Identification of Full and Partial Class Relevant Genes

Zexuan Zhu, Yew-Soon Ong, *Member, IEEE* and Jacek M. Zurada, *Fellow, IEEE*

Zexuan Zhu is with Bioinformatics Research Centre, School of Computer Engineering, Nanyang Technological University, Singapore

Yew-Soon Ong is with Division of Information Systems, School of Computer Engineering, Nanyang Technological University, Singapore

Jacek M. Zurada is with Computational Intelligence Laboratory, Electrical and Computer Engineering, University of Louisville, Louisville, United States

Abstract

Multiclass cancer classification on microarray data has provided the feasibility of cancer diagnosis across all of the common malignancies in parallel. Using multiclass cancer feature selection approaches, it is now possible to identify genes relevant to a set of cancer types. However, besides identifying the relevant genes for the set of all cancer types, it is deemed to be more informative to biologists if the relevance of each gene to specific cancer or subset of cancer types could be revealed or pinpointed. In this paper, we introduce two new definitions of multiclass relevancy features, i.e., full class relevant (FCR) and partial class relevant (PCR) features. Particularly, FCR denotes genes that serve as candidate biomarkers for discriminating all cancer types. PCR, on the other hand, are genes that distinguishes subsets of cancer types. Subsequently, a Markov blanket embedded memetic algorithm is proposed for the simultaneous identification of both FCR and PCR genes. Results obtained on commonly used synthetic and real world microarray datasets show that the proposed approach converges to valid FCR and PCR genes that would assist biologists in their research work. The identification of both FCR and PCR genes is found to generate improvement in classification accuracy on many microarray datasets. Further comparison study to existing state-of-the-art feature selection algorithms also reveals the effectiveness and efficiency of the proposed approach.

Index Terms

Bioinformatics, Microarray, Multiclass cancer classification, Feature/Gene selection, Memetic algorithm, Markov blanket.

I. INTRODUCTION

Microarray technology has provided a promising tool for cancer diagnosis. Various statistical and machine learning methods have since been successfully used for facilitating cancer identification and prediction. Most of these techniques were generally tailored for handling binary classification problems. Signal-to-noise ratio [1], t-test [2], between-groups to within-groups ratio [3], support vector machine (SVM) based on recursive feature elimination method [4], and least squares bound method [5], etc. are among those applicable mainly to two-class problems. Microarray data with multiclass labels remains to be a big challenge due to the significant statistical and analytical implications involved over the binary counterpart.

As a result, in recent years there has been increasing research activities in the field of bioinformatics focusing on new methods for handling microarray data containing multiple labels.

In particular, several studies on identifying the relevant genes in multiclass cancer data have been reported. Binary SVM with One-Versus-All (OVA) classification schema [6], partial least squares [7], genetic algorithm (GA) with maximum likelihood [8] or SVM hybridization [9], degree of differential prioritization [10], shrunken centroid [11], [12] and Bayesian model averaging [13], represent some of the core approaches that have emerged recently. For detailed comparative studies on multiclass gene selection problems, the reader is referred to the work of Li et al. [14] and Statnikov et al. [15], where various state-of-the-art feature selection methods and classification algorithms for solving multiclass microarray datasets are investigated.

Multiclass cancer classification using microarray data has provided the feasibility of cancer diagnosis across all of the common malignancies in parallel. Using multiclass cancer feature selection approaches, it is now possible to identify genes that are relevant to the set of cancer types considered. However, it is deemed to be more informative to biologists if the relevance of each gene to a specific cancer or subset of cancer types could be revealed or pinpointed. In this paper, we propose the concepts of multiclass relevancy features, namely, i) full class relevant (FCR) and ii) partial class relevant (PCR) features. FCR denotes genes that serve as candidate biomarkers for discriminating any cancer types. PCR, on the other hand, represents genes that serve to distinguish subsets of cancer types.

Further, for simultaneous identification of both FCR and PCR genes in multiclass microarray cancer data, we propose a Markov blanket [16]–[19] embedded multiobjective memetic algorithm (MBE-MOMA). In particular, MBE-MOMA searches on multiple objectives simultaneously where each objective corresponds to the search for an optimal PCR feature subset. An approximate Pareto optimal set [20] of solutions representing distinct options for solving the multiobjective optimization problem based on different tradeoffs of the objectives is obtained with a single simulation run of the MBE-MOMA. There exists one extremal solution for each objective, which represents the corresponding PCR feature subset. An FCR feature subset is formed as the intersection of all solutions.

With the uncovered FCR and PCR genes, two schemes for their possible synergies in multiclass prediction are investigated subsequently. The first ensemble scheme has been widely used in feature selection and classification [21], [22]. In this scheme, each individual FCR/PCR feature subset is used for classifying the multiclass labels, and the predictions of each respective trained classifier are subsequently aggregated based on a voting strategy. The second conjunction scheme

considers a simple union of all FCR and PCR feature subsets as a single feature subset. The resultant union feature subset is then used for the cancer types prediction.

The rest of this paper is organized as follows: Section II presents the notion of FCR and PCR features, including the details on the proposed MBE-MOMA to identify FCR and PCR features. Section III presents the experimental results and some discussions on both synthetic and microarray datasets. Finally, Section IV summarizes this study.

II. System and Methodology

In this section, we present the definitions of FCR and PCR features, search for optimal PCR/FCR feature subsets problem formulation, basic concepts of Pareto-based multiobjective evolutionary algorithm, followed by a detail description of the proposed Markov blanket embedded multiobjective memetic algorithm (MBE-MOMA).

A. Full and Partial Relevant Feature

To facilitate our definition of FCR/PCR features, we provide first some background information of relevant features. Let $C = \{c_1, \ldots, c_k\}$ be the class set to be considered, X be a full set of features, X_i be a feature, and $\overline{X}_i$ be the set of all features that excludes X_i, i.e., $\overline{X}_i = X - \{X_i\}$. Based on the definition given in [18], [23], features can be categorized as strongly relevant, weakly relevant, or irrelevant.

Strong Relevance: A feature X_i is strongly relevant if and only if $P(C|X_i, \overline{X}_i) \neq P(C|\overline{X}_i)$.

Weak Relevance: A feature X_i is weakly relevant if and only if $P(C|X_i, \overline{X}_i) = P(C|\overline{X}_i)$ and $\exists \overline{X}'_i \subseteq \overline{X}_i$ such that $P(C|X_i, \overline{X}'_i) \neq P(C|\overline{X}'_i)$.

A feature **relevant** to the learning classes C can pose as strongly relevant or weakly relevant, otherwise, it is regarded as **irrelevant** to C. To biologists, besides identifying genes of relevance to C in multiclass cancer classification problems, it is more effective to pinpoint the specific cancer or subset of cancer types a gene is relevant to. Inspired by this requirement, it is appropriate to extend the original view of relevant feature with new definitions that can differentiate between partial class relevant (PCR) from full class relevant (FCR) features in multiclass problems. Here, we propose the notions of PCR and FCR based on the concept of Subset-Versus-Subset. To proceed, we provide here the definitions of Subset-Versus-Subset, Full Class Relevant, and Partial Class Relevance.

Subset-Versus-Subset: A Subset-Versus-Subset (SVS) of classes C is defined as SVS(C)= $\{A, B\}$ where $A, B \subseteq C$ and $A \bigcap B = \varnothing$.

Full Class Relevance: A feature X_i is said to be full class relevant if and only if it is relevant to all possible SVSs.

Partial Class Relevance: A feature X_i is said to be partial class relevant if and only if it is relevant to only some of the SVSs and there exists a SVS C', such that X_i is irrelevant to C'.

FCR features are relevant to all SVSs, including all pairs of classes $\{\{c_i\}, \{c_j\}\}$ $(i, j \in (1, \ldots, k))$, they are important for distinguishing between any two classes. On the other hand, PCR features are relevant to only some of the SVSs, they are helpful for distinguishing only subset of classes but not all classes. An ideal solution for feature selection on multiclass problem would be a FCR feature subset that can distinguish between any pair of classes. In practice, it is generally hard to find such an ideal FCR feature subset that would perform well on all classes. A selected feature subset performing well on some SVSs would not be successful on other SVSs. Hence PCR features are indispensable for learning the multiclass problem.

B. Identification of FCR/PCR Features, Problem Formulation

It is worth noting that the search for true FCR and PCR features may pose to be computationally intractable due to the large number of possible SVSs. Hence, one key issue is to choose an approximate scheme that can generate a coverage of the classes C. The One-Versus-All (OVA) scheme is generally regarded as more effective than other competing strategies in the literature of multiclass cancer classification [15]. Hence, we demonstrate here the use of OVA scheme for generating the SVSs in the present study. For classes C, OVA scheme creates k pairwise two-class SVSs (labeled as **OVA sets** in this study), with each of them constructed as $\{c_i, \overline{c}_i\}$, where $i \in (1, \ldots, k)$ and $\overline{c}_i = C - c_i$.

The search for optimal PCR feature subsets of k OVA sets can naturally be casted as a multiobjective optimization problem (MOP) [20], [24]–[27] with each objective corresponding to the feature selection accuracy of each OVA set. The MOP considered is thus defined as:

$$min\ F(s) = (f_1(s), \ldots, f_k(s))\ subject\ to\ s \in S \tag{1}$$

$$f_i(s) = -Acc(s, \{c_i, \overline{c}_i\}), (i \in (1, \ldots, k))$$

where $F(s)$ is the objective vector, s is the candidate selected feature subset, k is the number of classes, and S is the feasible domain of s. The i-th objective $f_i(s)$ is then $-Acc(s, \{c_i, \bar{c}_i\})$, which gives the classification accuracy of the i-th OVA set for the selected feature subset s. Finding a single solution that optimizes all objectives is not always possible because a selected feature subset performing well on an OVA set would not be successful on other OVA sets.

The straightforward way to solve this MOP is to identify the optimal PCR feature subset of each OVA set individually. The intersection of PCR feature subsets then form the FCR subset. However, such a process would be computationally intensive. In this section, we describe the MBE-MOMA proposed for simultaneous identification of both FCR and PCR features in a single run.

C. Basic Concepts of Multiobjective Evolutionary Algorithm

Pareto-based multiobjective evolutionary algorithm (MOEA) [20] is one of the most popular approaches for handling multiobjective optimization problems, due to its ability to find multiple diverse solutions and approximating the Pareto optimal set in a single simulation run, which provides distinct options to solve the problem based on different tradeoffs of the objectives. To understand the notion of optimality in MOP, we introduce some basic concepts of Pareto dominance, Pareto optimal, and Pareto front:

Pareto Dominance: A vector $u = (u_1, \ldots, u_k)$ is said to dominate $v = (v_1, \ldots, v_k)$ if and only if u is partially less than v, i.e., $\forall i \in \{1, \ldots, k\} : u_i \leq v_i \; and \; \exists i \in \{1, \ldots, k\} : u_i < v_i$.

Pareto Optimal: A solution vector $s^* \in S$ is called Pareto optimal if and only if there is no $s \in S$ for which $u = F(s) = (f_1(s), \ldots, f_k(s))$ dominates $v = F(s^*) = (f_1(s^*), \ldots, f_k(s^*))$. The set of all Pareto optimal solutions is called **Pareto Optimal Set** (denoted by $\mathcal{P}^*$).

Pareto Front: For a given MOP $\min \; F(s)$ and Pareto optimal set $\mathcal{P}^*$, the Pareto Front ($\mathcal{PF}^*$) is defined as: $\mathcal{PF}^* = \{u = F(s^*) = (f_1(s^*), \ldots, f_k(s^*)) | s^* \in \mathcal{P}^*\}$

Solving the MOP defined in Equation (1) with the MOEA would lead to an approximate Pareto optimal set of solutions $(s_1, s_2, \ldots, s_p)$ with different tradeoffs of classification accuracy on the OVA sets. One core advantage of the MOP approach to classification accuracy on the OVA sets is that the user is able to choose a solution according to the problem at hand. Subsequently, based on the preferences of the user or decision maker on the objectives, different solutions in

the approximate Pareto optimal set could be selected in the final predictions. For instance, if a cancer type c_i is considered to generate greater fatalities and hence regarded as more crucial than others, it would be possible to select a solution performing superiorly on f_i, such that c_i is better distinguished from other cancers. In the experimental study, we consider equal importance of each cancer type and demonstrate the use of both PCR and FCR feature subsets in the final prediction (the details are provided later in Section II-F). Particularly, the optimal PCR feature subset for the i-th OVA set is presented as the extremal solution on f_i, i.e., $PCR_i = \arg\min_{s_j} f_i(s_j)$, and the FCR feature subset is defined as intersection of all solutions i.e., $FCR = \{s_1 \cap s_2 \cap, \ldots, \cap s_p\}$.

D. Markov Blanket Embedded Multiobjective Memetic Algorithm

In this subsection, we present the proposed Markov blanket embedded multiobjective memetic algorithm (MBE-MOMA), which is a synergy of MOEA [20] and Markov blanket-based local search [28] for simultaneous identification of FCR and PCR features by solving the MOP defined in Equation (1). The pseudo code of the MBE-MOMA is outlined in Fig. 1.

At the start of the MBE-MOMA search, an initial population of solutions is randomly generated with each chromosome encoding a candidate feature subset. In the present work, each chromosome is composed of a bit string of length equal to the total number of features in the feature selection problem of interest. Using binary encoding, a bit of '1' ('0') implies the corresponding feature is selected (excluded). The fitness of each chromosome is then obtained using an objective vector defined in Equation (1).

In each MBE-MOMA generation, an offspring population $P'(t)$ is created from the parent population $P(t)$ using the genetic operators, i.e., selection, crossover, and mutation. $P(t)$ and $P'(t)$ are merged together as a mating pool $\{P'(t) \bigcup P(t)\}$. Subsequently, a non-dominated sorting [20] is used to categorize the solutions of the mating pool into levels of Pareto fronts. The non-dominated solutions of $\{P'(t) \bigcup P(t)\}$ then undergo the Markov blanket based individual learning in the form of local search, which will be greater detailed in Subsection II-E. Our experiences in Markov blanket based local search reveal that it is capable of dealing with irrelevant and redundant features efficiently [28].

It is worth noting that two individual learning schemes are commonly considered in memetic algorithm. Lamarckian learning considers the inheritance of acquired traits by forcing the genotype to reflect the result of the locally improved individual in the population to compete for

Markov Blanket Embedded Multiobjective Memetic Algorithm

BEGIN

1) t = 0.
2) **Initialize:** Randomly generate an initial population $P(t)$ of feature subsets encoded with binary strings.
3) Evaluate fitness $F(s)$ of each solution in $P(t)$.
4) Rank $P(t)$ using Pareto dominance and calculate the crowding distance [24].
5) **While**(*Termination Criterion Not Fulfilled*)
6) Select a temporary population $P'(t)$ from $P(t)$ based on Pareto ranking and crowding distance.
7) Perform crossover and mutation on $P'(t)$.
8) Evaluate fitness $F(s)$ of each solution in $P'(t)$.
9) Rank $\{P'(t) \bigcup P(t)\}$ using Pareto dominance.
10) Apply **Markov blanket based local search** on the non-dominated solutions of $\{P'(t) \bigcup P(t)\}$ and generate an improved population $P''(t)$.
11) Rank the population $\{P(t) \bigcup P'(t) \bigcup P''(t)\}$ using Pareto dominance calculate the crowding distance.
12) Select solutions from $\{P(t) \bigcup P'(t) \bigcup P''(t)\}$ to create a new population $P(t+1)$ based on Pareto ranking and crowding distance.
13) t = t+1.
14) **End While**

END

Fig. 1. Outline of Markov blanket embedded Multiobjective Memetic Algorithm for Feature Selection

reproductive opportunities. Baldwinian learning, on the other hand, only alters the fitness of the individuals and the improved genotype is not encoded back into the population. In the present study, the locally improved solution s' of a selected solution s through individual learning is then archived in a temporary population $P''(t)$. Subsequently, a new population $P(t+1)$ of the

same size as $P(t)$ is generated from $\{P(t) \bigcup P'(t) \bigcup P''(t)\}$ in the spirit of Lamarckian learning. Elitism and diversity in $P(t+1)$ is maintained based on Pareto dominance and crowding distance [24].

In MBE-MOMA, the evolutionary operators include binary tournament selection [24], uniform crossover, and mutation operators [29]. Note that when two chromosomes are found having similar fitness on each objective (i.e., for a misclassification error of less than one data instance, the difference between their fitness on each objective is designed here to be less than a small value of $1/D$, where D is the number of instances), the one with the smaller number of selected features is given a higher chance of surviving to the next generation. On the other hand, like most existing work in the literature, if prior knowledge on the optimum number of features is available, it makes sense to constrain the number of bits '1' in each chromosome to a maximum of m in the evolutionary search process. In this case, specialized restrictive crossover and mutation [30] instead of the basic evolutionary operators are necessary, so that the number of bits '1' in each chromosome does not violate the constraint derived from the prior knowledge on m throughout the search.

E. Markov Blanket Based Local Search

MOEA has been widely used for feature selection [21], [31]–[34], which solves a MOP of minimizing the number of selected features while maximizing classification accuracy, simultaneously. Similar to single objective evolutionary algorithm, it has been shown that the hybridization of MOEAs and local search can lead to significant improvement in the searching ability of MOEAs [25], [26], [28], [30], [35]–[38]. Such a hybrid multiobjective algorithm is commonly also known as multiobjective memetic algorithm (MOMA).

The Markov blanket based local search [28] is used to fine-tune the MOEA solutions by adding relevant features and deleting redundant features. A weakly relevant feature having a Markov Blanket [16]–[19] in the selected feature subset is considered as a **redundant** feature. The Markov blanket of a feature X_i is defined as follows:

> **Markov Blanket:** Let M be a subset of features which does not contain X_i, i.e., $M \subseteq X$ and $X_i \notin M$. M is a Markov blanket of X_i if X_i is conditionally independent of $(X \cup C) - M - \{X_i\}$ given M, i.e., $P(X - M - \{X_i\}, C | X_i, M) = P(X - M - \{X_i\}, C | M)$.

If a feature X_i has a Markov blanket M within the currently selected feature subset, it suggests that X_i gives no more information beyond M about C and other selected features, therefore, X_i can be removed safely. However, since the computational complexity to determine the conditional independence of features is typically very high, the computationally efficient approximate Markov blanket proposed by Yu and Liu [18] is used in place of the original method, which uses only one feature to approximate the Markov blanket of X_i.

Approximate Markov Blanket: For two features X_i and $X_j (i \neq j)$, X_j is said to be an approximate Markov blanket of X_i if $SU_{j,C} \geq SU_{i,C}$ and $SU_{i,j} \geq SU_{i,C}$ where the symmetrical uncertainty SU [39] measures the correlation between features (including the class, C).The symmetrical uncertainty is defined as:

$$SU(X_i, X_j) = 2 \left[\frac{IG(X_i|X_j)}{H(X_i) + H(X_j)} \right] \tag{2}$$

where $IG(X_i|X_j)$ is the information gain between features X_i and X_j, $H(X_i)$ and $H(X_j)$ denote the entropies of X_i and X_j respectively. $SU_{i,C}$ denotes the correlation between feature X_i and the class C, and is named *C-correlation*.

The Markov blanket based local search used in MBE-MOMA is briefly outlined in Fig. 2. For a given candidate solution s encoded in a chromosome, MBE-MOMA apply the local search to improve a randomly selected objective $f_i(s)$ based on the *Add* and *Del* operators and the improvement first strategy [30]. Let $\mathcal{X}$ and $\overline{\mathcal{X}}$ represent the sets of selected and excluded features encoded in s, respectively. The purpose of the *Add* operator is to select a highly correlated feature $\overline{\mathcal{X}}_i$ from $\overline{\mathcal{X}}$ to $\mathcal{X}$ based on the *C-correlation* measure. The *Del* operator on the other hand selects highly correlated features $\mathcal{X}_i$ from $\mathcal{X}$ and removes other features that are covered by $\mathcal{X}_i$ using the approximate Markov blanket. If there is no feature in the approximate Markov blanket of $\mathcal{X}_i$, the operator then tries to delete $\mathcal{X}_i$ itself. The maximum numbers of both *Add* and *Del* operations are limited to l. Improvement is achieved when the new solution s' displays a higher fitness accuracy on the selected objective than original solution s, i.e., $f_i(s') > f_i(s)$. The work operations of *Add* and *Del* operators are depicted in Fig. 3 and 4, respectively.

Note that the *C-correlation* measure [18] of each feature in both memetic operators need only be calculated once in the MOMA search. This feature ranking information is then archived for use in any subsequent *Add* and *Del* operations, for fine-tuning the MOEA solutions throughout the search. We further illustrate the details of the *Add* and *Del* operations in Figure 5. For

Markov Blanket Based Local Search Using Improvement First Strategy

BEGIN

1) **For** Each non-dominated solution s.
2) Randomly select an objective $f_i(s)(i \in (1, \ldots, k))$ to improve.
3) **For** $j = 1$ **to** l^2
4) Generate a unique random pair $\{\#_{add}, \#_{del}\}$ where $0 \leq \#_{add}, \#_{del} < l$.
5) Apply $\#_{add}$ times *Add* and $\#_{del}$ times *Del* on s to generate a new chromosome s'.
6) Calculate fitness of modified chromosome $F(s')$.
7) **IF** $f_i(s') > f_i(s)$
8) Break local search and return s'.
9) **End IF**
10) **End For**
11) **End For**

END

Fig. 2. Markov blanket based local search in MBE-MOMA

instance, $F5$ and $F4$ represent the highest and lowest ranked features in $\overline{\mathcal{X}}$, while $F3$ and $F6$ are the highest and lowest ranked features in $\mathcal{X}$, respectively. In the *Add* operation, $F5$ is thus the most likely feature to be moved to $\mathcal{X}$. Further, since $F3$ is the approximate Markov blanket for $F1$ and $F6$, through the *Del* operation, $F1$ and $F6$ will be deleted from $\mathcal{X}$ since they are already blanketed by $F3$. For instance, the two most probable resultant chromosomes after undergoing the *Add* and *Del* operations are depicted in Fig. 5.

F. Synergy Between FCR and PCR Feature Subsets

The output of the MBE-MOMA is a set of non-dominated solutions. For k OVA sets, there exist k extremal solutions where each represents the optimal solution of the respective objective. In other words, each extremal solution encodes an optimal PCR feature subset for the corresponding

***Add* Operator:**

BEGIN

1) Rank the features in $\overline{\mathcal{X}}$ in a descending order based on *C-correlation* [18] measure.
2) Select a feature $\overline{\mathcal{X}}_i$ in $\overline{\mathcal{X}}$ using linear ranking selection [40] such that the larger the *C-correlation* of a feature in $\overline{\mathcal{X}}$ is the more likely it is to be selected.
3) Add $\overline{\mathcal{X}}_i$ to $\mathcal{X}$.

END

Fig. 3. *Add* operation

***Del* Operator:**

BEGIN

1) Rank the features in $\mathcal{X}$ in a descending order based on *C-correlation* [18] measure.
2) Select a feature $\mathcal{X}_i$ in $\mathcal{X}$ using linear ranking selection [40] such that the larger the *C-correlation* of a feature in $\mathcal{X}$ is the more likely it is to be selected.
3) Eliminate all features in $\mathcal{X} - \{\mathcal{X}_i\}$ which are in the approximate Markov blanket [18] of $\mathcal{X}_i$. If no feature is eliminated, try removing $\mathcal{X}_i$ itself.

END

Fig. 4. *Del* operation

OVA set. On the other hand, one FCR feature subset is formed by intersection of all non-dominated solutions. The next key issue to consider is how to combine these $k+1$ feature subset systematically for the final prediction.

In this study, we consider two schemes (as shown in Fig. 6) for synergizing the $k+1$ feature subsets. The first ensemble scheme is widely used and has been demonstrated to give better performances than the single model counterparts in feature selection and classification [21], [22]. An effective approach to construct an ensemble model of accurate and diverse base classifiers

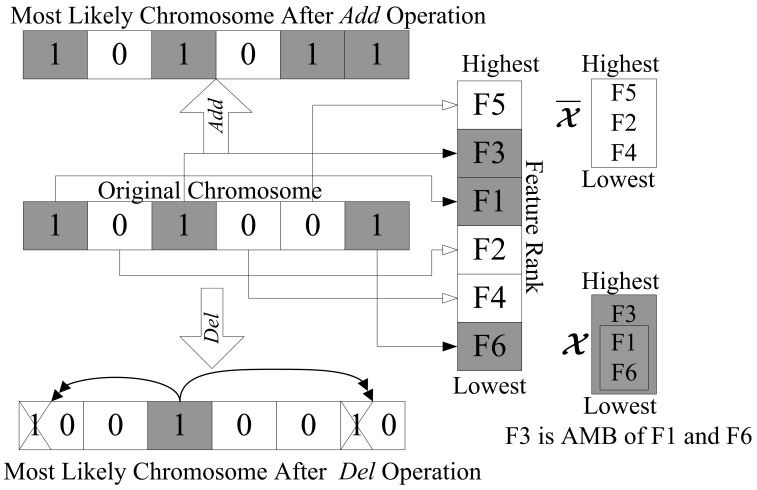


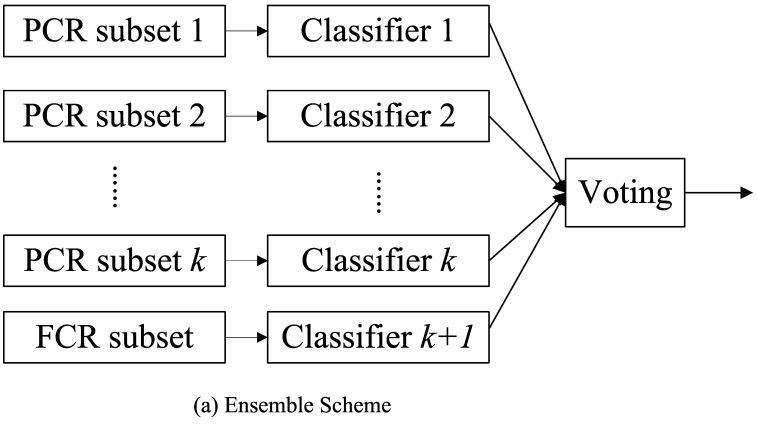
Fig. 5. Markov blanket based memetic operations (AMB denotes Approximate Markov Blanket)

is to use different feature subsets. In this study, the ensemble consists of multiple classifiers constructed based on the FCR and PCR feature subsets. In particular, each of the k+1 feature subsets is employed for classifying all classes, and the predictions of all trained classifiers are then aggregated based on voting. The second alternative considered is the conjunction scheme which is a union of all $k+1$ feature subsets as one feature subset. The newly formed union feature subset is then used in the classification accuracy prediction. It is worth noting that the number of selected features in MBE-MOMA is determined based on the number of unique features among all $k+1$ feature subsets.

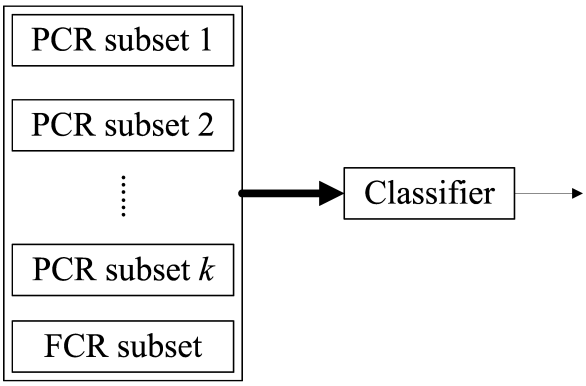
III. Empirical Study

In this section, we study the performance of the proposed method MBE-MOMA and compare with its single objective counterpart, i.e., the Markov blanket-embedded genetic algorithm (MBEGA) [28] and the non-local-search counterpart, i.e., the standard multiobjective evolutionary algorithm (MOEA) based on NSGAII [24]. The performance of the algorithms is then evaluated using both synthetic and real world microarray datasets. Further comparisons to existing state-of-the-art feature selection approaches for multiclass gene selection are also presented.

The parameters for MBEGA, MOEA, and MBE-MOMA are configured similarly with population size of 50, crossover probability of 0.6, and mutation rate of 0.1. The local search



(a) Ensemble Scheme



(b) Conjunction Scheme

Fig. 6. Synergy schemes of FCR and PCR features

range l in both MBEGA and MBE-MOMA is configured as in [28]. For fair comparison, the maximum computational budgets allowable for all these three algorithms are set to 2000 fitness functional calls. In most existing works on gene selection, prior knowledge is incorporated into the algorithm by constraining the upper limit of the gene size. Here, the maximum number of bit '1' or gene size allowable in each chromosome is constrained to 50 and 150 for synthetic and microarray datasets, respectively.

To evaluate the performances of the feature selection algorithms considered in this study, the average of 30 independent runs of external .632 bootstrap [41] are reported. A comparison of various error estimation methods on microarray classification [41] has suggested that .632 bootstrap is generally more appropriate than other estimators including re-substitution estimator, k-fold cross-validation, and leave-one-out estimation. For datasets with very few instances in

some classes, a stratified bootstrap sampling is used instead so that the distribution of each class in the sampled dataset is maintained consistent to the original dataset.

A. Synthetic Data

Here we consider first some synthetic multiclass datasets for studying the weaknesses and strengths of the feature selection algorithms as well as to illustrate the notion of FCR and PCR features. Three 3-class synthetic datasets (SD1, SD2, and SD3), with each class containing 25 samples, are generated based on the approach described in [42].

Each synthetic dataset consists of both relevant and irrelevant features. The relevant features in each dataset are generated from a multivariate normal distribution using the mean and covariance matrixes. And 4000 irrelevant features are added to each dataset. Among these 4000 features, 2000 are drawn from a normal distribution of N(0,1) and the other 2000 features are sampled with a uniform distribution of U[-1,1].

SD1 is designed to contain only 20 FCR and 4000 irrelevant features. The centroids of the three classes are located at (0,0), (3.7,1), and (1,3.7). Two groups of relevant genes are generated from a multivariate normal distribution, with 10 genes in each group. The variance of each relevant gene is 1. All these 20 relevant genes are generated with the following mean and covariance matrixes:

$$\mu = \begin{bmatrix} \mu_{\mathbf{0}} & \mu_{\mathbf{0}} \\ \mu_{\mathbf{3.7}} & \mu_{\mathbf{1}} \\ \mu_{\mathbf{1}} & \mu_{\mathbf{3.7}} \end{bmatrix} \quad \sigma = \begin{bmatrix} \mathbf{e} & \mathbf{0} \\ \mathbf{0} & \mathbf{e} \end{bmatrix}$$

where $\mu_{\mathbf{x}}$ is a 25×10 matrix with each element taking value x; $\mathbf{0}$ represents a 10×10 matrix with each element taking value 0; $\mathbf{e}$ is a 10×10 symmetrical matrix with each diagonal element having value of 1 while all other elements having value of 0.9, i.e,

$$\mathbf{e} = \begin{bmatrix} 1 & 0.9 & \dots & 0.9 \\ 0.9 & 1 & \dots & 0.9 \\ \vdots & \vdots & \ddots & \vdots \\ 0.9 & 0.9 & \dots & 1 \end{bmatrix}$$

The correlation between intra-group genes is 0.9, whereas the correlation between inter-group genes is 0. Genes in these two groups are FCR features, since they are relevant to any SVS sets of classes; and furthermore, genes in the same group are redundant with each other and the

optimal gene subset for distinguishing the three classes consists of any 2 relevant genes from different groups.

SD2 is designed to contain 10 FCR, 30 PCR, and 4000 irrelevant features. The centroids of the three classes are located at (0,1.18), (3.7,1.18), and (1,1.18) [1]. Four groups of relevant, i.e., FCR and PCR, genes (G0, G1, G2, and G3) are generated from a multivariate normal distribution, with 10 genes in each group. These 40 relevant genes are generated with the following mean and covariance matrixes:

$$\mu = \begin{bmatrix} \mu_{\mathbf{0}} & \mu_{\mathbf{1.18}} & \mu_{\mathbf{-1.18}} & \mu_{\mathbf{-1.18}} \\ \mu_{\mathbf{3.7}} & \mu_{\mathbf{-1.18}} & \mu_{\mathbf{1.18}} & \mu_{\mathbf{-1.18}} \\ \mu_{\mathbf{1}} & \mu_{\mathbf{-1.18}} & \mu_{\mathbf{-1.18}} & \mu_{\mathbf{1.18}} \end{bmatrix} \quad \sigma = \begin{bmatrix} \mathbf{e} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{e} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{e} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{e} \end{bmatrix}$$

Genes in the same group are redundant to each other. In this dataset only genes in G0 are FCR genes. While those genes in G1, G2, and G3 are PCR genes. Further, genes in G1 are drawn under different distributions for class 1 as compared to the other two classes, i.e., N(1.18,1) for class 1 and N(-1.18,1) for the remaining two classes, hence it is possible to distinguish class 1 from any other classes. However, since genes in G1 are drawn under the same distribution of N(-1.18,1) for samples in classes 2 and 3, it is difficult to distinguish between these two classes. The optimal gene subset to distinguish one class from others consists of one FCR gene from G0 and another PCR gene from the corresponding group. For instance, to correctly classify class 1, two genes each from G0 and G1 are required. The optimal gene subset to distinguish all the three classes thus consists of 4 genes, one FCR gene from G0 and three PCR genes each from G1, G2, and G3.

Last but not least, SD3 has been designed to contain only 60 PCR and 4000 irrelevant features. The centroids of the three classes are located at (-1.18,1.18), (1.18,-1.18), and (1.18,-1.18). Six groups of relevant genes (G0, G1, ..., G5) are generated from a multivariate normal distribution, with 10 genes in each group. These 60 relevant genes are generated with the following mean

[1]The two coordinates(features) of the centroids for each class may not be the same. For example, the centroid of class 1 is specified on G0 and G1, while the centroid of class 2 is specified on G0 and G2.

and covariance matrixes:

$$\mu = \begin{bmatrix} \mu_{-1.18} & \mu_{1.18} & \mu_{-1.18} & \mu_{1.18} & \mu_{-1.18} & \mu_{1.18} \\ \mu_{1.18} & \mu_{-1.18} & \mu_{1.18} & \mu_{-1.18} & \mu_{-1.18} & \mu_{1.18} \\ \mu_{1.18} & \mu_{-1.18} & \mu_{-1.18} & \mu_{1.18} & \mu_{1.18} & \mu_{-1.18} \end{bmatrix} \quad \sigma = \begin{bmatrix} \mathbf{e} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{e} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{e} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{e} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{e} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{e} \end{bmatrix}$$

Genes in the same group are designed to be redundant to each other. All the relevant features are PCR features and the optimal gene subset for distinguishing one class from the others consists of two PCR genes each from the corresponding group. For instance, two genes each from G0 and G1 form the optimal feature subset for separating class 1 and the other two classes. The optimal gene subset to distinguish all the three classes thus consists of 6 genes with one from each group.

1) Feature Selection Results on Synthetic Datasets: The MBEGA, MOEA, and MBE-MOMA algorithms are used to search on each of the three synthetic datasets (i.e., SD1, SD2, and SD3) and the list of corresponding selected features and classification accuracies obtained by all three algorithms are summarized in Tables I and II, respectively.

The results in Table I suggest that MBE-MOMA selects more features among the optimal subset than the other algorithms. On SD1, all three algorithms have successfully identified the 2 FCR features among the optimal subset. On the other two datasets (SD2 and SD3) which contain both FCR and PCR features, only MBE-MOMA manages to identify the majority of the FCR and PCR features belonging to the optimal set, while at the same time producing classification accuracy superior to both MBEGA and MOEA (see Table II). Further, the ensemble MBE-MOMA exhibits the best classification accuracies on all three datasets. In comparison, MBEGA fails to locate many of the important features belonging to the optimal subset. Note that since $k + 1$ feature subsets are used in MBE-MOMA, it is expected to select more features than MBEGA, where only a single optimal feature subset is considered. For the same computational budget, MOEA selects a larger number of features while arriving at lower classification accuracy on the datasets, since it lacks the ability to remove the irrelevant features.

2) FCR and PCR Features Obtained: In this section, we illustrate the FCR and PCR features that are discovered by the MBE-MOMA. The Pareto front obtained by MBE-MOMA on SD2

TABLE I

FEATURE SELECTED BY EACH ALGORITHM ON SYNTHETIC DATA

	Features	MBEGA	MOEA	MBE-MOMA
SD1	*(#)*	8.2	57.5	25.4
	OPT(2)	**2**	**2**	**2**
	Red	3.3	3.2	7.0
	Irr	2.9	52.3	16.4
SD2	*(#)*	6.6	61.3	19.3
	OPT(4)	3.3	3.8	**4**
	Red	2.4	3.1	7.2
	Irr	0.9	54.4	8.1
SD3	*(#)*	11.9	58.1	21.6
	OPT(6)	5.3	5.6	**5.9**
	Red	5.2	3.6	10.1
	Irr	1.4	48.9	4.6

(#): Number of Selected features; *OPT(x)*: Number of selected features with in the optimal subset, x indicates the optimal number of features; *Red*: Redundant Features; *Irr*: Irrelevant Features.

TABLE II

CLASSIFICATION ACCURACY BY EACH ALGORITHM ON SYNTHETIC DATA

	MBEGA	MOEA-E*	MOEA-C*	MBE-MOMA-E*	MBE-MOMA-C*
SD1	86.88	84.11	78.66	**90.83**	85.97
SD2	90.94	85.23	83.04	**94.55**	92.09
SD3	95.49	88.37	89.39	**98.26**	97.35

* -E and -C denote the algorithms using ensemble and conjunction scheme, respectively.

is plotted in Fig. 7. Each point in this front presents an approximate Pareto optimal solution, where at least one FCR feature can be observed, i.e., a feature belonging to group G0, selected in each approximate Pareto optimal solution.

The extremal solution in each objective dimension represents an optimal PCR feature subset of the corresponding OVA set. To further illustrate the selected PCR features, we re-plot the 3D Pareto front as pairs of objectives in Fig. 8. The fitness values of objective 1 (classification accuracy on $\{c_1, \overline{c}_1\}$) and objective 2 (classification accuracy on $\{c_2, \overline{c}_2\}$) for each approximate

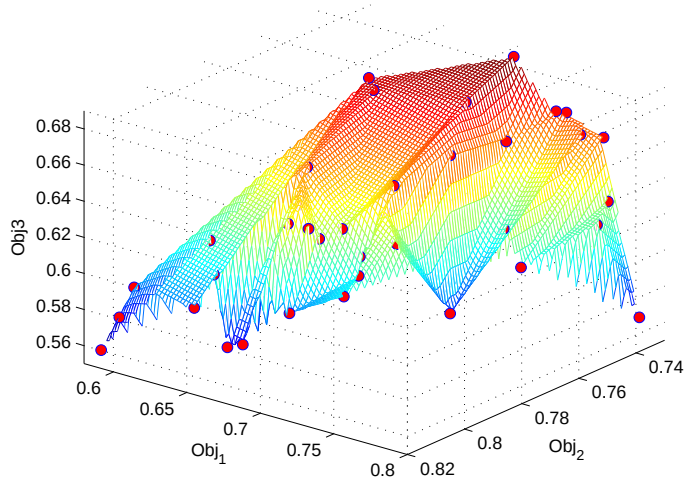


Fig. 7. The three-dimension Pareto front on the SD2

Pareto optimal solution are depicted in Fig. 8(a). In SD2, G1 and G2 are two groups of PCR genes relevant to objective 1 and 2, respectively. In Fig. 8(a), solutions consists of genes mostly from G2 are represented as "△", while solutions consists of genes mostly from G1 are denoted as "□". Solutions containing similar number of genes from G1 and G2 are marked with "○". The extremal "△" solution in dimension Obj_2 represents the optimal PCR feature subset for objective 2. Similarly, the extremal "□" solution in dimension Obj_1 represents the optimal PCR feature subset for objective 1. Fig. 8(b) and Fig. 8(c) illustrate the solutions with respect to 'objective 1 vs. objective 3' and 'objective 2 vs. objective 3', respectively.

B. Feature Selection Results on Microarray Data

Next, we extend our investigation on the efficiency of the proposed approach using 10 real world multiclass microarray datasets, which are briefly described in Table III.

1) Selected FCR and PCR Genes: One core strength of MBE-MOMA lies in the ability to identify both FCR and PCR genes simultaneously. In this subsection, we validate the FCR and PCR genes selected by MBE-MOMA. In particular, we focus on the ALL-AML-3 and Thyroid datasets as representatives datasets having high and low percentages of FCR genes, respectively.

ALL-AML-3 represents one of datasets that have been extensively studied by many re-

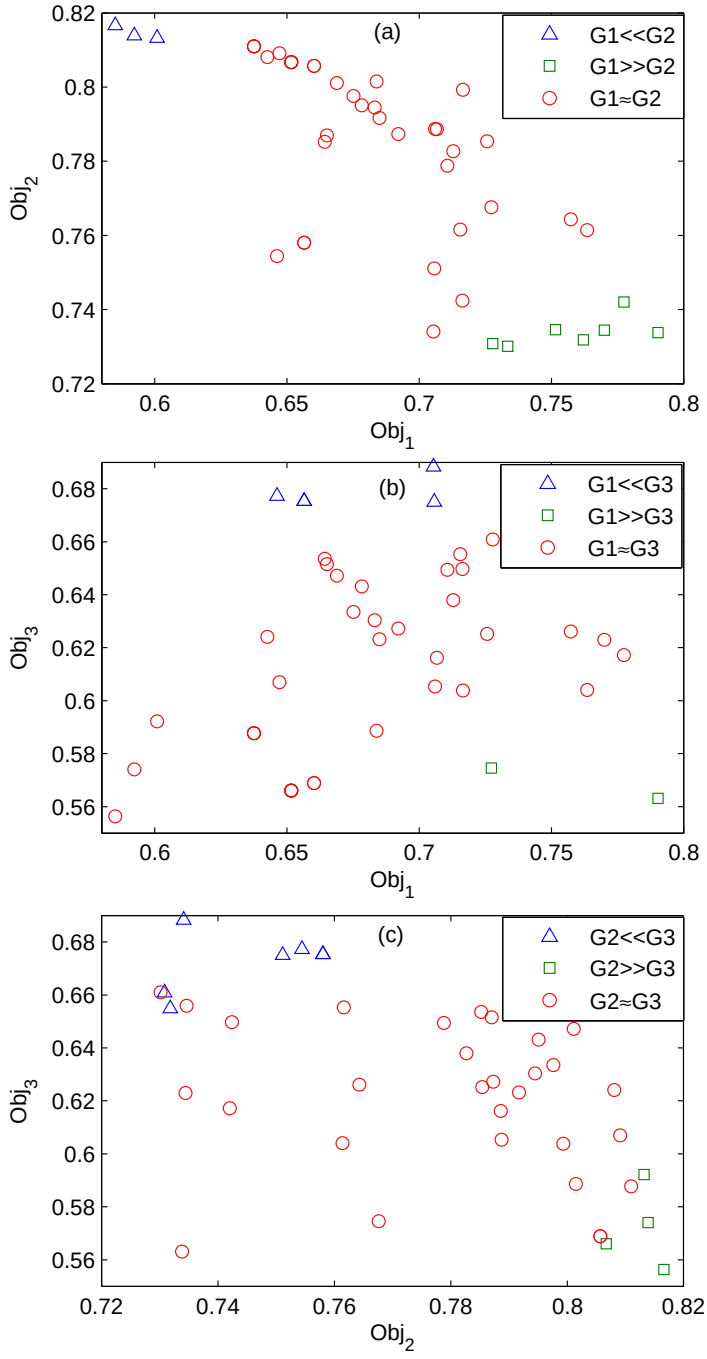


Fig. 8. Two-dimensional view of the Pareto front. (a) Objective 1 vs. Objective 2, (b) Objective 1 vs. Objective 3, (c) Objective 2 vs. Objective 3. $Gx \ll Gy$ $(x,y = 1, 2, 3)$ indicates more genes are selected from group y than x. $Gx \approx Gy$ $(x,y = 1, 2, 3)$ indicates similar number of genes are selected from group x and y,

TABLE III

DESCRIPTION OF 10 MULTICLASS MICROARRAY DATASETS

Dataset	Genes	Samples	Classes	Description and Reference
ALL-AML-3	7129	72	3	AML, ALL B-cell, and ALL T-Cell. [1][1]
ALL-AML-4	7129	72	4	AML-BM, AML-PB, ALL B-cell, and T-Cell. [1][2]
Lymphoma	4026	62	3	Three most prevalent adult lymphoid tumors. [43]
MLL	12582	72	3	AML, ALL, and mixed-lineage leukemia (MLL). [44]
SRBCT	2308	83	4	Small, round blue cell tumors of childhood. [45]
NCI60	6114	61	9	61 cell lines from 9 human cancer types. [8], [9], [46]
Brain	10367	50	4	4 malignant glioma types. [47]
Thyroid	2000	168	4	3 thyroid tumor types (FA, FC, and PC) and 1 normal tissues (N). [48][3]
ALL	12558	248	6	6 subtypes of ALL. [49]
Lung5c	12600	203	5	4 lung cancer types and 1 normal tissues. [50]

[1]AML: Acute Myelogenous Leukemia; ALL: Acute Lympboblastic Leukemia.

[2]BM: Bone Marrow samples; PB: Peripheral Blood samples.

[3]FA: Follicular adenoma; FC: Follicular carcinoma; PC: Papillary Carcinoma.

searchers. Hence, there is more information in the literature that may allow one to truly validate the selected genes. Table IV tabulates the 20 most selected FCR genes on ALL-AML-3 dataset. Note that 6 out of these 20 genes, i.e., U05259, X95735, M23197, M31523, M83652, and M84526, were identified as important biomarkers for distinguishing the classes AML and ALL, in [1]. X95739 and M84526 were also identified as distinguishing genes for AML and ALL by Zhou et al. [5]. Nevertheless, as MBE-MOMA considers a 3-class (AML, ALL T-cell, and ALL B-cell) multiobjective problem, as opposed to a two-class problem in [1], [5], the result obtained in the present study suggests that all 6 genes identified in [1] express differently in the subtypes of ALL, i.e., ALL T-cell and ALL B-cell. Besides these 6 genes, two other genes, X03934 and M27891, identified by MBE-MOMA are consistent to those also selected by Yeung et al. [13].

Besides FCR genes, we also identify the 10 most selected PCR genes for each of the 3 OVA sets on ALL-AML-3 dataset in Tables V to VII. Among the selected PCR genes, X77094 is found to be relevant to the first (ALL B-cell vs. others) and second (ALL T-Cell vs. others) OVA sets, suggesting it as a crucial biomarker for distinguishing ALL B-cell and ALL T-cell

TABLE IV

20 MOST SELECTED FCR GENES ON THE ALL-AML-3 DATASET

M84371_rna1_s	CD19 gene
U05259_rna1	MB-1 gene
D00749_s	T-cell antigen CD7 precursor
X00437_s	TCRB T-cell receptor, beta cluster
X95735	Zyxin
M23323_s	T-cell surface glycoprotein CD3 epsilon chain precursor
M37271_s	T-cell antigen CD7 precursor
U46499	Glutathione S-transferase, microsomal
U23852_s	GB DEF = T-lymphocyte specific protein tyrosine kinase p56lck (lck) abberant mRNA
M89957	IGB Immunoglobulin-associated beta (B29)
M23197	CD33 CD33 antigen (differentiation antigen)
M31523	TCF3 Transcription factor 3 (E2A immunoglobulin enhancer binding factors E12/E47)
D11327_s	PTPN7 Protein tyrosine phosphatase, non-receptor type 7
X03934	GB DEF = T-cell antigen receptor gene T3-delta
M83652_s	PFC Properdin P factor, complement
M84526	DF D component of complement (adipsin)
X58529	IGHM Immunoglobulin mu
L09209_s	APLP2 Amyloid beta (A4) precursor-like protein 2
X76223_s	GB DEF = MAL gene exon 4
M27891	CST3 Cystatin C (amyloid angiopathy and cerebral hemorrhage)

samples. Genes D88270 and L08895 are consistently found as relevant to the first (ALL B-cell vs. others) and third (AML vs. others) OVA sets, suggesting they represent important genes for studying AML and ALL B-cell cancer types. Indeed, they have been demonstrated to be potential biomarkers for discriminating subtypes of leukemia in [51], [52]. Last but not least, genes M31303 and M31211 listed in Table VII were also identified previously as biomarkers for distinguishing AML and ALL by Golub et al. [1].

Next we consider the Thyroid dataset, which is a relatively new dataset first published in 2006 [48]. It is hoped that this study on the Thyroid dataset will provide some useful prediction to assist biologists in the discovery of new biomarkers for thyroid cancer. Only 5 FCR (tabulated in Table VIII) are found by MBE-MOMA in this dataset. Ten most selected PCR genes for each of the 4 OVA sets are listed in Tables IX to XII. Among these selected PCR genes, LT55 is found

TABLE V

10 MOST SELECTED PCR GENES RELEVANT TO (ALL B-CELL VS. OTHERS) ON THE ALL-AML-3 DATASET

D88270	GB DEF = (lambda) DNA for immunoglobin light chain
L08895	MEF2C MADS box transcription enhancer factor 2, polypeptide C (myocyte enhancer factor 2C)
U31248	ZNF174 Zinc finger protein 174
M96326_rna1	Azurocidin gene
U09578	MAPKAP kinase (3pK) mRNA
X99920	S100 calcium-binding protein A13
X82240_rna1	TCL1 gene (T cell leukemia) extracted from H.sapiens mRNA for Tcell leukemia/lymphoma 1
M22960	PPGB Protective protein for beta-galactosidase (galactosialidosis)
X77094	P40phox
U51240	KIAA0085 gene, partial cds

TABLE VI

10 MOST SELECTED PCR GENES RELEVANT TO (ALL T-CELL VS. OTHERS) ON THE ALL-AML-3 DATASET

HG4128-HT4398	Anion Exchanger 3, Cardiac Isoform
X59871	TCF7 Transcription factor 7 (T-cell specific)
X00274	HLA class II histocompatibility antigen, DR alpha chain precursor
U14603	Protein tyrosine phosphatase PTPCAAX2 (hPTPCAAX2) mRNA
X62744	Class II histocompatibility antigen, M alpha chain precursor
U18548	GB DEF = GPR12 G protein coupled-receptor gene
U46006_s	GB DEF = Smooth muscle LIM protein (h-SmLIM) mRNA
Z35227	TTF mRNA for small G protein
X82240_rna1	TCL1 gene (T cell leukemia) extracted from H.sapiens mRNA for Tcell leukemia/lymphoma 1
X77094	P40phox

to be crucial for learning Follicular adenoma and Normal tissues, since it is always selected for distinguishing the OVA sets of (Follicular adenoma vs. Others) and (Normal vs. Others). For the same reason, gene GS2252 is noted to be crucial for studying Follicular carcinoma and Papillary Carcinoma tissues, since it is always chosen to be useful for distinguishing the OVA sets of (Follicular carcinoma vs. Others) and (Papillary Carcinoma vs. Others).

2) Classification Accuracy and Number of Selected Genes: Next, we also report the average classification accuracy and number of selected genes obtained by each feature selection algorithm

TABLE VII

10 MOST SELECTED PCR GENES RELEVANT TO (AML VS. OTHERS) ON THE ALL-AML-3 DATASET

D88270	GB DEF = (lambda) DNA for immunoglobin light chain
L08895	MEF2C MADS box transcription enhancer factor 2, polypeptide C (myocyte enhancer factor 2C)
M96326_rna1	Azurocidin gene
M63138	CTSD Cathepsin D (lysosomal aspartyl protease)
M31303_rna1	Oncoprotein 18 (Op18) gene
M92287	CCND3 Cyclin D3
M62762	ATP6C Vacuolar H+ ATPase proton channel subunit
M31211_s	MYL1 Myosin light chain (alkali)
D88422	CYSTATIN A
J05243	SPTAN1 Spectrin, alpha, non-erythrocytic 1 (alpha-fodrin)

TABLE VIII

5 SELECTED FCR GENES ON THE THYROID DATASET

LT36	keratin 19
GS421	Homo sapiens thymosin, beta 4, X chromosome (TMSB4X), mRNA.
GS8241	Homo sapiens protein phosphatase 1, regulatory (inhibitor) subunit15A (PPP1R15A), mRNA.
GS3295	Homo sapiens trefoil factor 3 (intestinal) (TFF3), mRNA.
GS3883	Homo sapiens cytochrome c oxidase subunit VIII (COX8), nuclear geneencoding mitochondrial protein, mRNA.

on the ten datasets over 30 runs of .632+ bootstraps in Table XIII. For each dataset, the percentage of FCR features against the total selected features made by MBE-MOMA are also tabulated.

In Table XIII, note that the classification accuracies obtained by all the algorithms are statistically not very significantly different on the first 5 microarray datasets considered. This suggests that the 5 datasets have higher percentages of FCR (except for the SRBCT), making them relatively easy problems such that most feature selection classification algorithms are capable of discriminating the classes accurately. On the latter 5 datasets, both MOEA and MBE-MOMA outperform MBEGA significantly, due to the low percentage of FCR features these datasets contain. Since the latter 5 datasets contain mainly PCR features, it becomes difficult for any single-objective approaches, which are incapable of identifying PCR features, including MBEGA

TABLE IX

10 MOST SELECTED PCR GENES RELEVANT TO (FOLLICULAR ADENOMA VS. OTHERS) ON THE THYROID DATASET

LT55	G0S3 mRNA
GS15425	Homo sapiens splicing factor, arginine/serine-rich 2 (SFRS2), mRNA.
GS2257	Homo sapiens thymosin, beta 10 (TMSB10), mRNA.
GS14741	Homo sapiens thyroid peroxidase (TPO), nuclear gene encodingmitochondrial protein, mRNA.
LT62	apolipoprotein D mRNA
GS4484	Homo sapiens KIAA0089 protein (KIAA0089), mRNA.
GS1213	Homo sapiens apoptosis inhibitor 5 (API5), mRNA.
GS13671	Homo sapiens metallothionein 1L (MT1L), mRNA.
LT38	adenosine A1 receptor (ADORA1) mRNA exons 1-6
GS2252	Homo sapiens serine (or cysteine) proteinase inhibitor, clade A(alpha-1 antiproteinase, antitrypsin), member 1(SERPINA1), mRNA.

TABLE X

10 MOST SELECTED PCR GENES RELEVANT TO (FOLLICULAR CARCINOMA VS. OTHERS) ON THE THYROID DATASET

GS2266	Homo sapiens transmembrane 4 superfamily member 1 (TM4SF1), mRNA.
GS2646	-
LT96	CBFB(core-binding factor, beta subunit, isoform 1)
GS6511	-
GS3909	-
GS6487	Homo sapiens fatty-acid-Coenzyme A ligase, long-chain 4 (FACL4),transcript variant 1, mRNA.
GS110	Homo sapiens ribosomal protein L21 (RPL21), mRNA.
GS7292	Homo sapiens activating transcription factor 3 (ATF3), mRNA.
GS14049	Homo sapiens paraneoplastic antigen MA1 (PNMA1), mRNA.
GS9988	Homo sapiens hypothetical protein FLJ21562 (FLJ21562), mRNA.

to locate the majority of important features. Further, it is worth noting that MBE-MOMA also arrives at competitive or improved classification accuracies than MOEA while selecting a smaller number of genes, due to the use of a memetic framework involving additional Markov blanket learning through local search. Overall, MBE-MOMA with the ensemble scheme yields the best average classification accuracy on the datasets.

3) Comparison to Other Feature Selection Studies in the Literature: Comparing the best results from this study with others published in the literature may not be most appropriate due

TABLE XI

10 MOST SELECTED PCR GENES RELEVANT TO (PAPILLARY CARCINOMA VS. OTHERS) ON THE THYROID DATASET

GS453	Homo sapiens cytochrome c oxidase subunit VIIc (COX7C), nucleargene encoding mitochondrial protein, mRNA.
LT57	thyroid peroxidase mRNA
GS4923	Homo sapiens tissue inhibitor of metalloproteinase 1 (erythroidpotentiating activity, collagenase inhibitor) (TIMP1),mRNA.
LT29	chitinase 3-like 1
GS2252	Homo sapiens serine (or cysteine) proteinase inhibitor, clade A(alpha-1 antiproteinase, antitrypsin), member 1(SERPINA1), mRNA.
GS3294	Homo sapiens short-chain dehydrogenase/reductase 1 (SDR1), mRNA.
LT156	stromelysin-3 precursor=MMP11
LT139	CTSH(cathepsin H isoform a preproprotein)
LT41	melanocyte-specific gene 1 (msg1)
GS6511	-

TABLE XII

10 MOST SELECTED PCR GENES RELEVANT TO (NORMAL VS. OTHERS) ON THE THYROID DATASET

LT55	G0S3 mRNA
GS1418	Homo sapiens cathepsin B (CTSB), transcript variant 5, mRNA.
LT41	melanocyte-specific gene 1 (msg1)
GS453	Homo sapiens cytochrome c oxidase subunit VIIc (COX7C), nucleargene encoding mitochondrial protein, mRNA.
GS4210	Homo sapiens Niemann-Pick disease, type C2 (NPC2), mRNA.
GS15881	Homo sapiens fibulin 2 (FBLN2), mRNA.
GS1329	Homo sapiens v-fos FBJ murine osteosarcoma viral oncogene homolog(FOS), mRNA.
GS2677	Homo sapiens NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 2,8kDa (NDUFB2), mRNA.
LT53	CL 100 mRNA for protein tyrosine phosphatase
LT62	apolipoprotein D mRNA

to the different data preprocessing, classifier, performance evaluation schema, experiment design etc. used in the different reported work. However, the reported results in the literature confirm that MBE-MOMA performs competitively or better than the previously published methods in terms of classification accuracy or the number of selected genes for the same datasets considered in Table XIV.

TABLE XIII

CLASSIFICATION ACCURACY AND NUMBER OF SELECTED GENES OF FEATURE SELECTION ALGORITHMS ON TEN MULTICLASS MICROARRAY DATASETS

	FCR%		MBEGA	MOEA-E	MOEA-C	MBE-MOMA-E	MBE-MOMA-C
ALL-AML-3	30.4	*Acc*	**96.64**	94.38	94.67	96.61	96.49
		#gene	18.1	164.0	164.0	30.7	30.7
ALL-AML-4	13.1	*Acc*	91.93	91.53	90.89	89.53	**93.02**
		#gene	26.2	247.7	247.7	49.1	49.1
Lymphoma	7.7	*Acc*	97.68	**98.74**	98.72	98.60	98.29
		#gene	34.3	190.6	190.6	81.6	81.6
MLL	25.0	*Acc*	94.33	95.14	93.82	**96.11**	95.45
		#gene	32.1	180.5	180.5	64.8	64.8
SRBCT	0.4	*Acc*	99.23	98.46	98.52	98.14	**99.43**
		#gene	60.7	274.4	274.4	110.1	110.1
NCI60	0	*Acc*	65.58	**73.56**	69.94	69.43	69.06
		#gene	24.3	488.9	488.9	255.3	255.3
Brain	1.5	*Acc*	72.53	79.93	77.65	**82.65**	77.75
		#gene	12.6	213.8	213.8	100.1	100.1
Thyroid	0.7	*Acc*	78.14	81.86	79.63	**81.87**	80.67
		#gene	51.7	263.9	263.9	105.4	105.4
ALL	0	*Acc*	95.25	94.76	94.95	97.24	**97.40**
		#gene	33.6	414.6	414.6	65.3	65.3
Lung5c	0.8	*Acc*	88.54	95.23	95.55	94.97	**95.83**
		#gene	55.6	259.6	259.6	54.2	54.2
		$\overline{Acc}$	87.99	90.36	89.43	**90.51**	90.34
		$\overline{\#gene}$	34.9	269.8	269.8	91.6	91.6

Acc: Classification accuracy; *#gene*: Number of selected genes; $\overline{Acc}$: Average of classification accuracy over all datasets; $\overline{\#gene}$: Average number of selected genes over all datasets.

Li et al. [14] reported their best investigation results on the ALL-AML-3, ALL-AML-4, MLL, ALL, Lymphoma, SRBCT, and NCI60 datasets using filter ranking feature selection algorithms,

TABLE XIV

COMPARISON OF BEST RESULTS BETWEEN MBE-MOMA AND OTHER FEATURE SELECTION STUDIES IN THE LITERATURE

	MBE-MOMA	Best results reported in the literature	
	Results	Results	Methods (Feature Selection+Classifier+Evaluation Scheme)
ALL-AML-3	96.61 (30.7)	∼95 (150)	Filter Ranking+SVM+4-fold CV, Li et al. [14]
		97.1 (60)	DDP+DAGSVM+F-splits, Ooi et al. [10]
ALL-AML-4	**93.02 (49.1)**	∼93 (150)	Filter Ranking+SVM+4-fold CV, Li et al. [14]
Lymphoma	98.60 (81.6)	**100 (150)**	Filter Ranking+SVM+4-fold CV, Li et al. [14]
MLL	96.11 (64.8)	**100 (150)**	Filter Ranking+SVM+4-fold CV, Li et al. [14]
		98.3 (60)	DDP+DAGSVM+F-splits, Ooi et al. [10]
SRBCT	**99.43 (110.1)**	∼95 (150)	Filter Ranking+SVM+4-fold CV, Li et al. [14]
		98.9 (80)	DDP+DAGSVM+F-splits, Ooi et al. [10]
NCI60	**69.43 (255.3)**	68 (150)	DDP+DAGSVM+F-splits, Ooi et al. [10]
NCI60_1000*	**74.47 (220.3)**	66.6 (150)	Filter Ranking+SVM+4-fold CV, Li et al. [14]
		65.79 (40)	GA+SVM+.632 bootstrap, Liu et al. [9]
Thyroid	**81.87 (105.4)**	79.8(-)	1A+Weighted-Voting+LOOCV, Yukinawa et al. [48]
ALL	97.40 (65.3)	**100 (150)**	Filter Ranking+SVM+4-fold CV, Li et al. [14]
Lung5c	**95.83 (54.2)**	94.1 (100)	DDP+DAGSVM+F-splits, Ooi et al. [10]

The values in the parentheses are the number of selected genes. *NCI60_1000 is NCI60 with 1000 pre-selected genes, which is preprocessed using the method described in [8].

SVM, and 4-fold cross validation[2]. The best classification accuracies obtained on the first 5 datasets are competitive with our results, however 150 genes were required to produce these competitive accuracies, which is significantly more than the average of 91.6 genes identified by MBE-MOMA. On the NCI60 dataset, MBE-MOMA obtains an average accuracy of 74.47%, which is also superior to the best accuracy of 66.66% reported in [14].

Ooi et al. [10] proposed a new feature selection method for multiclass gene selection based on the degree of differential prioritization (DDP) and F-splits evaluation procedure. On datasets, NCI60, Lung5c, SRBCT, MLL, and ALL-AML-3, their new method obtained best accuracies

[2]We would like to note that as no details on how the 4-fold cross validation was carried out in [14], i.e., whether repeated internal/external cross validation [53], [54] was used in producing the report results, a comparison to this early work may not be reliable. Nevertheless, we have included the comparison here for the sake of completeness to existing works where multi-classification on similar datasets has been studied.

of 68.0%, 93.8%, 98.9%, 97.9%, and 97.1%, respectively. In comparison, MBE-MOMA gives competitive accuracy on these five datasets, with higher accuracies on the first three datasets while fairing lower on the latter two.

Ooi et al. [8] and Liu et al. [9] proposed two GAs based methods, GA+maximum likelihood and GA+SVM, for multiclass gene selection, respectively. Both methods were applied to NCI60 dataset with 1000 pre-selected genes and the results are reported in [9]. GA+SVM was reported to outperform GA+maximum likelihood at a leave-one-out cross validation accuracy of 88.52% with 40 selected genes. Unfortunately, the results were obtained using an internal cross validation, which suffers from selection bias [53] or overfitting [54]. In [28], we have compared MBEGA with GA+SVM using the external .632 bootstrap. The single objective MBEGA was shown to attain a higher classification accuracy of 70.38% with 47.6 selected genes, while GA+SVM fares lower at 65.79% with 40 selected genes. Nevertheless, MBE-MOMA arrived at a significantly better accuracy of 74.47% with 220.3 genes on the same dataset.

On the Thyroid dataset, Yukinawa et al. [48] used a one class vs. any subset of other class scheme (1A) for constructing a multiclass predictor from binary classifier and reported the best accuracy of 79.8% using leave-one-out cross validation on the training data and 85.7% on the test data. However, due to the small number of instances, such a holdout approach is often regarded as unreliable in the machine learning community. MBE-MOMA on the other hand obtained an accuracy of 81.87% on the same dataset based on reliable .632 bootstrap.

IV. Conclusions

In this paper, we have defined two new types of relevant features, i.e., full class relevant (FCR) and partial class relevant (PCR), for multiclass problems, and also proposed a novel Markov blanket embedded multiobjective memetic algorithm (MBE-MOMA) for simultaneous identification of FCR and PCR features. Empirical study on 3 synthetic and 10 microarray datasets suggests that MBE-MOMA is efficient in identifying both FCR and PCR features and yields good classification accuracy. It performs better than the single objective counterpart, MBEGA, as well as standard MOEA without using local search. MBE-MOMA also produces superior or competitive performance over other gene selection methods published in the literature. Finally, the proposed MBE-MOMA can serve to enhance the process of candidate biomarkers discovery as well as assist researchers in analyzing the growing amounts of biological research

data, especially on multiclass problems.

V. Acknowledgment

This work has been funded in part under the A*STAR SERC Grant No. 052 015 0024 administered through the National Grid Office.

References

[1] T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, C. D. Bloomfield, and E. S. Lander, "Molecular classification of cancer: class discovery and class prediction by gene expression monitoring.," *Science*, vol. 286, no. 5439, pp. 531–537, 1999.

[2] D. V. Nguyen and D. M. Rocke, "Tumor classification by partial least squares using microarray gene expression data," *Bioinformatics*, vol. 18, no. 1, pp. 39–50, 2002.

[3] 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, no. 457, pp. 77–87, 2002.

[4] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine Learning*, vol. 46, no. 1-3, pp. 389–422, 2002.

[5] X. Zhou and K. Z. Mao, "Ls bound based gene selection for dna microarray data," *Bioinformatics*, vol. 21, no. 8, pp. 1559–1564, 2005.

[6] S. Ramaswamy, P. Tamayo, R. Rifkin, S. Mukherjee, C. Yeang, M. Angelo, C. Ladd, M. Reich, E. Latulippe, J. P. Mesirov, T. Poggio, W. Gerald, M. Loda, E. S. Lander, and T. R. Golub, "Multiclass cancer diagnosis using tumor gene expression signatures," *PNAS*, vol. 98, no. 26, pp. 15149–15154, 2001.

[7] D. V. Nguyen and D. M. Rocke, "Multi-class cancer classification via partial least squares with gene expression profiles," *Bioinformatics*, vol. 18, no. 9, pp. 1216–1226, 2002.

[8] C. H. Ooi and P. Tan, "Genetic algorithms applied to multi-class prediction for the analysis of gene expression data," *Bioinformatics*, vol. 19, no. 1, pp. 37–44, 2003.

[9] J. J. Liu, G. Cutler, W. Li, Z. Pan, S. Peng, T. Hoey, L. Chen, and X. B. Ling, "Multiclass cancer classification and biomarker discovery using ga-based algorithms," *Bioinformatics*, vol. 21, no. 11, pp. 2691–2697, 2005.

[10] C. H. Ooi, M. Chetty, and S. W. Teng, "Differential prioritization between relevance and redundancy in correlation-based feature selection techniques for multiclass gene expression data," *BMC Bioinformatics*, vol. 7, no. 320, 2006.

[11] R. Tibshirani, T. Hastie, B. Narasimhan, and G. Chu, "Diagnosis of multiple cancer types by shrunken centroids of gene expression," *PNAS*, vol. 99, no. 10, pp. 6567–6572, 2002.

[12] K. Y. Yeung and R. E. Bumgarner, "Multiclass classification of microarray data with repeated measurements: application to cancer," *Genome Biology*, vol. 4, no. 12, pp. R83, 2003.

[13] K. Y. Yeung, R. E. Bumgarner, and A. E. Raftery, "Bayesian model averaging: development of an improved multi-class, gene selection and classification tool for microarray data," *Bioinformatics*, vol. 21, no. 10, pp. 2394–2402, 2005.

[14] T. Li, C. Zhang, and M. Ogihara, "A comparative study of feature selection and multiclass classification methods for tissue classification based on gene expression.," *Bioinformatics*, vol. 20, no. 15, pp. 2429–2437, 2004.

[15] A. Statnikov, C. F. Aliferis, I. Tsamardinos, D. Hardin, and S. Levy, "A comprehensive evaluation of multicategory classification methods for microarray gene expression cancer diagnosis," *Bioinformatics*, vol. 21, no. 5, pp. 631–643, 2005.

[16] J. Pearl, *Probabilitic Reasoning in Intelligent Systems*, San Mateo: Morgan Kaufmann, 1988.

[17] D. Koller and M. Sahami, "Toward optimal feature selection," in *Proceedings of the 13th International Conference on Machine Learning*, 1996, pp. 284–292.

[18] L. Yu and H. Liu, "Efficient feature selection via analysis of relevance and redundancy," *Journal of Machine Learning Research*, vol. 5, pp. 1205–1224, 2004.

[19] I. Tsamardinos and C. F. Aliferis, "Towards principled feature selection: Relevance, filters, and wrappers," in *Proceedings of the 9th International Workshop on Artificial Intelligence and Statistics*, 2003.

[20] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, Reading, Mass., Addison-Wesley, 1989.

[21] L. S. Oliveira, M. Morita, and R. Sabourin, *Feature Selection for Ensembles Using the Multi-Objective Optimization Approach*, chapter 3,Yaochu Jin (editor), Multi-Objective Machine Learning. Springer, Berlin Heidelberg, 2006.

[22] A. Tsymbal, S. Puuronen, and D. W. Patterson, "Ensemble feature selection with the simple bayesian classification," *Information Fusion*, vol. 4, no. 2, pp. 87–100, 2003.

[23] R. Kohavi and G. H. John, "Wrapper for feature subset selection," *Artificial Intelligence*, vol. 97, no. 1-2, pp. 273–324, 1997.

[24] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: Nsga-ii," *IEEE Transaction on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.

[25] H. Ishibuchi and T. Murata, "A multi-objective genetic local search algorithm and its application to flowshop scheduling," *IEEE Trans. on Systems, Man, and Cybernetics -Part C: Applications and Reviews*, vol. 28, no. 3, pp. 392–403, 1998.

[26] H. Ishibuchi, T. Yoshida, and T. Murata, "Balance between genetic search and local search in memetic algorithm for multiobjective permutation flowshop scheduling," *IEEE Transaction on Evolutionary Computation*, vol. 7, no. 2, pp. 204–223, 2003.

[27] J. Handl, D. B. Kell, and J. Knowles, "Multiobjective optimization in bioinformatics and computational biology," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 4, no. 2, pp. 279–292, 2007.

[28] Z. Zhu, Y. S. Ong, and M. Dash, "Markov blanket-embedded genetic algorithm for gene selection," *Pattern Recognition*, vol. 40, no. 11, pp. 3236–3248, 2007.

[29] J. H. Holland, *Adaptation in natural artificial systems, 2nd edition*, MIT Press, 1992.

[30] Z. Zhu, Y. S. Ong, and M. Dash, "Wrapper-filter feature selection algorithm using a memetic framework," *IEEE Transactions On Systems, Man and Cybernetics - Part B*, vol. 37, no. 1, pp. 70–76, 2007.

[31] H. Ishibuchi and T. Nakashima, "Multi-objective pattern and feature selection by a genetic algorithm," in *Proceedings of Genetic and Evolutionary Computation Conference (Las Vegas, Nevada, U.S.A.)*, 2000, pp. 1069–1076.

[32] J. Liu and H. Iba, "Selecting informative genes using a multiobjective evolutionary algorithm," in *Proceedings 2002 Congress on Evolutionary Computation*, 2002.

[33] K. Deb and A. R. Reddy, "Reliable classification of two-class cancer data using evolutionary algorithms," *BioSystems*, vol. 72, pp. 111–129, 2003.

[34] M. Banerjee, S. Mitra, and A. Anand, *Feature Selection Using Rough Sets*, chapter 1,Yaochu Jin (editor), Multi-Objective Machine Learning. Springer, Berlin Heidelberg, 2006.

[35] Y. S. Ong, M. H. Lim, N. Zhu, and K. W. Wong, "Classification of adaptive memetic algorithms: A comparative study," *IEEE Transactions On Systems, Man and Cybernetics - Part B*, vol. 36, no. 1, pp. 141–152, 2006.

[36] Y. S. Ong and A. J. Keane, "Meta-lamarckian in memetic algorithm," *IEEE Transaction on Evolutionary Computation*, vol. 8, no. 2, pp. 99–110, 2004.

[37] A. Jaszkiewicz, "Do multiple-objective metaheuristics deliver on their promise? a computational experiment on the set-covering problem," *IEEE Transcation on Evolutionary Computation*, vol. 7, no. 2, pp. 133–143, 2003.

[38] J. D. Knowles and D. W. Corne, "M-paes: A memetic algorithm for multiobjective optimization," Proceedings of Congress of Evolutionary Computation CEC2000, 2000.

[39] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*, Cambridge University Press,Cambridge, 1998.

[40] J. E. Baker, "Adaptive selection methods for genetic algorithms," in *Proceedings of the 1st International Conference on Genetic Algorithms*, 1985, pp. 101–111.

[41] U. M. Braga-Neto and E. R. Dougherty, "Is cross-validation valid for small-sample microarray classification?," *Bioinformatics*, vol. 20, no. 3, pp. 374–380, 2004.

[42] R. Diaz-Uriarte and S. A. de Andres, "Gene selection and classification of microarray data using random forest," *BMC Bioinformatics*, vol. 7, no. 3, 2006.

[43] A. A. Alizadeh, M. B. Eisen, and E. E. Davis et al., "Distinct types of diffuse large b-cell lymphoma identified by gene expression profiling," *Nature*, vol. 403, no. 6769, pp. 503–511, 2000.

[44] S. A. Armstrong, J. E. Staunton, L. B. Silverman, R. Pieters, M. L. den Boer, M. D. Minden, S. E. Sallan, E. S. Lander, T. R. Golub, and S. J. Korsmeyer, "Mll translocations specify a distinct gene expression profile that distinguishes a unique leukemia," *Nature Genetics*, vol. 30, no. 1, pp. 41–47, 2002.

[45] 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.," *Nat Med*, vol. 7, no. 6, pp. 673–679, 2001.

[46] D. T. Ross, U. Scherf, and M. B. Eisen et al., "Systematic variation in gene expression patterns in human cancer cell lines," *Nature Genetics*, vol. 24, no. 3, pp. 208–209, 2000.

[47] C. L. Nutt, D. R. Mani, and R. A. Betensky et al., "Gene expression-based classification of malignant gliomas correlates better with survival than histological classification," *Cancer Research*, vol. 63, no. 7, pp. 1602–1607, 2003.

[48] N. Yukinawa, S. Oba, K. Kato, K. Taniguchi, K. Iwao-Koizumi, Y. Tamaki, S. Noguchi, and S. Ishii, "A multi-class predictor based on a probabilistic model: application to gene expression profiling-based diagnosis of thyroid tumors," *BMC Genomics*, vol. 7, no. 190, 2006.

[49] E. J. Yeoh, M. E. Ross, and S. A. Shurtleff et al., "Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling," *Cancer Cell*, vol. 1, no. 2, pp. 109–110, 2002.

[50] A. Bhattacherjee, W. G. Richards, and J. Stauton et al., "Classification of human lung carcinomas by mrna expression profiling reveals distinct adenocarcinoma subclasses," *PNAS*, vol. 98, no. 24, pp. 13790–13795, 2001.

[51] S. R. Bauer, H. Kubagawa, I. Maclennan, and F. Melchers, "Vpreb gene expression in hematopoietic malignancies: a lineage- and stage-restricted marker for b-cell precursor leukemias," *Blood*, vol. 78, no. 6, pp. 1581–1588, 1991.

[52] A. V. Krivtsov, D. Twomey, Z. Feng, M. C. Stubbs, Y. Wang, J. Faber, J. E. Levine, J. Wang, W. C. Hahn, D. G. Gilliland, and T. R. Golub adn S. A. Armstrong, "Transformation from committed progenitor to leukemia stem cell initiated by mll-af9," *Nature*, vol. 442, no. 7104, pp. 818–822, 2006.

[53] C. Ambroise and G. J. McLachlan, "Selection bias in gene extraction on the basis of microarray gene-expression data," *PNAS*, vol. 99, no. 10, pp. 6562–6566, 2002.

[54] J. Reunanen, "Overfitting in making comparisons between variable selection methods," *Journal of Machine Learning Research*, vol. 3, pp. 1371–1382, 2003.

PLACE PHOTO HERE

Zexuan Zhu received the BS degree in 2003 from the Department of Computer Science and Engineering, Fudan University, China, and the PhD degree in 2008 from the School of Computer Engineering, Nanyang Technological University, Singapore. His research interests lie in bioinformatics, evolutionary computation and machine learning.

Yew-Soon Ong is currently an Associate Professor with the School of Computer Engineering, Nanyang Technological University, Singapore, and he is Deputy Director of the Emerging Research Laboratory. He received his B.Eng and M.Eng degrees in Electrical and Electronics Engineering from Nanyang Technology University, Singapore. He then joined the Computational Engineering and Design Center, University of Southampton, UK, where he completed his Ph.D. degree. His current research interest lies in Natural Computing that spans across Memetic algorithms, Evolutionary algorithms, Robust and Complex Design Optimization, Informatics, and Grid computing. He currently serves as co-editor-in-chief of Memetic Computing Journal by Springer, Associate Editor of IEEE Transactions on Systems, Man and Cybernetics (Part B), International Journal of Systems Science and the Soft Computing Journal. He is also member of the Emergent Technologies Technical Committee (ETTC) and Evolutionary Computation Technical Committee (ECTC) of the IEEE Computational Intelligence Society and also Chair of the Task Force on Memetic Algorithms in IEEE ETTC. Previously, he has been guest editor of IEEE Transactions on Systems, Man and Cybernetics, Genetic Programming and Evolvable Machines Journal, Soft Computing Journal and has co-edited several volumes on Advances in Natural Computation, Evolutionary Computation in Dynamic and Uncertain Environments in the field of Computational Intelligence.

Jacek M. Zurada received his MS and PhD degrees (with distinction) in electrical engineering from the Technical University of Gdansk, Poland in 1968 and 1975, respectively. Since 1989 he has been a Professor, and since 1993 a distinguished Samuel T. Fife Professor with the Electrical and Computer Engineering Department at the University of Louisville, Kentucky. He was Department Chair from 2004 to 2006. Dr. Zurada was an Associate Editor of IEEE Transactions on Circuits and Systems, Part I and Part II, and served on the Editorial Board of the Proceedings of IEEE. From 1998 to 2003 he was the Editor-in-Chief of IEEE Transactions on Neural Networks. He is an Associate Editor of Neurocomputing, Schedae Informaticae, International Journal of Applied Mathematics and Computer Science, Advisory Editor of International Journal of Information Technology and Intelligent Computing, and Editor of Springer Natural Computing Book Series. He has served the profession and the IEEE in various elected capacities, including as President of IEEE Computational Intelligence Society in 2004-05. He has been member and Chair of various IEEE CIS and IEEE TAB committees, including the IEEE TAB Periodicals Committee and IEEE TAB Periodicals Review Committee. He was Chair of the Plenary Symposium on Critical and Emerging Issues in Computational Intelligence held at the IEEE WCCI in Orlando, Florida, and was the Founding Chair of the NNC Neural Networks Technical Committee. He is a Distinguished Speaker of IEEE CIS and IEEE Fellow.