An Evolutionary Algorithm Approach to Optimal Ensemble Classifiers for DNA Microarray Data Analysis

Kyung-Joong Kim, *Member, IEEE*, and Sung-Bae Cho, *Senior Member, IEEE*

***Abstract*—In general, the analysis of microarray data requires two steps: feature selection and classification. From a variety of feature selection methods and classifiers, it is difficult to find optimal ensembles composed of any feature-classifier pairs. This paper proposes a novel method based on the evolutionary algorithm (EA) to form sophisticated ensembles of features and classifiers that can be used to obtain high classification performance. In spite of the exponential number of possible ensembles of individual feature-classifier pairs, an EA can produce the best ensemble in a reasonable amount of time. The chromosome is encoded with real values to decide the weight for each feature-classifier pair in an ensemble. Experimental results with two well-known microarray datasets in terms of time and classification rate indicate that the proposed method produces ensembles that are superior to individual classifiers, as well as other ensembles optimized by random and greedy strategies.**

***Index Terms*—Classification, DNA microarray, ensemble, evolutionary algorithm (EA), feature selection, real-valued encoding.**

I. Introduction

DNA MICROARRAYS measure the expression levels of thousands of genes simultaneously [1]. This measurement process consists of either monitoring each gene multiple times or using a single time point in different states (for example, when dealing with diseases or types of tumors) [2]. It is important to identify functionally related genes or to classify samples by using informative genes. In this paper, we focus on the latter case: the classification of DNA microarray data.

In this kind of classification, classifiers receive input vectors from the feature selection step to make decisions. However, it is difficult to choose appropriate feature selection methods and classifiers because there are so many candidates. Cho and Won explored seven feature selection methods and four classifiers in three benchmark datasets to systematically evaluate the performance of both the feature selection methods, as well as the machine learning classifiers [3]. The researchers used to test different feature-classifier pairs, using various datasets, in order to construct robust classification systems. An artificial neural network with principal component analysis (PCA) feature reduction was used on the small, round blue-cell tumor (SRBCTs) dataset [4], and it also used a well-known feature selection method (signal-to-noise ratio) based on [5] with a support vector machine (SVM) on the ovarian tissue dataset [6].

The ensemble classifier enabled us to obtain a more reliable solution than when using a single feature-classifier alone [7]. However, because not all ensembles yielded good classification performance, it was necessary to find the optimal ensembles in order to classify the samples accurately. In the neural network domain, Zhou *et al.* reported that ensembles with many classifiers were better than those with all classifiers [8]. Therefore, forming an ensemble of all the feature-classifier pairs did not prove to be a good heuristic. A straightforward method of finding the optimal ensemble is to compare all the ensembles and simply select the best one. However, the possible number would be too large. Therefore, we used 42 feature-classifier pairs, which resulted in 2^{42} possible ensembles. It would have been almost impossible to enumerate all the ensembles with even the most powerful computer.

In this paper, we propose a sophisticated method based on the evolutionary algorithm (EA) approach. This is a robust search method that requires little information to search effectively in a large or poorly understood search space [9]. Kuncheva and Jain used the genetic algorithm (GA) approach to design the classifier fusion system which was tested on a nonbiological benchmark dataset. In their experiment, they fixed the number of classifiers and used the GA approach for generating disjoint feature subsets (no feature overlapping), finding the best type of classifiers, and determining the inputs of each classifier [10]. Also, there have been some studies that have used GA to select optimal genes for analyzing DNA microarray datasets [11], [12]. These studies fixed the classification algorithm and evolved a subset of the genes.

In our experiments, the number of members in the ensemble is not represented as a fixed number. Instead, we use randomly selected initial chromosomes to represent ensembles of various population sizes and parameters, and then search for the optimal ensemble by using genetic operations. Real-value encoding is proposed to improve the performance of binary encoding. The approach can be used to determine the degree of participation by each feature-classifier pair in an ensemble. The results are then systematically evaluated in terms of performance and time efficiency using two well-known benchmark datasets: the lymphoma cancer dataset and the colon cancer dataset.

Manuscript received March 21, 2006; revised October 9, 2006, January 22, 2007, and June 1, 2007. This work was supported by MIC, Korea, under ITRC IITA-2007-(C1090-0701-0046).

K. J. Kim was with the Department of Computer Science, Yonsei University, Seoul 120-749, Korea (e-mail: kjkim@cs.yonsei.ac.kr). He is now with the Department of Mechanical and Aerospace Engineering, Cornell University, Ithaca, NY 14853 USA.

S.-B. Cho is with the Department of Computer Science, Yonsei University, Seoul 120-749, Korea (e-mail: sbcho@cs.yonsei.ac.kr).

Digital Object Identifier 10.1109/TEVC.2007.906660

II. Related Works

A. DNA Microarrays

In general, DNA microarrays are composed of thousands of individual DNA sequences. Using a robotic arrayer, these sequences can be printed in the form of a high-density array on a glass microscope slide. Usually, the relative abundance of these spotted DNA sequences (in two samples: DNA and RNA) may be assessed by monitoring the differential hybridization of the two samples to the sequences on the array. For the mRNA samples, the two samples can be reverse-transcribed into cDNA and then labeled with mixed fluorescent dyes (red-fluorescent dye Cy5 and green-fluorescent dye Cy3). After the hybridization of these samples with the arrayed DNA probes, the slides can be imaged using a scanner that produces fluorescence measurements for each dye. The log ratio between the two intensities of each dye can be used as the gene expression data

$$\mathrm{gene\_expression} = \log_2 \frac{\mathrm{Int(Cy5)}}{\mathrm{Int(Cy3)}} \tag{1}$$

where Int(Cy5) and Int(Cy3) are the intensities of the red and green colors. Since many genes can be put on the DNA microarray, it is possible to investigate the genome-wide information in a short time.

B. The Ensemble Approach for DNA Microarrays

Many researchers have evaluated the performance of up-to-date classification methods in their microarray experiments, and they have provided guidelines for finding the most appropriate classification tools in various situations. Dudoit *et al.* compared the performance of different discrimination methods for the classification of tumors based on gene expression data [13]. They reported that simple classifiers such as linear discriminant analysis and the nearest neighbor method performed remarkably well compared with more sophisticated methods such as aggregated classification trees. However, they did not include certain popular classifiers such as neural networks or support vector machines.

Lee *et al.* extended this kind of work with more gene selection techniques (3 methods), more classification methods (21 methods), and more datasets (7 datasets) [14]. This study showed that more sophisticated classifiers produced better performance than conventional methods such as the K-nearest neighbor (KNN), diagonal LDA (DLDA), and diagonal quadratic discriminant analysis (DQDA). Also, the choice of gene selection method had a large effect on performance.

Cho and Ryu proposed a classification framework that was able to combine a pair of classifiers trained with mutually exclusive features [15]. They conducted a thorough quantitative comparison of 42 combinations of features (7 methods) and classifiers (6 classifiers) in three benchmark datasets [3]. They combined 3, 5, and 7 classifiers among the 42 classifiers using the majority voting, weighted voting, and Bayesian combination methods. Their results showed that the ensemble classifiers surpassed the basic classifiers.

Also, Tan and Gilbert observed that ensemble learning (bagged and boosted decision trees) often performed better than single decision trees when used with several datasets [16]. Dettling proposed a novel algorithm called BagBoosting, the combination of two ensemble schemes (bagging and boosting), which consistently lowered the misclassification error of plain boosting and bagging [17]. Dettling and Buhlmann proposed modifications and extensions of boosting classifiers for microarray gene expression data from several tissue or cancer types [18]. Valentini *et al.* showed that bagged ensembles of SVMs were more reliable and achieved equal or better classification accuracy against a single SVM [19]. Table I summarizes the ensemble-based methods for DNA microarray classification.

Most of the ensemble-based studies have attempted to find multiple subsets of feature and classifier pairs, but studies mainly been conducted on ensembles for gene selection. Li *et al.* constructed an ensemble of the feature subsets extracted from decision trees [20]. This method was used to select relevant genes based on the ensemble decision. Xu and Zhang proposed a method to select genes from the bootstraps of a training set [21]. Compared with the single classifier method, the selection of informative gene subsets for each member of the ensemble classifier group is not a trivial problem because it is possible that the gene subsets could be different. Although most previous methods used one feature selection method for the ensemble, we assign suitable different feature selection methods to each member classifier automatically.

C. Evolutionary Algorithms (EAs) for the Ensemble Approach

EAs have been used to develop a neural network ensemble from a trained pool of neural networks [32]. Stefano *et al.* used GAs to determine the threshold values for selecting participation in a majority vote. In general, for each combination of classifier and class, there is a threshold that determines the participation of the classifier [33]. GAs can be also used to adjust the weights of neural networks in weighted voting [34], [35]. Sirlantzis *et al.* used GAs to find the optimal combination of classifier ensembles and combination methods [36]. Liu *et al.* used negative correlation learning and evolutionary learning to address the issues of automatic determination of the number of individual neural networks in an ensemble [37]. EAs are used in many different areas of multiple classifier systems. To the best of our knowledge, EAs have not been applied to determine optimal ensembles of feature selection and classifier pairs. Also, most applications have been limited to handwritten digit recognition and benchmark problems.

III. Method

The proposed method of searching for the optimal ensemble with the EA approach is illustrated in Fig. 1. First, the gene expression data are normalized between 0 and 1 (by rescaling the minimum value as 0 and the maximum value as 1) for feature selection. The preprocessed data are then divided into training, validation and test data. The training dataset is used for training the classifiers and the validation dataset is used for finding the optimal ensemble. For gene selection, informative genes are selected by M (seven in our case) different feature selection methods, which are entered to N (six in our case) different classifiers. The individual feature-classifier pairs are then exploited to search for the best ensemble obtained by the EA with the validation dataset.

The main reason that we adopt only a filter approach is to increase the number of candidates for the number of ensembles and the possibility of synergism. There are $N \times M$ pos-

TABLE I
RELEVANT METHODS FOR CLASSIFICATION OF DNA MICROARRAY DATA USING THE ENSEMBLE METHOD. (CNS: CENTRAL NERVOUS SYSTEM, BSS/WSS: BETWEEN-GROUPS TO WITHIN-GROUPS SUMS OF SQUARES, PC: PEARSON CORRELATION, ED: EUCLIDEAN DISTANCE, IG: INFORMATION GAIN, SN: SIGNAL-TO-NOISE RATIO, SP: SPEARMAN, CC: COSINE COEFFICIENT, MI: MUTUAL INFORMATION, LDA: LINEAR DISCRIMINANT ANALYSIS, FLDA: FISHER'S LDA, FDA: FLEXIBLE DISCRIMINANT ANALYSIS, MDA: MIXTURE DISCRIMINANT ANALYSIS, DT: DECISION TREE, MLP: MULTI-LAYER PERCEPTRON, SASOM: STRUCTURE ADAPTIVE SOM, LOGISTIC: LOGISTIC REGRESSION, GPLS: GENERALIZED PARTIAL LEAST SQUARES, NN: NEURAL NETWORK, PAM: PREDICTIVE ANALYSIS OF MICROARRAY, PDA: PENALIZED DISCRIMINANT ANALYSIS)

Researcher	Feature selection	Classifier	Ensemble	Data
Dudoit *et al.* (2002) [13]	Ratio of BSS/WSS	KNN, LDA, DT	Bagging, Boosting	Lymphoma [22], Leukemia [5], NCI 60 [23]
Cho and Ryu (2002) [15]	PC, SP, ED, CC, IG, MI, SN	NN	Ensemble with mutually exclusive features	Leukemia [5]
Dettling and Buhlmann (2003) [18]	Wilcoxon's rank-based statistics	DT	Boosting	Leukemia [5], Colon [24], Estrogen and Nodal [25], Lymphoma [22], NCI 60 [23]
Cho and Won (2003) [3]	PC, SP, ED, CC, IG, MI, SN	NN, SASOM, SVM, KNN	Exhaustive search of all possible ensembles (fixed number of members)	Leukemia [5], Colon [24], Lymphoma [22]
Tan and Gilbert (2003) [16]	An entropy-based discretization method	DT	Bagging, Boosting	Leukemia [5], Breast cancer [26], CNS embryonal tumor [27], Colon [24], Lung [28], Prostate [29]
Lee *et al.* (2005) [14]	Ratio of BSS/WSS, Wilcoxon's rank-based statistics, Soft-thresholding	FLDA, DLDA, DQDA, LOGISTIC, GPLS, KNN, DT, NN, SVM, FDA, PDA, MDA, PAM	Bagging, Boosting, RandomForest, LogitBoost	Leukemia [5], Lymphoma [22], NCI 60 [23], Colon [24], Lung cancer [30], SRBCT [4], Yeast [31]
Dettling (2004) [17]	Wilcoxon's rank-based statistics	DT	BagBoost	Leukemia [5], Colon [24], Prostate [29], Lymphoma [22], SRBCT [4], CNS embryonal tumor [27]
Valentini et al. (2004) [19]	Correlation analysis	SVM	Bagging	Leukemia [5], Colon [24]
Li *et al.* (2004) [20]	Ensemble genes with forest	SVM	Bagging	Leukemia [5], Colon [24]

sible candidates in the filter approach. However, if we used the wrapper approach, performance may have improved individually but there would have been only N candidates. Though the filter-based approach generally shows relatively low performance compared with the wrapper-based approach in individual classifiers, there is more possibility of performance improvement in the ensemble.

When the best ensemble is found, performance is verified by using a separate test dataset. As mentioned in [38], due to the small number of samples available, leave-one-out cross validation (LOOCV) is conducted to measure the generalization performance of the proposed method. Although the LOOCV method shows a low level of bias in general, it can be highly variable. To reduce the level of variability, the LOOCV method is repeated ten times.

The ensemble method is then used to solve the problems of the single classifier. In general, an ensemble consists of a set of individually trained classifiers whose predictions are combined when classifying novel samples, for example

$$f_e = \Phi(f_1, f_2, \ldots, f_n) \tag{2}$$

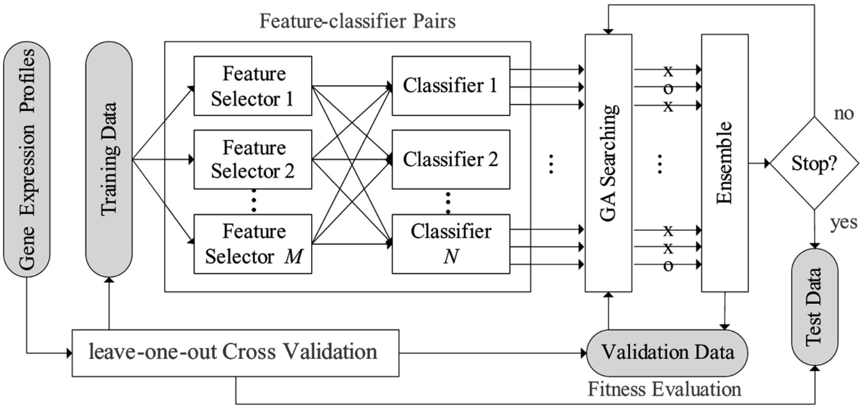

Fig. 1. Flowchart of the proposed method of searching for the optimal ensemble classifier.

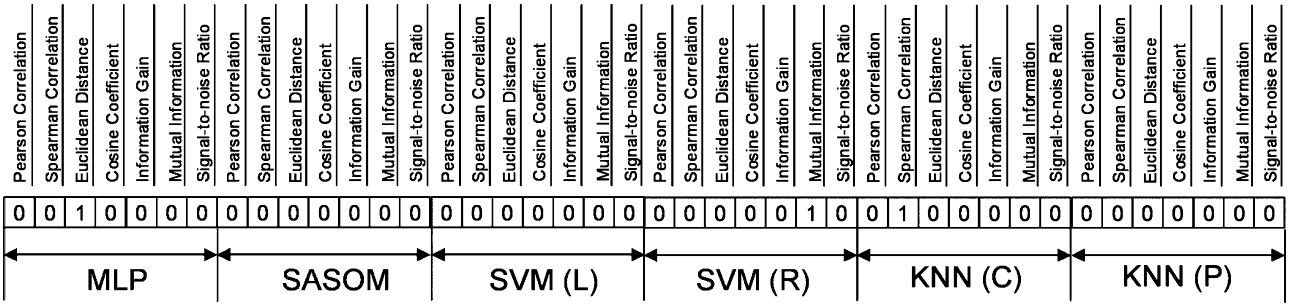

Fig. 2. Structure of a chromosome. Each bit indicates whether the corresponding feature-classifier is part of the ensemble or not. In this example, an ensemble is composed of three feature-classifier pairs: a MLP with Euclidean distance, a SVM (RBF) with mutual information, and a KNN (Cosine) with Spearman correlation.

where f_e is the output of an ensemble, f_i is the output of the ith feature-classifier, and Φ is an ensemble function. In this paper, Φ refers to a majority voting function.

It has been reported that the ensemble is generally more accurate than any of the individual classifiers belonging to the ensemble [39]. Both theoretical and empirical studies have demonstrated that a good ensemble can be obtained with the individual classifiers that make their errors on different parts of the input space [40]. If the individual feature-classifiers in the ensemble produce their errors mutually exclusively, they can complement each other, so the optimal ensemble can be produced. Therefore, it is important to find complementary feature-classifier sets. We use 42 feature-classifiers from 7 feature selection methods and 6 classifiers, yielding 2^{42} (about 4.3×10^{12}) combinations of the different ensemble classifiers. It may have been possible to find the optimal ensemble by enumerating and comparing all the combinations, but this would have taken too long. Since a more efficient method is required, this paper exploits the EA to solve the problem.

The GA can be applied effectively in order to solve combinatorial optimization problems. Each classifier is generally named after their feature selection method and the classification algorithm used. For example, b_{11} means that the base classifier uses the first feature selection method to reduce the dimensionality of data and the first classification algorithm to classify the reduced samples.

The structure of the chromosome used to find the optimal ensemble is shown in Fig. 2 (binary encoding). Each chromosome consists of a 42-bit string, each part of which indicates whether the corresponding feature-classifier pair is joined to the ensemble or not. Each bit corresponds to a specific feature-classifier pair; for example, the first bit corresponds to an MLP with Pearson correlation, the second bit to an MLP with Spearman correlation, and so on. Fig. 2 shows an ensemble of the third (an MLP with Euclidean distance), the 27th (a SVM (RBF) with mutual information), and the 30th (KNN (Cosine) with Spearman correlation) feature-classifier pairs. If any new feature-classifier pairs were to be added, the same process would have to be repeated after changing the chromosome structure with the number of individual feature-classifier pairs.

To determine possible performance differences, a real-valued GA was also used. In this representation, each bit represents the weight (real value, 0.0–1.0) for each feature-classifier pair in an ensemble, which is multiplied to the decision of the classifier for the final output. If the output of each feature-classifier pair is p_i (0 or 1 in binary classification) and the weight for each pair is w_i, the final decision is as follows:

$$\begin{aligned}
&\text{weighted\_sum} = \sum_{i=1}^{N\times M} p_i \times w, \quad \max = \sum_{i=1}^{N\times M} p_i \\
&\text{if weighted\_sum} > \frac{\max}{2} \text{ then class } 1 \\
&\text{if weighted\_sum} < \frac{\max}{2} \text{ then class } 0 \\
&\text{if weighted\_sum} = \frac{\max}{2} \text{ then reject.}
\end{aligned} \tag{3}$$

In binary encoding, the fitness of each given chromosome was evaluated by the accuracy of the corresponding ensemble and the number of feature-classifiers in the ensemble. For simplicity, we favored the ensemble with smaller number of feature-classifiers. Hence, the fitness of a chromosome was defined as follows:

$$C_{\text{fitness}} = \text{ENS}_{\text{accuracy}} - kN_p \tag{4}$$

where k is a constant, N_p is the number of feature-classifiers in the ensemble, and $\text{ENS}_{\text{accuracy}}$ is the accuracy of the corresponding ensemble

$$\text{ENS}_{\text{accuracy}} = \frac{\#\text{ of correlectly classified samples}}{\#\text{ of total validation samples}} \times 100.0 \tag{5}$$

In the real-valued GA, the fitness was defined using the weighted sum of each pair's output and the maximum value of the weighted sum

$$\begin{aligned} &C_{\text{fitness}} = \sum_{i=0}^{\text{V}} C_i \\ &\text{if true class is } 1, \text{then } C_i = \text{weighted\_sum} \\ &\text{if true class is } 0, \text{then } C_i = \text{max}-\text{weighted\_sum} \\ &\text{V} = \text{The number of total validation samples.} \end{aligned} \tag{6}$$

In binary encoding, first, chromosomes were selected at random. Each chromosome was evaluated with the fitness function as shown in the equation. In general, chromosomes go to a mating pool for genetic operations. Good chromosomes have a higher chance of being selected than poor ones. We used two different strategies for selection: the roulette wheel and rank-based strategies. In the roulette wheel strategy, the probability of being selected is in direct proportion to the fitness of each chromosome. In the rank-based strategy, the probability of being selected is directly proportional to each chromosome's order of fitness. In the mating pool, pairs of chromosomes were selected, and then the chromosome information for each part was exchanged by means of a crossover process, known as one-point crossover. Some bits were also mutated according to the mutation probability. The mutation operations consist of addition or deletion of the pair. Through these genetic operations, the chromosomes eventually evolve into optimal ensembles. Fig. 3 illustrates the operation of the genetic operators. After finishing this stage, the chromosomes were evaluated and the whole steps were repeated until a satisfactory solution was found.

In the real-valued GA, the crossover operation is similar to that of binary encoding with which the chromosome information was exchanged by the point of crossover selected at random. In mutation, the value selected was randomly changed with new one (0.0–1.0).

Yao *et al.* also used a GA to search for a near-optimal subset of classifiers for an ensemble [41]. In their GA, a chromosome (binary string) represents the inclusion of classifiers to the ensemble and its fitness is defined by recursive linear square (RLS) algorithm as the mean square error of linear combination of them. In our real-valued GA, a chromosome (real-value vector) represents the weights of feature-classifier pairs and a new fitness function for optimizing the weights is derived. Instead of calculating the distance between actual and predicted outputs, we consider the value ranged from the maximum weighted sum of ensemble (every classifier yields 1) and 0 (every classifier yields 0).

Although there are many sophisticated methods of combining multiple classifiers, we used the majority voting method to combine the feature-classifier pairs, because we wanted to focus our attention on verifying the proposed method of searching for the optimal ensemble. In future work, it would be interesting to incorporate the combining methods into the chromosome structure to be searched.

IV. Experimental Results

To evaluate the performance of the proposed method for classification of the expression patterns from the DNA microarray, we performed a series of experiments. To avoid overfitting, a large number of experiments were conducted with LOOCV which guaranteed fair judgment of the results. As mentioned previously, the data were divided into three parts: training data for training the classifiers, validation data for searching for the optimal ensemble using EA and test data for evaluating the proposed method. The process of training, validating, and testing was repeated n (the number of samples) times, and the prediction errors were averaged. Both the base classifier and the EA ensemble used training and validation data in the training phase and were tested on the single test sample with leave-one-out-cross-validation.

To compare the performance of the single classifiers and the ensemble classifier optimized by the EA, the single classifiers were trained using both the training and validation data and tested using one single test sample. On the other hand, the ensemble classifier optimized by EA used classifiers which were only trained with training data. The selection of the ensemble was based on the performance of the validation data. Finally, the performance of the ensemble was tested on a single test sample.

It was expected that the ensemble would exploit more genes than the single classifiers, and in that sense, the comparison with the single classifiers would not be fair. However, it is well-known in this field that more genes do not necessarily lead to performance improvement [3], [5]. It might be possible to use our approach to select the kind of informative genes that may help induce performance improvement. Comparison with the single classifiers was simply done to provide a general idea about performance improvement when using the ensemble approach, and we also compared the proposed method with other ensemble approaches.

As mentioned previously, we did not pour the best efforts to optimize the base classifiers. Some parameters were chosen according to the preliminary experimental results [3], so that some of these parameters may have been suboptimal. Since it is hard to obtain optimal single classifiers in general, we used an approximated solution that showed the best published results of single classifiers and bagging in the published literature for the colon cancer dataset. The parameters of the SVM were then determined by cross-validation methods.

A. Datasets

We used two well-known gene expression profiles: the lymphoma cancer dataset and colon cancer dataset. The diffuse large B cell lymphoma (DLBCL) group was a heterogeneous group of tumors, based on significant variations in morphology, clinical

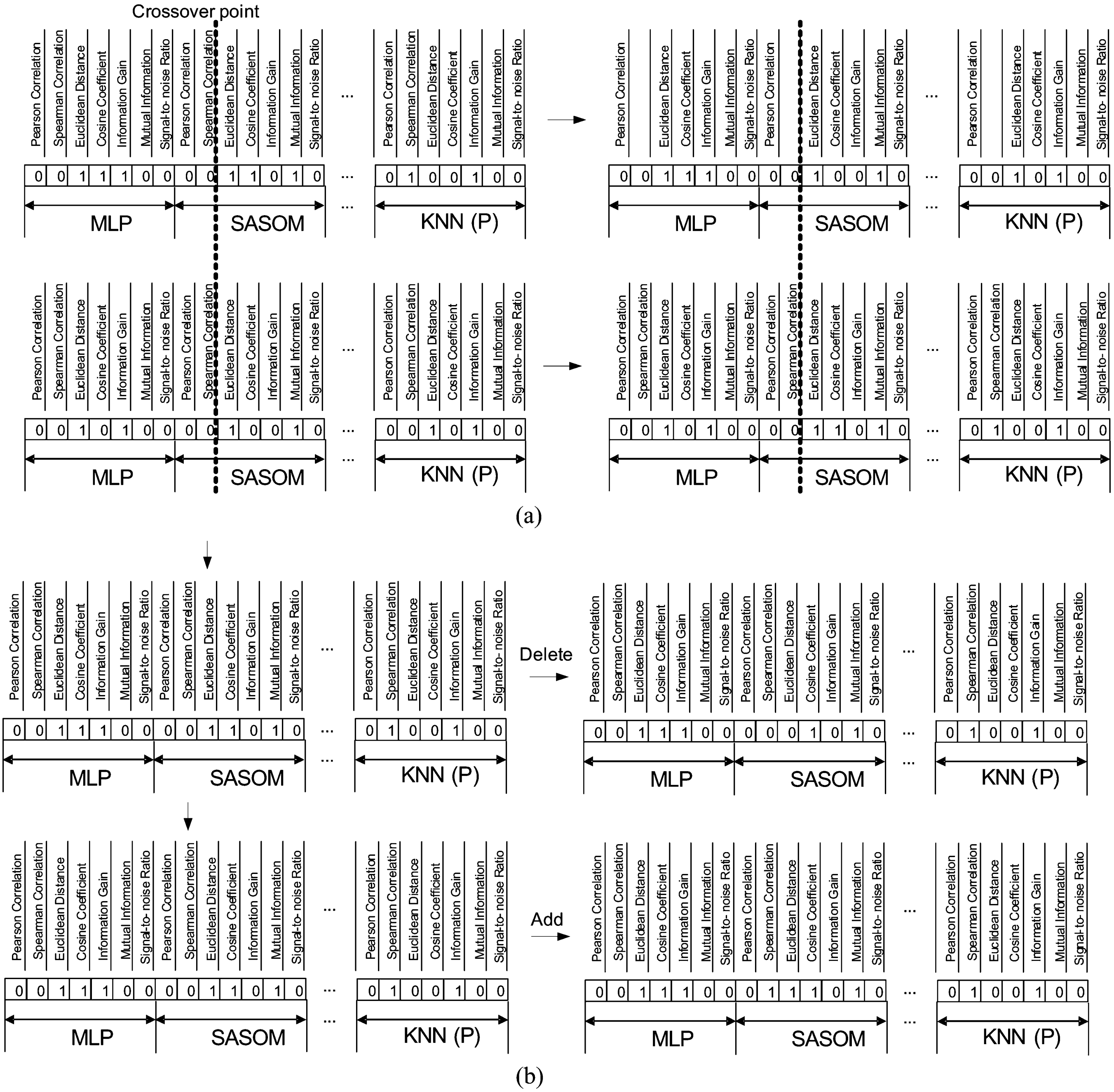

Fig. 3. Genetic operations in binary encoding. (a) Crossover. (b) Mutation (deletion and addition).

presentation, and response to treatment. Gene expression profiling has revealed two distinct tumor subtypes of the DLBCL group: germinal center B cell-like and activated B cell-like [22]. This lymphoma cancer dataset consisted of 24 GC B-like samples and 23 activated B-like samples (http://genome-www.stanford.edu/lymphoma) with 4026 genes.

The colon cancer dataset consisted of 62 samples of colon epithelial cells taken from colon cancer patients [24]. Each sample was taken both from regions with tumors and from normal healthy parts of the colons of the same patients and measured using high-density oligonucleotide arrays. Each sample contained 2000 gene expression levels. Although the original dataset consisted of 6000 gene expression levels, 4000 out of 6000 levels were removed due to insufficient confidence in the measured expression levels. Forty of the 62 samples came from regions with tumors and the remaining samples came from normal regions (http://microarray.princeton.edu/oncology/affydata/index.html).

We then separated the lymphoma cancer dataset into 23 training samples, 23 validation samples, and 1 test sample for LOOCV. Similarly, we divided the colon cancer dataset into 31 training samples, 30 validation samples, and 1 test sample. The separation of samples was done randomly.

For gene selection, the genes were ranked according to their feature scores that were calculated using the feature selection functions (correlation analysis, distance measure, entropy-based measure, and signal-to-noise ratio). Twenty-five high score genes were selected for classification, since a pre-

TABLE II
THE GENES OVERLAPPED BY PEARSON'S CORRELATION COEFFICIENT, THE COSINE COEFFICIENT, AND THE EUCLIDEAN DISTANCE

Colon	HEAT SHOCK PROTEIN HSP 90-BETA (HUMAN)
	SPLICING FACTOR SC35 (Homo sapiens)
	Human tyrosine kinase (HTK) mRNA, complete cds.
	Human serine kinase mRNA, complete cds.
	26S PROTEASE REGULATORY SUBUNIT 7 (Homo sapiens)
	Homo sapiens PP2A B56-gamma1 mRNA, 3' end of cds.
	Human aspartyl-tRNA synthetase alpha-2 subunit mRNA, complete cds.
	COLLAGEN ALPHA 2(XI) CHAIN (Homo sapiens)
Lymphoma	SLAP=src-like adapter protein; Clone=815774
	(Unknown UG Hs.169081 ets variant gene 6 (TEL oncogene); Clone=1355435)
	PTP-1B=phosphotyrosyl-protein phosphatase; Clone=472182
	PTP-1B=phosphotyrosyl-protein phosphatase; Clone=685177
	BCL-2; Clone=342181
	(Unknown UG Hs.59368 ESTs; Clone=1353778)
	Unknown; Clone=1372162
	(CXCR5=BLR1=B-cell homing chemokine receptor=L1; Clone=31)
	(CXCR5=BLR1=B-cell homing chemokine receptor=L1; Clone=31)
	PRK=putative serine/threonine protein kinase; Clone=739192
	Cyclin D2/KIAK0002=3' end of KIAK0002 cDNA; Clone=366412
	MCL1=myeloid cell differentiation protein; Clone=711870
	Potassium voltage-gated channel, shaker-related subfamily, member 3; Clone=1337856
	T-cell protein-tyrosine phosphatase=Protein tyrosine phosphatase, non-receptor type 2; Clone=665903
	core binding factor alpha1b subunit=CBF alpha1=PEBP2aA1 transcription factor =AML1 Proto-oncogene=translocated in acute myeloid leukemia; Clone=263251
	(DEC1=basic helix-loop-helix protein; Clone=469297)
	MAPKAP kinase (3pK); Clone=342647
	(Unknown UG Hs.209764 EST, Weakly similar to [H.sapiens]; Clone=1241842)

liminary study yielded the optimal number of genes as 25–30 [3]. The feature selection methods were performed on the combined set of training data only.

B. Experimental Design

We designed four separate experiments: Performance of individual classifiers, performance of the EA as an ensemble optimization method, comparison of ensemble optimization and individual classifiers, and generalization performance of both the ensembles and the individual classifiers.

In general, the "greedy ensemble" process sequentially adds the feature-classifier pair that maximizes the accuracy of the ensemble. The procedure is as follows.

1) Create an empty ensemble.
2) Add the best feature-classifier with the highest accuracy on the validation data.
3) Score all possible single changes (using the addition of one feature-classifier pair) of the ensemble. The score is based on the performance of the ensemble on the validation data.
4) If there is a performance increase, add one with the best increase to the ensemble and go back to step 3). If not, terminate the process.

In a random search, a number of random ensembles were evaluated with the validation data and the best one was selected. The degree of error was assessed with the test data. The number of random ensembles refers to the multiplication of the population size and maximum generation of the EA. We conducted statistical significance tests (student's t-tests) for all the results. N is the number of experiments. The degrees of freedom are defined as $(N + N - 2)$. If the t-value is larger than the value in t-table, it is statistically significant. μ_A and σ_A are the mean and standard deviation values of group A. The t-value between A and B is as follows:

$$t = \frac{|\mu_A - \mu_B|}{\sqrt{\frac{\sigma_A^2}{N} + \frac{\sigma_B^2}{N}}}. \tag{7}$$

For classification, we used a two-layered MLP with eight hidden nodes, two output nodes, a learning rate of 0.01–0.50, a momentum of 0.3–0.9, and 500 maximum iterations. The back-propagation algorithm stopped training when it reached 98% training accuracy, which should be reconsidered with respect to overtraining. In the case of kNN, we set k as 6, and used the Pearson correlation coefficients [KNN(P)] or the cosine coefficients [KNN(C)] for the similarity measure. We used SVMs

TABLE III
AVERAGE ACCURACY (%) OF THE INDIVIDUAL CLASSIFIERS FOR THE COLON DATASET (PC: PEARSON CORRELATION, SC: SPEARMAN CORRELATION, ED: EUCLIDEAN DISTANCE, CC: COSINE COEFFICIENT, IG: INFORMATION GAIN, MI: MUTUAL INFORMATION, SN: SIGNAL-TO-NOISE RATIO). TRAINED ON THE COMBINED TRAINING AND VALIDATION SETS, AND TESTED ON THE SINGLE TEST SAMPLE WITH LEAVE-ONE-OUT-CROSS-VALIDATION

	MLP	KNN(C)	KNN(P)	SASOM	SVM(L)	SVM(R)	Avg
CC	66.93±1.08	72.58±0.0	74.41±0.0	74.83±2.41	64.51±0.0	64.51±0.0	69.62±0.58
ED	70.80±0.86	79.03±0.0	75.80±0.0	70.64±2.95	64.51±0.0	64.51±0.0	70.88±0.63
IG	67.74±1.90	66.12±0.0	75.80±0.0	72.09±4.56	72.58±0.0	70.96±0.0	70.88±1.07
MI	66.93±1.65	66.12±0.0	75.80±0.0	70.48±3.75	72.58±0.0	70.96±0.0	70.47±0.9
PC	71.77±0.80	74.19±0.0	72.58±0.0	73.06±3.30	64.51±0.0	64.51±0.0	70.10±0.68
SN	74.19±0.0	74.19±0.0	72.58±0.0	60.32±3.83	64.51±0.0	64.51±0.0	68.38±0.63
SC	**79.19±0.48**	75.80±0.0	74.19±0.0	72.25±2.25	64.51±0.0	64.51±0.0	71.74±0.45
Avg	71.08±0.97	72.58±0.0	74.88±0.0	70.52±3.29	66.82±0.0	66.35±0.0	70.37±0.71

TABLE IV
AVERAGE ACCURACY (%) OF THE INDIVIDUAL CLASSIFIERS FOR THE LYMPHOMA DATASET. TRAINED ON THE COMBINED TRAINING AND VALIDATION SETS, AND TESTED ON THE SINGLE TEST SAMPLE WITH LEAVE-ONE-OUT-CROSS-VALIDATION

	MLP	KNN(C)	KNN(P)	SASOM	SVM(L)	SVM(R)	Avg
CC	75.53±1.06	72.34±0.0	78.72±0.0	65.74±2.41	65.95±0.0	70.21±0.0	70.91±0.57
ED	68.93±1.95	68.08±0.0	68.08±0.0	60.0±2.82	61.70±0.0	61.70±0.0	64.74±0.79
IG	89.36±2.34	**90.85±0.0**	**90.85±0.0**	87.23±1.59	**90.85±0.0**	**90.85±0.0**	89.99±0.65
MI	84.04±1.06	78.72±0.0	74.46±0.0	68.08±4.46	72.34±0.0	70.21±0.0	74.64±0.92
PC	62.12±0.85	78.72±0.0	76.59±0.0	61.06±4.15	68.08±0.0	65.95±0.0	68.75±0.83
SN	70.63±1.59	74.46±0.0	76.59±0.0	62.12±3.90	70.21±0.0	61.70±0.0	69.28±0.91
SC	63.82±0.0	70.21±0.0	61.70±0.0	57.44±2.85	63.82±0.0	61.70±0.0	63.11±0.47
Avg	73.70±1.26	76.20±0.0	75.28±0.0	66.47±3.17	70.42±0.0	68.90±0.0	71.83±0.73

with linear [SVL(L)] or RBF [SVM(R)] kernel functions. The penalty factor used over the slack variables was 10 (the default value of the SVM software). In the SASOM, we used an initial rectangular 4×4 map.

When searching for the optimal ensemble using the EA, we used the roulette wheel and rank-based methods as the selection operators. In rank-based selection, we assigned a higher rank to the chromosomes whose number of 1 s was smaller than the others for the tie-break. In the real-valued GA, roulette wheel selection is used. Since the preliminary results showed some convergence to a local minimum when the GA was run with fewer than 100 chromosomes, we set the population size as 100. Experiments were conducted with different crossover rates of 0.3, 0.5, 0.7, and 0.9, and mutation rates of 0.01 and 0.05. The GA stopped when it found the perfect ensemble (100% accuracy) on the validation dataset, or when the generation exceeded 10 000. The constant k level in the fitness function was 0.01. Because the constant k was used to provide a small advantage to that with a simpler structure, it was determined to be a relatively small value. If it was too large, we might have lost the one with the highest accuracy.

C. Analysis of Results

Table II shows the IDs of some common genes that were overlapped by Pearson's correlation coefficient, the cosine coefficient, and the Euclidean distance for the two datasets. Among them, there were some genes which were also overlapped by several feature selection methods. The number of overlapped genes in the lymphoma dataset was 18. The number of overlapped genes in the colon dataset was 9. These overlapped genes proved to be very informative.

Table III shows the average accuracy (the percentage of correct classification) of the individual classifiers in the colon dataset. Table IV shows the average accuracy (the percentage of correct classification) of the individual classifiers in the lymphoma dataset. These results reflect an average of ten runs. Since the LOOCV group required high computational costs, the number of runs was limited to ten.

In the colon dataset, the EA ensemble with real-value encoding (88.87 ± 1.41) showed better performance than the GA with binary encoding (84.19 ± 8.37) and they performed better than the greedy ensemble [81.00 $\pm$ 7.19, the difference with real-valued GA is significant $(t = 3.4 > 2.8784, p = 0.01, \text{degrees of freedom} = 18)$], random ensemble (70.80 ± 3.01), and the best individual classifier (79.19 ± 0.48) methods. "Random ensemble" means the best among 1 000 000 ensembles randomly chosen. In the lymphoma dataset, the GA with real-value encoding (97.78 ± 0.0) showed better performance than that with binary encoding (91.60 ± 6.79) and they also performed better than the greedy ensemble [90.26 $\pm$ 7.46, the difference with real-valued GA is significant $(t = 3.19 > 2.8784, p = 0.01, \text{degrees of freedom} = 18)$], random ensemble (80.21 ± 4.76), and the best individual classifier (90.85 ± 0.0) methods.

In general, bagging is a method of reducing variance, not bias. It is different from both greedy selection and EA which specifically aim to reduce bias. In our experiments, this is why both greedy selection and bagging were needed for

TABLE V
COMPARISON WITH OTHER STUDIES (CV =CROSS VALIDATION). (a) COLON, (b) LYMPHOMA

(a)

	Accuracy	Classifier	Feature Selection	Data Separation
Lee *et al.* (2005) [14]	88	SVM-RAD	Soft-Thresholding	2:1 CV
	86	LBoost		
	85	Boost		
	84	Bagging		
Dettling (2004)[17]	88.1	PAM	wilcoxon test statistic	2:1 Random partitions
	83.9	Bagboosting		
	80.86	Boosting		
Dettling and Buhlmann (2003) [18]	85.48	LogitBoost	wilcoxon test statistic	LOOCV
	82.26	1-NN		
	80.65	Classification Tree		
Proposed Method	88.87	6 classification algorithms	7 feature selection methods	LOOCV

(b)

	Accuracy	Classifier	Feature Selection	Data Separation
Li *et al.* [11]	84.6	KNN	GA	5:2 Random partitions
Valentini [44]	91.3	Linear Programming	Heuristic	10-fold CV
Kim *et al.* [45]	92.0	MLP	Correlation analysis	1:1 Random partitions
Cho *et al.* [46]	93.0	MLP	Multiple significant gene subsets	10-fold CV
Hong *et al.* [47]	97.6	Genetic Programming	Pearson correlation	5-fold CV
Proposed Method	97.78	6 classification algorithms	7 feature selection methods	LOOCV

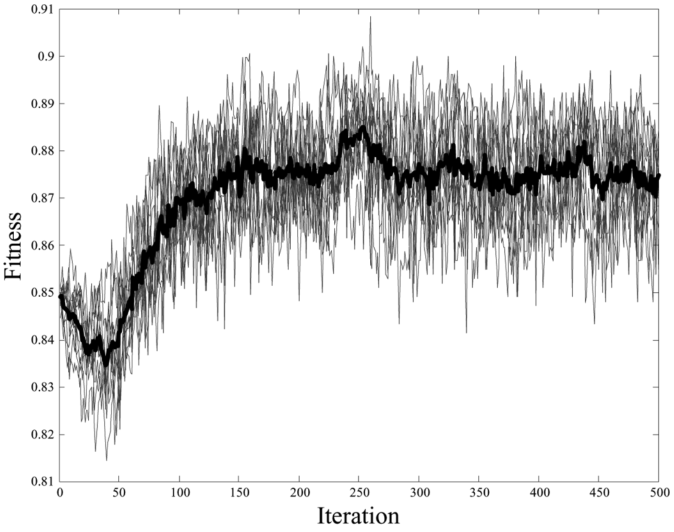

Fig. 4. Change of fitness with respect to the iterations (thick line shows an average).

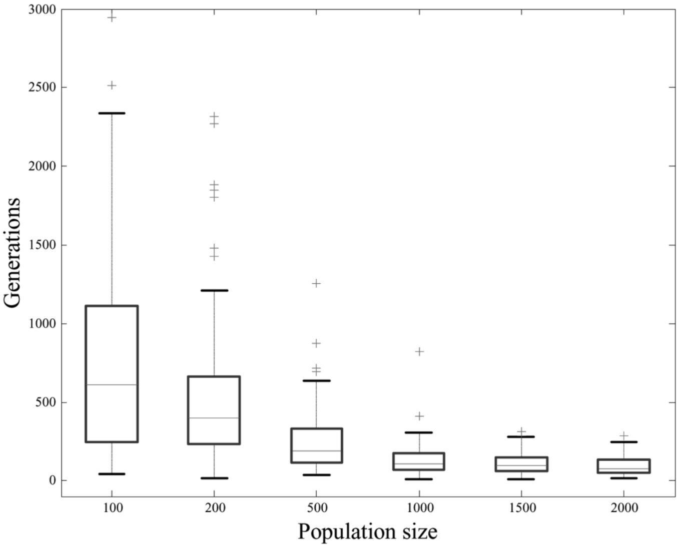

Fig. 5. Comparison of generations that found the optimal ensembles with respect to the population sizes for the lymphoma cancer dataset.

comparison purposes. Bagging showed $78.10 \pm 4.20(t = 7.69 > 2.8784, p = 0.01, \text{degrees of freedom} = 18)$ for the colon dataset, and $86.30 \pm 6.20(t = 5.86 > 2.8784, p = 0.01, \text{degrees of freedom} = 18)$ for the lymphoma dataset. Further analysis of the best solutions showed that the optimal ensemble size was not constant. In the lymphoma dataset, the ensemble size was relatively high (1–14) compared with that of the colon dataset (1–7). The optimized ensembles for the two datasets were not the same. In the colon dataset, SVM classifiers were not used, but in the lymphoma dataset, the SVM classifiers were common components.

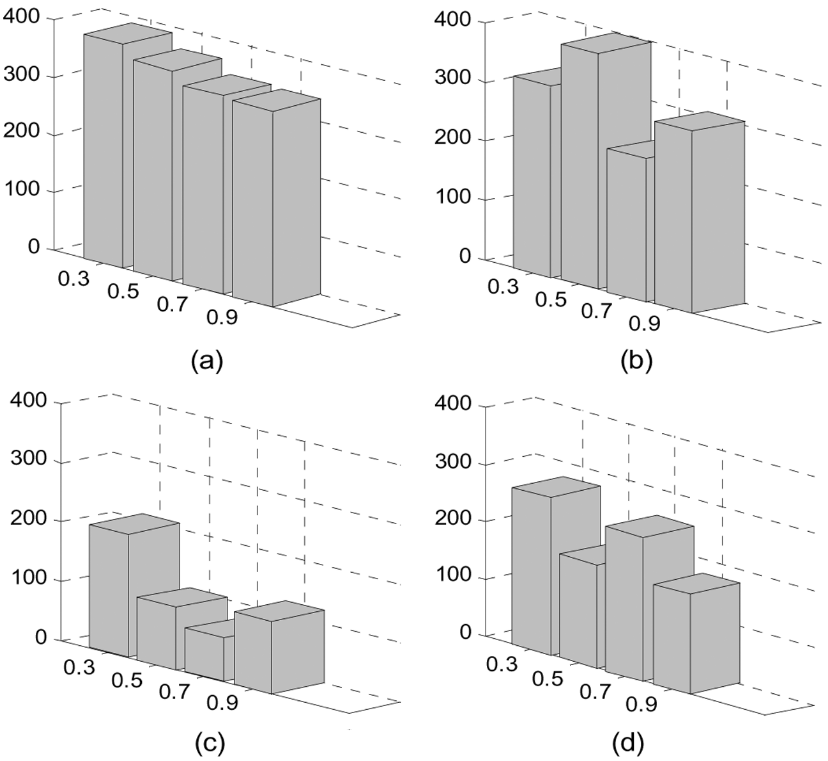

Fig. 6. Comparison of generations that found the optimal ensembles with different parameters: selection strategies, crossover rates, and mutation rates. X axis: crossover rate and y axis: generation. (a) Roulette wheel selection and 0.01 of mutation rate. (b) Roulette wheel selection and 0.05 of mutation rate. (c) Rank-based selection and 0.01 of mutation rate. (d) Rank-based selection and 0.05 of mutation rate.

Table V summarizes the performance comparison with other studies. Although Dettling and Buhlmann conducted the LOOCV method for the colon dataset [18], we could find no experiments that used bagging with the LOOCV method in the literature. However, a comparison confirmed that the proposed method (real-valued GA) performed better than the 1-NN, logitboost and classification tree methods, which assured that the proposed method was competitive. Although the data partitioning and other conditions are different, the real-valued GA outperforms the state-of-the art performance on the colon dataset (88.1%). Also, in the lymphoma dataset, the performance of real-valued GA is higher than the best results (97.6%) in the literature.

Determining the number of features and classifiers is a very important issue in ensemble research [42]. In general, the experiment designer needs to determine the maximum number of feature selection and classification algorithms for the ensemble. In our experiments, the maximum number of feature selection methods was seven and that of the classifiers was six. The EA searches for the optimal ensembles were composed of only subsets, and the final solutions discovered by the proposed method showed that only the subsets of all the feature selection methods and classifiers were used for the ensemble. The numbers of feature selection methods and classifiers for the optimal ensemble were determined by the EA.

We found that the average fitness level for the lymphoma cancer dataset increased as iteration went on, as shown in Fig. 4. After about 150 generations, they seemed to converge. The fact that we were not able to obtain satisfactory performance with the individual feature-classifiers (even after significant effort) indicates that there is room for improvement by using ensembles.

We also ran experiments using various parameters of the proposed method. Figs. 5 and 6 show satisfactory ensembles in most cases. Fig. 5 shows the generations that discovered the optimal ensemble with respect to the population size for the lymphoma cancer dataset, suggesting that larger population sizes identified optimal ensembles faster than smaller populations on average. Fig. 6 shows the generations that discovered the optimal ensemble with respect to various crossover rates, mutation rates, and selection strategies. In most cases, the EA found the optimal ensemble robustly. On average, the rank-based selection method Fig. 6(c) and (d) found the optimal ensemble using fewer generations than the roulette wheel selection method Fig. 6(a) and (b).

The practical usefulness of the proposed method lies in its time-efficiency. When we experimented with all the ensembles composed of only 25 feature-classifier pairs (about 3×10^7 ensembles), it took about 1 h. This means that we were able to evaluate 9320 ensembles per second. It would have been impossible to try evaluating all the ensembles of 42 feature-classifier pairs (about 4.3×10^{12} ensembles). On the other hand, it took less than 5 min using the EA, when we used 2000 chromosomes with 1000 generations. We were usually able to find the optimal ensemble within 100 generations with 2000 chromosomes, and 900 generations with 100 chromosomes, as summarized in Table VI.

TABLE VI
COMPARISON OF THE TIME REQUIRED TO SEARCH FOR THE OPTIMAL ENSEMBLE (IN THE GA, THE ROULETTE WHEEL SELECTION METHOD AND THE MAJORITY VOTING METHOD)

Searching method	The number of ensembles	Estimated required time
Exhaustive (25 feature-classifier pairs)	3×10^7	1 hour
Exhaustive (42 feature-classifier pairs)	4.3×10^{12} (All cases)	1 year
GA (900 iterations, 100 chromosomes)	1×10^5	15 seconds
GA (100 iterations, 2000 chromosomes)	1×10^6	3 minutes

V. CONCLUDING REMARKS

This paper deals with a novel application of EA to a fascinating field of bioinformatics, and reports promising, state-of-the-art results [18] in analyzing DNA microarray datasets. The EA is improved with better encoding scheme for chromosomes and a statistical t-test is conducted to show the superiority to other approaches for ensembles. By comparing the performance with previous publications that used the same datasets, we confirmed that the proposed method provided the competitive, state-of-the-art results. The EA-based ensemble classifier could produce better performance if there was enough diversity in the classification algorithms and feature selection methods. Also, this method can perform well without any prior knowledge about feature selection and classification algorithms.

Furthermore, the proposed method has the following biological implications.

- With the proposed scheme, even a novice could build the optimal ensemble composed of many different algorithms and feature extraction methods given the dataset. Also, the proposed scheme could provide minimal ensembles without redundancy or irrelevant components.
- The informative genes extracted from the ensemble can be used by biologists for future analysis to find relevant gene subsets. This would provide insights into building ensembles manually for very complicated problems.

In this paper, only binary classification problems were considered for the experiments but multiclass problems will be investigated in the future. There are some works for multiclass classification of gene expression data [43]. More exploration on the possibility of new genetic operators, selection methods, and combination methods needs to be done in future work.

REFERENCES

[1] Y. Moreau, F. de Smet, G. Thijs, K. Marchal, and B. de Moor, "Functional bioinformatics of microarray data: From expression to regulation," *Proc. IEEE*, vol. 90, no. 11, pp. 1722–1743, 2002.

[2] W. P. Kuo, E. Y. Kim, J. Trimarchi, T. K. Jenssen, S. A. Vinterbo, and L. Ohno-Machado, "A primer on gene expression and microarrays for machine learning researchers," *J. Biomed. Informatics*, vol. 37, no. 4, pp. 293–303, 2004.

[3] S.-B. Cho and H. H. Won, "Data mining for gene expression profiles from DNA microarray," *J. Softw. Eng. Knowl. Eng.*, vol. 13, no. 6, pp. 593–608, 2003.

[4] 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, pp. 673–679, 2001.

[5] T. Golub, D. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. Mesirov, H. Coller, M. Loh, J. Downing, M. Caligiuri, C. Bloomfield, and E. Lander, "Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring," *Science*, vol. 286, pp. 531–537, 1999.

[6] T. S. Furey, 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, pp. 906–914, 2000.

[7] L. I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*. New York: Wiley, 2004.

[8] Z. H. Zhou, J. Wu, and W. Tang, "Ensembling neural networks: Many could be better than all," *Artif. Intell.*, vol. 137, no. 1–2, pp. 239–263, 2002.

[9] S. Chatterjee, C. Carrera, and L. A. Lynch, "Genetic algorithms and traveling salesman problems," *Eur. J. Oper. Res.*, vol. 93, no. 3, pp. 490–510, 1996.

[10] L. I. Kuncheva and L. C. Jain, "Designing classifier fusion systems by genetic algorithms," *IEEE Trans. Evol. Comput.*, vol. 4, no. 4, pp. 327–336, 2000.

[11] 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.

[12] L. Li, W. Jiang, X. Li, K. L. Moser, Z. Guo, L. Du, Q. Wang, E. J. Topol, Q. Wang, and S. Rao, "A robust hybrid between genetic algorithm and support vector machine for extracting an optimal feature gene subset," *Genomics*, vol. 85, pp. 16–23, 2005.

[13] S. Dudoit, J. Fridlyand, and P. Speed, "Comparison of discrimination methods for classification of tumors using gene expression data," *J. Amer. Statist. Assoc.*, vol. 97, pp. 77–87, 2002.

[14] J. W. Lee, J. B. Lee, M. Park, and S. H. Song, "An extensive comparison of recent classification tools applied to microarray data," *Comput. Statist. Data Anal.*, vol. 48, pp. 869–885, 2005.

[15] S.-B. Cho and J. Ryu, "Classifying gene expression data of cancer using classifier ensemble with mutually exclusive features," *Proc. IEEE*, vol. 90, no. 11, pp. 1744–1753, 2002.

[16] A. C. Tan and D. Gilbert, "Ensemble machine learning on gene expression data for cancer classification," *Appl. Bioinformatics*, vol. 2, pp. s75–s83, 2003.

[17] M. Dettling, "Bagboosting for tumor classification with gene expression data," *Bioinformatics*, vol. 20, no. 18, pp. 3583–3593, 2004.

[18] M. Dettling and P. Buhlmann, "Boosting for tumor classification with gene expression data," *Bioinformatics*, vol. 19, no. 9, pp. 1061–1069, 2003.

[19] G. Valentini, M. Muselli, and F. Ruffino, "Cancer recognition with bagged ensembles of support vector machines," *Neurocomputing*, vol. 56, pp. 461–466, 2004.

[20] X. Li, S. Rao, Y. Wang, and B. Gong, "Gene mining: A novel and powerful ensemble decision approach to hunting for disease genes using microarray expression profiles," *Nucleic Acids Res.*, vol. 32, no. 9, pp. 2685–2694, 2004.

[21] X. Xu and A. Zhang, "Boost feature selection: A new gene selection algorithm for microarray dataset," in *Proc. 6th Int. Conf. Comput. Sci.: Workshop Bioinformatics Res. Appl.*, 2006, pp. 670–677.

[22] A. A. Alizadeh, M. B. Eisen, R. E. Davis, C. Ma, I. S. Lossos, A. Rosenwald, J. C. Boldrick, H. Sabet, T. Tran, X. Yu, J. I. Powell, L. Yang, G. E. Marti, T. Moore, J. Hudson, Jr., L. Lu, D. B. Lewis, R. Tibshirani, G. Sherlock, W. C. Chan, T. C. Greiner, D. D. Weisenburger, J. O. Armitage, R. Warnke, R. Levy, W. Wilson, M. R. Grever, J. C. Byrd, D. Botstein, P. O. Brown, and L. M. Staudt, "Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling," *Nature*, vol. 403, pp. 503–511, 2000.

[23] D. T. Ross, U. Scherf, M. B. Eisen, C. M. Perou, P. Spellman, V. Iyer, S. S. Jeffrey, M. Van de Rijn, M. Waltham, A. Pergamenschikov, J. C. F. Lee, D. Lashkari, D. Shalon, T. G. Myers, J. N. Weinstein, D. Bostein, and P. O. Brown, "Systematic variation in gene expression patterns in human cancer cell lines," *Nature Genetics*, vol. 24, pp. 227–234, 2000.

[24] U. Alon, N. Barkai, D. A. Notterman, K. Gish, S. Y. D. Mack, and J. Levine, "Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays," *Proc. Nat. Acad. Sci.*, vol. 96, pp. 6745–6750, 1999.

[25] M. West, C. Blanchette, H. Dressman, E. Huang, S. Ishida, R. Spang, H. Zuzan, J. Olson, J. Marks, and J. Nevins, "Predicting the clinical status of human breast cancer by using gene expression profiles," *Proc. Nat. Acad. Sci.*, vol. 98, pp. 11462–11467, 2001.

[26] L. J. van't Veer, H. Dai, M. J. van de Vijver, Y. D. He, A. A. Hart, M. Mao, H. L. Peterse, K. van der Kooy, M. J. Marton, A. T. Witteveen, G. J. Schreiber, R. M. Kerkhoven, C. Roberts, P. S. Linsle, R. Bernards, and S. H. Friend, "Gene expression profiling predicts clinical outcome of breast cancer," *Nature*, vol. 415, pp. 530–536, 2002.

[27] S. L. Pomeroy, D. Tamayo, M. Gaasenbeek, L. M. Sturla, M. Angelo, M. E. McLaughlin, J. Y. Kim, L. C. Goumnerova, P. M. Black, C. Lau, J. C. Allen, D. Zagzag, J. M. Olson, T. Curran, C. Wetmore, J. A. Biegel, T. Poggio, S. Mukherjee, R. Rifkin, A. Califano, G. Stolovitzky, D. N. Louis, J. P. Mesirov, E. S. Lander, and T. R. Golub, "Prediction of central nervous system embryonal tumor outcome based on gene expression," *Nature*, vol. 415, pp. 436–432, 2002.

[28] G. J. Gordon, R. V. Jensen, L. L. Hsiao, S. R. Gullans, J. E. Blumenstock, S. Ramaswamy, W. G. Richard, D. J. Sugarbaker, and R. Bueno, "Translation of microarray data into clinically relevant cancer diagnostic tests using gene expression ratios in lung cancer and mesothelioma," *Cancer Res.*, vol. 62, pp. 4963–4967, 2002.

[29] D. Singh, P. G. Febbo, K. Ross, D. G. Jackson, J. Manola, C. Ladd, P. Tamayo, A. A. Renshaw, A. V. D'Amico, J. P. Richie, E. S. Lander, M. Loda, P. W. Kantoff, T. R. Golub, and W. R. Sellers, "Gene expression correlates of clinical prostate cancer behavior," *Cancer Cell*, vol. 1, no. 2, pp. 203–209, 2002.

[30] M. E. Garber, O. G. Troyanskaya, K. Schluens, S. Petersen, Z. Thaesler, M. Pacyna-Gengelbach, M. V. De Rijn, G. D. Rosen, C. M. Perou, R. I. Whyte, R. B. Altman, P. O. Brown, D. Bostein, and I. Petersen, "Diversity of gene expression in adenocarcinoma of the lung," *Proc. Nat. Acad. Sci.*, vol. 98, pp. 13784–13789, 2001.

[31] M. B. Eisen, P. T. Spellman, P. O. Brown, and D. Bostein, "Clustering analysis and display of genome-wise expression patterns," *Proc. Nat. Acad. Sci.*, vol. 95, pp. 14863–14868, 1998.

[32] S. Dondeti, K. Kannan, and R. Manavalan, "Genetic algorithm optimized neural networks ensemble for estimation of mefenamic acid and paracetamol in tablets," *Acta Chimica Slovenica*, vol. 52, pp. 440–449, 2005.

[33] C. D. Stefano, A. D. Cioppa, and A. Marcelli, "Exploiting reliability for dynamic selection of classifiers by means of genetic algorithms," in *Proc. 7th Int. Conf. Document Anal. Recogn.*, 2003, pp. 671–675.

[34] Z. Q. Shen and F. S. Kong, "Optimizing weights by genetic algorithm for neural network ensemble," *Lecture Notes in Computer Science*, vol. 3173, pp. 323–331, 2004.

[35] S.-B. Cho, "Pattern recognition with neural networks combined by genetic algorithm," *Fuzzy Sets Syst.*, vol. 103, no. 2, pp. 339–347, 1999.

[36] K. Sirlantzis, M. C. Fairhust, and R. M. Guest, "An evolutionary algorithm for classifier and combination rule selection in multiple classifier systems," in *Proc. 16th Int. Conf. Pattern Recogn.*, 2002, vol. 2, pp. 771–774.

[37] Y. Liu, X. Yao, and T. Higuchi, "Evolutionary ensembles with negative correlation learning," *IEEE Trans. Evol. Comput.*, vol. 4, no. 4, pp. 380–387, 2000.

[38] D. V. Nguyen and D. M. Rocke, "Partial least squares proportional hazard regression for application to DNA microarray survival data," *Bioinformatics*, vol. 18, no. 12, pp. 1625–1632, 2002.

[39] D. Opitz and R. Maclin, "Popular ensemble methods: An empirical study," *J. Artif. Intell. Res.*, vol. 11, pp. 169–198, 1999.

[40] L. Hansen and P. Salamon, "Neural network ensembles," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, pp. 993–1001, 1990.

[41] X. Yao and Y. Liu, "Making use of population information in evolutionary artificial neural networks," *IEEE Trans. Syst., Man, Cybern.—Part B*, vol. 28, no. 3, pp. 417–425, 1998.

[42] X. Hu, "Using rough set theory and database operations to construct a good ensemble of classifier for data mining application," in *Proc. IEEE Int. Conf. Data Mining*, 2001, pp. 233–240.

[43] D. V. Nguyen and D. M. Rocke, "Multi-class cancer classification via partial least squares with gene expression profiles," *Bioinformatics*, vol. 18, pp. 1216–1226, 2002.

[44] G. Valentini, "Gene expression data analysis of human lymphoma using support vector machines and output coding ensembles," *Artif. Intell. Med.*, vol. 26, pp. 281–304, 2002.

[45] K.-J. Kim and S.-B. Cho, "Ensemble classifiers based on correlation analysis for DNA microarray classification," *Neurocomputing*, vol. 70, no. 1–3, pp. 187–199, 2006.

[46] S.-B. Cho and H.-H. Won, "Cancer classification using ensemble of neural networks with multiple significant gene subsets," *Appl. Intell.*, vol. 26, pp. 243–250, 2007.

[47] J.-H. Hong and S.-B. Cho, "The classification of cancer based on DNA microarray data that uses diverse ensemble genetic programming," *Artif. Intell. Med.*, vol. 36, pp. 43–58, 2006.

Kyung-Joong Kim (M'07) received the B.S., M.S., and Ph.D. degrees in computer science from Yonsei University, Seoul, Korea, in 2000, 2002, and 2007, respectively.

Since 2007, he has been a Postdoctoral Researcher with the Department of Mechanical and Aerospace Engineering, Cornell University, Ithaca, NY. His research interests include evolutionary computation, evolving analog circuits, robotics, DNA microarray analysis, and artificial life.

Sung-Bae Cho (SM'06) received the B.S. degree in computer science from Yonsei University, Seoul, Korea, in 1988, and the M.S. and Ph.D. degrees in computer science from Korea Advanced Institute of Science and Technology (KAIST), Taejeon, Korea, in 1990 and 1993, respectively.

He was a Member of the Research Staff at the Center for Artificial Intelligence Research, KAIST, from 1991 to 1993. He was an Invited Researcher of Human Information Processing Research Laboratories, Advanced Telecommunications Research (ATR) Institute, Kyoto, Japan, from 1993 to 1995, and a Visiting Scholar at University of New South Wales, Canberra, Australia, in 1998. He was also a Visiting Professor at the University of British Columbia, Vancouver, BC, Canada, from 2005 to 2006. Since 1995, he has been a Professor with the Department of Computer Science, Yonsei University. His research interests include neural networks, pattern recognition, intelligent man-machine interfaces, evolutionary computation, and artificial life.

Dr. Cho is a Member of the Korea Information Science Society, ACM, the IEEE Computer Society, the IEEE Systems, Man, and Cybernetics Society, and the IEEE Computational Intelligence Society. He was awarded outstanding paper prizes from the IEEE Korea Section in 1989 and 1992, and another award from the Korea Information Science Society in 1990. He was also the recipient of the Richard E. Merwin Prize from the IEEE Computer Society in 1993. He was listed in *Who's Who in Pattern Recognition* from the International Association for Pattern Recognition in 1994, and received the Best Paper Award at the International Conference on Soft Computing in 1996 and 1998. Also, he received the Best Paper Award at the World Automation Congress in 1998, and is listed in *Marquis Who's Who in Science and Engineering* in 2000 and in *Marquis Who's Who in the World* in 2001.