f-Information Measures for Efficient Selection of Discriminative Genes From Microarray Data

Pradipta Maji

***Abstract*—Among the great amount of genes presented in microarray gene expression data, only a small fraction is effective for performing a certain diagnostic test. In this regard, mutual information has been shown to be successful for selecting a set of relevant and nonredundant genes from microarray data. However, information theory offers many more measures such as the f-information measures that may be suitable for selection of genes from microarray gene expression data. This paper presents different f-information measures as the evaluation criteria for gene selection problem. To compute the gene–gene redundancy (respectively, gene-class relevance), these information measures calculate the divergence of the joint distribution of two genes' expression values (respectively, the expression values of a gene and the class labels of samples) from the joint distribution when two genes (respectively, the gene and class label) are considered to be completely independent. The performance of different f-information measures is compared with that of the mutual information based on the predictive accuracy of naive Bayes classifier, K-nearest neighbor rule, and support vector machine. An important finding is that some f-information measures are shown to be effective for selecting relevant and nonredundant genes from microarray data. The effectiveness of different f-information measures, along with a comparison with mutual information, is demonstrated on breast cancer, leukemia, and colon cancer datasets. While some f-information measures provide 100% prediction accuracy for all three microarray datasets, mutual information attains this accuracy only for breast cancer dataset, and 98.6% and 93.6% for leukemia and colon cancer datasets, respectively.**

***Index Terms*—Classification, feature selection, gene selection, microarray analysis, mutual information.**

I. Introduction

RECENT advancement and wide use of high-throughput technology are producing an explosion in using gene expression phenotype for identification and classification in a variety of diagnostic areas. An important application of gene expression data in functional genomics is to classify samples according to their gene expression profiles, such as to classify cancer versus normal samples or to classify different types or subtypes of cancer [1].

A microarray gene expression dataset can be represented by an expression table, $\mathcal{T} = \{w_{ij} | i = 1, \ldots, m, j = 1, \ldots, n\}$, where $w_{ij} \in \Re$ is the measured expression level of gene $\mathbb{G}_i$ in the jth sample, and m and n represent the total number of genes and samples, respectively. Each row in the expression table corresponds to one particular gene and each column to a sample [1].

Manuscript received May 19, 2008; revised July 15, 2008. First published September 16, 2008; current version published May 6, 2009.

The author is with the Machine Intelligence Unit, Indian Statistical Institute, Kolkata 700 108, India (e-mail: pmaji@isical.ac.in).

However, for most gene expression data, the number of training samples is still very small compared to the large number of genes involved in the experiments [1]. When the number of genes is significantly greater than the number of samples, it is possible to find biologically relevant correlations of gene behavior with the sample categories [2].

However, among the large amount of genes, only a small fraction are effective for performing a certain task. Also, a small subset of genes is desirable in developing gene-expression-based diagnostic tools for delivering precise, reliable, and interpretable results. With the gene selection results, the cost of biological experiment and decision can be greatly reduced by analyzing only the marker genes. Hence, identifying a reduced set of most relevant genes is the goal of gene selection. The small number of training samples and a large number of genes make gene selection a more relevant and challenging problem in gene-expression-based classification. This is an important problem in machine learning and referred to as feature selection [3], [4].

Conventional methods of gene (feature) selection involve evaluating different gene subsets using some index and selecting the best among them. Depending on the way of computing the gene (feature) evaluation index, gene selection methods are generally divided into two broad categories: filter approach [3], [5]–[8] and wrapper approach [3], [4], [9]. In filter approach, the algorithms do not perform classification of the data in the process of gene evaluation. Before application of the actual learning algorithm, the best subset of genes is selected in one pass by evaluating some predefined criteria, which are independent of the actual generalization performance of the learning machine. Hence, the filter approach is computationally less expensive and more general [3], [5]–[8]. On the other hand, in its most general formulation, the wrapper approach consists of using the prediction performance of a given learning machine to assess the relative usefulness of different subsets of genes. Since the wrapper approach uses the learning machine as a black box, it generally outperforms the filter approach in the aspect of final predictive accuracy of the learning machine. However, it is computationally more expensive than that of filter approach [3], [4], [9].

In a gene selection process, an optimal gene subset is always relative to a certain criterion. In general, different criteria may lead to different optimal gene subsets. However, every criterion tries to measure the discriminating ability of a gene or a subset of genes to distinguish different class labels. To measure the gene-class relevance, different statistical and information theoretic measures such as the F-test, t-test [5], [6], entropy, information gain, mutual information [5], [7], and normalized mutual information [8] are typically used, and the same or a different

metric like mutual information, the L_1 distance, Euclidean distance, Pearson's correlation coefficient, etc., [5], [7], [10] are employed to calculate the gene–gene redundancy. However, as the F-test, t-test, Euclidean distance, Pearson's correlation, etc., depend on the actual gene expression values of the microarray data, they are very sensitive to noise or an outlier of the dataset [5], [7], [10], [11]. On the other hand, as information measures depend only on the probability distribution of a random variable rather than on its actual values, they are more effective to evaluate both gene-class relevance and gene–gene redundancy [7], [8].

However, measures of the distance between a joint probability distribution and product of the marginal distributions are information measures [12], [13]. Information measures constitute a subclass of the divergence measures, which are measures of the distance between two arbitrary distributions. A specific class of information (divergence) measures, of which mutual information is a member, is formed by f-information (f-divergence) measures [12], [13]. In this paper, several f-information measures are compared with the mutual information by applying them to the selection of genes from microarray data. The performance of different information measures is studied using the predictive accuracy of naive Bayes (NB) classifier, K-nearest neighbor (K-NN) rule, and support vector machine (SVM). The effectiveness of different f-information measures, along with a comparison with mutual information, is demonstrated on breast cancer, leukemia, and colon cancer datasets.

II. Gene Selection Using f-Information Measures

In a microarray data analysis, the dataset may contain a number of redundant genes with low relevance to the classes. The presence of such redundant and nonrelevant genes leads to a reduction in the useful information. Ideally, the selected genes should have high relevance with the classes while the redundancy among them would be as low as possible. The genes with high relevance are expected to be able to predict the classes of the samples. However, the prediction capability is reduced if many redundant genes are selected. In contrast, a dataset that contains genes not only with high relevance with respect to the classes but also with low mutual redundancy is more effective in its prediction capability. Hence, to assess the effectiveness of the genes, both relevance and redundancy need to be measured quantitatively. In this paper, the maximum relevance-minimum redundancy framework of [5] and [7] is used to select a set of relevant and nonredundant genes from microarray gene expression datasets.

A. Maximum Relevance-Minimum Redundancy Framework

Let $\mathbb{G} = \{\mathbb{G}_1, \ldots, \mathbb{G}_i, \ldots, \mathbb{G}_j, \ldots, \mathbb{G}_m\}$ denote the set of m genes of a given microarray dataset and $\mathbb{S}$ is the set of selected genes. Define $\hat{f}(\mathbb{G}_i, \mathbb{C})$ as the relevance of the gene $\mathbb{G}_i$ with respect to the class label $\mathbb{C}$, while $\tilde{f}(\mathbb{G}_i, \mathbb{G}_j)$ is the redundancy between two genes $\mathbb{G}_i$ and $\mathbb{G}_j$. The total relevance of all selected genes is, therefore, given by

$$\mathcal{J}_{\text{relev}} = \sum_{\mathbb{G}_i \in \mathbb{S}} \hat{f}(\mathbb{G}_i, \mathbb{C}) \tag{1}$$

while the total redundancy among the selected genes is

$$\mathcal{J}_{\text{redun}} = \sum_{\mathbb{G}_i, \mathbb{G}_j \in \mathbb{S}} \tilde{f}(\mathbb{G}_i, \mathbb{G}_j). \tag{2}$$

Therefore, the problem of selecting a set $\mathbb{S}$ of relevant and nonredundant genes from the whole set $\mathbb{G}$ of m genes is equivalent to maximize $\mathcal{J}_{\text{relev}}$ and minimize $\mathcal{J}_{\text{redun}}$, i.e., to maximize the objective function $\mathcal{J}$, where

$$\mathcal{J} = \mathcal{J}_{\text{relev}} - \mathcal{J}_{\text{redun}} = \sum_i \hat{f}(\mathbb{G}_i, \mathbb{C}) - \sum_{i,j} \tilde{f}(\mathbb{G}_i, \mathbb{G}_j). \tag{3}$$

To solve the previous problem, a greedy algorithm, as follows, is widely used [5], [7].

1) Initialize $\mathbb{G} \leftarrow \{\mathbb{G}_1, \ldots, \mathbb{G}_i, \ldots, \mathbb{G}_j, \ldots, \mathbb{G}_m\}, \mathbb{S} \leftarrow \emptyset$.
2) Calculate the relevance $\hat{f}(\mathbb{G}_i, \mathbb{C})$ of each gene $\mathbb{G}_i \in \mathbb{G}$.
3) Select gene $\mathbb{G}_i$ as the most relevant gene that has highest relevance $\hat{f}(\mathbb{G}_i, \mathbb{C})$. In effect, $\mathbb{G}_i \in \mathbb{S}$ and $\mathbb{G} = \mathbb{G} \setminus \mathbb{G}_i$.
4) Repeat the following two steps until the desired number of genes are selected.
5) Calculate the redundancy between selected genes of $\mathbb{S}$ and each of the remaining genes of $\mathbb{G}$.
6) From the remaining genes of $\mathbb{G}$, select gene $\mathbb{G}_j$ that maximizes

$$\hat{f}(\mathbb{G}_j, \mathbb{C}) - \frac{1}{|\mathbb{S}|} \sum_{i \in \mathbb{S}} \tilde{f}(\mathbb{G}_i, \mathbb{G}_j). \tag{4}$$

As a result of that, $\mathbb{G}_j \in \mathbb{S}$ and $\mathbb{G} = \mathbb{G} \setminus \mathbb{G}_j$.

B. f-Information Measures for Gene Selection

In this paper, different f-information measures are reported to compute both gene-class relevance and gene–gene redundancy for selection of genes from microarray data. The f-information measures calculate the distance between a given joint probability p_{ij} and the joint probability when the variables are independent $p_i p_j$. In the following analysis, it is assumed that all probability distributions are complete, i.e., $\sum_i p_i = \sum_j p_j = \sum_{i,j} p_{ij} = 1$.

The extent to which two probability distributions differ can be expressed by a so-called measure of divergence. Such a measure will reach a minimum value when the two probability distributions are identical and the value increases with increasing disparity between the two distributions. A specific class of divergence measures is the set of f-divergence measures [12], [13]. For two discrete probability distributions $P = \{p_i | i = 1, 2, \ldots, n\}$ and $Q = \{q_i | i = 1, 2, \ldots, n\}$, the f-divergence is defined as

$$f(P\|Q) = \sum_i q_i f\left(\frac{p_i}{q_i}\right). \tag{5}$$

The demands on the function f are that: 1) $f : [0, \infty) \rightarrow (-\infty, \infty]$; 2) f is continuous and convex on $[0, \infty)$; 3) finite on $(0, \infty)$; and 4) strictly convex at some point $x \in (0, \infty)$. The

following definition completes the definition of f-divergence for the two cases for which (5) is not defined:

$$q_i f\left(\frac{p_i}{q_i}\right) = \begin{cases} 0, & \text{if } p_i = q_i = 0 \\ p_i \lim_{x\to\infty} \frac{f(x)}{x}, & \text{if } p_i > 0, q_i = 0. \end{cases} \quad (6)$$

A special case of f-divergence measures is the f-information measures. These are defined similarly to f-divergence measures, but apply only to specific probability distributions, namely, the joint probability of two variables P and their marginal probabilities' product $P_1 \times P_2$. Thus, the f-information is a measure of dependence: it measures the distance between a given joint probability and the joint probability when the variables are independent [12], [13]. The frequently used functions that can be used to form f-information measures include V-information, I_α-information, M_α-information, and χ^α-information. On the other hand, the Renyi's distance measure does not fall in the class of f-divergence measures as it does not satisfy the definition of f-divergence. However, it is divergence measure in the sense that it measures the distance between two distributions, and it is directly related to f-divergence.

1) V-Information: One of the simplest measures of dependence can be obtained using the function $V = |x - 1|$, which results in the V-information [12], [13]

$$V(P||P_1 \times P_2) = \sum_{i,j} |p_{ij} - p_i p_j| \quad (7)$$

where $P_1 = \{p_i | i = 1, 2, \ldots, n\}$, $P_2 = \{p_j | j = 1, 2, \ldots, n\}$, and $P = \{p_{ij} | i = 1, 2, \ldots, n; j = 1, 2, \ldots, n\}$ represent two marginal probability distributions and their joint probability distribution, respectively. That is, the V-information calculates the absolute distance between joint probability of two variables and their marginal probabilities' product.

2) I_α-Information: The I_α-information is defined as

$$I_\alpha(P||P_1 \times P_2) = \frac{1}{\alpha(\alpha-1)} \left(\sum_{i,j} \frac{(p_{ij})^\alpha}{(p_i p_j)^{\alpha-1}} - 1 \right) \quad (8)$$

for $\alpha \neq 0, \alpha \neq 1$. The class of I_α-information includes mutual information, which equals I_α for the limit $\alpha \to 1$, i.e.,

$$I_1(P||P_1 \times P_2) = \sum_{i,j} p_{ij} \log\left(\frac{p_{ij}}{p_i p_j}\right) \text{ for } \alpha \to 1. \quad (9)$$

3) M_α-Information: The M_α-information, defined by Matusita [12], [13], is as follows:

$$M_\alpha(x) = |x^\alpha - 1|^{\frac{1}{\alpha}}, \quad 0 < \alpha \leq 1. \quad (10)$$

When applying this function in the definition of an f-information measure, the resulting M_α-information measures are

$$M_\alpha(P||P_1 \times P_2) = \sum_{i,j} |(p_{ij})^\alpha - (p_i p_j)^\alpha|^{1/\alpha} \quad (11)$$

for $0 < \alpha \leq 1$. These constitute a generalized version of V-information, i.e., the M_α-information is identical to V-information for $\alpha = 1$.

4) χ^α-Information: The class of χ^α-information measures, proposed by Liese and Vajda [12], is as follows:

$$\chi^\alpha(x) = \begin{cases} |1 - x^\alpha|^{1/\alpha}, & \text{for } 0 < \alpha \leq 1 \\ |1 - x|^\alpha, & \text{for } \alpha > 1. \end{cases} \quad (12)$$

For $0 < \alpha \leq 1$, this function equals to the M_α function. The χ^α- and M_α-information measures are, therefore, also identical for $0 < \alpha \leq 1$. For $\alpha > 1$, χ^α-information can be written as

$$\chi^\alpha(P||P_1 \times P_2) = \sum_{i,j} \frac{|p_{ij} - p_i p_j|^\alpha}{(p_i p_j)^{\alpha-1}}. \quad (13)$$

5) Renyi Distance: The Renyi distance, a measure of information of order α [12], [13], can be defined as

$$R_\alpha(P||P_1 \times P_2) = \frac{1}{\alpha - 1} \log \sum_{i,j} \frac{(p_{ij})^\alpha}{(p_i p_j)^{\alpha-1}}$$

for $\alpha \neq 0, \alpha \neq 1$. It reaches its minimum value when p_{ij} and $p_i p_j$ are identical, in which case, the summation reduces to $\sum p_{ij}$. As a complete probability distribution is assumed, the sum is 1 and the minimum value of the measure is, therefore, equal to 0. The limit of Renyi's measure for α approaching 1 equals $I_1(P||P_1 \times P_2)$, which is the mutual information.

6) Discretization: In microarray gene expression datasets, the class labels of samples are represented by discrete symbols, while the expression values of genes are continuous. Hence, to measure both gene-class relevance of a gene with respect to class labels and gene–gene redundancy between two genes using information theoretic measures such as mutual information [5], [7], normalized mutual information [8], and f-information measures, the continuous expression values of a gene are divided into several discrete partitions. The *a prior* (marginal) probabilities and their joint probabilities are then calculated to compute both gene-class relevance and gene–gene redundancy using the definitions for discrete cases. In this paper, the discretization method reported in [5] and [7] is employed to discretize the continuous gene expression values. The expression values of a gene are discretized using mean μ and standard deviation σ computed over n expression values of that gene: any value larger than $(\mu + \sigma/2)$ is transformed to state 1; any value between $(\mu - \sigma/2)$ and $(\mu + \sigma/2)$ is transformed to state 0; any value smaller than $(\mu - \sigma/2)$ is transformed to state -1. These three states correspond to the overexpression, baseline, and underexpression of genes.

III. Experimental Results and Discussion

The performance of different f-information measures is extensively compared with that of mutual information and normalized mutual information. Based on the argumentation given in Section II-B, the following information measures are chosen to include in the study:

1) I_α- and R_α-information measures for $\alpha \neq 0$ and $\alpha \neq 1$;
2) mutual information ($I_{1.0}$- and $R_{1.0}$-information);
3) M_α-information measure for $0 < \alpha \leq 1$;
4) χ^α-information measure for $\alpha > 1$;
5) normalized mutual information U.

A. Gene Expression Datasets

In this paper, three public datasets of cancer microarrays are used. Since binary classification is a typical and fundamental issue in diagnostic and prognostic prediction of cancer, different f-information measures are compared using the following binary-class datasets.

1) Breast Cancer Dataset: The breast cancer dataset contains expression levels of 7129 genes in 49 breast tumor samples [14]. The samples are classified according to their estrogen receptor (ER) status: 25 samples are ER positive while the other 24 samples are ER negative.

2) Leukemia Dataset: It is an affymetrix high-density oligonucleotide array that contains 7070 genes and 72 samples from two classes of leukemia: 47 acute lymphoblastic leukemia and 25 acute myeloid leukemia [1].

3) Colon Cancer Dataset: The colon cancer dataset contains expression levels of 40 tumor and 22 normal colon tissues. Only the 2000 genes with the highest minimal intensity were selected by [15].

B. Class Prediction Methods

The following three classifiers are used to evaluate the performance of different f-information measures.

1) NB Classifier: The NB classifier [16] is one of the oldest classifiers. It is obtained by using the Bayes rule and assuming that features (variables) are independent of each other given its class. For the jth sample s_j with m gene expression levels $\{w_{1j}, \ldots, w_{ij}, \ldots, w_{mj}\}$ for the m genes, the posterior probability that s_j belongs to class c is

$$p(c \mid s_j) \propto \prod_{i=1}^{m} p(w_{ij} \mid c) \tag{14}$$

where $p(w_{ij} \mid c)$ are conditional tables (or conditional density) estimated from training examples.

2) SVM: The SVM [17] is a margin classifier that draws an optimal hyperplane in the feature vector space; this defines a boundary that maximizes the margin between data samples in different classes, therefore leading to good generalization properties. A key factor in the SVM is to use kernels to construct nonlinear decision boundary. In this paper, linear kernels are used.

3) K-NN Rule: The K-NN rule [16] is used for evaluating the effectiveness of the reduced gene set for classification. It classifies samples based on closest training samples in the feature space. A sample is classified by a majority vote of its K-neighbors, with the sample being assigned to the class most common among its K-NNs. The value of K, chosen for the K-NN, is the square root of the number of samples in training set.

C. Performance Analysis

The experimental results on three microarray datasets are presented in Tables I–IX. Subsequent discussions analyze the results with respect to the prediction accuracy of the NB, SVM, and K-NN classifiers. Tables I, IV, VII and Tables II, V, VIII provide the performance of different f-information measures using the NB and SVM, respectively, while Tables III, VI, IX, show the results using the K-NN. The values of α for f-information measures investigated are 0.2, 0.5, 0.8, 1.5, 2.0, and 3.0. Some measures resemble mutual information for $\alpha = 1.0$ (I_α and R_α) and some resemble another measure ($M_{1.0}$ and $\chi^{1.0}$ equal V). To compute the prediction accuracy of the NB, SVM, and K-NN, the leave-one-out cross validation is performed on each gene expression dataset. The number of genes selected ranges from 2 to 30.

Tables I, IV, and VII show that, for three microarray datasets, genes selected by $I_{0.2}$-, $M_{0.2}$-, and $R_{0.2}$-information measures lead to higher classification accuracy than those selected by mutual information and other f-information measures. With

TABLE I
PERFORMANCE ON BREAST CANCER DATASET USING NB CLASSIFIER

f	Number of Selected Genes							
	2	5	8	10	15	20	25	30
$I_{0.2}$	95.9	98.0	98.0	98.0	100	100	100	100
$I_{0.5}$	95.9	98.0	98.0	98.0	100	100	100	98.0
$I_{0.8}$	95.9	100	95.9	98.0	98.0	98.0	95.9	93.9
$I_{1.0}$	95.9	98.0	95.9	100	98.0	93.9	93.9	89.8
$I_{1.5}$	95.9	98.0	95.9	93.9	93.9	91.8	91.8	89.8
$I_{2.0}$	95.9	95.9	95.9	93.9	91.8	91.8	91.8	87.8
$I_{3.0}$	95.9	95.9	95.9	93.9	91.8	91.8	89.8	87.8
$M_{0.2}$	85.7	95.9	95.9	98.0	100	100	100	100
$M_{0.5}$	95.9	98.0	98.0	98.0	100	100	100	98.0
$M_{0.8}$	95.9	93.9	95.9	98.0	93.9	91.8	91.8	87.8
$M_{1.0}$	87.8	89.8	83.7	85.7	89.8	87.8	87.8	83.7
$\chi^{1.5}$	95.9	98.0	95.9	98.0	93.9	89.8	89.8	85.7
$\chi^{2.0}$	95.9	95.9	95.9	93.9	91.8	91.8	91.8	87.8
$\chi^{3.0}$	95.9	95.9	95.9	93.9	93.9	93.9	93.9	89.8
$R_{0.2}$	95.9	98.0	98.0	98.0	100	100	100	100
$R_{0.5}$	95.9	98.0	98.0	98.0	100	100	100	98.0
$R_{0.8}$	95.9	100	95.9	95.9	98.0	98.0	95.9	93.9
$R_{1.0}$	95.9	98.0	95.9	100	98.0	93.9	93.9	89.8
$R_{1.5}$	95.9	98.0	95.9	93.9	91.8	91.8	91.8	89.8
$R_{2.0}$	95.9	91.8	95.9	95.9	91.8	91.8	91.8	89.8
$R_{3.0}$	93.9	89.8	93.9	93.9	93.9	91.8	91.8	91.8
U	95.9	98.0	98.0	100	95.9	93.9	93.9	91.8

TABLE II
PERFORMANCE ON BREAST CANCER DATASET USING SVM

f	Number of Selected Genes							
	2	5	8	10	15	20	25	30
$I_{0.2}$	81.6	100	95.9	98.0	98.0	100	95.9	95.9
$I_{0.5}$	81.6	100	100	100	95.9	95.9	100	95.9
$I_{0.8}$	81.6	98.0	100	100	98.0	95.9	95.9	98.0
$I_{1.0}$	81.6	98.0	100	100	98.0	95.9	95.9	93.9
$I_{1.5}$	85.7	91.8	98.0	100	98.0	100	95.9	95.9
$I_{2.0}$	85.7	95.9	98.0	100	100	100	95.9	95.9
$I_{3.0}$	85.7	95.9	98.0	100	100	95.9	95.9	95.9
$M_{0.2}$	77.6	95.9	91.8	89.8	87.8	93.9	93.9	95.9
$M_{0.5}$	81.6	100	100	100	95.9	95.9	100	95.9
$M_{0.8}$	85.7	89.8	93.9	89.8	93.9	95.9	93.9	93.9
$M_{1.0}$	83.7	81.6	87.8	91.8	87.8	83.7	83.7	83.7
$\chi^{1.5}$	85.7	87.8	91.8	89.8	93.9	91.8	95.9	95.9
$\chi^{2.0}$	85.7	95.9	98.0	100	100	100	95.9	95.9
$\chi^{3.0}$	85.7	89.8	100	95.9	98.0	95.9	98.0	93.9
$R_{0.2}$	81.6	100	95.9	98.0	98.0	98.0	95.9	95.9
$R_{0.5}$	81.6	100	100	100	95.9	95.9	100	95.9
$R_{0.8}$	81.6	98.0	100	100	98.0	95.9	95.9	98.0
$R_{1.0}$	81.6	98.0	100	100	98.0	95.9	95.9	93.9
$R_{1.5}$	85.7	91.8	98.0	100	98.0	100	95.9	95.9
$R_{2.0}$	85.7	89.8	95.9	95.9	98.0	100	95.9	95.9
$R_{3.0}$	87.8	87.8	100	100	93.9	95.9	93.9	95.9
U	81.6	98.0	100	100	98.0	95.9	98.0	95.9

TABLE III
PERFORMANCE ON BREAST CANCER DATASET USING K-NN RULE

f	Number of Selected Genes							
	2	5	8	10	15	20	25	30
$I_{0.2}$	89.8	93.9	93.9	95.9	98.0	95.9	95.9	93.9
$I_{0.5}$	89.8	93.9	95.9	95.9	98.0	98.0	95.9	95.9
$I_{0.8}$	89.8	98.0	95.9	95.9	98.0	95.9	95.9	98.0
$I_{1.0}$	89.8	98.0	100	98.0	100	95.9	93.9	98.0
$I_{1.5}$	85.7	93.9	100	100	98.0	95.9	93.9	95.9
$I_{2.0}$	85.7	91.8	100	100	98.0	95.9	95.9	98.0
$I_{3.0}$	85.7	91.8	100	100	98.0	93.9	95.9	98.0
$M_{0.2}$	83.7	93.9	85.7	83.7	87.8	89.8	85.7	85.7
$M_{0.5}$	89.8	93.9	95.9	95.9	98.0	98.0	95.9	95.9
$M_{0.8}$	85.7	89.8	85.7	91.8	95.9	95.9	95.9	93.9
$M_{1.0}$	71.4	89.8	89.8	89.8	89.8	91.8	91.8	91.8
$\chi^{1.5}$	85.7	89.8	93.9	91.8	95.9	93.9	95.9	93.9
$\chi^{2.0}$	85.7	91.8	100	100	98.0	95.9	95.9	98.0
$\chi^{3.0}$	85.7	91.8	95.9	98.0	98.0	98.0	100	98.0
$R_{0.2}$	89.8	93.9	93.9	95.9	98.0	98.0	95.9	93.9
$R_{0.5}$	89.8	93.9	95.9	95.9	98.0	98.0	95.9	95.9
$R_{0.8}$	89.8	98.0	95.9	95.9	98.0	95.9	95.9	98.0
$R_{1.0}$	89.8	98.0	100	98.0	100	95.9	93.9	98.0
$R_{1.5}$	85.7	93.9	100	100	95.9	95.9	95.9	93.9
$R_{2.0}$	85.7	91.8	98.0	95.9	95.9	95.9	95.9	93.9
$R_{3.0}$	91.8	98.0	100	95.9	93.9	95.9	98.0	100
U	89.8	98.0	98.0	98.0	98.0	95.9	93.9	93.9

TABLE IV
PERFORMANCE ON LEUKEMIA DATASETT USING NB CLASSIFIER

f	Number of Selected Genes							
	2	5	8	10	15	20	25	30
$I_{0.2}$	97.2	98.6	100	100	97.2	97.2	95.8	95.8
$I_{0.5}$	98.6	97.2	97.2	95.8	95.8	95.8	94.4	93.1
$I_{0.8}$	98.6	98.6	97.2	95.8	94.4	93.1	90.3	87.5
$I_{1.0}$	98.6	98.6	95.8	95.8	94.4	90.3	87.5	86.1
$I_{1.5}$	94.4	97.2	95.8	95.8	91.7	87.5	84.7	84.7
$I_{2.0}$	94.4	97.2	95.8	95.8	91.7	88.9	84.7	84.7
$I_{3.0}$	94.4	97.2	97.2	94.4	94.4	87.5	84.7	83.3
$M_{0.2}$	87.5	95.8	100	100	100	100	100	100
$M_{0.5}$	98.6	97.2	97.2	95.8	95.8	95.8	94.4	93.1
$M_{0.8}$	100	97.2	97.2	95.8	91.7	88.9	86.1	86.1
$M_{1.0}$	94.4	95.8	94.4	94.4	88.9	90.3	87.5	84.7
$\chi^{1.5}$	94.4	97.2	97.2	95.8	90.3	87.5	86.1	86.1
$\chi^{2.0}$	94.4	97.2	95.8	95.8	91.7	88.9	84.7	84.7
$\chi^{3.0}$	94.4	97.2	97.2	94.4	90.3	87.5	84.7	84.7
$R_{0.2}$	97.2	98.6	100	100	98.6	97.2	95.8	95.8
$R_{0.5}$	98.6	97.2	97.2	95.8	95.8	95.8	94.4	93.1
$R_{0.8}$	98.6	97.2	95.8	95.8	94.4	91.7	90.3	87.5
$R_{1.0}$	98.6	98.6	95.8	95.8	94.4	90.3	87.5	86.1
$R_{1.5}$	98.6	98.6	97.2	97.2	93.1	88.9	86.1	86.1
$R_{2.0}$	98.6	97.2	95.8	94.4	93.1	88.9	86.1	84.7
$R_{3.0}$	98.6	97.2	95.8	94.4	91.7	88.9	87.5	84.7
U	98.6	97.22	95.8	94.4	93.1	90.3	88.9	87.5

TABLE V
PERFORMANCE ON LEUKEMIA DATASET USING SVM

f	Number of Selected Genes							
	2	5	8	10	15	20	25	30
$I_{0.2}$	94.4	94.4	94.4	94.4	93.1	94.4	97.2	97.2
$I_{0.5}$	94.4	95.8	95.8	95.8	97.2	95.8	95.8	97.2
$I_{0.8}$	94.4	94.4	95.8	95.8	95.8	98.6	97.2	98.6
$I_{1.0}$	94.4	94.4	93.1	95.8	95.8	98.6	98.6	98.6
$I_{1.5}$	93.1	97.2	97.2	95.8	100	97.2	97.2	97.2
$I_{2.0}$	93.1	95.8	95.8	94.4	98.6	97.2	97.2	97.2
$I_{3.0}$	93.1	95.8	97.2	95.8	97.2	97.2	97.2	95.8
$M_{0.2}$	93.1	97.2	97.2	95.8	95.8	94.4	93.1	94.4
$M_{0.5}$	94.4	95.8	95.8	95.8	97.2	95.8	95.8	97.2
$M_{0.8}$	90.3	97.2	94.4	95.8	100	98.6	98.6	95.8
$M_{1.0}$	90.3	97.2	98.6	98.6	98.6	100	97.2	97.2
$\chi^{1.5}$	93.1	95.8	97.2	98.6	97.2	97.2	95.8	97.2
$\chi^{2.0}$	93.1	95.8	95.8	94.4	98.6	97.2	97.2	97.2
$\chi^{3.0}$	93.1	97.2	97.2	95.8	98.6	94.4	98.6	98.6
$R_{0.2}$	94.4	95.8	94.4	94.4	93.1	94.4	97.2	97.2
$R_{0.5}$	94.4	95.8	95.8	95.8	97.2	95.8	95.8	97.2
$R_{0.8}$	94.4	95.8	95.8	95.8	95.8	98.6	97.2	98.6
$R_{1.0}$	94.4	94.4	93.1	95.8	95.8	98.6	98.6	98.6
$R_{1.5}$	94.4	94.4	98.6	97.2	97.2	98.6	97.2	97.2
$R_{2.0}$	94.4	93.1	95.8	94.4	97.2	98.6	97.2	97.2
$R_{3.0}$	94.4	93.1	95.8	97.2	97.2	98.6	95.8	95.8
U	94.4	95.8	95.8	95.8	95.8	95.8	98.6	97.2

TABLE VI
PERFORMANCE ON LEUKEMIA DATASET USING K-NN RULE

f	Number of Selected Genes							
	2	5	8	10	15	20	25	30
$I_{0.2}$	91.7	94.4	95.8	94.4	97.2	97.2	97.2	97.2
$I_{0.5}$	94.4	94.4	95.8	95.8	95.8	95.8	97.2	97.2
$I_{0.8}$	94.4	94.4	94.4	94.4	97.2	97.2	97.2	97.2
$I_{1.0}$	94.4	94.4	94.4	94.4	97.2	97.2	97.2	95.8
$I_{1.5}$	93.1	95.8	94.4	95.8	97.2	95.8	97.2	97.2
$I_{2.0}$	93.1	93.1	95.8	95.8	97.2	98.6	97.2	97.2
$I_{3.0}$	93.1	93.1	95.8	94.4	97.2	97.2	97.2	97.2
$M_{0.2}$	93.1	94.4	95.8	95.8	95.8	95.8	97.2	97.2
$M_{0.5}$	94.4	94.4	95.8	95.8	95.8	95.8	97.2	97.2
$M_{0.8}$	90.3	95.8	94.4	95.8	97.2	97.2	98.6	98.6
$M_{1.0}$	88.9	94.4	98.6	97.2	97.2	97.2	97.2	98.6
$\lambda^{1.5}$	93.1	93.1	95.8	95.8	97.2	98.6	95.8	98.6
$\lambda^{2.0}$	93.1	93.1	95.8	95.8	97.2	98.6	97.2	97.2
$\lambda^{3.0}$	93.1	95.8	95.8	95.8	97.2	97.2	97.2	95.8
$R_{0.2}$	91.7	94.4	95.8	94.4	97.2	97.2	97.2	97.2
$R_{0.5}$	94.4	94.4	95.8	95.8	95.8	95.8	97.2	97.2
$R_{0.8}$	94.4	94.4	94.4	95.8	95.8	95.8	97.2	97.2
$R_{1.0}$	94.4	94.4	94.4	94.4	97.2	97.2	97.2	95.8
$R_{1.5}$	94.4	94.4	94.4	95.8	98.6	97.2	97.2	97.2
$R_{2.0}$	94.4	93.1	94.4	93.1	97.2	97.2	97.2	97.2
$R_{3.0}$	94.4	94.4	94.4	97.2	94.4	97.2	97.2	97.2
U	94.4	94.4	94.4	94.4	95.8	95.8	97.2	98.6

the NB, a classification accuracy of 100% is obtained for $I_{0.2}$- and $R_{0.2}$-information measures considering 15 or more genes in case of breast cancer data, eight or ten genes in case of leukemia data, and 25 or more genes in case of colon cancer data, while in case of $M_{0.2}$-information measure, 15 or more genes for breast cancer data, 8 or more genes for leukemia data, and 30 or more genes for colon cancer data are required to achieve this accuracy. Similarly, 100% accuracy for breast cancer data is obtained for $I_{0.8}$- and $R_{0.8}$-information measures using only five genes. However, both mutual information and normalized mutual information provide maximum 98.6% accuracy for leukemia data, 93.6% and 95.2% accuracy for colon cancer data, and 100% for breast cancer data using ten genes.

The results reported in Tables II, V, and VIII are based on the predictive accuracy of the SVM. The results show that in case of breast cancer dataset, the f-information measures, along with mutual information and normalized mutual information, achieve 100% classification accuracy. While at least eight genes are required for mutual information and normalized mutual information to attain this accuracy, $I_{0.2}$-, $I_{0.5}$-, $M_{0.5}$-, $R_{0.2}$-, and $R_{0.5}$-information measures need only five genes. On the other hand, both mutual information and normalized mutual information provide maximum 98.6% accuracy for leukemia data using 25 genes, while $I_{1.5}$-, $M_{0.8}$-, and V- (i.e., $M_{1.0}$-) information measures give 100% accuracy using only 15, 15, and 20 genes, respectively. Similarly, for colon cancer dataset, while mutual

TABLE VII
PERFORMANCE ON COLON CANCER DATASET USING NB CLASSIFIER

f	Number of Selected Genes							
	2	5	8	10	15	20	25	30
$I_{0.2}$	72.6	85.5	90.3	91.9	96.8	98.4	100	100
$I_{0.5}$	83.9	93.6	93.6	91.9	95.2	93.6	93.6	96.8
$I_{0.8}$	83.9	90.3	93.6	91.9	93.6	93.6	93.6	93.6
$I_{1.0}$	83.9	90.3	90.3	91.9	93.6	93.6	93.6	91.9
$I_{1.5}$	83.9	90.3	91.9	90.3	93.6	93.6	93.6	93.6
$I_{2.0}$	83.9	93.6	90.3	90.3	93.6	93.6	93.6	91.9
$I_{3.0}$	83.9	93.6	91.9	91.9	93.6	91.9	91.9	91.9
$M_{0.2}$	72.6	80.7	91.9	91.9	95.2	95.2	98.4	100
$M_{0.5}$	83.9	93.6	93.6	91.9	95.2	93.6	93.6	96.8
$M_{0.8}$	85.5	85.5	90.3	88.7	88.7	90.3	91.9	93.6
$M_{1.0}$	85.5	85.5	90.3	93.6	87.1	90.3	88.7	90.3
$\chi^{1.5}$	77.4	88.7	91.9	91.9	91.9	90.3	95.2	91.9
$\chi^{2.0}$	83.9	93.6	90.3	90.3	93.6	93.6	93.6	91.9
$\chi^{3.0}$	83.9	93.6	91.9	91.9	93.6	93.6	95.2	93.6
$R_{0.2}$	72.6	85.5	90.3	93.6	95.2	98.4	100	100
$R_{0.5}$	83.9	93.6	93.6	91.9	95.2	93.6	95.2	96.8
$R_{0.8}$	83.9	90.3	93.6	91.9	93.6	93.6	93.6	93.6
$R_{1.0}$	83.9	90.3	90.3	91.9	93.6	93.6	93.6	91.9
$R_{1.5}$	83.9	90.3	91.9	90.3	93.6	93.6	93.6	93.6
$R_{2.0}$	83.9	93.6	90.3	91.9	91.9	93.6	93.6	93.6
$R_{3.0}$	83.9	93.6	91.9	91.9	90.3	95.2	93.6	91.9
U	83.9	90.3	87.1	90.3	93.6	93.6	93.6	95.2

TABLE VIII
PERFORMANCE ON COLON CANCER DATASET USING SVM

f	Number of Selected Genes							
	2	5	8	10	15	20	25	30
$I_{0.2}$	83.9	80.7	80.7	82.3	80.7	80.7	80.7	80.7
$I_{0.5}$	83.9	79.0	87.1	85.5	87.1	80.7	82.3	80.7
$I_{0.8}$	83.9	83.9	85.5	83.9	80.7	80.7	80.7	79.0
$I_{1.0}$	83.9	83.9	83.9	88.7	80.7	82.3	75.8	79.0
$I_{1.5}$	83.9	80.7	87.1	87.1	88.7	88.7	83.9	79.0
$I_{2.0}$	83.9	87.1	87.1	87.1	90.3	87.1	82.3	80.7
$I_{3.0}$	83.9	87.1	88.7	85.5	88.7	80.7	80.7	75.8
$M_{0.2}$	83.9	75.8	77.4	87.1	82.3	80.7	80.7	77.4
$M_{0.5}$	83.9	79.0	87.1	85.5	87.1	80.7	82.3	80.7
$M_{0.8}$	79.0	83.9	87.1	82.3	82.3	88.7	79.0	75.8
$M_{1.0}$	79.0	83.9	87.1	85.5	91.9	83.9	82.3	75.8
$\chi^{1.5}$	77.4	91.9	85.5	85.5	90.3	83.9	82.3	79.0
$\chi^{2.0}$	83.9	87.1	87.1	87.1	90.3	87.1	82.3	80.7
$\chi^{3.0}$	83.9	87.1	85.5	88.7	85.5	82.3	87.1	75.8
$R_{0.2}$	83.9	83.9	74.2	79.0	82.3	80.7	80.7	75.8
$R_{0.5}$	83.9	79.0	87.1	85.5	87.1	80.7	83.9	80.7
$R_{0.8}$	83.9	83.9	85.5	83.9	80.7	80.7	80.7	79.0
$R_{1.0}$	83.9	83.9	83.9	88.7	80.7	82.3	75.8	79.0
$R_{1.5}$	83.9	80.7	87.1	87.1	88.7	88.7	83.9	79.0
$R_{2.0}$	83.9	87.1	87.1	88.7	91.9	87.1	80.7	77.4
$R_{3.0}$	83.9	87.1	88.7	88.7	85.5	83.9	83.9	72.6
U	83.9	83.9	77.4	85.5	83.9	82.3	82.3	85.5

information and normalized mutual information attain maximum 88.7% and 85.5% accuracy, respectively, $M_{1.0}$-, $\chi^{1.5}$-, and $R_{2.0}$-information measures provide maximum 91.9% accuracy, and both $I_{2.0}$- and $\chi^{2.0}$-information measures provide maximum 90.3% accuracy.

For breast cancer dataset using the K-NN, 100% accuracy is obtained in case of mutual information as well as I_α- ($\alpha = 1.5, 2.0, 3.0$), χ^α- ($\alpha = 2.0, 3.0$), R_α- ($\alpha = 1.5, 3.0$) information measures, although the normalized mutual information provides maximum 98.0% accuracy. For the K-NN, while mutual information and normalized mutual information achieve maximum 97.2% and 98.6% accuracy using at least 15 and 30 genes in case of leukemia dataset, the V- or $M_{1.0}$-information measure provides 98.6% accuracy using only eight genes. Similarly, the $\chi^{3.0}$-information measure achieves 90.3% predictive

TABLE IX
PERFORMANCE ON COLON CANCER DATASET USING K-NN RULE

f	Number of Selected Genes							
	2	5	8	10	15	20	25	30
$I_{0.2}$	82.3	82.3	77.4	74.2	79.0	82.3	75.8	75.8
$I_{0.5}$	83.9	77.4	75.8	79.0	83.9	87.1	88.7	85.5
$I_{0.8}$	83.9	74.2	85.5	85.5	85.5	85.5	87.1	87.1
$I_{1.0}$	83.9	74.2	82.3	88.7	87.1	87.1	87.1	87.1
$I_{1.5}$	83.9	87.1	85.5	85.5	88.7	88.7	88.7	88.7
$I_{2.0}$	83.9	85.5	85.5	85.5	88.7	88.7	88.7	88.7
$I_{3.0}$	83.9	85.5	87.1	85.5	87.1	88.7	87.1	87.1
$M_{0.2}$	82.3	80.7	79.0	77.4	79.0	80.7	74.2	75.8
$M_{0.5}$	83.9	77.4	75.8	79.0	83.9	87.1	88.7	85.5
$M_{0.8}$	83.9	83.9	85.5	85.5	83.9	88.7	82.3	87.1
$M_{1.0}$	83.9	83.9	88.7	87.1	87.1	87.1	85.5	83.9
$\chi^{1.5}$	79.0	88.7	88.7	85.5	87.1	85.5	83.9	85.5
$\chi^{2.0}$	83.9	85.5	85.5	85.5	88.7	88.7	88.7	88.7
$\chi^{3.0}$	83.9	85.5	85.5	88.7	90.3	88.7	88.7	88.7
$R_{0.2}$	82.3	75.8	82.3	82.3	74.2	79.0	75.8	75.8
$R_{0.5}$	83.9	77.4	75.8	79.0	83.9	83.9	88.7	87.1
$R_{0.8}$	83.9	74.2	85.5	85.5	85.5	85.5	87.1	87.1
$R_{1.0}$	83.9	74.2	82.3	88.7	87.1	87.1	87.1	87.1
$R_{1.5}$	83.9	87.1	85.5	85.5	88.7	88.7	88.7	88.7
$R_{2.0}$	83.9	85.5	85.5	88.7	87.1	88.7	87.1	87.1
$R_{3.0}$	83.9	85.5	87.1	88.7	87.1	88.7	87.1	87.1
U	83.9	74.2	82.3	83.9	85.5	87.1	85.5	87.1

accuracy for colon cancer dataset, while mutual information and normalized mutual information provide maximum 88.7% and 87.1% accuracy, respectively. However, in case of the K-NN-based results for both leukemia and colon cancer datasets, the majority of f-information measures produces results similar to those of mutual information and normalized mutual information.

From the results reported here, it is seen that, for a particular number of selected genes, the predictive accuracy for some f-information measures is higher as compared to that of mutual information and normalized mutual information, irrespective of the classification models and microarray datasets used. Also, the I_α-, M_α-, and R_α-information measures attain 100% prediction accuracy using the NB for $\alpha = 0.2$ in all three datasets. In all cases, the $M_{0.5}$- and $I_{0.5}$-information measures provide same results as well as the $\chi^{2.0}$- and $I_{2.0}$-information measures show exactly same performance as they are related by the following relations:

$$I_{0.5}(P||P_1 \times P_2) = 2M_{0.5}(P||P_1 \times P_2) \tag{15}$$

$$\chi^{2.0}(P||P_1 \times P_2) = 2I_{2.0}(P||P_1 \times P_2). \tag{16}$$

For colon cancer and leukemia datasets, top ranked genes selected by $I_{0.2}$-, $M_{0.2}$- and $R_{0.2}$-information measures are available.[1]

D. Analysis on Class Separability

In case of leukemia and colon cancer datasets, I_α-, M_α- and R_α-information measures provide significantly better results for $\alpha = 0.2$ as compared to mutual information and normalized mutual information. In order to analyze the results of these measures further, the class separability index is used next. The class separability index $\mathcal{S}$ [3] of a dataset is defined as $\mathcal{S} = \text{trace}(S_b^{-1} S_w)$, where S_w and S_b are the within and between

[1] http://www.isical.ac.in/~pmaji/generesults.html

TABLE X
CLASS SEPARABILITY ANALYSIS

Dataset	$I_{1.0}/R_{1.0}$	U	$I_{0.2}$	$R_{0.2}$	$M_{0.2}$
Breast cancer	3.30	3.31	3.29	3.31	3.30
Leukemia	2.00	2.19	1.91	1.88	1.87
Colon cancer	0.72	0.72	0.46	0.45	0.55

class scatter matrices, defined as follows:

$$S_w = \sum_{j=1}^{C} p_j E\{(X-\mu_j)(X-\mu_j)^T | c_j\} = \sum_{j=1}^{C} p_j \Sigma_j$$

$$S_b = \sum_{j=1}^{C} (\mu_j - M_0)(\mu_j - M_0)^T ; M_0 = E\{X\} = \sum_{j=1}^{C} p_j \mu_j$$

where C is the number of classes, p_j is *a priori* probability that a pattern belongs to class c_j, X is a feature vector, μ_j is the sample mean vector for the entire data points, Σ_j is the sample covariance matrix of class c_j, and $E\{\cdot\}$ is the expectation operator. A lower value of the separability criteria ensures that the classes are well separated by their scatter means. Table X shows that the class separability index $\mathcal{S}$ obtained using $I_{0.2}$-, $M_{0.2}$-, and $R_{0.2}$-information measures are better than those obtained using $I_{1.0}$ (mutual information) and U (normalized mutual information) for breast cancer, leukemia, and colon cancer datasets.

IV. CONCLUSION

This paper introduces different f-information measures in order to identify discriminative genes from high-dimensional gene expression data. It presents the results of selecting relevant and nonredundant genes from microarray data using different measures from information theory. The popular and extensively researched measure of mutual information is compared with V-, I_α-, χ^α-, and R_α-information measures. The maximum relevance-minimum redundancy framework is used here as the gene selection method for different f-information measures. The performance of different measures is evaluated by the predictive accuracy of NB classifier, K-NN rule, and SVM.

For all datasets, significantly better results are found for several measures as compared to mutual information. The results obtained on real datasets demonstrate that the proposed f-information measures can bring a remarkable improvement on gene selection problem, and therefore, the f-information measures can be a promising alternative to mutual information for gene selection. They are capable of identifying discriminative genes that may contribute to revealing underlying class structures, providing a useful tool for the exploratory analysis of biological data. In order to address the problem of multiplicity of marker genes, a detailed analysis of the biological relevance of the selected genes will be conducted in future. The gene interactions will be studied in detail to see whether the incorporation of gene interaction information can improve the diagnostic test.

ACKNOWLEDGMENT

The author would like to thank reviewers and J. Chatterjee of the National Institute for the Orthopaedically Handicapped, India, for providing valuable comments and criticisms.

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] J. G. Liao and K.-V. Chin, "Logistic regression for disease classification using microarray data: Model selection in a large p and small n case," *Bioinformatics*, vol. 23, no. 15, pp. 1945–1951, 2007.

[3] P. A. Devijver and J. Kittler, *Pattern Recognition: A Statistical Approach*. Englewood Cliffs, NJ: Prentice-Hall, 1982.

[4] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artif. Intell.*, vol. 97, no. 1/2, pp. 273–324, 1997.

[5] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," in *Proc. Comput. Syst. Bioinformatics*, Stanford, CA, 2003, pp. 523–528.

[6] J. Li, H. Su, H. Chen, and B. W. Futscher, "Optimal search-based gene subset selection for gene array cancer classification," *IEEE Trans. Inf. Technol. Biomed.*, vol. 11, no. 4, pp. 398–405, Jul. 2007.

[7] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.

[8] X. Liu, A. Krishnan, and A. Mondry, "An entropy based gene selection method for cancer classification using microarray data," *BMC Bioinformatics*, vol. 6, no. 76, pp. 1–14, 2005.

[9] R. Blanco, P. Larranaga, I. Inza, and B. Sierra, "Gene selection for cancer classification using wrapper approaches," *Int. J. Pattern Recog. Artif. Intell.*, vol. 18, no. 8, pp. 1373–1390, 2004.

[10] D. Jiang, C. Tang, and A. Zhang, "Cluster analysis for gene expression data: A survey," *IEEE Trans. Knowl. Data Eng.*, vol. 16, no. 11, pp. 1370–1386, Nov. 2004.

[11] L. J. Heyer, S. Kruglyak, and S. Yooseph, "Exploring expression data: Identification and analysis of coexpressed genes," *Genome Res.*, vol. 9, no. 11, pp. 1106–1115, 1999.

[12] I. Vajda, *Theory of Statistical Inference and Information*. Dordrecht, The Netherlands: Kluwer, 1989.

[13] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "f-information measures in medical image registration," *IEEE Trans. Med. Imag.*, vol. 23, no. 12, pp. 1508–1516, Dec. 2004.

[14] M. West, C. Blanchette, H. Dressman, E. Huang, S. Ishida, R. Spang, H. Zuzan, J. A. Olson, J. R. Marks, and J. R. Nevins, "Predicting the clinical status of human breast cancer by using gene expression profiles," *Proc. Nat. Acad. Sci. USA*, vol. 98, no. 20, pp. 11-462–11-467, 2001.

[15] U. Alon, N. Barkai, D. A. Notterman, K. Gish, S. Ybarra, D. Mack, and A. J. Levine, "Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays," *Proc. Nat. Acad. Sci.*, USA, vol. 96, no. 12, pp. 6745–6750, 1999.

[16] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification and Scene Analysis*. New York: Wiley, 1999.

[17] V. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.

Pradipta Maji received the B.Sc. (Hons.) degree in physics, in 1998, the M.Sc. degree in electronics science, in 2000, and the Ph.D. degree in computer science in 2005, all from Jadavpur University, Calcutta, India.

Currently, he is an Assistant Professor in the Machine Intelligence Unit, Indian Statistical Institute, Kolkata, India. He is also associated with the Center for Soft Computing Research, A National Facility, Indian Statistical Institute. His current research interests include pattern recognition, computational biology and bioinformatics, medical image processing, cellular automata, soft computing, and so forth. He is the author or coauthor of more than 50 papers published various in international journals and conferences. He is also a reviewer of many international journals.