

©2009 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

An Integrated Approach of Particle Swarm Optimization and Support Vector Machine for Gene Signature Selection and Cancer Prediction

¹C.W. Yeung, *Member, IEEE*, ¹F.H.F Leung, *Senior Member, IEEE*,
²K.Y.Chan, *Member, IEEE*, and ³S.H. Ling, *Member, IEEE*

Abstract— To improve cancer diagnosis and drug development, the classification of tumor types based on genomic information is important. As DNA microarray studies produce a large amount of data, expression data are highly redundant and noisy, and most genes are believed to be uninformative with respect to the studied classes. Only a fraction of genes may present distinct profiles for different classes of samples. Classification tools to deal with these issues are thus important. These tools should learn to robustly identify a subset of informative genes embedded in a large dataset that is contaminated with high dimensional noises. In this paper, an integrated approach of support vector machine (SVM) and particle swarm optimization (PSO) is proposed for this purpose. The proposed approach can simultaneously optimize the selection of feature subset and the classifier through a common solution coding mechanism. As an illustration, the proposed approach is applied to search the combinational gene signatures for predicting histologic response to chemotherapy of osteosarcoma patients. Cross-validation results show that the proposed approach outperforms other existing methods in terms of classification accuracy. Further validation using an independent dataset shows misclassification of only one out of fourteen patient samples, suggesting that the selected gene signatures can reflect the chemoresistance in osteosarcoma.

I. INTRODUCTION

Cancer is one of the leading causes of death in developed countries [34]. Studies reveal that many cancer patients were diagnosed at a late stage, which were more difficult to treat. Cancer develops mainly in epithelial cells (carcinomas), connecting/muscle tissue (sarcomas), and bone marrow cells (leukaemia and lymphomas). Successive mutations in the normal cells lead to DNA damages and impair the replication mechanism, ultimately causing significant gene features that contribute to a cancerous state.

A microarray chip can simultaneously interrogate thousands of genes, which provides an extremely powerful tool for genomic studies of cancer. A few key genes (typically involving oncogenes and tumour suppressor genes), when mutated, will cause deregulation of the transcription and translation of other genes through complicated signalling pathways to initiate oncogenesis, and ultimately leading to derangement of the cellular phenotype and the clinical manifestation of cancer [31]. Analyzing microarray data could help discover some significant cancer genes and their mutual interactions, which can be used to generate hypothesis for the identification and validation of genetic biomarkers for diagnostic and therapeutic purposes [8, 17, 19, 21].

Significance based methods (e.g. T-test, confidence intervals, etc.) [10], which aim at finding statistically significant genes in differentiating various patient groups, have been extensively used. However, the common characteristic of these methods is to evaluate each individual gene one by one, thus neglecting the intrinsic interactions among genes. Multivariate methods, instead, are highly desirable to identify a group of genes that can cooperatively regulate tumor patterns. One of the commonly used approaches is to iteratively combine different genes as predictors for tumor classification. Gene predictors that account for specific clinical phenotypes should classify tumor samples into different groups in a robust manner. In other words, the performance of classification of a predictor can implicate significances of genes. Classification problems have been addressed by different machine learning tools like neural network [5,15,23], k -nearest neighborhood [25], decision tree, self-organizing maps [21], graph theory [22], and support vector machine (SVM) [4,12,18,25,30,32,35], etc. However, the choice and implementation of these tools are rather an empirical and experimental work. In fact, the data-driven machine learning techniques suffer from the “curse of dimensionality”, which is a typical problem when analyzing microarray data. On the other hand, the recent method [30] often solves the problems of feature selection and classifier optimization separately in tandem. Gene features are selected first according to prior knowledge or their statistical significances. Then the classifier is optimized based on the selected feature set. There is no guarantee that the selected gene features can achieve optimal classification.

In this paper, an approach that integrates support vector machine [4, 35] with particle swarm optimization (PSO) [20] is used to build a more reliable classifier. This

The work described in this paper was substantially supported by a research grant of the Hong Kong Polytechnic University (Project Account Code: RGYG).

¹Authors are with the Centre for Signal Processing, Dept of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hong Kong

²Author is with the Digital Ecosystems and Business Intelligence Institute, Curtin University of Technology, Perth, Australia

³Author is with University of Technology, Sydney, NSW, Australia

approach incorporates gene features and SVM parameters into one common solution code. The problem is solved by simultaneously optimizing the selection of gene features and the parameters of the SVM classifier. The optimal solution achieved by this procedure may reach the best gene features and the corresponding optimal classifier. We shall focus on osteosarcoma, which is the most common malignant bone tumor in children and accounts for 60% of malignant bone tumors diagnosed in the first two decades of life [26]. The current standard care procedure is to give neo-adjuvant chemotherapy after initial diagnosis is made, followed by definitive surgery and postoperative chemotherapy. After surgery, the tumor will be categorized to be a good responder (>90% necrosis) or a poor responders (<90% necrosis) according to its response to pre-operative chemotherapy, and based on which a proper postoperative chemotherapy can be determined. Unfortunately the overall survival rate of the poor responders did not significantly improve [14, 26]. It is possible that resistant tumor cells have additional time to either metastasize to the lungs or evolve further during the period when ineffective preoperative chemotherapy is given. Therefore it is necessary to identify at the time of initial diagnosis the patients who are likely to have a poor response to standard preoperative therapy and therefore a poor outcome eventually. The DNA microarray data were collected from the tumor samples of patients enrolled in a clinical trial at the Texas Children's Hospital. Expression profiling by DNA microarray was done on tumor tissues before and after chemotherapy treatments. The classifier was initially developed with the integrated approach to distinguish good responders and poor responders based on the sampled data after chemotherapy. The results were compared with the existing methods tested on the same dataset [27] to verify its performance and robustness. The derived classifier was further validated through predicting the potential chemo-response of biopsy samples acquired before chemotherapy.

II. DATA DESCRIPTION

We use a set of osteosarcoma microarray data [27], which were generated through institutional review board-approved protocols at four centers (Texas Children's Hospital/Baylor College of Medicine, Cook Children's Medical Center, Pediatric Branch of the National Cancer Institute, and University of Oklahoma Health Science Center). A total of 34 samples (14 initial biopsies and 20 definitive surgery specimens) were included [27], which were obtained from 28 individual patients with 18 males and 10 females. Six patients contributed two samples each, both initial biopsies and definitive surgery specimens, whereas the remaining 22 patients contributed one sample each, either an initial biopsy or a definitive surgery specimen. The initial biopsy samples were obtained at the time of diagnosis before the initiation of preoperative chemotherapy. The definitive surgery samples were collected after the completion of preoperative chemotherapy. The drug responses are centrally reviewed

by one pathologist after definitive surgery. Good response is defined as more than 90 percent necrosis in tumor, and poor response with less than 90 percent necrosis.

It is clinically useful to monitor drug response after chemotherapy so that further therapeutic measures could be adjusted to optimize the chance of survival of the poor responders as revealed at the time of definitive surgery. Another more significant approach is to predict the drug response before chemotherapy is initiated, which provides clinicians the option to customize therapy based on the predicted response. Therefore the gene arrays of the 20 definitive surgery specimens can be used to train the classifier, and the gene arrays of the 14 initial biopsies can be used to verify the trained classifier.

To process the gene data, the raw quantification outputs of all array experiments were pre-processed and filtered by removing spots with low signal intensity and low sample variance ($P > 0.01$), as well as those that were missing in the experiments. A total of 1,934 genes remained after pre-processing and filtering. The intensities were then normalized by an intensity dependent local weighted regression method. After normalization, the intensity ratios were log-transformed before further analysis.

III. CLASSIFICATION AND FEATURE SELECTION WITH SUPPORT VECTOR MACHINE

Let a gene microarray dataset D be denoted by $\{(X_i, Y_i)\}_{i=1}^l$, where $X_i \in R^m$ is the gene expression level of the i -th patient, $Y_i \in \{-1, 1\}$ is the condition label for a binary classification problem, and m is the number of gene features. The dataset after performing gene selection is defined as $\ell(D) = \{(\ell(X_i), Y_i)\}_{i=1}^l$ with $\ell \in R^{m'}$, where the function $\ell(\cdot)$ selects m' ($< m$) gene features among all the m gene features from the gene expression data set D . With SVM, all training samples are mapped to the feature space by a non-linear function $\phi(\ell(X_i))$. A separating hyperplane in the feature space can be expressed as $w = \sum_{i=1}^l \alpha_i y_i \phi(\ell(X_i))$, $\alpha_i \geq 0$ and $w \in R^{m'}$. The optimal separating hyperplane is defined as a linear classifier which can separate the two classes of training samples with the largest marginal width, and the optimal parameters $\alpha = [\alpha_i]$ are obtained by maximizing the following function:

$$W(\alpha) = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l [\alpha_i \alpha_j y_i y_j \phi(\ell(X_i)) \cdot \phi(\ell(X_j))] \quad (1)$$

with $0 \leq \alpha_i \leq C$, $i = 1, \dots, l$ in which C is the regularization parameter controlling the tradeoff between model complexity and empirical risk. In the hyperplane expression, only those items with $\alpha_i > 0$ can remain. The samples that lie along the margins of the decision boundary (by Kuhn-Tucker theorem) are called support vectors, which satisfy the equation:

$$w = \sum \alpha_i y_i \phi(\ell(X_i)) \text{ where } \alpha_i \geq 0 \quad (2)$$

To avoid the computation of the inner product $\langle \phi(\ell(X_i)), \phi(\ell(X_j)) \rangle$ in the high dimensional space during the optimization of (1), the kernel function that can satisfy the Mercer's condition is introduced:

$$K(\ell(X_i), \ell(X_j)) = \langle \phi(\ell(X_i)), \phi(\ell(X_j)) \rangle \quad (3)$$

Typical kernel functions like linear, sigmoid, radial basis function (RBF) can be used. RBF is widely used as the kernel function in gait classification studies [12, 35]. As there is no analytical study for choosing the optimal kernel function, it is usually chosen by trial and error, which depends on the problem nature. As a result, the selected kernel to the problem may not be a good one. An integrated kernel that combines different kernels is proposed in this paper. The parameters in the integrated kernel could be found by some tuning algorithm.

$$K = \sum_{i=1}^3 \omega_i K_i \quad (4)$$

$$\sum_{i=1}^3 \omega_i = 1 \quad (5)$$

where K_1, K_2, K_3 are the linear, sigmoid and radial basis functions respectively, and ω_1 is the weight for each kernel. For a new test sample X , a decision function of an SVM classifier with a kernel K can then be defined based on the selected gene subset:

$$f(X, D, \ell) = \text{sgn} \left(\sum_{\substack{\text{SUPPORT} \\ \text{VECTOR}}} \alpha_i y_i K(\ell(X_i), \ell(X)) + b \right) \quad (6)$$

To develop a robust SVM model based on the training set in a practical way, the leave-one-out cross-validation (LOOCV) can be applied. Under this approach, one sample is left out as a testing sample, and the remaining $l-1$ samples are employed as training samples. Let $\overline{D}_k = \{(X_i, Y_i)\}_{i=1, i \neq k}^l$ represent the training set, the overall accuracy is reflected by:

$$J(X, D, \ell) = \frac{1}{2l} \left(\sum_{k=1}^l \left| f(X, \overline{D}_k, \ell) - y_k \right| \right) \quad (7)$$

where $J(X, D, \ell)$ is the total error of the classifier. Now the integrated gene feature selection and SVM classification problem are transformed to optimizing the above single objective function (7) by searching the feature selection l and the SVM parameters in $f(\cdot)$.

IV. PSO WITH WAVELET MUTATION (WPSO)

Consider a swarm $X(t)$ at the t -th iteration. Each particle $\mathbf{x}^p(t) \in X(t)$ contains κ elements $x_j^p(t) \in \mathbf{x}^p(t)$ at the t -th iteration, where $p = 1, 2, \dots, \gamma$ and $j = 1, 2, \dots, \kappa$; γ denotes the number of particles in the swarm. First, the particles of the swarm are initialized and then evaluated based on a defined fitness values. The objective of PSO is to minimize the fitness value (cost value) of a particle through iterative steps. The swarm evolves from iteration t to $t+1$ by repeating the procedures as shown in Fig. 1. The velocity $v_j^p(t)$ (corresponding to the flight speed in a search space) and the position $x_j^p(t)$ of the j -th element of the p -th particle at the t -th generation can be calculated using the following formulas :

$$v_j^p(t) = (w \cdot v_j^p(t-1) + \varphi_1 \cdot \text{rand}_j^p()) \cdot (pbest_j^p - x_j^p(t-1)) + \varphi_2 \cdot \text{rand}_j^p() \cdot (gbest_j - x_j^p(t-1)) \quad (8)$$

$$x_j^p(t) = x_j^p(t-1) + k \cdot v_j^p(t) \quad (9)$$

where $pbest^p = [pbest_1^p \ pbest_2^p \ \dots \ pbest_\kappa^p]$ and $gbest = [gbest_1 \ gbest_2 \ \dots \ gbest_\kappa]$. The best previous position of the p -th particle is recorded and represented as $pbest^p$; the position of the best particle among all the particles is represented as $gbest$; w is an inertia weight factor; φ_1 and φ_2 are acceleration constants; $\text{rand}_j^p()$ returns a uniform random number in the range of $[0,1]$; k is a constriction factor derived from the stability analysis of (8) to ensure the system converges but not prematurely [6]. PSO utilizes $pbest^p$ and $gbest$ to modify the current search point in order to prevent the particles from moving in the same direction, but to converge gradually toward $pbest^p$ and $gbest$. A suitable selection of the inertia weight w provides a balance between the global and local explorations. Generally, w can be dynamically set with the following equation [7]:

$$w = w_{\max} - \frac{w_{\max} - w_{\min}}{T} \times t \quad (10)$$

where t is the current iteration number, T is the total number of iteration, w_{\max} and w_{\min} are the upper and lower limits of the inertia weight, and are set to 1.2 and 0.1 respectively in this paper. In (8), the particle velocity is limited by a maximum value v_{\max} . The parameter v_{\max} determines the resolution, or fitness, of regions between the present position and the target position to be searched. This limit enhances the local exploration of the problem space, and depicts the incremental changes of learning.

```

begin
  t→0          // iteration number
  Initialize X(t) // X(t): Swarm for iteration t
  Evaluate f(X(t)) // f(·): fitness function
  while (not termination condition) do
    begin
      t→t+1
      // Process of SPSO //
      Update velocity v(t) and position of each
      particle x(t) based on (8) and (9) respectively
      if v(t) > vmax
        v(t) = vmax
      end
      if v(t) < -vmax
        v(t) = -vmax
      end
      Perform mutation operation with pm
      Reproduce a new X(t)
      Evaluate f(X(t))
    end
  end

```

Fig.1. Pseudo code for WPSO

Before generating a new $X(t)$, the wavelet mutation operation is performed. The resulting WPSO has the following mutation operation: Every particle of the swarm will have a chance to mutate governed by a probability of mutation $\mu_m \in [0, 1]$, which is defined by the user. For each particle, a random number between 0 and 1 is generated such that if it is less than or equal to μ_m , the mutation will take place on that element of the particle. For instance, if $\mathbf{x}^p(t) = [x_1^p(t), x_2^p(t), \dots, x_\kappa^p(t)]$ is the selected p -th particle and the particle element $x_j^p(t)$ is randomly selected for mutation (the value of $x_j^p(t)$ is inside the element's boundaries $[para_{\min}^j, para_{\max}^j]$), the resulting particle is given by $\bar{\mathbf{x}}^p(t) = [\bar{x}_1^p(t), \bar{x}_2^p(t), \dots, \bar{x}_\kappa^p(t)]$,

$$\bar{x}_j^p(t) = \begin{cases} x_j^p(t) + \sigma \times (para_{\max}^j - x_j^p(t)) & \text{if } \sigma > 0 \\ x_j^p(t) + \sigma \times (x_j^p(t) - para_{\min}^j) & \text{if } \sigma \leq 0 \end{cases} \quad (11)$$

where $j \in 1, 2, \dots, \kappa$. Using the Morlet wavelet as the mother wavelet, we have

$$\sigma = \frac{1}{\sqrt{a}} e^{-\left(\frac{\phi}{a}\right)^2 / 2} \cos\left(5\left(\frac{\phi}{a}\right)\right) \quad (12)$$

The value of the dilation parameter a is set to vary with the value of t/T in order to meet the fine-tuning purpose, where T is the total number of iteration and t is the current number of iteration. In order to perform a local search when t is large, the value of a should increase as t/T

increases so as to reduce the significance of the mutation. Hence, a monotonic increasing function governing a and t/T is proposed as follows.

$$a = e^{-\ln(g) \times \left(1 - \frac{t}{T}\right)^{\zeta_{wm}} + \ln(g)} \quad (13)$$

where ζ_{wm} is the shape parameter of the monotonic increasing function, g is the upper limit of the parameter a .

V. FORMULATION OF WPSO OPTIMIZATION SYSTEM

The number of features in microarrays is significantly larger than the number of samples. In order to improve the performance of the classification, reducing the dimension of the data set is necessary, and the filter technique has been adopted for this purpose.

As shown in Fig. 2, the two-sample t -test is performed to pre-screen the noisy genes, and the remaining unfiltered genes are the more significant genes. The t -value of the gene feature can be determined by the following equation:

$$t = \frac{\mu_+ - \mu_-}{\sqrt{\frac{Var_+}{n_+} + \frac{Var_-}{n_-}}} \quad (14)$$

where n_+ and n_- are the numbers of patients with positive and negative conditions. μ_+ and Var_+ are the mean value and variance of gene feature in the positive group, and μ_- and Var_- are the mean value and variance of gene feature in the negative group. The gene features with low t -values are filtered out, and the ones with high t -values are passed to the WPSO optimization unit, which can search for the most significant gene features and the SVM parameters for classification. Practically, the gene features that have t -values higher than 2.5 are passed to the WPSO optimization unit and the rest of the gene features are not considered.

In the WPSO selection, the algorithm provides a binary encoded particle where each bit represents a gene. If a bit value is 1, that gene is kept in the subset; and the bit value of 0 indicates that the gene is not included in the subset. Therefore, the particle length is equal to the number of genes in the subset. Based on the selected gene features from the binary coded representation, the SVM classifier is then trained in Step 2b with real coded parameters through leave-one-out-cross-validation (LOOCV), during which the cost function is used to evaluate the fitness of a solution

$$fitness = error = J(X, D, \ell, \sigma, C) \quad (15)$$

The objective of the WPSO is to minimize the fitness function by searching the optimal gene feature. It can be seen that the best fitness value is zero if there is no any error. On realizing the WPSO, the following simulation conditions are used:

- Shape parameter of the wavelet mutation (ζ_{wm}): 0.2
- Probability of mutation (μ_m): 0.2
- Acceleration constant ϕ_1 : 2.05
- Acceleration constant ϕ_2 : 2.05
- Constriction factor k : 0.005
- Parameter g : 1000
- Swarm size: 20
- Number of iteration: 1000
- Initial population $X(0)$: all the particles are initialized randomly
- Initial global best g_{best} : it is generated uniformly at random
- Initial previous best p_{best}^p : it is generated uniformly at random.

VI. RESULT AND ANALYSIS

The SVM classifier was first trained by the sample data training set to classify the tumour to be a good responder or a poor responder. The best gene combination can be derived according to its cross-validation accuracy. Then the gene subset and the trained SVM classifier will be verified by using the data test set. The results show the relevance of using the selected gene combination as the signature of chemoresistance, and the ability of this gene signature to predict at the time of initial diagnosis the histologic response as determined at the time of definitive surgery.

The result of the proposed algorithm was used to compare with the results of various supervised classification algorithms (compound covariate predictor CCP, linear discriminant analysis LDA, 1-nearest neighbor 1-NN, 3-nearest neighbor 3-NN, nearest centroid NC and SVM) collected from [27]. For all algorithms, LOOCV was used to fairly evaluate the performance of each classifier, and all of them can only achieve 60-70% classification accuracies. In the training, the proposed method can achieved 80% classification accuracy, as shown in Table 1. That means out of the 20 samples of training sets, there are 16 samples that can be classified successfully. Comparing with other methods, the proposed method is a more robust technique for gene signature selection and classification.

The weights of kernels ω_1 , ω_2 and ω_3 were found to be 0.533, 0.1663 and 0.3007 respectively. In this paper, there are 937 genes selected as the predictor genes; however, there is not enough information to explain the cooperative relationship among genes in the subsets.

tumor ID of the training sets	Histologic response	concordance of classification with histologic response					
		CCP	LDA	1-NN	3-NN	NC	proposed method
300	Good	No	No	No	No	No	No
308	Good	Yes	Yes	Yes	Yes	Yes	Yes
386	Good	No	No	No	No	No	No
394	Good	No	No	No	No	No	No
452	Good	Yes	Yes	Yes	Yes	Yes	Yes
759	Good	Yes	Yes	Yes	Yes	Yes	Yes
771	Good	Yes	Yes	Yes	Yes	Yes	Yes
221	poor	Yes	Yes	Yes	Yes	Yes	Yes
236	poor	Yes	Yes	Yes	Yes	Yes	Yes
241	poor	Yes	Yes	Yes	Yes	Yes	Yes
252	poor	Yes	Yes	Yes	Yes	Yes	Yes
311	poor	Yes	Yes	Yes	Yes	Yes	Yes
342	poor	Yes	Yes	Yes	Yes	Yes	Yes
392	poor	Yes	Yes	Yes	Yes	Yes	Yes
483	poor	Yes	Yes	Yes	Yes	Yes	Yes
591	poor	Yes	Yes	Yes	Yes	Yes	Yes
680	poor	No	No	No	No	No	Yes
691	poor	No	No	No	No	No	Yes
760	poor	No	Yes	Yes	Yes	No	Yes
761	poor	No	No	No	No	No	No
% correctly classified		65%	70%	70%	70%	65%	80%

Table 1. LOOCV of 20 definitive surgery osteosarcoma samples

The optimal SVM classifier with the selected gene subset was then respectively applied to predict the chemo-response of the 14 initial biopsy samples for further validation purpose. The test data are completely independent of the training data and process. In the testing, the corresponding SVM classifier misclassified one sample among the 14 samples, with a correct classification rate of 92.9%, as shown in Table 2. The results in [27] also misclassified Tumor ID 410 of a clinically good responder as a poor responder. This finding is consistent with the study in [27], where it was pointed out that this patient initially presented with localized disease but eventually developed recurrence in the lungs 25 months after the completion of therapy, suggesting that there were resistant cells present in the initial biopsy, which might have metastasized to the lungs before definitive surgery and subsequently gave rise to the recurrent tumor. This result further indicates that the gene expression signature of the resistant cells in the definitive surgery samples was already present in the initial biopsy samples at the time of diagnosis.

Tumor ID	Histologic response	Classified by proposed algorithm
410	Good	No
197	Poor	Yes
207	Poor	Yes
278	Good	Yes
289	Poor	Yes
345	Good	Yes
204	Poor	Yes
274	Poor	Yes
299	Good	Yes
464	Poor	Yes
479	Poor	Yes
481	Poor	Yes
545	Good	Yes
654	Good	Yes

Table 2. Classification results of the proposed algorithm.

VII. CONCLUSION

In this paper, we have proposed an integrated approach of WPSO and support vector machine to select a compact gene subset and optimize parameters simultaneously. Prediction of chemo-response using an independent dataset of 14 initial biopsy samples achieved 92.9% of accuracy, reflecting that the proposed algorithm is promising in selecting gene signatures that the derived genes are consistently expressed in initial biopsy samples and thus have predicting value for drug response. The results suggest that the proposed algorithm can be used to generate hypothesis for the identification and validation of genetic biomarkers for diagnostic and therapeutic purposes. In the future, we will further validate the proposed algorithm in solving other similar classification problems with large data sets of nasopharyngeal carcinoma and lung cancer.

REFERENCES

- [1] U. Alon, N. Barkai, D.A. Notterman, K. Gish, S. Ybarra, D. Mack, and A.J. Levine, "Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays," *Proceedings of the National Academy of Sciences of the United States of America*, pp. 6745-6750, 1999.
- [2] J. Bins and B. Draper, "Feature selection from huge features sets," in *Proc. Int. Conf. Computer Vision*, vol. 2, 2001, pp. 159-165.
- [3] M. Bittner, P. Meltzer, Y. Chen, "Molecular classification of cutaneous malignant melanoma by gene expression profiling," *Nature*, vol. 406(6795), pp. 536-540, 2000.
- [4] B. Boser, I. Guyon, and V. Vapnik, "An training algorithm for optimal margin classifiers," in *Proc. 5th Annual Workshop on Computational Learning Theory*, 1992, pp. 144-152.
- [5] F.Z. Brill, D.E. Brown, and W.N. Martin, "Fast genetic selection of features for neural network classifiers," *IEEE Trans. Neural Networks*, vol. 3, no. 2, pp. 324-328, 1998.
- [6] R.C. Eberhart and Y. Shi, "Comparing inertia weights and constriction factors in particle swarm optimization," in *Proc. Congress on Evolutionary Computing*, vol. 1, Jul. 2000, pp.84-88.
- [7] J. Kennedy and R. Eberhart, *Swarm Intelligence*. Morgan Kaufmann Publishers, 2001.
- [8] M. Daly and R. Ozol, "The search for predictive patterns in ovarian cancer: proteomics meets bioinformatics," *Cancer Cell*, vol. 1, issue 2, 111-112, 2002.
- [9] K.A. De Jong, *The analysis and behaviour of a class of genetic adaptive systems*. Ph.D. thesis, University of Igan, 1975.
- [10] S. Dudoit, J. Fridlyand, and T.P. Speed, "Comparison of discrimination methods for the classification of tumors using gene expression data," *Journal of the American Statistical Association*, vol. 97(457), pp. 77-87, 2002.
- [11] S. Dudoit and N. Friedman, "Classification in microarray experiments," in *Statistical analysis of gene expression microarray data*. Boca Raton: Chapman and Hall/CRC, pp. 93-158, 2003.
- [12] V. Vapnik and O. Chapelle, "Bounds on error expectation for support vector machines," *Neural Computation*, vol. 12, pp. 2013-2036, 2000.
- [13] I. Guyon, J. Weston, and S. Barnhill, "Gene selection for cancer classification using support vector machines," *Machine Learning*, vol. 46, pp. 389-422, 2002.
- [14] P.A. Meyers, G. Heller, J. Healey, A. Huvos, J. Lane, R. Marcove, A. Applewhite, V. Vlamis and G. Rosen, "Chemotherapy for nonmetastatic osteogenic sarcoma: the Memorial Sloan-Kettering experience," *Journal of Clinical Oncology*, vol. 10, pp. 5-15, 1992.
- [15] J.H. Hong and S.B. Cho, "Efficient huge-scale feature selection with speciated genetic algorithm," *Pattern Recognition Letters*, vol. 27, pp. 143-150, 2006.
- [16] T.M. Huang and V. Kecman, "Gene extraction for cancer diagnosis by support vector machines - An improvement," *Artificial Intelligence in Medicine*, vol. 35, mpp. 185-194, 2005.
- [17] M.B. Eisen, P.T. Spellman, P.O. Brown, and D. Bostein, "Cluster analysis and display of genome-wide expression patterns," *Proceedings of the National Academy of Science*, vol. 95, no. 25, pp. 14863-14868, 1998.
- [18] T.S. Furer, N. Cristianini, N. Duffy, D.W. Bednarski, M. Schummer, and D. Haussler, "Support vector machine classification and validation of cancer tissue samples using microarray expression data," *Bioinformatics*, vol. 16, no. 10, pp. 906-914, 2000.
- [19] L.M. Fu and E.S. Youn, "Improving reliability of gene selection from microarray functional genomics data," *IEEE Trans. Information Technology in Biomedicine*, vol. 7, no. 3, pp. 191-196, 2003.
- [20] D.E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison Wesley, 1989.
- [21] T.R. Golub, D.K. Slonim, P. Tamayo, C. Hurd, M. GassenBeek, J.P. Mesirov, H. Coller, M.L. Loh, J.R. Downing, M.A. Caligiuri, C.D. Blomfield, and E.S. Lander, "Molecular classification of cancer: class discovery and class prediction by gene-expression monitoring," *Science*, vol. 286, pp. 531-537, 1999.
- [22] E. Hartuv, A. Schmitt, J. Lange, S. Meier-Ewert, H. Lehrach, and R. Shamir, "An algorithm for clustering cDNA fingerprints," *Genomics*, vol. 66, no. 3, pp. 249-256, 2000.
- [23] J. Khan, J.S. Wei, M. Ringner, L.H. Saal, M. Ladanyi, F. Westermann, F. Berthold, M. Schwab, C.R. Antonescu, C. Peterson, and P.S. Meltzer, "Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks," *Nature Medicine*, vol. 7, no. 6, pp. 673-679, 2001.
- [24] R. Kohavi and G. John, "Wrappers for feature subset selection," *Artificial Intelligence*, vol. 97, pp. 273-324, 1997.
- [25] L. Li, C.R. Weinberg, T.A. Darden, and L.G. Pedersen, "Gene selection for sample classification based on gene expression data: study of sensitivity to choice of parameters of the GA/KNN method," *Bioinformatics*, vol. 17, no. 12, pp. 1131-1142, 2001.
- [26] M.P. Link, M.C. Gebhardt, and P.A. Meyers, *Principles and Practice of Pediatric Oncology*. Philadelphia, PA, Lippincott, Williams & Wilkins, pp. 1051-1089, 2002.
- [27] T.K. Man, M. Chintagumpala, J. Visvanathan, J. Shen, L. Perlaky, J. Hicks, M. Johnson, N. Davino, J. Murray, L. Helman, W. Meyer, T. Triche, K.K. Wong, and C.C. Lau, "Expression profiles of osteosarcoma that can predict response to chemotherapy," *Cancer Research*, vol. 65, no. 18, pp. 8142-8150, 2005.
- [28] H. Muhlenbein and D.S. Voosen, "Predictive models for the breeder genetic algorithm: I. Continuous parameter optimization," *Evolutionary Computation*, vol. 1, no. 1, pp. 25-49, 1993.
- [29] I.M. Oliver, D.J. Smith, and C.R.J. Holland, "A study of permutation crossover operators on the traveling salesman problem," in *Proc. 2nd Int. Conf. Genetic Algorithms*, 1987, pp. 224-230.
- [30] S. Peng, Q. Xu, X.B. Ling, X. Peng, W. Du, and L. Chen, "Molecular classification of cancer types from microarray data using the combination of genetic algorithms and support vector machines," *FEBS Letters*, vol. 555, pp. 358-362, 2003.
- [31] M.D. Radmacher, L.M. McShane, and R. Simon, "A paradigm for class prediction using gene expression profiles," *Journal of Computational Biology*, vol. 9, pp. 505-511, 2002.
- [32] Y.D. Zhao, C. Pinilla, D. Valmori, R. Martin, and R. Simon, "Application of support vector machines for T-cell epitopes prediction," *Bioinformatics*, vol. 19, no. 15, pp. 1978-1984, 2003.
- [33] Single Nucleotide Polymorphisms for Biomedical Research, The SNP Consortium Ltd., 2003. Available at <http://snp.cshl.org>.
- [34] A. Vander, J. Sherman, and D. Luciano, *Human Physiology*. McGraw-hill, New York, 2001.
- [35] V.N. Vapnik, *Statistical Learning Theory*. Wiley Interscience, 1998.

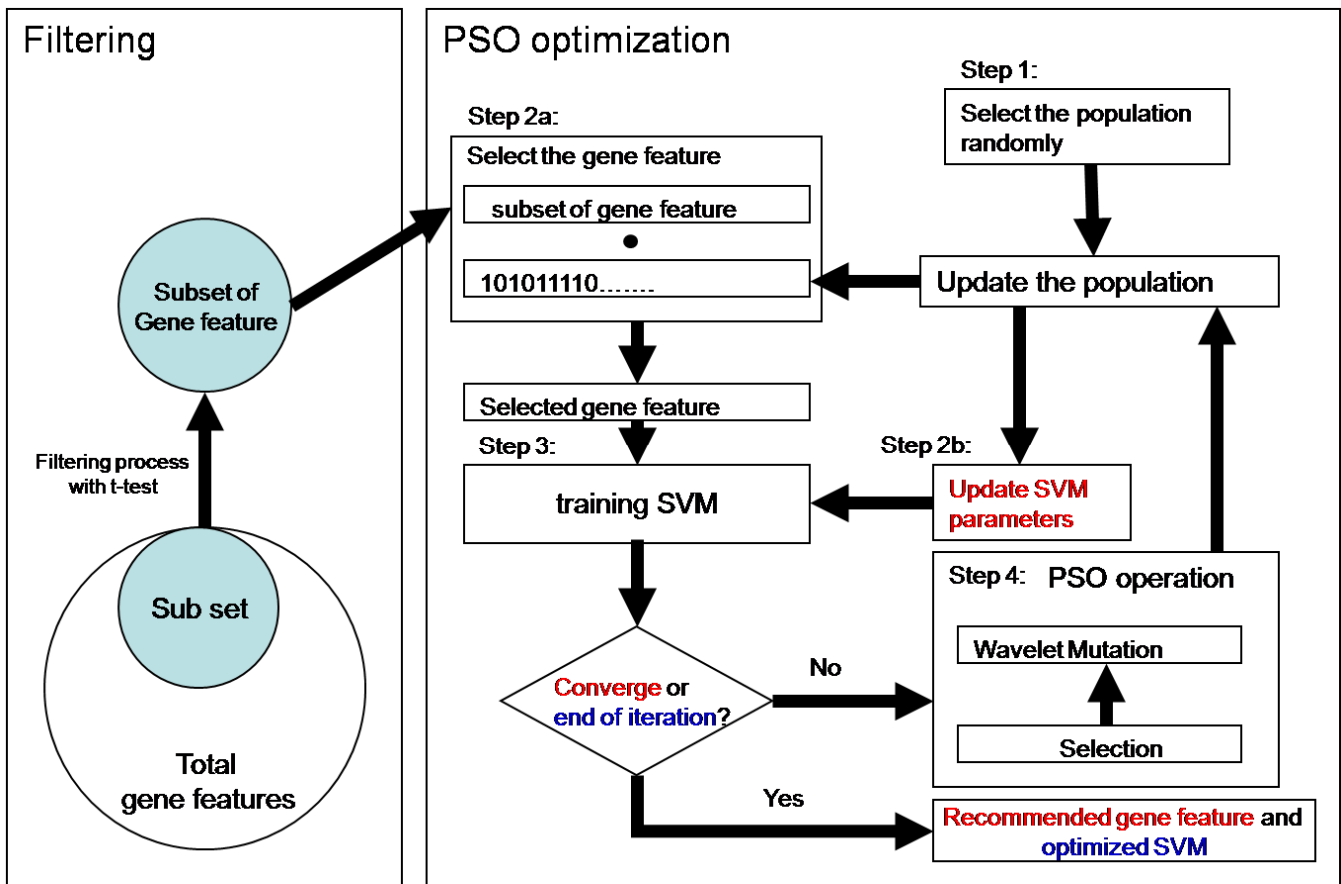


Fig. 2. Illustration for the proposed algorithm.