

A Swarm Intelligence Based Algorithm for Proteomic Pattern Detection of Ovarian Cancer

Yan Meng

Department of Electrical and Computer Engineering
Stevens Institute of Technology
Hoboken, New Jersey, USA
Email: yan.meng@stevens.edu

Abstract - The advanced protein profiling technologies can simultaneously resolve and analyze multiple proteins. Evaluating multiple proteins will be essential to establish signature proteomic patterns that distinguish cancer from non-cancer. It is desirable to have complex and intelligent analytical tools to detect the changes in protein expression and their correlation to diseases conditions. This paper proposed a swarming-agent based intelligence algorithm using a hybrid ant colony optimization/particle swarm optimization (ACO/PSO) algorithm to identify the diagnostic proteomic patterns of biomarkers for early detection of ovarian cancer. The experimental results demonstrated that the proposed system has high predictive accuracy and better diagnostic performance.

I. INTRODUCTION

An ovarian cancer victim's chances for long-term survival are greatly improved by early detection, and early detection in turn is enhanced by an accurate diagnosis. The choice of appropriate treatments immediately following surgery is largely influenced by prognosis, that is, the expected long-term behavior of the disease. This needs is especially urgent in women who have a high risk of ovarian cancer due to family or personal history of cancer, and for women with a genetic predisposition to cancer due to abnormalities in predisposition genes such as BRCA1 and BRCA2. There are no effective screening options for this population.

Ovarian cancer presents at a late clinical stage in more than 80% of patients, and is associated with a 5-year survival of 35% in this population. By contrast, the 5-year survival for patients with stage I ovarian cancer exceeds 90%, and most patients are cured of their disease by surgery alone [1][2]. Therefore, increasing the number of women diagnosed with stage I disease should have a direct effect on the mortality and economics of this cancer without the need to change surgical or chemotherapeutic approaches.

Cancer antigen 125 (CA125) is the most widely used biomarker for ovarian cancer [1] [2]. Although concentrations of CA 125 are abnormal in about 80% of patients with late stage disease, they are increased in only 50-60% of patients with stage I ovarian cancers.

It is important that additional diagnostic biomarkers be identified to reduce ovarian cancer. However, because of the robust molecular and cellular heterogeneity, it is likely that a combination or a panel of biomarkers will be required to improve the early detection of ovarian cancer.

Certain protein profiling reflecting the pathological stage of organs will definitely aid in the early detection of ovarian cancer. The mass spectrometric method using surface-enhanced laser desorption/ionization time-of-flight (SELDI-TOF) technology allows sensitive and high-throughput protein profiling of complex biological specimens [3] [4] [5]. These profilings can contain thousands of data points, requiring complex and intelligent analytical tools to detect the changes in protein expression and their correlation to disease conditions, thereby hastening the identification of novel biomarkers with better diagnostic performance.

Vlahou et. al. [4] proposed a decision tree algorithm for cluster data analysis. The decision tree classification splits up a data set into two bins or nodes, using one rule at a time in the form of a question. The splitting decision is defined by presence or absence and the intensity levels of one peak. This splitting process continues until terminal nodes or leaves are produced or further splitting has no gain. Classification of terminal nodes is determined by the group of samples representing the majority of samples in that node. The limitation of this binary decision tree algorithm is that if the error exists in the classification rule of the parent nodes, this error will be propagated to all its children nodes.

Since one set of data samples may generate multiple different decision trees based on different classification rules, it is hard to tell which classification tree is better than the other. Furthermore, if the new samples are applied, the decision tree algorithm has to start over again from the scratch. All the learning experiences accumulated from the previous preliminary samples are discarded. In other words, the diagnosis system using the decision tree algorithm lacks the capability of accumulating the learning experience over the time.

Some researchers have explored data analysis using genetic algorithm [5] for pattern matching to identify cancer. The genetic algorithm in [5] starts by randomly selecting many proteomic patterns within the training data for analysis. Each chosen pattern is tested to see how well it can discriminate affected from unaffected in the training set, by cluster analysis. Successful proteomic patterns are kept and recombined, where unsuccessful patterns are discarded. Ultimately, a best pattern emerges after many successive iterations by the algorithm. This pattern, which best segregates the training sets, is used to classify diagnostically unknown samples. The point in space for the unknown patient is compared for its proximity to clusters in N-dimensional

space corresponding to the location of individuals in the preliminary set – an established pattern-recognition principle. The preliminary training population for comparison becomes enriched as the system learns over time from the accumulated cases, and the predictive accuracy will be increased eventually.

As we know, use of potential multiple biomarkers could offer increased statistical power for superior predictive value and greater utility in diagnosis, patient stratification and patient monitoring. Therefore, a diagnosis system which can automatically identify the patterns of multiple biomarkers with highly predictive accuracy is extremely desirable. Instead of using genetic algorithm, in this paper, we turn our attention to a swarming-agent based intelligence algorithm for pattern detection.

The swarm intelligence design approach adapts robust, self-organizing coordination mechanism observed in distributed natural systems (e.g., social insect colonies) to engineered systems. Ant Colony Optimization (ACO) algorithms involve simple ants that cooperate with one another to achieve an emergent unified behavior for the system as a whole, producing a robust system capable of finding high-quality solutions for problems with a large search space. In the context of rule discovery, an ACO algorithm has the ability to perform a flexible robust search for a good combination of terms involving values of the predictor attributes. Dorigo et. al. [6] [7] proposed a new meta-heuristic based on the ACO to solve combinatorial optimization problem, which has been shown to be robust and versatile. ACO optimization algorithm has also been adopted recently for biological data mining applications [8] where an ant-colony-based data miner (Ant-Miner) was proposed to extract classification rules from data.

Another popular collective intelligence is Particle Swarm Optimization (PSO), developed in 1995 by James Kennedy and Russel Eberhart [9]. The PSO is a biologically-inspired algorithm motivated by a social analogy, such as flocking, herding, and schooling behavior in animal populations. The PSO algorithm is population-based: a set of potential solutions evolves to approach a convenient solution (or set of solutions) for a problem. Being an optimization method, the aim is finding the global optimum of a real-valued function (fitness function) defined in a given space (search space).

Inspired by the swarming agents algorithm proposed by Brueckner and Parunak [10] for distributed data pattern detection, a hybrid ACO/PSO algorithm based on the swarming-agent architecture is proposed in this paper. Basically, large amount of mobile agents are generated in the search space. Two collective and coordination processes for the mobile agents are proposed, one is based on the ACO algorithm, where the agents' movements are guided by pheromones in the shared environment locally, the other one is based on PSO algorithm, where a global maximum of the attribute values can be obtained through the random interaction between the agents.

The objective of this paper is to develop a hybrid ACO/PSO algorithm to analyze the large amount of SELDI protein profiling of serum using samples from women with a known diagnosis to define an optimum discriminatory

proteomic pattern. Then, the detected discriminatory pattern are applied to predict the diagnosis of the testing samples, predictive accuracy, sensitivity, and specificity of the detected proteomic pattern are evaluated.

II. PROTEOMIC DATA ANALYSIS

The ASCII files of proteomic spectra generated by SELDI-TOF are input to the analysis system, where each spectrum is composed of 15200 M/Z values on the x axis with a corresponding amplitude on the y-axis. The objective of the data analysis is to discover the most fit discriminatory pattern that can best segregates serum from patients with ovarian cancer from the unaffected populations. Each pattern is composed of the amplitude values at selected M/Z values. Two phases are conducted, Phase I is to detect the candidate patterns from the preliminary training samples, Phase II evaluate the detected patterns by the testing samples.

100 mass spectra preliminary samples are collected, 50 patients with cancer and 50 unaffected patients. Initially, large amount of candidate patterns, which consist of small sets of exact M/Z values selected along the x-axis in the SELDI-TOF mass spectra, are randomly generated. Each candidate pattern contains N numbers of the 15200 potential x-axis values that define the spectra, and is formed by the combined y-axis amplitudes of the candidate set of key M/Z values in N -dimensional space, where N is the number of M/Z values in the candidate pattern.

Then each candidate pattern is evaluated by the fitness function for its capability to distinguish the two preliminary populations. The fitness function is defined as the product of specificity and sensitivity, e.g. $\text{fitness} = \text{sensitivity} * \text{specificity}$ [11]. The sensitivity is defined as the ratio of the number of correctly classified disease samples to the total number of disease samples. The specificity is defined as the ratio of the number of nondisease sample correctly classified to the total number of nondisease samples. Specifically, the fitness is defined as

$$F = \frac{TP}{TP + FN} \cdot \frac{TN}{FP + TN} \quad (1)$$

where TP (true positive) is the number of correctly classified disease samples, FN (false negative) is the number of incorrectly classified disease samples, TN (true negative) is the number of correctly classified nondisease samples, and FP (false positive) is the number of incorrectly classified nondisease samples.

Since the use of potential multiple biomarkers or pattern of multiple biomarkers could offer increased statistical power for superior predictive value and greater utility in diagnosis, patient stratification and patient monitoring, the discriminatory patterns consisting of a set of individual biomarkers are desirable. The best fit candidate pattern can be selected based on the fitness value defined in (1). Since each pattern consists of a set of key M/Z values, which can be treated as the attributes of the pattern, or the individual biomarkers.

It is possible that some attributes inside a pattern has the higher fitness value, while others may have much lower

fitness value due to the random attribute generation. Or some combinations of the attributes from different patterns may gain higher fitness value than other combinations or original combinations in each pattern.

In [3], a genetic algorithm is applied to search for the survival of the fittest subset of data in a sample, where the successful proteomic patterns are kept and recombined ("mated"), whereas unsuccessful patterns are discarded. Ultimately, a best pattern emerges after many successive iterations by the algorithm. In this paper, we apply the swarming-agent based hybrid ACO/PSO algorithm to search for the global optimized discriminatory pattern.

III. ACO ALGORITHM AND PSO ALGORITHM

An ACO algorithm is essentially a system that simulate the natural behavior of ants, including mechanisms of cooperation and adaptation. ACO algorithms are based on the following ideas. First, each path followed by an ant is associated with a candidate solution for a given problem. Second, when an ant follows a path, the amount of pheromone deposit on that path is proportional to the quality of the corresponding candidate solution for the target problem. Third, when an ant has to choose between two or more paths, the path(s) with a larger amount of pheromone are more attractive to the ant. After some iterations, eventually, the ants will converge to a short path, which is expected to be the optimum or a near-optimum solution for the target problem.

The PSO algorithm is population-based: a set of potential solutions evolves to approach a convenient solution (or set of solutions) for a problem. Being an optimization method, the aim is finding the global optimum of a real-valued function (fitness function) defined in a given space (search space). Rather than simply a social simulation, PSO can be treated as a powerful new search algorithm, capable of optimizing a wide range of N-dimensional problems.

The social metaphor that led to this algorithm can be summarized as follows: the individuals that are part of a society hold an opinion that is part of a "belief space" (the search space) shared by every possible individual. Individuals may modify this "opinion state" based on three factors:

- The knowledge of the environment (its fitness value)
- The individual's previous history of states (its memory)
- The previous history of states of the individual's neighborhood

An individual's neighborhood may be defined in several ways, configuring somehow the "social network" of the individual. Following certain rules of interaction, the individuals in the population adapt their scheme of belief to the ones that are more successful among their social network. Over the time, a culture arises, in which the individuals hold opinions that are closely related.

In the PSO algorithm each individual is called a "particle", and is subject to a movement in a multidimensional space that represents the belief space. Particles have memory, thus

retaining part of their previous state. There is no restriction for particles to share the same point in belief space, but in any case their individuality is preserved. Each particle's movement is the composition of an initial random velocity and two randomly weighted influences: individuality, the tendency to return to the particle's best previous position, and sociality, the tendency to move towards the neighborhood's best previous position.

The velocity and position of the particle at any iteration is updated based on the following equations:

$$\begin{aligned} v_{id}^{t+1} &= w \cdot v_{id}^t + c_1 \cdot \varphi_1 \cdot (p_{id}^t - x_{id}^t) + c_2 \cdot \varphi_2 \cdot (p_{gd}^t - x_{id}^t) \\ x_{id}^{t+1} &= x_{id}^t + v_{id}^{t+1} \end{aligned} \quad (2)$$

where v_{id}^t is the component in dimension d of the i th particle velocity in iteration t , x_{id}^t is the component in dimension d of the i th particle position in iteration t , c_1, c_2 are constant weight factors, p_i is the best position achieved by particle i , p_g is the best position found by the neighbors of particle i , φ_1, φ_2 are random factors in the $[0,1]$ interval, and w is the inertia weight.

The PSO algorithm requires tuning of some parameters: the individual and sociality weights c_1, c_2 , and the inertia factor w .

IV. A HYBRID ACO/PSO ALGORITHM USING SWARMING AGENTS

An *agent* is an independent processing entity that interacts with the external environment and the other agents to pursue its particular set of goals. By using the ACO algorithm, the agents in the systems coordinate their behaviors and communicate their results through pheromone, which is a shared dynamic environment. The self-organizing and self-adapting system-level behaviors can be obtained through these pheromone interactions. Each agent can only observe the limited environment within its local sensors. In other words, the agent can only sense the pheromone which is close to its current position, while it has no idea of the existence of other pheromone far away. Without a central host, it is possible that most agents may easily get locked in a local maximum.

On the other hand, agents using PSO algorithm coordinates their behaviors through the random interaction with other agents. By following certain rules of interaction, the agents in the population adapt their scheme of belief to the ones that are more successful among their social network. Over the time, a global optimization can be obtained.

Therefore, a hybrid ACO/PSO algorithm is proposed in the paper for the swarming-agent based diagnosis system, where two collective and coordination processes are proposed, one is based on the ACO algorithm, where the agents' movements are guided by pheromones in the shared environment locally, the other one is based on PSO algorithm, where a global maximum of the attribute values can be obtained through the random interaction between the agents.

A. Swarming Agents Architecture

Initially, 200 proteomic patterns are generated randomly, more specifically for each pattern, 10 key M/Z values are randomly selected from 200-15000 Da, and the corresponding y-axis at each selected M/Z value are also generated randomly. The search space is defined as a 20x10 rectangular grid, where the 200 patterns reside. One grid is corresponding to one pattern. The pattern is allocated to the first row of the grid in the order of the generation. After all the grids in the first row are filled up, go to next row, and this procedure continues until all the grids are filled up by the patterns.

We call the spectra at key M/Z value as attribute for each pattern. Therefore, the structure of a pattern is defined as the following equation:

$$\mathbf{P}_i = \{a_{i1}(x_{i1}, y_{i1}), a_{i2}(x_{i2}, y_{i2}), \dots, a_{iN}(x_{iN}, y_{iN})\} \quad (3)$$

where $i = 1, 2, \dots, n$.

Where \mathbf{P}_i denotes the generated pattern i , $a_{i1}(x_{i1}, y_{i1})$ represents the attribute 1 with M/Z value x_{i1} and the corresponding amplitude y_{i1} , N is the number of key M/Z values selected in each pattern, and n is the number of the generated patterns.

For each pattern, an associated *pattern agent* is created, which is fixed to each pattern. 200 mobile agents are generated, whose task is to detect the most fit proteomic pattern from the randomly generated patterns, and recombine some of the selected patterns together to build a pattern with more fitness value. We call these mobile agents *particle agents*, which can move from one pattern to another in the search space and interact with other particle agents dynamically. Initially, the particle agents are uniformly distributed in the search space. After many iterations of the algorithm, eventually the particle agents will converge to the most fit pattern.

The particle agents are allowed to deposit pheromones and sense local attributes inside each pattern. The pattern agent is responsible to execute the dynamics of pheromone aggregation, dispersion and evaporation. It provides topological information of its neighbor patterns. The particle agents request local topology information from their current pattern agent at the time of their movement. In return, they receive a list of direct neighbors of the pattern (spatial neighborhood) and they are only permitted to move to these locations.

There are two levels of pheromones, one is for pattern pheromone, and the other is attribute pheromone inside the pattern. If the value of the pheromone for the pattern is high, it has two meanings. One is that this pattern has higher probability to be the potential best fit discriminatory pattern; the second one is that some of the attributes inside this pattern has the higher potential to be included in the best fit pattern. Either case would attract more particle agents to move to this pattern. The pheromone will be evaporated over time if no agent deposits more pheromone on it. Once an agent enters into a pattern, it deposits pheromone on the selected attribute(s), and deposits the pheromone on the pattern, both

are proportional to their fitness values evaluated through the preliminary samples.

Based on the above description, the structure of the pattern agent and particle agent are defined as follows:

$$\text{Pattern_agent}_i = \{\text{pattern_ph}_i, \text{attrib_ph}_{ij}, \text{neighbor}_k\},$$

where $i = 1, 2, \dots, n$, $j = 1, 2, \dots, N$, $k = 1, 2, \dots, 8$. (4)

where pattern_ph_i represents the pattern pheromone for pattern i . attrib_ph_{ij} represent the attribute pheromone at attribute j in the pattern i , neighbor_k represent the vector of all the neighbors of the pattern i where the maximum number of each pattern is 8, and local_agent_list keeps the total number of the particle agents which are currently located in the pattern i .

The structure of the particle agent is defined as:

$$\text{particle_agent}_i = \{\text{past_position}_{ij}, f_{p_i}, \text{last_attrib}(1, \dots, N), f_{a_i}(1, \dots, N)\}$$

where $i = 1, 2, \dots, m$, $j = 1, 2, \dots, M$. (5)

where the $\text{past_position}_{ij}$ represents the short term memory of the particle agent i in the last j iterations, f_{p_i} and $\text{last_attrib}(1, \dots, N)$ denote the attributes of last position the particle agent i visited, and $f_{a_i}(1, \dots, N)$ are the fitness values of the attribute of last position.

The individual movement of the particle agents are guided by the pheromones they can sensed locally as well as the pheromones information gained through the interaction of other particle agents.

To obtain a global optimization of the distribution of the particle agents, the PSO algorithm is applied, where the current position, the best past position, and the neighbors' best position are applied to adjust its movement. More precisely, the pheromone update by the particle agents should consider not only the fitness values of the pattern/attributes it is currently located, but also the fitness value of its best past pattern/attribute it has visited as well as its neighbors best fitness value. The updating equation is as follows:

$$d_i = w * d_i + c_1 * \varphi_1 * F_c + c_2 * \varphi_2 * F_p + c_3 * \varphi_3 * F_n \quad (6)$$

where d_i is the pheromone deposited on the pattern i , c_1, c_2 are constant weight factors, F_c, F_p, F_n are the fitness value of the current pattern, the best past pattern, and the neighbors' best pattern, respectively, $\varphi_1, \varphi_2, \varphi_3$ are random factors in the [0,1] interval, and w is the inertia weight. The pheromone update for pattern attributes is similar to (6). Every iteration of mobile agent movement, a normalization of the pheromones for patterns and attributes is conducted.

B. A Hybrid ACO/PSO Algorithm

The pseudo code of the swarming-agent based hybrid ACO/PSO algorithm is shown as the followings.

SWARM ALGORITHM:

```

Training set = {preliminary training samples};
Randomly generate 200 candidate discriminatory patterns  $P_i$  ;
Locate the pattern to the search grid in the order of the
generation;
Generate 200 particle agents randomly uniformly in the search
grid, where each particle agent is located in one grid.
WHILE ( iteration < max_no_iteration)
  /* Use ACO algorithm to deposit the pheromone to the
  pattern and attributes */
  FOR i = 1 to (no of candidate patterns)
    Update the pheromone of the pattern selected by the
    particle agents; otherwise, decrease the pheromone in
    the pattern;

    FOR j = 1 to no of attributes in each pattern
      Update the pheromone of all attribute trails by
      increasing pheromone in the trails followed by
      agent_i (proportional to the fitness value) and
      decreasing pheromone in the other trails
      (simulating pheromone evaporation).
    END FOR_LOOP(j)
  END FOR_LOOP(i)

  /* Use the modified PSO algorithm to adjust the movement
  of the particle agent in this iteration. */
  FOR i = 1 to (no of particle agents)
    Update the pattern and attribute pheromones using
    equation (6).
    IF pattern_fitness value of current position < their
    neighbors' pattern fitness value
      Move to the neighbor's pattern;
      IF the fitness value of the attribute(s) in the last
      position is (or are) higher than the new position
        Replace those lower-fitness-value attributes with
        the higher-fitness-value attribute from last position;
      ELSE
        Leave the attributes of the new pattern as it is.
      ELSE
        Stay in the current pattern position;
      END FOR_LOOP (i)
    END WHILE
  
```

Initially, 200 candidate discriminatory patterns are generated by randomly picking up 10 key M/Z values on x-axis and the corresponding amplitude on y-axis. These patterns are filled into the grids in the search space, where each pattern is corresponding to one grid based on the order of their generation. 200 particle agents are generated and uniformly distributed to the search space, where each particle agent occupies one grid.

For each iteration, first each particle agent evaluates the fitness value of the pattern and all the attributes inside the pattern where it is currently located. If the fitness value of the pattern or attribute(s) is higher than the threshold, the particle

agent deposits the pheromone on the pattern or the attribute(s). The amount of the pheromone deposited is proportional to the fitness value.

Based on the ACO algorithm, pheromones are deposited on each attribute based on their different fitness value by the particle agents. Multiple deposits at the same attribute with specific value aggregate in strength. Other agents who reside the same pattern may change their behavior by sensing pheromones of a particular flavor, following pheromone trail to the potential pattern attributes. Pheromone concentrations in the environment disperse in space and evaporate over time.

For those patterns with lower fitness values may have one or more attributes with higher fitness values. In other words, there are some key M/Z values who can efficiently segregate the preliminary samples. To integrate these attribute to other patterns, the agents should memorize those high-level attributes, and integrate those attribute into its neighbor pattern with high pheromones. However, to avoid the local maximum, a global threshold is necessary for the agents to evaluate if those attributes are higher enough to be selected. Therefore, global optimization through PSO algorithm is applied through the interaction of the particle agents, where the population of agents that identifies the attributes, not the agent themselves. Pheromone updates are conducted using (6).

Due to the fixed topology for the pattern agents in the search grid, the maximum number of the neighbors is 8. It is assumed that each particle agent can only interact to its neighbor agents only. Since the agents are moving around from one pattern to another, if the new position has a higher quality than any position the particle has ever occupied then it is set as the agent's past best position (e.g, the past best pattern detected). Therefore, each agent should have short term memory of its past trajectory and the patterns it has visited. Due to the randomly interaction with other detector agents locally, the best neighbour's position (e.g. the best neighbor's pattern detected) can be obtained. Therefore, a modified PSO algorithm can be applied to update the attribute values of the patterns. The big difference from the standard PSO is not only the past best and its best neighbour's positions are used to guide the movement of the agent, but also its current position are used. The quality of these three positions, multiplied by individual random learning factors as usual in PSO, are added to the values in the appropriate entries in the pheromone matrices of the current particle agent.

The basic difference in our algorithm from the standard PSO algorithm is that the particle not only moves to the new position based on the updated rule, but also brings the last position's attributes and its associated fitness values along with its movement to the new position. And update the new position with better quality of the attribute(s), and discard those low-quality attribute of the new position. The outcome is a new combined pattern with higher quality than both original ones. During next iteration, the newly built pattern will be evaluated by the agents and deposit the pheromone as appropriate. After many iterations, eventually, the most strong pheromone trail will be the most fit discriminatory pattern.

It is worth noting that another advantage of this algorithm is that when the new preliminary training samples are available, the new detection pattern can be built based on the

previously detected pattern, thus all the previously learning experiences can be accumulated over time. Eventually, in theory, the predictive accuracy will be increased over time.

V. EXPERIMENTAL RESULTS

The experimental phases of proteomic data analysis is shown in Fig. 1.

100 preliminary control samples are collected from the National Ovarian Cancer Early Detection Program database, while 50 are preliminary sample with cancer, and 50 are unaffected people. 50 samples are used as the preliminary training samples and the other 50 testing samples are used for verification. The ASCII files of proteomic spectra generated by SELDI-TOF are input to the analysis system, where each spectrum is composed of 15200 M/Z values on the x axis with the corresponding amplitude on the y-axis.

The classification results of the testing samples are listed in TABLE I. The proposed algorithm correctly classified 22 of 25 (88%) of the samples which are not cancer, and correctly classified 23 of 25 (92%) of the testing samples which are cancer. The sensitivity of this diagnosis system is 92% and the specificity of the system is 88%.

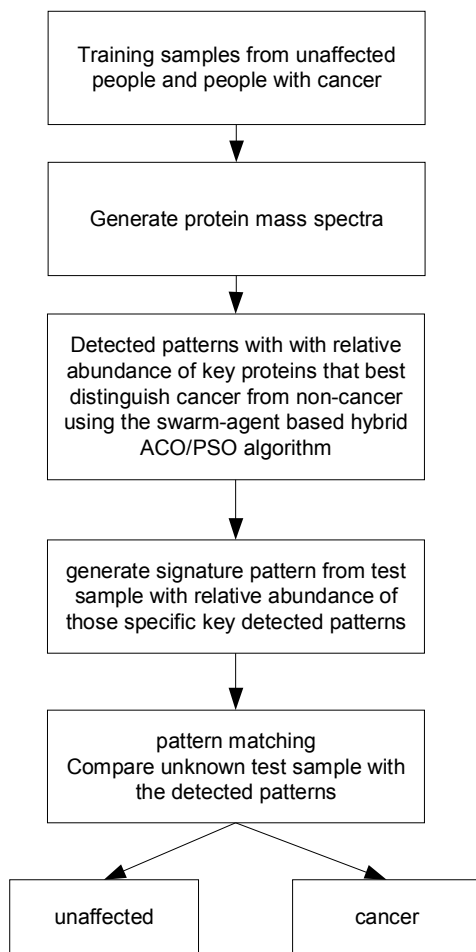


Fig.1. Phases of proteomic data analysis

TABLE I
CLASSIFICATION RESULTS OF SERUM SAMPLES FROM THE TESTING BY PROTEOMIC PATTERN

	cancer	unaffected
Affected women	3/25	22/25
Women with cancer		
Stage I	9/25	1/25
Stage II, III, IV	14/25	1/25

VI. CONCLUSIONS AND FUTURE WORKS

In this paper, a swarm-agent based hybrid ACO/PSO algorithm is proposed to automatically detect the discriminatory pattern of multiple biomarkers which can best segregate the preliminary samples with cancer from the unaffected populations. Two collective and coordination processes are proposed, one is based on the ACO algorithm, where the agents' movements are guided by pheromones in the shared environment locally, the other one is based on PSO algorithm, where a global maximum of the attribute values can be obtained.

Compared with other classifiers, such as neural networks and biostatistical methods, the protein peaks used in the pattern classifier are easily attainable by matching the detected discriminatory biomarker patterns. Multiple biomarker patterns might be obtained from the same training set with the similar fitness value.

The accuracy, sensitivity and specificity of the proposed classification algorithms have been evaluated by the test set. The detected discriminatory pattern with multiple biomarkers can be applied to conduct diagnosis and prognosis with new uncontrolled samples. With the self-learning capability, the proposed system will automatically identify the patterns of potential biomarkers and accumulate the learning experience over time, contributing to the system evolving one with a highly predictive accuracy, especially with noise-insensitive approach.

In addition, a patient-specific differential diagnosis and prognosis will be tailored by adjusting the attribute parameters of the pattern based on the patient's own history database. The prognosis can be realized by a patient-specific diagram, providing the physician and the patient better information with which to plan a more accurate treatment. Moreover, remote login to the system will be developed to extend the system practice to distance operation.

The ability to automatically discover potential patterns formed by multiple biomarkers will greatly improve the sensitivity and specificity of clinical proteomics for predictive medicine, leading to a highly predictive diagnostic test. This experience-accumulated clinical platform has the potential to significantly improve the early detection and differential diagnosis of ovarian cancer. Furthermore, appropriate treatments for an individual patient can be provided via a patient-specific system and remote access through internet will

allow the system to be used with a larger population of patients.

Our future work will focus on developing a diagnosis and prognosis system for proteomic pattern detection by using machine learning technologies to mimic the capabilities of specialized physicians whose long-term accumulated experiences lead to the effective and accurate diagnosis for cancer. Specific aims include: 1) enhancing the initial model with novel noise-insensitive capabilities; and 2) accommodating the system to be patient-specific and remote-accessible.

REFERENCES

- [1] L. S. Cohen, P. R. Escobar, C. Scharm, B. Flimco, D. A. Fishman, "Three-dimensional power Doppler ultrasound improves the diagnostic accuracy for ovarian cancer prediction", *Gynecol Oncol*, vol. 82, 2001, pp:40-48.
- [2] U. Menon and I. Jacobs, "Tumor marders", In: Hoskins WJ, Perez CA, Young RC, eds. *Principles and practice of gynecologic oncology*. Philadelphia: Lippincott Williams & Wilkins, 2000: 165-82.
- [3] Z. Zhang, R. C. Bast, Y. Yu, et al, "Three biomarkers identified from serum proteomic analysis for the detection of early stage of ovarian cancer", *Cancer Research*, vol. 64, 2004, pp: 5882-5890.
- [4] A. Vlahou, J. O. Schorge, B. W. Gregory, and R. L. Coleman, "Diagnosis of ovarian cancer using decision tree classification of mass spectral data", *J. Biomed Biotechnol.*, vol. 5, 2003, pp: 308-314.
- [5] E. Petricion, A. Ardekani, B. A. Hitt, P. J. Levine, V. A. Fusaro, S. M. Steinberg, G. B. Mills, C. Simone, D. A. Fishman, E. C. Kohn, and L. A. Liotta, "Use of proteomic patterns in serum to identify ovarian cancer", *THE LANCET*, vol. 359, 2002, pp:572-577.
- [6] M. Dorigo, A. Coloni, and V. Maniezzo, "The ant system: Optimization by a colony of cooperating agents", *IEEE Trans. on Syst. Man Cybern. B*, vol.26, pp.29-41, Feb. 1996.
- [7] M. Dorigo and G. DI Caro, "The ant colony optimization meta-heuristic", In *New Ideas in Optimization*, D. Corne, M. Dorigo, and F. Glover, Eds. New York: McGraw-Hill, 1999, pp. 11-32.
- [8] R. S. Parpinelli, H. S. Lopes, and A. A. Freitas, Data ming with an ant colony optimization algorithm, *IEEE Trans. on Evolutionary Computing*, Vol. 6, No. 4, August 2002.
- [9] Eberhart, R. C., and Kennedy, J. 1995: A New Optimizer Using Particles Swarm Theory, *Proc. Sixth International Symposium on Micro Machine and Human Science* (Nagoya, Japan), IEEE Service Center, Piscataway, NJ, 39-43.
- [10] S. A. Brueckner and H. V.D. Parunak, "Swarming agents for distributed pattern detection and classification", *AAMAS 2002*, July 15-19, 2002, Bologna, Italy.
- [11] H. S. Lopes, M. S. Coutinho, and W. C. Lima, "An evolutionary approach to simulate cognitive feedback learning in medical domain," in *Genetic Algorithm and Fuzzy Logic Systems: Soft Computing Perspectives*, E. Sanchez, T. Chibata, and L. Zadeh, Eds, Singapore: World Scientific, 1998, pp. 193-207.