Algorithm for Solving Quadratic Assignment Problem yi-Liang Xu,Meng-Hiot Lim,Yew-Soon Ong,Jing Tang.Intelligent System Center,Research Techno Plaza Nanyang Technological University
- [9] B. Liu, H.A. Abbass, and B. McKay, "Classification rule discovery with ant colony optimization," in *Proceedings of IEEE/WIC International Conference of Intelligent Agent Technology*, pp. 83–88, 2003.
- [10] W. Shahzad and A.R. Baig, "Compatibility as a heuristic for construction of rules by artificial ants," *Journal of Circuits, Systems, and Computers*, Vol. 19, No. 1,pp. 297-306, Feb. 2010.
- [11]B. Liu, H.A. Abbass, and B. McKay, "Density-based heuristic for rule discovery with ant-miner," in *Proceedings of 6th Australia-Japan Joint Workshop on Intelligent Evolutionary Systems*. Canberra, Australia, 2002, pp. 180–184.
- [12] B. Liu, H.A. Abbass, and B. McKay, "Classification rule discovery with ant colony optimization," *IEEE Computational Intelligence Bulletin*, Vol. 3, No. 1, Feb. 2004.