Received: 19 September 2018 | Revised: 5 December 2018 | Accepted: 24 December 2018

DOI: 10.4218/etrij.2018-0522

ORIGINAL ARTICLE

WILEY ETRI Journal

An enhanced feature selection filter for classification of microarray cancer data

Dilwar Hussain Mazumder[1] | **Ramachandran Veilumuthu**[2]

[1]Department of Computer Science and Engineering, National Institute of Technology Nagaland, Chumukedima, Nagaland, India

[2]Department of Computer Science and Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, Tamil Nadu, India

Correspondence
Dilwar Hussain Mazumder, Department of Computer Science and Engineering, National Institute of Technology Nagaland, Chumukedima, Dimapur, Nagaland, India.
Email: dilwar2k4@yahoo.co.in

The main aim of this study is to select the optimal set of genes from microarray cancer datasets that contribute to the prediction of specific cancer types. This study proposes the enhancement of the feature selection filter algorithm based on Joe's normalized mutual information and its use for gene selection. The proposed algorithm is implemented and evaluated on seven benchmark microarray cancer datasets, namely, central nervous system, leukemia (binary), leukemia (3 class), leukemia (4 class), lymphoma, mixed lineage leukemia, and small round blue cell tumor, using five well-known classifiers, including the naive Bayes, radial basis function network, instance-based classifier, decision-based table, and decision tree. An average increase in the prediction accuracy of 5.1% is observed on all seven datasets averaged over all five classifiers. The average reduction in training time is 2.86 seconds. The performance of the proposed method is also compared with those of three other popular mutual information–based feature selection filters, namely, information gain, gain ratio, and symmetric uncertainty. The results are impressive when all five classifiers are used on all the datasets.

KEYWORDS
cancer classification, feature selection, gene expression, microarray, normalized mutual information

1 | INTRODUCTION

The advent of microarray technology has encompassed a new dimension to cancer research in recent years. Accordingly, microarray-based gene expression data analysis has emerged as a proficient method for the classification, diagnosis, and treatment of cancer. A microarray gene expression dataset consists of thousands of features called genes. Such a dataset includes records (samples or instances) from a few patients only. The availability of a limited number of samples in comparison to the large number of genes is often referred to as the "curse of dimensionality" problem. The existence of this problem in combination with irrelevant and noisy gene data hinders effective classification processes. These characteristics of microarray gene expression data lead to (a) an observable reduction in the classification accuracy and to (b) a significant increase in training time for any classification task applied on these data. Moreover, these challenging issues hinder the extraction of useful information from the dataset. Therefore, the reduction in the number of genes and the selection of highly informative genes become vital in microarray data analyses.

Feature selection, also called attribute selection or gene selection in the context of microarray data classification, aims to select a small subset of features from the huge feature space. Accordingly, by removing irrelevant and redundant

features, feature selection improves the classification accuracy and reduces the classification time [1]. Feature selection algorithms can be broadly classified into four types: filter, wrapper, embedded, and hybrid. In the filter approach [2], individual features are ranked, and a subset is selected without the use of a learning algorithm, whereas the wrapper approach [3] uses a learner to evaluate the feature subset to be selected. Filter methods are faster, while wrapper methods lead to higher classification accuracies for classifiers with higher computational costs. Conversely, the embedded approach [2] performs the feature selection process during the training phase of a specific classifier. In the hybrid approach [4], both filter and wrapper methods are combined to benefit from the best attributes of both methods.

In addition to the application of traditional feature selection methods, in recent years, many new feature selection algorithms have been proposed in the literature and have been applied to microarray cancer datasets for the selection of influential genes to enhance the classification performance. Many prominent algorithms follow the filter approach based on mutual information, whereas several other methods have been based on the hybrid approach with the use of filters and wrappers. Wang et al. [5] proposed a weighting-scheme–based feature selection algorithm known as maximum weight minimum redundancy, which achieved better classification performance on microarray datasets. Liu et al. [6] proposed a discrete biogeography-based optimization (BBO) supported technique to perform feature selection in molecular signatures. When tested on breast cancer datasets, their method produced convincing results. A constrained version of BBO was used by Samaneh et al. [7] for gene selection, and good results were achieved when it was applied to breast cancer datasets. Regarding gene selection from microarray data, a statistical dependence–based feature selection method was presented in [8] that also led to an outstanding performance when it was applied on binary datasets. Mayer et al. [9] introduced an information theoretic filter approach for feature selection in microarray data which selected genes by fine tuning the mutual information of a data subset. A feature selection method based on dependence maximization between selected features and class labels proposed in [10] also performed well when it was applied to microarray data. Li and Yin [11] proposed a binary encoding scheme to design multiobjective binary BBO for performing gene selection. Their experiments were conducted on 10 microarray datasets and demonstrated that the produced outcomes were better and comparable with those obtained from the use of particle swarm optimization (PSO) and support vector machines. A wrapper-based gene selection technique known as successive feature selection was developed in [12] that partitioned the features into blocks, and combined the best features from each block according to their classification performances to form the best feature subset. This method attained increased classification accuracy on a number of microarray datasets. Thawkar and Ingolikar [13] used BBO to select a subset of features for classification of masses in digital mammograms. Selected features were evaluated using an adaptive neuro-fuzzy system and ANN. Results from 651 mammograms showed that the BBO-based method had produced excellent outcomes in mammogram classification. Mohamad et al. [14] proposed an improved binary PSO to perform gene selection on 10 microarray datasets and achieved significant improvements in classification performances on most of the datasets. In general, the use of evolutionary algorithms are prominent in wrapper and hybrid feature selection methods, while many filter algorithms use mutual information for feature selection. This is because of the established fact that mutual information plays vital role in formulating the relationship between the features and the class labels in an efficient way and is easy to implement.

A filter-based feature selector evaluates feature goodness either individually or based on feature subsets. Individual feature ranking algorithms, such as Relief [15], rank the features based on their relevance to the target class. Feature subsets are evaluated based on consistency (Dash et al. [16]) and correlation measures (Hall [17]) to evaluate the goodness of feature subsets. Yu and Liu [18] proposed a fast correlation-based filter which performed feature selection by identifying relevancy and redundancy among features without the need to conduct pairwise correlation analysis. This method works in two phases. First, it selects relevant features based on symmetrical uncertainty (SU) [19] using a predefined threshold. From this selected list, redundant features are removed and only the predominant features are maintained. SU normalizes the mutual information values in the interval [0, 1]. The value of zero indicates that the two variables are independent, while the value of one indicates they are fully dependent. In the case of a perfect functional dependence between the two variables, SU does not necessarily take the value of one. This was observed by Joe [20] who defined a new version of mutual information. The new version also normalizes the values in the interval [0, 1] and ensures to take the value of one if and only if a perfect functional dependence exists between the two variables. This normalized mutual information has been utilized to propose an enhanced feature selection filter.

The theoretical background with respect to the mutual information-based feature selection filters are reviewed in Section 2 along with a discussion on other existing works related to the proposed method. The proposed algorithm is also presented in this section, and ranks features based on Joe's normalized mutual information (JNMIF). Section 3 outlines the details of the empirical study, including the experimental setup and descriptions of the used datasets. In Section 4, the performance of the proposed method is evaluated in terms of its classification accuracy and training time using five well-known classifiers applied on seven benchmark microarray

datasets. Additionally, in this section, a comparative study of the performance of the proposed method is conducted with those of three other popular mutual information-based feature selection methods. Section 5 outlines the conclusions and proposes a possible extension of the study reported herein.

2 | THEORETICAL BACKGROUND OF MUTUAL INFORMATION–BASED FEATURE SELECTION

The entropy of a random variable is a measure of its uncertainty. Higher entropy reflects a higher uncertainty of the value of that random variable and vice versa [21].

Definition 1 The entropy $H(X)$ of a random discrete variable X with a probability function $p(x)$ is defined as,

$$H(X) = -\sum_{x} p(x) \log p(x). \tag{1}$$

Definition 2 The joint entropy $H(X, Y)$ of a pair of discrete random variables X and Y with a joint distribution $p(x, y)$ is defined as follows,

$$H(X,Y) = -\sum_{x}\sum_{y} p(x,y) \log p(x,y). \tag{2}$$

Definition 3 The conditional entropy $H(Y|X)$ is defined as,

$$H(Y|X) = \sum_{x} p(x) H(Y|X = x). \tag{3}$$

The conditional entropy of Y on X refers to the average entropy of Y conditioned on the value of X averaged over all possible values of X.

Theorem 1 *The chain rule of joint entropy is defined as,*

$$H(X, Y) = H(X) + H(Y|X). \tag{4}$$

The chain rule for joint entropy states that the total uncertainty of the values of X and Y is equal to the uncertainty of X plus the average uncertainty of Y once X is known.

Definition 4 The mutual information $I(X, Y)$ between two random variables X and Y measures how much on average the realization of Y tells about the realization of X. Correspondingly, it is defined as,

$$I(X,Y) = \sum_{x}\sum_{y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}, \tag{5}$$

where $p(x, y)$ is the joint probability mass function of X and Y, and $p(x)$ and $p(y)$ are the marginal probability functions of X and Y, respectively. Equation (5) can also be written in terms of entropies according to (6), which states how much the entropy of X is reduced if the realization of Y is known.

$$I(X, Y) = H(X) - H(X|Y). \tag{6}$$

Mutual information is symmetric. In this sense, X provides the same exact information on Y as the information provided by Y on X.

Theorem 2 *The symmetry of mutual information is defined as,*

$$I(X, Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) = I(Y, X). \tag{7}$$

The mutual information can be expressed by applying the chain rule of Theorem 1 as follows,

$$\begin{aligned} I(X, Y) &= H(Y) - H(Y|X) = H(Y) - (H(X,Y) - H(X)) \\ &= H(X) + H(Y) - H(X, Y). \end{aligned} \tag{8}$$

Given that the mutual information ranking criteria can efficiently formulate the relationships between the features and the target class, and given that they are easy to implement, many of the filter algorithms reported in the literature have adopted these criteria.

2.1 | Prior studies

A normalized version of mutual information, known as asymmetric uncertainty coefficient $U(X, Y)$, formulates the relative decrease in uncertainty of X given Y, and is expressed as follows,

$$U(X, Y) = \frac{I(X,Y)}{H(X)}. \tag{9}$$

The SU coefficient [19] is a symmetric version of $U(X, Y)$, and is defined as follows,

$$SU(X, Y) = \frac{I(X, Y)}{\frac{1}{2}[H(X) + H(Y)]}. \tag{10}$$

SU(X, Y) normalizes the mutual information values within the interval [0, 1]. Accordingly, the value of zero indicates that X and Y are independent, while the value of one indicates that X and Y are fully dependent. However, when there is a perfect functional dependence between X and Y, it does not necessarily take the value of one. This observation was made by Joe [20] who defined another version of mutual information.

2.2 | Proposed work—Joe's normalized mutual information–based filter approach

The new version of normalized mutual information defined by Joe [20] is as follows,

$$\mathrm{JNMI}(X, Y) = \frac{I(X, Y)}{\min\left[H(X), H(Y)\right]}. \tag{11}$$

JNMI(X, Y) also normalizes the values in the interval [0, 1]. Moreover, JNMI(X, Y) is equal to one if and only if X and Y are functionally dependent [20]. This fact is exploited in the proposed method of feature selection. Based on the formulation of (11), the JNMIF algorithm is developed.

Algorithm: JNMIF
Input: Dataset **D** with n features $\mathbf{F} = \{\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3, \ldots, \mathbf{f}_n\}$ and class levels **C**
Output: FS, the set of selected features
Steps:
1. Choose threshold k
2. $\mathbf{FS} = \boldsymbol{\phi}$
3. for $i = 1$ to n do
4. Calculate JNMI($\mathbf{f}_i$, **C**) (using (11))
5. end for
6. count = 1
7. while count $\leq k$ do
8. Select the feature $\mathbf{f}_i$ with the maximum JNMI($\mathbf{f}_i$, **C**) value
9. $\mathbf{FS} = \mathbf{FS}\ \mathbf{U}\ \{\mathbf{f}_i\}$
10. $\mathbf{F} = \mathbf{F} - \{\mathbf{f}_i\}$
11. count = count + 1
12. end while
13. return **FS**

The algorithm identifies the JNMI($\mathbf{f}_i$, **C**) value between each feature $\mathbf{f}_i$ and class label **C**. In each iteration, the algorithm then picks up the feature with the highest JNMI($\mathbf{f}_i$, **C**) value, includes it in the selected set, and removes it from the original set, until the top k features have been selected.

3 | EXPERIMENTAL SETUP AND DATASETS

The proposed method (JNMIF) is implemented in the WEKA framework [22]. All experiments were performed within the WEKA framework on a standalone PC with Intel i3 CPU (two cores with two threads each) and a 3 GB RAM. To evaluate the performance of JNMIF, WEKA implementations of five popular classifiers, namely, naive Bayes (NB), RBF network (RBFN), instance-based classifier (IB1), decision table (DTB), and decision tree (J48), are chosen with their default parameter settings.

Performance matrices, such as the classification accuracy and classification time, are considered to evaluate the performance of JNMIF. Although the classification accuracy is a popular metric, in some situations, classifiers are biased and set implicit cutoffs so that the peak-point accuracy is highlighted. To avoid this, a more stable metric known as the area under the receiver operating characteristic (ROC) curve (AUC) is used to confirm the performance. The three performance matrices are briefly described below.

3.1 | Classification accuracy

Accuracy is expected to measure how well the test predicts the categories. It represents how many samples in total are correctly classified to their respective classes. Accuracy is generally expressed as a percentage and is calculated according to (12).

$$\text{Accuracy} = \frac{\text{TrueNegative} + \text{TruePositive}}{\text{TrueNegative} + \text{TruePositive} + \text{FalseNegative} + \text{FalsePositive}} \times 100\% \tag{12}$$

where *TrueNegative* is the count of the samples classified as negative class samples that actually belong to a negative class and *TruePositive* represents the count of the samples classified as positive class samples that actually belong to a positive class. *FalseNegative* indicates the total number of samples that actually belong to a positive class but are classified as negative, and *FalsePositive* is the number of samples that are erroneously classified as positive but they belong to the negative class.

3.2 | Classification Time

It is the total CPU time needed to build the classifier model with training and the time required for the subsequent prediction of the output of the test data.

3.3 | AUC

This is a scalar value between zero and one that summarizes the analysis of ROC. It is calculated according to (13).

$$\begin{aligned} \mathrm{AUC} &= \int_{\infty}^{-\infty} \mathrm{TPR}(T)(-\mathrm{FPR}'(T))\,\mathrm{d}T \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I(T' > T)\mathbf{f}_1(T')\mathbf{f}_0(T)\,\mathrm{d}T'\,\mathrm{d}T = P(X_1 > X_0) \end{aligned} \tag{13}$$

where T is the threshold in which the instance X is classified as "positive" if $X > T$, and "negative" otherwise. Additionally, X_1 is the score for a positive instance, and X_0 is the score for a negative instance. TPR = *TruePositive*/

(*TruePositive* + *FalseNegative*) and FPR = *FalsePositive*/(*TrueNegative* + *TruePositive*) are the true positive and the false positive rates, respectively. An AUC value close to one indicates a better performance for the method. Unlike accuracy, AUC does not depend on the cutoff chosen by the classifier or on the class distribution of the samples in the dataset. Hence, it is a more robust metric for performance evaluation.

To compare the performance of JNMIF, WEKA implementations of three other feature selection algorithms based on mutual information are used, namely, information gain (IG), gain ratio (GR), and SU, with their default parameter settings. As suggested by Li et al. [23], the top-ranked 150 features from the results obtained following the application of the feature selection algorithm from all the experimental datasets are used for classification. Owing to the class imbalance associated with microarray data, the classifiers are more biased toward the major classes. To tackle this issue, a 10-fold cross validation mechanism is used to record the classification accuracy, AUC, and the times before and after the application of the feature selection algorithms for all the classifiers and datasets used. This ensures that the classifier model is trained on most parts (90%) of the dataset. Hence, the training data have almost the same underlying distributions as the entire dataset. Moreover, to ensure the stability of the models, all the 10-fold cross-validation experiments with all the features and with all the feature selection methods are repeated five times independently on all the datasets with all the mentioned classifiers. The average accuracy, AUC, and times of these five independent runs are presented as the final results. The corresponding standard deviations are also recorded. To validate the improvements in terms of the performance matrices used, a one-tailed paired *t* test at a 5% level of significance is performed to assess significant differences in the values obtained from the five runs between any two methods under comparison, and for a particular performance metric.

Seven benchmark microarray gene expression datasets experimented by Zhu et al. [24] are used in the proposed JNMIF for analysis. A summary of these datasets is listed in Table 1. The central nervous system (CNS) dataset includes 60 samples. It consists of two classes: (a) the survivor class represents the patients who are alive after treatment (21 samples), and (b) the rest are represented by the failure class (39 samples). The leukemia (binary) dataset contains gene expression profiles for two classes of leukemia, namely, acute lymphoblastic leukemia (ALL) and acute myeloblastic leukemia (AML). This dataset is referred to as Leukemia_2c in later parts of this article. The part associated with ALL

TABLE 1 Summary of datasets

Sl. No.	Name	#Features	#Instances	#Classes	Class distribution	
					Class Labels	#Samples
1	CNS	7,129	60	2	Survivors	21
					Failures	39
2	Leukemia_2c	7,129	72	2	ALL	47
					AML	25
3	Leukemia_3c	7,129	72	3	B-cell	38
					T-cell	9
					AML	25
4	Leukemia_4c	7,129	72	4	B-cell	38
					T-cell	9
					BM	21
					PB	4
5	Lymphoma	4,026	66	3	DLBCL	46
					FL	9
					CLL	11
6	MLL	12,582	72	3	ALL	24
					MLL	20
					AML	28
7	SRBCT	2,308	83	4	EWS	29
					BL	11
					NBL	18
					RMS	25

consists of two types, namely, B- and T-cells, whereas the bone marrow (BM) and peripheral blood (PB) samples are two types in AML. Accordingly, this dataset has three-class (B cell, T cell, and AML) and four-class (B cell, T cell, AML-BM and AML-PB) versions which are referred in this article as Leukemia_3c and Leukemia_4c, respectively. All the three versions of the leukemia datasets contain 72 samples in total. Leukemia_2c includes 47 ALL and 25 AML samples. Leukemia_3c has 38 B-cell, 9T-cell, and 25 AML samples. In Leukemia_4c, there are 38 B-cell samples, 9T-cell, 21 BM, and 4 PB samples. The lymphoma dataset includes three classes of lymphoid malignancies with 66 samples. The three classes are diffuse large B-cell lymphoma (DLBCL), follicular lymphoma (FL), and chronic lymphocytic lymphoma (CLL), with 46, 9, and 11 samples, respectively. The mixed lineage leukemia (MLL) dataset also consists of 72 sample data from the three classes of ALL, MLL, and AML, with 24, 20, and 28 samples in each class, respectively. The small round blue cell tumor (SRBCT) dataset is consisted of small round blue cell tumors. It has four classes that include 83 samples. The Ewing's sarcoma (EWS) class has 29 samples and the Burkitt's lymphoma (BL), neuroblastoma (NBL), and rhabdomyosarcoma (RMS) classes have 11, 18, and 25 numbers of samples, respectively.

4 | RESULTS AND DISCUSSIONS

The classification accuracies, AUC, and times needed to build the model from all the five classifiers applied on the seven datasets before and after the application of JNMIF are shown in Tables 2 and 3, respectively. Tables 4 and 5 present the improvements in classification accuracies, and AUC and classification times, respectively. It can be observed from Tables 2 and 4 that the classification accuracies for all the five classifiers are significantly improved.

The CNS dataset has received an average classification accuracy of 61.82% and an average AUC of 0.597 with the use of all the features when classified with the NB classifier. The average classification time by the NB classifier with all features on this dataset is 0.47 seconds.

When JNMIF is applied, the average accuracy of the CNS dataset with the NB classifier increased to 74.08%, the AUC to 0.739, and the classification time reduced to 0.0003 seconds. The margins were 12.26%, 0.142, and 0.47 seconds, for accuracy, AUC, and time, respectively. These are well-defined positive margins and are proved to be statistically significant based on one-tailed paired *t* tests at the 5% level of significance. With the IB1 classifier, the CNS dataset has received a statistically significant improvement of 18.15% in terms of accuracy, and a value of 0.141 for AUC, while a reduction in the classification time of 0.09 seconds was observed. However, these outcomes were found to be insignificant according to the *t* test. Similarly, with the DTB and J48 classifiers, considerable improvements were observed for the CNS dataset. With the RBFN classifier, a negative margin of −6.27% in accuracy was achieved which proved to be a significant loss according to the *t* test outcome. However, in terms of the AUC, the margin was 0.008, and it was found to be statistically insignificant. Overall, the CNS dataset received an average increase of 6.67% in terms of accuracy, 0.071 in AUC, and a reduction of 2.86 seconds in the classification time after the application of JNMIF when averaged over all the five classifiers used. These improvements were found to be highly significant. With the use of all the features, the Leukemia_2c dataset has the highest average accuracy of 98.76% and an AUC value of 0.973 with the NB classifier, while the lowest accuracy of 83.48% was evoked with the DTB and J48 classifiers. The AUC was 0.829 when all features were used with the DTB classifier. It can be observed that when JNMIF is applied to the Leukemia_2c dataset, the highest margin of accuracy and an AUC of 9.26% and 0.099 were, respectively, achieved with the IB1 and DTB classifiers. These represent considerable and statistically significant improvements. Similarly, good improvements of 8.74% and 4.62% in accuracy, and 0.099 and 0.062 in AUC, were, respectively, obtained with the DTB and J48 classifiers. While RBFN has received a positive margin of 0.4% in accuracy, the NB classifier yielded a negative margin of −0.79% but both were found to be statistically insignificant. Interestingly, in terms of AUC, NB has received a positive margin of 0.004, although this was insignificant. By contrast, RBFN led to a statistically significant improvement of 0.028 in the value of AUC. When averaged over the five classifiers, the Leukemia_2c dataset gained 4.45% in terms of accuracy, 0.056 in AUC, and 1.17 seconds in classification time. These represent considerable and statistically significant improvements. If the results on the Leukemia_3c dataset are considered, it can be observed that with the use of all the classifiers, it achieved positive margins for both the accuracy and AUC when applied to JNMIF. With the IB1 classifier, the highest improvement was achieved with a 10.65% accuracy margin and a 0.137 gain in AUC. Similarly, RBFN also led to a considerable improvement of 8.73% in accuracy and an improvement in AUC of 0.008, which was also significant. In the case of the NB classifier, the accuracy improvement was 4.77% while the AUC gain was 0.038. Both of these changes were significant. Regarding the J48 classifier, a gain of 0.45% in accuracy was found to be statistically insignificant, while a gain of 0.025 in AUC was proved to be significant. The overall improvement in the classification accuracy in the Leukemia_3c dataset was 5.83% which was further confirmed by the overall gain of 0.058 in AUC. The average reduction in classification time when all the classifiers

TABLE 2 Classification performances (mean ± SD) for accuracies (in percentage) and AUC for the five classifiers with the use of all the features and after the application of the JNMIF

	Classifiers							
	NB				RBFN			
	ALL Features		JNMIF		ALL Features		JNMIF	
Datasets	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC
CNS	61.82 ± 4.37	0.597 ± 0.023	74.08 ± 0.91	0.739 ± 0.021	72.02 ± 3.21	0.673 ± 0.037	65.75 ± 2.31	0.665 ± 0.027
Leukemia_2c	98.76 ± 1.79	0.973 ± 0.019	97.97 ± 0.89	0.977 ± 0.015	97.57 ± 2.23	0.963 ± 0.021	97.97 ± 1.01	0.991 ± 0.007
Leukemia_3c	94.59 ± 2.31	0.951 ± 0.026	99.36 ± 0.73	0.989 ± 0.011	90.63 ± 5.13	0.893 ± 0.035	99.36 ± 0.61	0.981 ± 0.011
Leukemia_4c	87.65 ± 4.03	0.881 ± 0.032	96.58 ± 1.01	0.959 ± 0.016	89.24 ± 4.21	0.901 ± 0.026	96.58 ± 1.52	0.953 ± 0.021
Lymphoma	92.57 ± 3.76	0.919 ± 0.021	99.23 ± 0.67	0.981 ± 0.009	92.77 ± 4.79	0.898 ± 0.031	97.72 ± 0.98	0.979 ± 0.017
MLL	95.98 ± 2.97	0.971 ± 0.027	97.97 ± 0.97	0.986 ± 0.012	93.4 ± 3.94	0.947 ± 0.024	97.97 ± 1.12	0.973 ± 0.018
SRBCT	98.01 ± 1.81	0.961 ± 0.011	99.91 ± 0.08	0.995 ± 0.006	90.07 ± 5.26	0.889 ± 0.036	99.97 ± 0.03	0.989 ± 0.011
	Classifiers							
	IB1				DTB			
	ALL Features		JNMIF		ALL Features		JNMIF	
Datasets	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC
CNS	57.02 ± 5.31	0.602 ± 0.031	75.17 ± 1.09	0.743 ± 0.019	75.15 ± 2.17	0.739 ± 0.028	77.22 ± 0.79	0.793 ± 0.019
Leukemia_2c	87.85 ± 6.93	0.881 ± 0.022	97.11 ± 0.97	0.969 ± 0.023	83.48 ± 4.39	0.829 ± 0.021	92.22 ± 1.12	0.928 ± 0.011
Leukemia_3c	82.29 ± 5.32	0.819 ± 0.029	92.94 ± 1.79	0.956 ± 0.017	84.87 ± 3.17	0.851 ± 0.019	89.44 ± 0.98	0.853 ± 0.012
Leukemia_4c	83.68 ± 2.31	0.841 ± 0.036	90.17 ± 2.12	0.916 ± 0.026	77.93 ± 2.98	0.781 ± 0.032	83.88 ± 1.17	0.846 ± 0.017
Lymphoma	97.32 ± 1.73	0.959 ± 0.023	96.98 ± 1.13	0.953 ± 0.018	77.42 ± 3.12	0.781 ± 0.029	82.37 ± 0.97	0.819 ± 0.010
MLL	85.07 ± 4.06	0.863 ± 0.029	92.94 ± 1.89	0.941 ± 0.027	83.48 ± 2.87	0.807 ± 0.027	81.1 ± 0.89	0.798 ± 0.023
SRBCT	84.69 ± 3.91	0.853 ± 0.032	98.5 ± 0.87	0.993 ± 0.005	67.62 ± 5.17	0.669 ± 0.037	72.84 ± 2.21	0.697 ± 0.027
	Classifier							
	J48							
	ALL Features				JNMIF			
Datasets	Accuracy		AUC		Accuracy		AUC	
CNS	58.48 ± 5.37		0.573 ± 0.036		65.6 ± 2.32		0.598 ± 0.025	
Leukemia_2c	83.48 ± 3.39		0.841 ± 0.021		88.1 ± 1.59		0.903 ± 0.016	
Leukemia_3c	95.98 ± 1.92		0.946 ± 0.019		96.43 ± 0.92		0.971 ± 0.019	
Leukemia_4c	87.65 ± 2.01		0.869 ± 0.022		89.49 ± 1.21		0.889 ± 0.024	

(Continues)

TABLE 2 (Continued)

	Classifier			
	J48			
	ALL Features		JNMIF	
Datasets	Accuracy	AUC	Accuracy	AUC
Lymphoma	92.57 ± 1.98	0.912 ± 0.020	99.08 ± 0.67	0.987 ± 0.009
MLL	84.87 ± 4.24	0.821 ± 0.027	81.15 ± 2.09	0.829 ± 0.023
SRBCT	84.49 ± 3.98	0.851 ± 0.031	87.35 ± 1.51	0.883 ± 0.015

TABLE 3 Classification times (mean ± SD) (in seconds) for the five classifiers with the use of all the features and after the application of the JNMIF

	Classifiers									
	NB		RBFN		IB1		DTB		J48	
Datasets	ALL Features	JNMIF	ALL Features	JNMIF	ALL Features	JNMIF	ALL Features	JNMIF	ALL Features	JNMIF
CNS	0.47 ± 0.019	0.0003 ± 0.0001	1.73 ± 0.23	0.09 ± 0.022	0.09 ± 0.026	0.0005 ± 0.0002	11.09 ± 3.18	0.21 ± 0.007	1.26 ± 0.32	0.05 ± 0.002
Leukemia_2c	0.27 ± 0.026	0.12 ± 0.012	1.34 ± 0.16	0.23 ± 0.017	0.03 ± 0.009	0.0003 ± 0.0001	6.91 ± 1.89	0.55 ± 0.016	1.01 ± 0.13	0.13 ± 0.017
Leukemia_3c	0.25 ± 0.032	0.0002 ± 0.0001	1.32 ± 0.37	0.84 ± 0.12	0.12 ± 0.089	0.0007 ± 0.0003	12.62 ± 4.02	0.28 ± 0.011	1.08 ± 0.21	0.05 ± 0.009
Leukemia_4c	0.25 ± 0.071	0.09 ± 0.053	1.85 ± 0.52	0.06 ± 0.017	0.09 ± 0.015	0.0029 ± 0.018	15.56 ± 4.85	0.28 ± 0.008	1.25 ± 0.36	0.05 ± 0.011
Lymphoma	0.14 ± 0.093	0.0001 ± 0.0001	0.65 ± 0.09	0.07 ± 0.002	0.06 ± 0.022	0.0017 ± 0.002	5.75 ± 1.23	0.22 ± 0.006	0.36 ± 0.01	0.03 ± 0.001
MLL	0.41 ± 0.082	0.0014 ± 0.0002	2.87 ± 0.91	0.04 ± 0.016	0.03 ± 0.007	0.0052 ± 0.013	25.72 ± 5.12	0.28 ± 0.014	1.82 ± 0.42	0.04 ± 0.002
SRBCT	0.42 ± 0.057	0.0007 ± 0.0003	1.43 ± 0.41	0.03 ± 0.006	0.07 ± 0.011	0.0025 ± 0.009	5.28 ± 1.07	0.3 ± 0.009	0.73 ± 0.26	0.07 ± 0.012

were used on the Leukemia_3c dataset was 2.84 seconds. In the Leukemia_4c dataset, four among the five classifiers achieved considerable improvements in terms of accuracy, while a similar pattern of increased improvements was also observed for AUC. In the case of the J48 classifier, the improvements were not high but these were statistically significant. The average gain was 6.11% in accuracy, 0.058 in AUC, and 3.7 seconds in classification time when the outcomes from all five classifiers were averaged. Similarly, the overall performances on the lymphoma dataset for all the classifiers were found to be adequate. Although a negative margin of −0.034 in accuracy and a margin of −0.006 in AUC were observed with the IB1 classifier, both were proved to be statistically insignificant based on the one-tailed paired *t* test at the 5% level of significance. A mixed pattern of margins was observed in the case of the MLL dataset. Considerable improvements were also observed with the IB1 and RBFN classifiers in terms of accuracy and AUC, while acceptable and significant improvements were also observed with the NB classifier. The MLL dataset suffered with the use of the DTB and J48 classifiers, whereby a statistically significant loss in accuracy was observed. In terms of the AUC with DTB classifiers, a negative margin was obtained, while the use of the J48 classifier led to a small positive margin, but both were found to be statistically insignificant. In the case of the SRBCT dataset, considerable improvements were gained with the IB1, RBFN, and DTB classifiers both in terms of accuracy and AUC, whereas adequate and statistically significant gains were obtained with the NB and J48 classifiers. If the average improvements for all the five classifiers are considered, it can be observed that the SRBCT dataset can lead to the highest improvement of 6.74% in classification accuracy, and to the second highest improvement—following the CNS dataset—in terms of AUC with a value of 0.067.

If classifier-wise average performances over all the datasets are considered, it can be observed from the last rows of Tables 4 and 5 that IB1 is the highest performer both in terms accuracy and AUC with respective improvements of 9.41% and 0.093. NB was the next highest performed and yielded improvements of 5.10% and 0.053 on average over all the seven datasets in terms of accuracy and AUC after the application of the JNMIF. The performances of RBFN and DTB were comparable, while J48 led to the lowest improvement of 2.81% in accuracy and 0.035 in AUC. In terms of classification time, DTB yielded the highest improvement of 11.54 seconds, while IB1 had the lowest average margin of 0.07 seconds averaged over all the seven datasets after the application of JNMIF.

The average improvements in terms of classification accuracies, AUC, and classification time, obtained by the five classifiers averaged over the seven datasets were found to be 5.15%, 0.055, and 2.86 seconds, respectively. These

TABLE 4 Improvements in classification performances (mean ± SD) in terms of accuracies (in percentage) and AUC values for the five classifiers with the use of all the features and after the application of the JNMIF

Datasets	Classifiers											
	NB		RBFN		IB1		DTB		J48		Average	
	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC
CNS	12.26 ± 5.96	0.142 ± 0.036	−6.27 ± 4.78	−0.008 ± 0.001	18.15 ± 7.32	0.141 ± 0.052	2.07 ± 0.98	0.054 ± 0.015	7.12 ± 2.31	0.025 ± 0.008	6.67 ± 9.38	0.071 ± 0.068
Leukemia_2c	−0.79 ± 0.32	0.004 ± 0.001	0.4 ± 0.21	0.028 ± 0.008	9.26 ± 3.91	0.088 ± 0.027	8.74 ± 2.32	0.099 ± 0.036	4.62 ± 0.89	0.062 ± 0.024	4.45 ± 4.62	0.056 ± 0.04
Leukemia_3c	4.77 ± 1.21	0.038 ± 0.006	8.73 ± 5.89	0.088 ± 0.017	10.65 ± 5.16	0.137 ± 0.049	4.57 ± 1.57	0.002 ± 0.001	0.45 ± 0.23	0.025 ± 0.010	5.83 ± 3.98	0.058 ± 0.054
Leukemia_4c	8.93 ± 2.67	0.078 ± 0.012	7.34 ± 3.87	0.052 ± 0.010	6.49 ± 4.22	0.075 ± 0.029	5.95 ± 2.33	0.065 ± 0.027	1.84 ± 0.93	0.02 ± 0.009	6.11 ± 2.64	0.058 ± 0.024
Lymphoma	6.66 ± 3.58	0.062 ± 0.010	4.95 ± 2.97	0.081 ± 0.015	−0.34 ± 0.16	−0.006 ± 0.001	4.95 ± 1.98	0.038 ± 0.017	6.51 ± 2.11	0.075 ± 0.038	4.55 ± 2.85	0.05 ± 0.035
MLL	1.99 ± 0.89	0.015 ± 0.002	4.57 ± 1.98	0.026 ± 0.003	7.87 ± 2.19	0.078 ± 0.023	−2.38 ± 1.03	−0.009 ± 0.001	−3.72 ± 1.67	0.008 ± 0.002	1.67 ± 4.81	0.024 ± 0.033
SRBCT	1.90 ± 0.71	0.034 ± 0.007	9.90 ± 4.77	0.1 ± 0.032	13.81 ± 4.93	0.14 ± 0.043	5.22 ± 2.11	0.028 ± 0.009	2.86 ± 1.21	0.032 ± 0.011	6.74 ± 5.02	0.067 ± 0.051
Average	5.10 ± 4.53	0.053 ± 0.047	4.23 ± 5.59	0.052 ± 0.029	9.41 ± 5.82	0.093 ± 0.053	4.16 ± 3.49	0.04 ± 0.037	2.81 ± 3.76	0.035 ± 0.024	5.15 ± 1.79	0.055 ± 0.015

TABLE 5 Improvements in classification times (mean ± SD) (in seconds) for the five classifiers with the use of all the features and after the application of the JNMIF

	Classifiers					
Datasets	**NB**	**RBFN**	**IB1**	**DTB**	**J48**	**Average**
CNS	0.47 ± 0.197	1.64 ± 0.968	0.09 ± 0.017	10.88 ± 4.028	1.21 ± 0.898	2.86 ± 4.53
Leukemia 2c	0.15 ± 0.027	1.11 ± 0.763	0.03 ± 0.009	6.36 ± 2.971	0.88 ± 0.673	1.71 ± 2.64
Leukemia 3c	0.25 ± 0.058	0.48 ± 0.068	0.12 ± 0.018	12.34 ± 5.893	1.03 ± 0.989	2.84 ± 5.32
Leukemia 4c	0.16 ± 0.031	1.79 ± 0.884	0.09 ± 0.011	15.28 ± 9.013	1.2 ± 0.893	3.7 ± 6.51
Lymphoma	0.14 ± 0.027	0.58 ± 0.097	0.06 ± 0.009	5.53 ± 2.108	0.33 ± 0.071	1.33 ± 2.36
MLL	0.41 ± 0.039	2.83 ± 1.013	0.02 ± 0.003	25.44 ± 14.37	1.78 ± 0.959	6.1 ± 10.87
SRBCT	0.42 ± 0.035	1.40 ± 0.897	0.07 ± 0.005	4.98 ± 1.998	0.66 ± 0.058	1.51 ± 2.00
Average	0.29 ± 0.144	1.4 ± 0.802	0.07 ± 0.035	11.54 ± 7.24	1.01 ± 0.46	2.86 ± 1.668

outcomes are very impressive and statistically significant in the context of cancer prediction.

Table 6 lists the classification accuracies and AUC values of the five classifiers after the application of the feature selection methods IG, GR, and SU, on all the seven datasets. The corresponding values have already been presented in Table 2 after the application of JNMIF. The comparison of the performance margins (mean ± SD) of accuracy and AUC among JNMIF and the other three feature selection methods using the five classifiers according to the values listed in Tables 2 and 6 are computed and presented in Table 7. Statistical significance was assessed using one-tailed paired *t* tests. The symbols "↑," "~," and "↓," in Table 7 indicate statistically significant "better," "equally good," or "worse" performances for the proposed method in reference to the compared method according to the one-tailed paired *t* test at the 5% level of significance. The last row of Table 7 shows counts of wins/ties/losses for the proposed method. With the IB1 classifier, JNMIF wins with larger margins compared to GR. Moreover, a statistically significant number of wins are observed in the case of the proposed method with good margins compared to the three other methods in most of the cases. The only exception was the case of the RBFN classifier, whereby IG defeated JNMIF with a good accuracy margin. This loss was also validated by the statistically significant margin in AUC.

In a few cases, the performances of JNMIF led to small positive margins compared to other methods, while in a few other cases, the performances of JNMIF yielded small negative margins. These small differences were found to be statistically insignificant and were thus considered as ties in terms of performance. For the CNS dataset, the total number of wins/ties/losses of JNMIF compared to IG, GR, and SU, were, respectively, 3/1/1, 4/1/0, and 4/1/0, in terms of accuracy, and 4/0/1, 4/1/0, and 5/0/0, in terms of AUC. In the case of the Leukemia_2c dataset, JNMIF won compared to other methods with good margins in all the cases except the case of the NB classifier where the small positive margin in AUC was found to be a tie with that from IG. Among the wins, considerable margins were yielded compared to SU with the use of NB, compared to IG, and compared to GR and SU with the RBFN classifiers, both in accuracy and AUC. The Leukemia_3c dataset had three higher margin wins for JNMIF compared to IG, and four higher margin wins compared to SU. Two small positive and another two small negative margins were considered as ties in the comparison of JNMIF with the other methods based on this dataset. In the other cases considered in reference to the Leukemia_3c dataset, JNMIF won compared to the three methods with an adequate margin. Similar win/tie/loss patterns were observed in the case of JNMIF in the lymphoma and MLL dataset cases. In the case of the Leukemia_4c dataset, wins were observed for JNMIF compared to IG and SU with NB, RBFN, and DTB, while an increased margin of defeat was also obtained in the case of IG with the IB1 classifier. The total number of wins/ties/losses of JNMIF over IG, GR, and SU, on the Leukemia_4c dataset, were, respectively, 3/0/2, 4/1/0, and 3/2/0, in terms of accuracy, and 2/2/1, 5/0/0, and 2/3/0, in terms of the AUC.

An intense competition among the other methods was observed with the proposed method in the case of the SRBCT dataset with the NB, RBFN, and IB1 classifiers. Small positive and negative margins of accuracy and AUC values were observed with this dataset with these three classifiers, but all were found to be ties based on the *t* test. With the other two classifiers of DTB and J48, wins of JNMIF were observed compared to SU both in terms of accuracy as well as in terms of AUC values. Wins with good margins were achieved by JNMIF compared to IG with DTB, and compared to GR with both DTB and J48. A small negative margin in the AUC on the SRBCT dataset with the DTB classifier was found as a tie with IG.

Overall, from the last row of Table 7 it can be observed that the total number of wins/ties/losses of JNMIF over IG, GR,

TABLE 6 Classification accuracies and AUC (mean ± SD) of the five classifiers after application of the feature selection methods IG, GR, and SU

		Feature Selection Methods					
		IG		GR		SU	
Dataset	Classifier	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC
CNS	NB	71.01 ± 2.58	0.698 ± 0.029	71.36 ± 2.73	0.711 ± 0.029	70.87 ± 3.01	0.701 ± 0.033
	RBFN	67.68 ± 3.21	0.701 ± 0.030	66.36 ± 2.18	0.638 ± 0.033	65.21 ± 2.87	0.631 ± 0.029
	IB1	71.01 ± 2.11	0.697 ± 0.033	71.36 ± 2.33	0.721 ± 0.021	70.87 ± 2.96	0.689 ± 0.034
	DTB	77.68 ± 3.03	0.763 ± 0.023	74.7 ± 1.98	0.749 ± 0.019	74.21 ± 2.43	0.752 ± 0.026
	J48	61.01 ± 2.73	0.586 ± 0.037	63.03 ± 3.21	0.603 ± 0.032	62.54 ± 3.45	0.573 ± 0.037
Leukemia_2c	NB	96.29 ± 1.96	0.971 ± 0.019	95.25 ± 2.12	0.961 ± 0.011	94.76 ± 2.97	0.953 ± 0.021
	RBFN	94.90 ± 2.78	0.957 ± 0.026	93.86 ± 1.87	0.963 ± 0.022	94.76 ± 2.68	0.957 ± 0.016
	IB1	94.90 ± 1.98	0.951 ± 0.017	95.25 ± 2.43	0.944 ± 0.018	94.76 ± 1.98	0.937 ± 0.024
	DTB	89.35 ± 1.79	0.898 ± 0.031	91.08 ± 1.75	0.901 ± 0.029	90.59 ± 2.21	0.911 ± 0.029
	J48	82.4 ± 2.18	0.873 ± 0.029	85.53 ± 2.83	0.864 ± 0.019	83.65 ± 2.67	0.851 ± 0.031
Leukemia_3c	NB	96.29 ± 1.93	0.953 ± 0.018	96.64 ± 1.87	0.977 ± 0.015	96.15 ± 1.69	0.951 ± 0.018
	RBFN	96.29 ± 2.01	0.956 ± 0.021	96.64 ± 2.06	0.987 ± 0.021	96.15 ± 2.32	0.973 ± 0.013
	IB1	92.12 ± 2.93	0.921 ± 0.027	93.01 ± 2.49	0.937 ± 0.031	91.98 ± 2.95	0.912 ± 0.026
	DTB	81.01 ± 3.02	0.790 ± 0.029	86.92 ± 2.96	0.851 ± 0.029	85.04 ± 2.87	0.861 ± 0.030
	J48	93.51 ± 1.57	0.926 ± 0.015	93.86 ± 1.43	0.945 ± 0.017	93.37 ± 1.96	0.929 ± 0.019
Leukemia_4c	NB	92.12 ± 2.11	0.898 ± 0.026	93.86 ± 1.97	0.931 ± 0.026	93.37 ± 2.77	0.929 ± 0.023
	RBFN	93.51 ± 2.63	0.961 ± 0.013	93.86 ± 2.32	0.931 ± 0.019	93.37 ± 3.01	0.946 ± 0.018
	IB1	93.51 ± 3.13	0.921 ± 0.028	89.7 ± 3.68	0.903 ± 0.025	90.37 ± 3.33	0.912 ± 0.027
	DTB	76.85 ± 2.97	0.771 ± 0.033	81.36 ± 2.58	0.823 ± 0.029	78.09 ± 2.96	0.783 ± 0.029
	J48	91.45 ± 1.45	0.915 ± 0.027	86.92 ± 3.97	0.871 ± 0.033	90.37 ± 3.21	0.891 ± 0.024
Lymphoma	NB	97.68 ± 1.33	0.989 ± 0.012	96.51 ± 1.45	0.955 ± 0.027	96.02 ± 2.45	0.973 ± 0.019
	RBFN	96.16 ± 2.23	0.959 ± 0.024	95.01 ± 2.13	0.961 ± 0.023	96.02 ± 1.69	0.955 ± 0.026
	IB1	90.1 ± 2.98	0.912 ± 0.031	96.51 ± 2.89	0.967 ± 0.019	96.02 ± 2.87	0.959 ± 0.018
	DTB	76.47 ± 3.19	0.773 ± 0.019	81.36 ± 3.45	0.793 ± 0.028	79.36 ± 3.01	0.801 ± 0.029
	J48	96.16 ± 2.21	0.971 ± 0.011	96.51 ± 1.99	0.973 ± 0.014	96.02 ± 2.26	0.959 ± 0.017
MLL	NB	93.51 ± 2.58	0.941 ± 0.026	95.25 ± 2.64	0.947 ± 0.021	91.98 ± 2.43	0.936 ± 0.026
	RBFN	93.51 ± 2.35	0.945 ± 0.023	95.25 ± 2.24	0.949 ± 0.024	93.37 ± 1.89	0.937 ± 0.028
	IB1	93.51 ± 1.97	0.929 ± 0.029	92.47 ± 2.41	0.923 ± 0.033	94.76 ± 2.12	0.95 ± 0.019
	DTB	82.4 ± 2.93	0.831 ± 0.019	85.53 ± 3.19	0.843 ± 0.029	76.71 ± 3.33	0.773 ± 0.030
	J48	83.79 ± 3.01	0.835 ± 0.021	78.58 ± 3.33	0.773 ± 0.024	80.87 ± 2.58	0.824 ± 0.021
SRBCT	NB	99.37 ± 0.47	0.994 ± 0.009	99.97 ± 0.21	0.998 ± 0.008	99.01 ± 0.51	0.991 ± 0.011
	RBFN	99.01 ± 0.28	0.991 ± 0.018	99.97 ± 0.33	0.999 ± 0.005	99.37 ± 0.23	0.985 ± 0.016
	IB1	99.37 ± 0.38	0.989 ± 0.027	99.37 ± 0.46	0.983 ± 0.011	99.01 ± 0.76	0.987 ± 0.018
	DTB	71.17 ± 2.13	0.701 ± 0.031	70.32 ± 3.13	0.673 ± 0.034	67.42 ± 3.43	0.652 ± 0.033
	J48	83.22 ± 1.98	0.857 ± 0.025	84.78 ± 1.78	0.851 ± 0.028	83.08 ± 2.01	0.842 ± 0.024

and SU, on all the seven datasets were respectively equal to 24/7/4, 25/9/1, and 26/8/1, in terms of accuracy, and 23/9/3, 27/6/2, and 24/11/0, in terms of the AUC values. These analyses confirmed that the performance of the proposed JNMIF method was better or at least comparable with those of the other three methods.

5 | CONCLUSIONS

In this study, an enhanced feature selection filter based on Joe's normalized mutual information, named as JNMIF, was introduced. The proposed algorithm was implemented, and

TABLE 7 Performance margins (mean ± SD) of accuracy and AUC of JNMIF and other methods using the five classifiers according to the values listed in Tables 2 and 6. The statistical significance is tested using one-tailed paired *t* tests. The symbols "↑," "~," and "↓," indicate that the proposed method respectively achieves statistically significant "better," "equal," or "worse" performances than the compared method based on the one-tailed paired *t* test at the 5% level of significance. The last row shows a count of wins/ties/losses associated with the proposed method

		Feature selection methods					
		IG		GR		SU	
Dataset	Classifier	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC
CNS	NB	3.07 ± 1.69↑	0.041 ± 0.003↑	2.72 ± 1.23↑	0.028 ± 0.002↑	3.21 ± 1.73↑	0.038 ± 0.003↑
	RBFN	−1.93 ± 0.73↓	−0.036 ± 0.002↓	−0.61 ± 0.47~	0.027 ± 0.002↑	0.54 ± 0.43~	0.034 ± 0.002↑
	IB1	4.16 ± 2.11↑	0.046 ± 0.003↑	3.81 ± 1.89↑	0.022 ± 0.001↑	4.30 ± 2.23↑	0.054 ± 0.004↑
	DTB	−0.46 ± 0.18~	0.03 ± 0.001↑	2.52 ± 1.11↑	0.044 ± 0.003↑	3.01 ± 1.97↑	0.041 ± 0.001↑
	J48	4.59 ± 1.93↑	0.012 ± 0.001↑	2.57 ± 1.32↑	−0.005 ± 0.001~	3.06 ± 1.86↑	0.025 ± 0.001↑
Leukemia_2c	NB	1.68 ± 0.97↑	0.006 ± 0.001~	2.72 ± 1.45↑	0.016 ± 0.001↑	3.21 ± 2.01↑	0.024 ± 0.001↑
	RBFN	3.07 ± 1.21↑	0.034 ± 0.003↑	4.11 ± 2.79↑	0.028 ± 0.002↑	3.21 ± 1.69↑	0.034 ± 0.003↑
	IB1	2.21 ± 1.01↑	0.018 ± 0.001↑	1.86 ± 0.98↑	0.025 ± 0.002↑	2.35 ± 1.26↑	0.032 ± 0.002↑
	DTB	2.87 ± 1.32↑	0.03 ± 0.002↑	1.14 ± 0.75↑	0.027 ± 0.002↑	1.63 ± 0.97↑	0.017 ± 0.001↑
	J48	5.7 ± 2.86↑	0.03 ± 0.001↑	2.57 ± 1.19↑	0.039 ± 0.003↑	4.45 ± 2.13↑	0.052 ± 0.004↑
Leukemia_3c	NB	3.07 ± 1.36↑	0.036 ± 0.002↑	2.72 ± 1.76↑	0.012 ± 0.001↑	3.21 ± 1.16↑	0.038 ± 0.003↑
	RBFN	3.07 ± 2.01↑	0.025 ± 0.001↑	2.72 ± 2.01↑	−0.006 ± 0.001~	3.21 ± 1.52↑	0.008 ± 0.001~
	IB1	0.82 ± 0.37~	0.035 ± 0.002↑	−0.07 ± 0.03~	0.019 ± 0.001↑	0.96 ± 0.63↑	0.044 ± 0.003↑
	DTB	8.43 ± 3.87↑	0.063 ± 0.005↑	2.52 ± 0.96↑	0.002 ± 0.001~	4.4 ± 2.19↑	−0.008 ± 0.001~
	J48	2.92 ± 1.12↑	0.045 ± 0.003↑	2.57 ± 1.35↑	0.026 ± 0.002↑	3.06 ± 1.86↑	0.042 ± 0.003↑
Leukemia_4c	NB	4.46 ± 2.31↑	0.061 ± 0.005↑	2.72 ± 1.79↑	0.028 ± 0.002↑	3.21 ± 1.62↑	0.03 ± 0.002↑
	RBFN	3.07 ± 1.15↑	−0.008 ± 0.001~	2.72 ± 1.11↑	0.022 ± 0.001↑	3.21 ± 1.37↑	0.007 ± 0.001~
	IB1	−3.34 ± 1.01↓	−0.005 ± 0.001~	0.47 ± 0.27~	0.013 ± 0.001↑	−0.20 ± 0.18~	0.004 ± 0.003
	DTB	7.03 ± 2.83↑	0.075 ± 0.006↑	2.52 ± 1.37↑	0.023 ± 0.001↑	5.79 ± 3.61↑	0.063 ± 0.005↑
	J48	−1.96 ± 0.58↓	−0.026 ± 0.001↓	2.57 ± 1.19↑	0.018 ± 0.001↑	−0.88 ± 0.59~	−0.002 ± 0.001~
Lymphoma	NB	1.55 ± 0.61↑	−0.008 ± 0.001~	2.72 ± 2.01↑	0.026 ± 0.002↑	3.21 ± 1.73↑	0.008 ± 0.001~
	RBFN	1.56 ± 0.51↑	0.02 ± 0.001↑	2.72 ± 1.98↑	0.018 ± 0.001↑	1.7 ± 0.91↑	0.024 ± 0.001↑
	IB1	6.88 ± 2.67↑	0.041 ± 0.003↑	0.47 ± 0.29~	−0.014 ± 0.001↓	0.96 ± 0.68~	−0.006 ± 0.001~
	DTB	5.9 ± 2.37↑	0.046 ± 0.004↑	1.01 ± 0.83~	0.026 ± 0.002↑	3.01 ± 1.58↑	0.018 ± 0.001↑
	J48	2.92 ± 1.17↑	0.016 ± 0.001↑	2.57 ± 1.41↑	0.014 ± 0.001↑	3.06 ± 1.32↑	0.028 ± 0.002↑
MLL	NB	4.46 ± 2.08↑	0.045 ± 0.003↑	2.72 ± 1.67↑	0.039 ± 0.003↑	5.99 ± 3.21↑	0.05 ± 0.004↑
	RBFN	4.46 ± 2.31↑	0.028 ± 0.001↑	2.72 ± 1.29↑	0.024 ± 0.001↑	4.6 ± 2.43↑	0.036 ± 0.003↑
	IB1	−0.57 ± 0.23~	0.012 ± 0.001↑	0.47 ± 0.31~	0.018 ± 0.001↑	−1.82 ± 0.91↓	−0.009 ± 0.001~
	DTB	−1.3 ± 0.39~	−0.033 ± 0.002↓	−4.43 ± 2.02↓	−0.045 ± 0.003↓	4.39 ± 2.17↑	0.025 ± 0.002↑
	J48	−2.64 ± 0.95↓	−0.006 ± 0.001~	2.57 ± 1.61↑	0.056 ± 0.004↑	0.28 ± 0.13~	0.005 ± 0.001~
SRBCT	NB	0.54 ± 0.35~	0.001 ± 0.001~	−0.06 ± 0.03~	−0.003 ± 0.001~	0.90 ± 0.84~	0.004 ± 0.001~
	RBFN	0.96 ± 0.27~	−0.002 ± 0.001~	0.01 ± 0.01~	−0.01 ± 0.001~	0.60 ± 0.46~	0.004 ± 0.001~
	IB1	−0.36 ± 0.22~	0.004 ± 0.001~	−0.36 ± 0.12~	0.01 ± 0.001~	0.01 ± 0.01~	0.006 ± 0.001~
	DTB	1.67 ± 0.98↑	−0.004 ± 0.001~	2.52 ± 1.43↑	0.024 ± 0.001↑	5.42 ± 3.06↑	0.045 ± 0.004↑
	J48	4.13 ± 2.31↑	0.026 ± 0.001↑	2.57 ± 1.55↑	0.032 ± 0.002↑	4.27 ± 2.69↑	0.041 ± 0.003↑
#Win/ties/losses		24/7/4	23/9/3	25/9/1	27/6/2	26/8/1	24/11/0

its performance was evaluated in terms of classification accuracy, AUC, and time, based on extensive experiments using five classification algorithms that included NB, RBFN, IB1, DTB, and J48, applied on seven high-dimensional benchmark microarray gene expression datasets. The classification accuracies and AUC values of these classifiers were compared after the application of JNMIF with three other mutual information–based feature selection techniques, namely, IG, GR,

and SU. The proposed method demonstrated excellent improvement in classification accuracy and AUC values with reductions of classification time in all the five classifier cases on all the seven datasets used. Furthermore, the method generated compatible results to those produced by the other three feature selection methods. The overall performance of the proposed method was shown to be impressive when all the five classifiers were used on all the seven datasets. The use of the proposed feature selection method may be extended by taking the feature-feature redundancy into consideration.

ORCID

Dilwar Hussain Mazumder https://orcid.org/0000-0002-6316-2977

REFERENCES

1. M. Dash and H. Liu, *Feature selection for classifications*, Intell. Data Anal. **1** (1997), 131–156.
2. I. Guyon and A. Elisseeff, *An introduction to variable and feature selection*, J. Mach. Learn. Res. **3** (2003), 1157–1182.
3. A. L. Blum and P. Langley, *Selection of relevant features and examples in machine learning*, Artif. Intell. **97** (1997), 245–271.
4. H. H. Hsu, C. W. Hsieh and M. D. Lu, *Hybrid feature selection by combining filters and wrappers*, Expert Syst. Appl. **38** (2011), 8144–8150.
5. J. Wang et al., *Maximum weight and minimum redundancy: a novel framework for feature subset selection*, Pattern Recognit. **46** (2013), 1616–1627.
6. B. Liu et al., *Discrete biogeography based optimization for feature selection in molecular signatures*, Mol. Inf. **34** (2015), 197–215.
7. Y. Samaneh, J. Shanbehzadeh, and E. Aminian, *Feature subset selection using constrained binary/integer biogeography based optimization*, ISA Trans. **52** (2013), 383–390.
8. V. Bolón-Canedo et al., Statistical dependence measure for feature selection in microarray datasets, in Proc. Eur. Symp. Artif. Neural Netw. -ESANN, Bruges, Belgium, Apr. 27–29, 2011, pp. 23–28.
9. P. Meyer, C. Schretter, and G. Bontempi, *Information-theoretic feature selection in microarray data using variable complementarity*, IEEE J. Sel. Top. Signal Process. **2** (2008), 261–274.
10. L. Song et al., *Feature selection via dependence maximization*, J. Mach. Learn. Res. **13** (2012), 1393–1434.
11. X. Li and M. Yin, *Multi-objective binary biogeography based optimization for feature selection using gene expression data*, IEEE Trans. Nano Biosci. **12** (2013), 343–353.
12. A. Sharma, S. Imoto, and S. Miyano, *A top-r feature selection algorithm for microarray gene expression data*, IEEE/ACM Trans. Comput. Biol. Bioinform. (TCBB) **9** (2012), 754–764.
13. S. Thawkar and R. Ingolikar, *Classification of masses in digital mammograms using Biogeography-based optimization technique*, J. King Saud Univ. Comp. Inf. Sci. (2018), https://doi.org/10.1016/j.jksuci.2018.01.004.
14. M. S. Mohamad et al., *A modified binary particle swarm optimization for selecting the small subset of in-formative genes from gene expression data*, IEEE Trans. Inf. Technol. Biomed. **15** (2011), 813–822.
15. K. Kira and L. Rendell, *The feature selection problem: Traditional methods and a new algorithm*, in Proc. Tenth Natl Conf, Artif. Intell., AAAI Press/The MIT Press, Menlo Park, 1992, pp. 129–134.
16. M. Dash, H. Liu, and H. Motoda, *Consistency based feature selection*, in Proc. Fourth Pacific Asia Conf. Knowl. Discov. Data Min., Springer-Verlag, 2000, pp. 98–109.
17. M. Hall, *Correlation based feature selection for machine learning*, Ph.D. Thesis, Univ. Waikato, Dept. Comp. Sci. (1999).
18. L. Yu and H. Liu, *Feature selection for high-dimensional data: a fast correlation-based filter solution*, in Proc. Twentieth Int. Conf. Mach. Learning ICML, Washington, DC, USA, Aug. 21–24, 2003, pp. 856–863.
19. C. E. Sarndal, *A comparative study of association measures*, Psychometrika **39** (1974), 165–187.
20. H. Joe, *Relative entropy measures of multivariate dependence*, J. Am. Stat. Assoc. **84** (1989), 157–164.
21. C. A. Shannon, *A mathematical theory of communication*, Bell Syst. Tech. J. **27** (1948), 379–423.
22. I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools with Java Implementations*, Morgan Kaufmann, San Francisco, CA, 2000.
23. T. Li, C. Zhang, and M. Ogihara, *A comparative study of feature selection and multiclass classification methods for tissue classification based on gene expression*, Bioinformatics **20** (2004), 2429–2437.
24. Z. Zhu, Y. S. Ong, and M. Dash, *Markov blanket-embedded genetic algorithm for gene selection*, Pattern Recognit. **49** (2007), 3236–3248.

AUTHOR BIOGRAPHIES

Dilwar Hussain Mazumder received his BE and MTech degrees in computer science and engineering from Jorhat Engineering College, Assam, India and Rajiv Gandhi University, Arunachal Pradesh, India, in 2008 and 2012, respectively. He is pursuing his PhD degree in computer science and engineering at National Institute of Technology Nagaland, India. He is currently an assistant professor at the Department of Computer Science and Engineering, National Institute of Technology Nagaland, India. His research interests include computational methods for gene selection in cancer prediction, such as biogeography-based optimizers, particle swarm optimizers, hybrid approaches, genetic algorithms, support vector machines, and neural networks.

Ramachandran Veilumuthu received his MS and PhD degrees in electrical engineering from College of Engineering Guindy, Anna University, Chennai, India. He worked in various capacities in the Department of Information Science and Technology, College of Engineering, Anna University, Chennai. He has 35 years of teaching experience and he is currently working as the Vice Chancellor of the Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, Tamil Nadu, India. His research interests include cloud computing, web technologies, and the internet-of-things.