

## Research Article

# Robust Feature Selection from Microarray Data Based on Cooperative Game Theory and Qualitative Mutual Information

Atiyeh Mortazavi<sup>1</sup> and Mohammad Hossein Moattar<sup>2</sup>

<sup>1</sup>Department of Computer Engineering, Imam Reza International University, Mashhad, Iran

<sup>2</sup>Department of Software Engineering, Islamic Azad University, Mashhad Branch, Mashhad, Iran

Correspondence should be addressed to Mohammad Hossein Moattar; moattar@mshdiau.ac.ir

Received 28 November 2015; Revised 20 February 2016; Accepted 22 February 2016

Academic Editor: Pietro H. Guzzi

Copyright © 2016 A. Mortazavi and M. H. Moattar. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

High dimensionality of microarray data sets may lead to low efficiency and overfitting. In this paper, a multiphase cooperative game theoretic feature selection approach is proposed for microarray data classification. In the first phase, due to high dimension of microarray data sets, the features are reduced using one of the two filter-based feature selection methods, namely, mutual information and Fisher ratio. In the second phase, Shapley index is used to evaluate the power of each feature. The main innovation of the proposed approach is to employ Qualitative Mutual Information (QMI) for this purpose. The idea of Qualitative Mutual Information causes the selected features to have more stability and this stability helps to deal with the problem of data imbalance and scarcity. In the third phase, a forward selection scheme is applied which uses a scoring function to weight each feature. The performance of the proposed method is compared with other popular feature selection algorithms such as Fisher ratio, minimum redundancy maximum relevance, and previous works on cooperative game based feature selection. The average classification accuracy on eleven microarray data sets shows that the proposed method improves both average accuracy and average stability compared to other approaches.

## 1. Introduction

In feature selection, the features of the data sets are selected which are effective for predicting the target class. By eliminating additional features from data sets, the efficiency of the learning models dramatically increases. Genetic data sets have high dimensions and small size and are usually imbalanced. Increasing the high dimensions leads to classification complexity and it can reduce the classification accuracy. The small size of the data sets is another challenge [1]. Robustness issue is often ignored in feature selection. Increasing and decreasing the training samples in a nonrobust feature selection algorithm will lead to different results [2].

Feature selection methods are classified as filter methods, wrapper methods, and embedded methods [3]. Filtering methods are independent from learning algorithms. They are statistical tests which rely on the basic features of the training data and have much lower computational complexity compared with wrapper methods [4]. These methods

use some measurements. Among these measurements are distance measurements [5, 6], rough set theory [7], and information theoretic measures [8]. A common distance measure is Euclidean distance, which was applied in Relief method by Kira and Rendell [5], which uses Euclidean distance to assign weights to each feature. Then, [6] developed Relief algorithm which adds the ability of dealing with multiclass problems. Peng et al. [9] proposed minimum redundancy maximum relevance (mRMR) approach, which selects features that have the highest relevance with the target class and are minimally redundant. Wang et al. [10] proposed a new filtering algorithm based on the maximum weight minimum redundancy (MWMR) criterion. The weight of the feature shows its importance and redundancy represents the correlation between the features. Nguyen et al. [11] proposed a hierarchical approach called Modified Analytic Hierarchy Process (MAHP) which uses five individual gene ranking methods including  $t$ -test, entropy, Receiver Operating

Characteristics (ROC) curve, Wilcoxon test, and Signal to Noise Ratio (SNR) for feature ranking and selection.

In the wrapper methods, each subset is evaluated using a specific learning algorithm, and optimal features are selected to improve classification performance [3]. For example, Inza et al. [12] proposed classical wrapper algorithms for this purpose (i.e., sequential forward and backward selection, floating selection, and best-first search) and evaluated them on three microarray data sets. Another wrapper approach for gene selection from microarray data is proposed in [13] using modified ant colony optimization. Another heuristic based wrapper approach is proposed in [14] which is based on Genetic Bee Colony (GBC) optimization. More referencing materials on wrapper methods can be found in [15–18].

Similar to wrapper methods, embedded feature selection methods are classifier dependent but this relationship is stronger in embedded approaches. Guyon et al. [19] proposed Support Vector Machine (SVM) based on Recursive Feature Elimination (SVM-RFE) for feature selection in cancer classification. Canul-Reich et al. [20] introduced and applied Iterative Feature Perturbation (IFP) method, as an embedded gene selector, on four microarray data sets.

Robust feature selection algorithms are another source of researches. Yang and Mao [2] tried to improve the robustness of feature selection algorithms, with an ensemble method called Multicriterion Fusion-Based Recursive Feature Elimination (MCF-RFE). Also, Yassi and Moattar [1] presented robust and stable feature selection by integrating ranking methods and wrapper technique for genetic data classification.

The authors of [3, 21] tried to overcome the weaknesses of feature selection methods using cooperative game. They introduced a framework based on cooperative game theory to evaluate the power of each feature. To evaluate the weight of each of the features, the Banzhaf power index and the Shapley value index are used. Paper [3] used the Banzhaf power index in game theory for evaluating the weight of each feature, while paper [21] used the Shapley value index in game theory for evaluating the weight of each feature. Reference [22] presented a novel feature selection approach called Neighborhood Entropy-Based Cooperative Game Theory (NECGT) which was based on information theoretic approaches. The results of the evaluation of some UCI data sets showed that the approach yields better performance compared to classical methods in terms of accuracy.

Stability and robustness are important specially when the data is scarce and classes are imbalanced. In this paper, cooperative game theory is used for robust feature selection. In the proposed method, a cooperative game theory framework based on Shapley value is introduced for evaluating the power of each feature. To score the features in Shapley value index, we propose Qualitative Mutual Information criterion to achieve more stable results even in the presence of class imbalance and data scarcity. This criterion improves the robustness of feature selection algorithm. In this criterion, we use the Fisher ratio as the utility function in calculating Qualitative Mutual Information and this criterion leads to better robustness of the feature selection algorithm. The rest of this paper is organized as follows. Section 2 introduces

the fundamental materials and methods. In Section 3, our proposed method is described. In Section 4, the simulation results and evaluation of the proposed method are discussed. Finally, conclusions and future work are described in Section 5.

## 2. Mutual Information, Qualitative Mutual Information, and Conditional Mutual Information

The entropy,  $H(X)$ , of discrete random variable  $X$  with  $p(x) = \Pr(X = x)$  as its probability density function is defined as follows:

$$H(X) = - \sum_{x \in X} \Pr(X) \log \Pr(X). \quad (1)$$

Also, the mutual information (MI) between two random variables  $X$  and  $Y$  with a joint probability distribution  $\Pr(x, y)$  is formulated as follows [22]:

$$\begin{aligned} \text{MI}(X; Y) &= \sum_{Y_j \in Y} \sum_{X_i \in X} \Pr(X_i, Y_j) \log \left( \frac{\Pr(X_i, Y_j)}{\Pr(X_i) \Pr(Y_j)} \right), \quad (2) \\ &0 \leq \text{MI}(X; Y) \leq 1. \end{aligned}$$

Hence, Qualitative Mutual Information (QMI) is defined by multiplying a utility function  $U(X_i, Y_j)$  to mutual information formula; the formula is as follows [23, 24]:

$$\begin{aligned} \text{QMI}(X; Y) &= \sum_{Y_j \in Y} \sum_{X_i \in X} U(X_i, Y_j) \Pr(X_i, Y_j) \\ &\cdot \log \left( \frac{\Pr(X_i, Y_j)}{\Pr(X_i) \Pr(Y_j)} \right). \quad (3) \end{aligned}$$

Different informative functions can be applied as the utility function. In the proposed approach, we use the Fisher ratio [2]. Also, Conditional Mutual Information (CMI) noted as  $\text{CMI}(X; Y | Z)$  denotes the amount of information between  $X$  and  $Y$ , when  $Z$  is given and is formulated as follows [21]:

$$\begin{aligned} \text{CMI}(X; Y | Z) &= \sum_{x \in S_x} \sum_{y \in S_y} \sum_{z \in S_z} \Pr(X, Y, Z) \\ &\cdot \log \left( \frac{\Pr(X, Y | Z)}{\Pr(X | Z) \Pr(Y | Z)} \right). \quad (4) \end{aligned}$$

## 3. Proposed Approach

The proposed method consists of three phases, which are described as follows. Figure 1 depicts the flowchart of the proposed feature selection algorithm.

*3.1. Filter Approaches for Dimension Reduction.* Due to high dimension of features of microarray data, in the first phase of the proposed method, the features are reduced by using

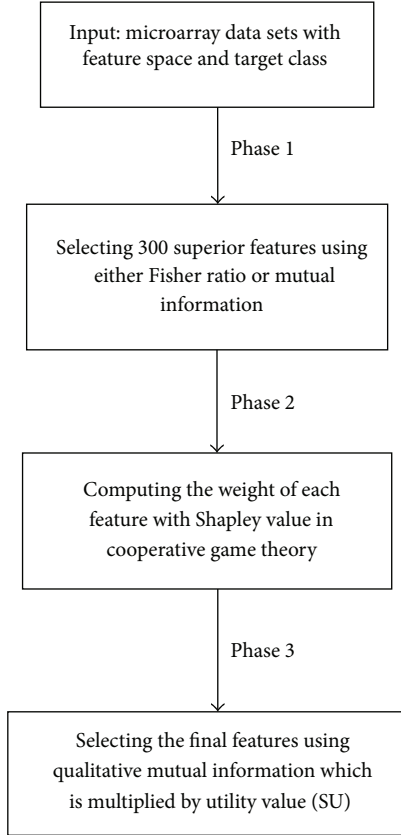


FIGURE 1: Flowchart of the proposed feature selection algorithm.

one of the two filter-based feature selection methods, namely, mutual information and Fisher ratio. Fisher's ratio is an individual feature evaluation criterion. Fisher's ratio measures the discriminative power of a feature  $j$  by the ratio of interclass to intraclass variability. This relationship is as follows [2]:

$$FR(j) = \frac{(\hat{\mu}_{j1} - \hat{\mu}_{j2})^2}{\sigma_{j1}^2 - \sigma_{j2}^2}, \quad (5)$$

where  $\hat{\mu}_{jc}$  is the sample mean of feature  $j$  in class  $c$  and  $\sigma_{jc}^2$  is variance of feature  $j$  in  $c$ .

**3.2. Feature Evaluation.** In large data sets, there are intrinsic correlations among features. However, most of the filter-based feature selection algorithms discard redundant features that are highly correlated with the selected ones [21]. Each feature is weighted using information theory measures such as CMI and QMI [21].

Theoretically, the more relevancy of a feature means that it shares more information with the target class [21, 25]. Feature  $f_i$  is to be redundant with feature  $f_j$  if the following form is accepted [21]:

$$CMI(f_j; \text{class} | f_i) \leq QMI(f_j; \text{class}). \quad (6)$$

Two features  $f_i$  and  $f_j$  are interdependent if the following form is accepted [21, 26]:

$$CMI(f_j; \text{class} | f_i) > QMI(f_j; \text{class}). \quad (7)$$

The relevance criterion that is introduced by Peng et al. [9] is the relevance criterion of set  $K$  with the target class; we have

$$D(K, \text{class}) = \frac{1}{|K|} \sum_{f_j \in K} QMI(f_j; \text{class}). \quad (8)$$

The change of the relevance of set of  $K$  features with the target class considering the feature  $f_i$  ( $f_i \notin k$ ) which is introduced by Jain et al. [27] is measured by the following formula:

$$\begin{aligned} CMI(K; \text{class}; f_i) \\ \approx \frac{1}{|K|} \sum_{f_j \in K} CMI(f_j; \text{class} | f_i) - QMI(f_j; \text{class}). \end{aligned} \quad (9)$$

**3.3. Feature Evaluation via Shapley Value.** In the second phase of the proposed method, after filtering the features, the weight of each feature is obtained using the Shapley value [28, 29] for all features. In this work, we used QMI in all our formulations. The Shapley value is an efficient way for estimating the importance of features weight. The Shapley value of the  $i$ th feature is denoted by  $\phi_i(v)$  as is formulated as follows:

$$\phi_i(v) = \sum_k \Delta_i(k) \frac{|k|!(n - |k| - 1)!}{n!}, \quad (10)$$

where  $n$  denotes the total number of features and  $\Delta_i(k)$  is defined as follows:

$$\begin{aligned} \Delta_i(k) \\ = \begin{cases} 1, & CMI(k; \text{class}; f_i) \geq 0, \sum_{f_j \in K} \psi(i, j) \geq \frac{|K|}{2}, \\ 0, & \text{else.} \end{cases} \end{aligned} \quad (11)$$

And  $\psi(i, j)$  denotes the interdependence index and is defined as follows:

$$\begin{aligned} \psi(i, j) \\ = \begin{cases} 1, & CMI(f_j; \text{class} | f_i) > QMI(f_j; \text{class}), \\ 0, & \text{else} \end{cases} \end{aligned} \quad (12)$$

which means that the feature will be appropriate if it increases the relevance of subset  $K$  on the target class and is interdependent with at least half of the features [21]. Figure 2 shows the flowchart of feature evaluation. The output is a weight vector indicating the Shapley value of feature  $f_j$ . In all the above relationships, we used QMI criterion for more robustness of selected features.

In step 2 of the proposed algorithm, for each feature  $i$ , we should calculate all subsets of features, so that these subsets do not contain feature  $i$ . So at first we calculate

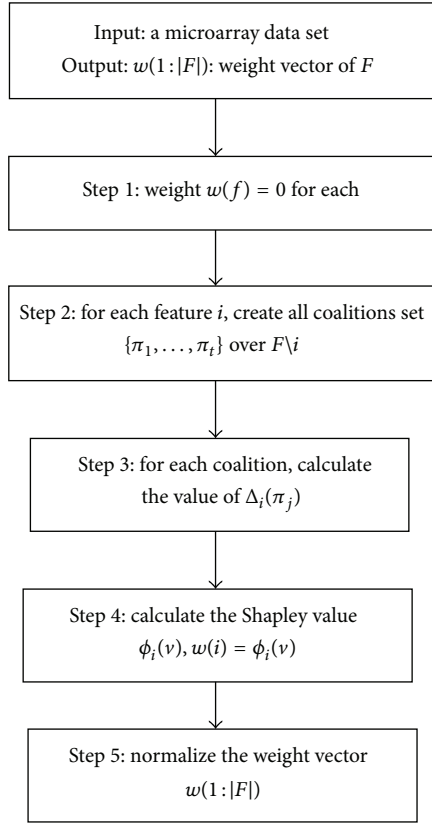


FIGURE 2: Feature evaluation flowchart.

all two member subsets of features. Then, using (9), Conditional Mutual Information measure between set  $k$  and target class is calculated. Then, we calculate the Qualitative Mutual Information measure between each feature and class, namely,  $QMI(f_j; \text{class})$ , and set these two measures in the interdependence index  $\psi(i, j)$ . Finally, Shapley value (weight) of each feature is computed according to (10).

**3.4. Final Feature Selection Phase.** In the third phase that is the final phase of the proposed method, the feature selection algorithm is performed using the weight of each feature. Accordingly, we use the weights of features to reevaluate the features. For this purpose, the features with the highest weights are used. A straightforward optimization is used by employing any information criterion, such as BIF [26], SU [30], and mRMR [9]. In this paper, we used SU criterion. The formula for obtaining this criterion is provided as follows:

$$SU(f_j, \text{class}) = 2 \frac{MI(f_j | \text{class})}{H(\text{class}) + H(f_j)}, \quad (13)$$

where  $H$  indicates the entropy and  $MI$  is the mutual information between feature and class. The flowchart of this phase is specified in Figure 3.

To select the optimal feature in each iteration, the victory criterion  $V(f)$  is defined to evaluate the superiority of each feature on others. The criterion function  $g(f_i)$  (i.e., SU) is used to select feature  $f_i$  by feature relevance or redundancy

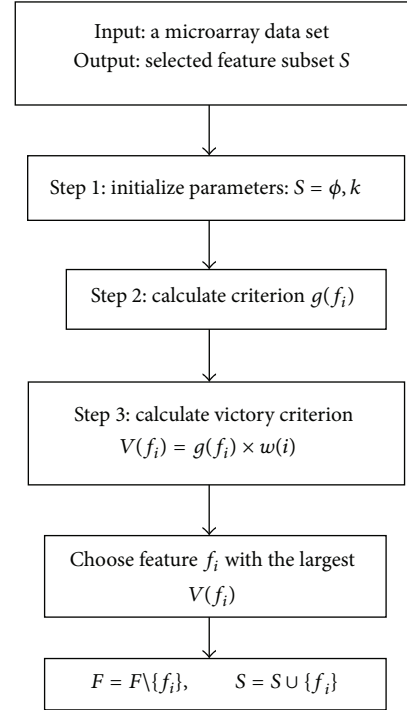


FIGURE 3: The flowchart of the final feature selection.

TABLE 1: Characteristics of the evaluation data sets.

| Number | Data set       | Features | Samples | Classes |
|--------|----------------|----------|---------|---------|
| 1      | 11_Tumors      | 12533    | 174     | 11      |
| 2      | 14_Tumors      | 15009    | 308     | 26      |
| 3      | 9_Tumors       | 5726     | 60      | 9       |
| 4      | Brain_Tumor1   | 5920     | 90      | 5       |
| 5      | Brain_Tumor2   | 10367    | 50      | 4       |
| 6      | Leukemia1      | 5327     | 72      | 3       |
| 7      | Leukemia2      | 11225    | 72      | 3       |
| 8      | Lung_Cancer    | 12601    | 203     | 5       |
| 9      | SRBCT          | 2309     | 83      | 4       |
| 10     | Prostate_Tumor | 12600    | 102     | 2       |
| 11     | DLBCL          | 7129     | 77      | 2       |

analysis. The weight  $w(i)$ , which denotes the impact of feature  $f_i$  on the whole feature space, is used to regulate the relative importance of evaluation value  $g(f_i)$  in feature selection. Finally, we choose the features with the largest victory measure.

## 4. Experimental Results and Discussion

**4.1. Data Set Descriptions.** In this paper, we used the cancer microarray data sets from the Plymouth University [31]. For better performance and better evaluation of the proposed approach, we applied 11 high dimensional microarray data. These eleven data sets are briefly summarized in Table 1.

**4.2. Evaluation Criteria.** The criteria used to evaluate the proposed method consist of accuracy, classification precision,

$F$ -measure, and stability criterion for feature selection algorithm.

**4.2.1. Accuracy of Classification.** Classification accuracy is the main criterion for evaluating the classification and predicting the samples. The classification accuracy is as follows:

$$\text{accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{TN} + \text{FN} + \text{FP})}. \quad (14)$$

Accuracy is described in terms of true positives (TP), true negatives (TN), false negatives (FN), and false positives (FP).

**4.2.2. Precision, Recall, and  $F$ -Measure.** The precision in (15) and recall in (16) are two other evaluation criteria.  $F$ -measure criterion in (17) is used to integrate precision and recall into a single criterion for the comparison:

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (15)$$

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (16)$$

$$F\text{-measure} = \left( \frac{2 \times \text{recall} \times \text{precision}}{\text{recall} + \text{precision}} \right). \quad (17)$$

**4.2.3. Robustness Measure.** The robustness (stability) of a feature selection algorithm can be evaluated based on the ability to select repeated features, given various categories under the same distribution [2]. Let  $S_i$  and  $S_o$  denote feature subsets selected using the  $i$ th categories of resampled data and the full data, respectively. For measuring the similarity between two feature subsets, similarity index  $JC \in [0, 1]$  is used which is defined as follows:

$$JC_i(k) = \frac{|S_i \cap S_o| + SC_i}{k}, \quad (18)$$

where  $|S_i \cap S_o|$  is the number of common features between  $S_i$  and  $S_o$ ,  $SC_i$  is the sum of the absolute correlation values between different features from  $S_i$  and  $S_o$ , and  $k$  is set size of  $S_i$  and  $S_o$ . Assume  $m$  categories of data are produced by resampling and  $m$  feature subsets are selected. The robustness or stability measure of the feature selection algorithm is calculated by (19) as follows:

$$\overline{JC}(k) = \frac{\sum_{i=1}^m JC_i(k)}{m}. \quad (19)$$

**4.3. Classification Results.** First, in preprocessing phase, the features of data sets are normalized, but the SRCBT data is not normalized because its features are small. Microarray data sets are classified using KNN, SVM, Naive Bayes (NB), and Classification and Regression Tree (CART) classifiers. The Gaussian kernel function and standard deviation equal to 20 are used for SVM classifier in this study. We used  $K = 3$  for KNN classifier, but since the classification result for  $K = 3$  is not good for Lung\_Cancer data, we used  $K = 1$  for this data. Among 11 data sets, only two Prostate\_Tumor and DLBCL

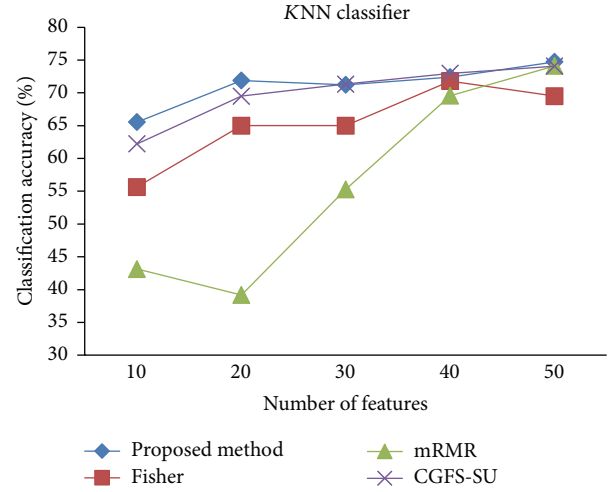


FIGURE 4: KNN classification accuracy for 11\_Tumors.

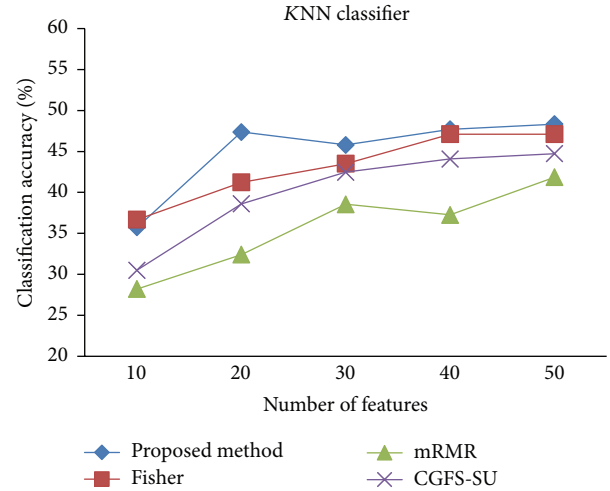


FIGURE 5: KNN classification accuracy for 14\_Tumors.

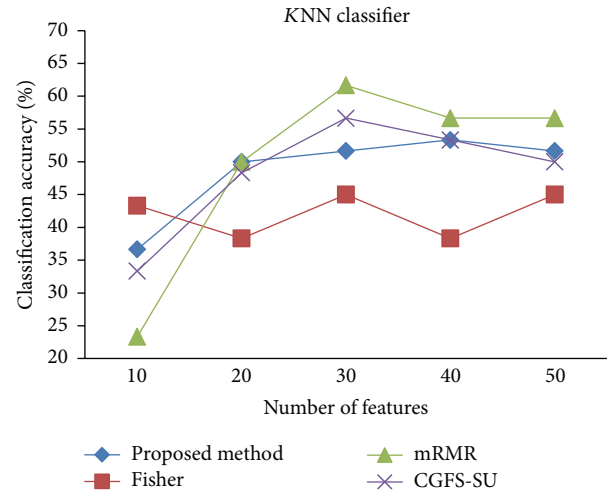


FIGURE 6: KNN classification accuracy for 9\_Tumors.

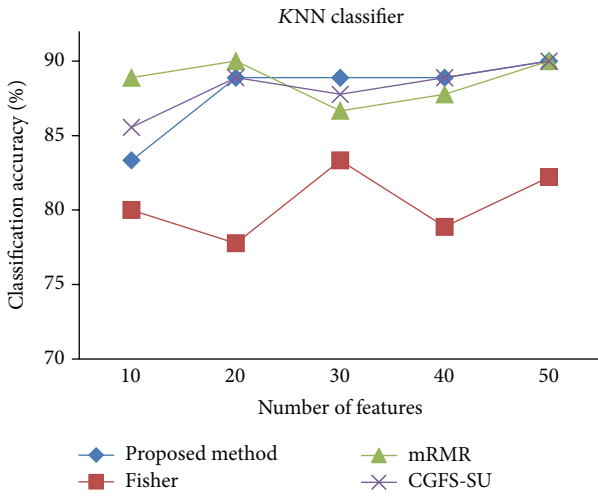


FIGURE 7: KNN classification accuracy for Brain\_Tumor1.

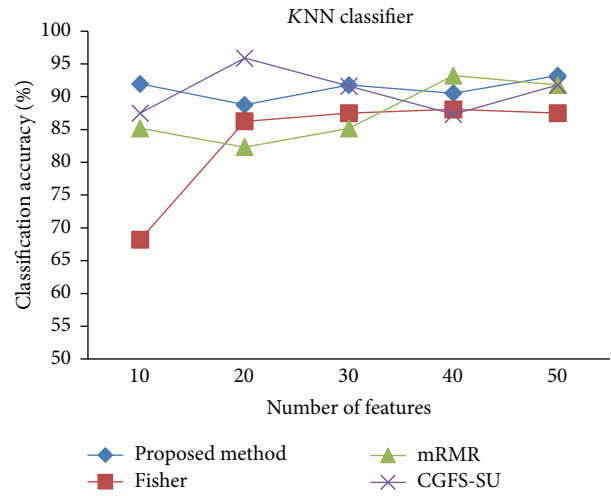


FIGURE 10: KNN classification accuracy for Leukemia2.

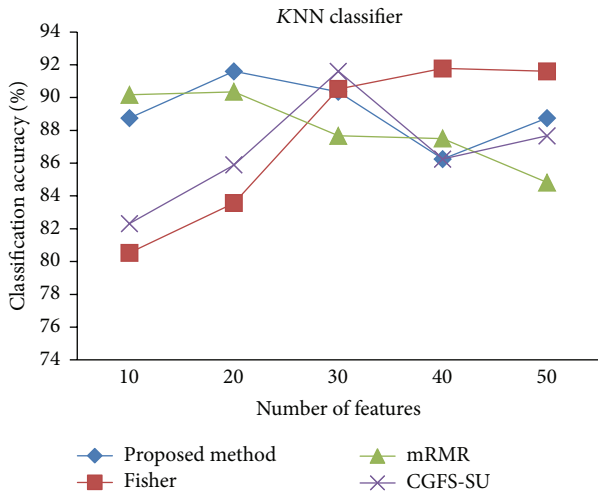


FIGURE 8: KNN classification accuracy for Leukemia1.

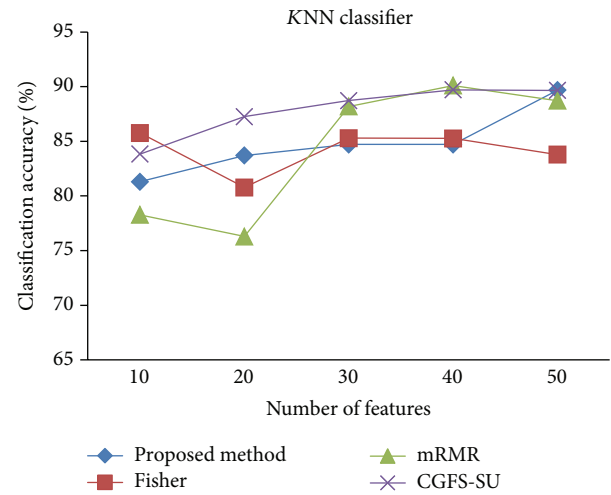


FIGURE 11: KNN classification accuracy for Lung\_Cancer.

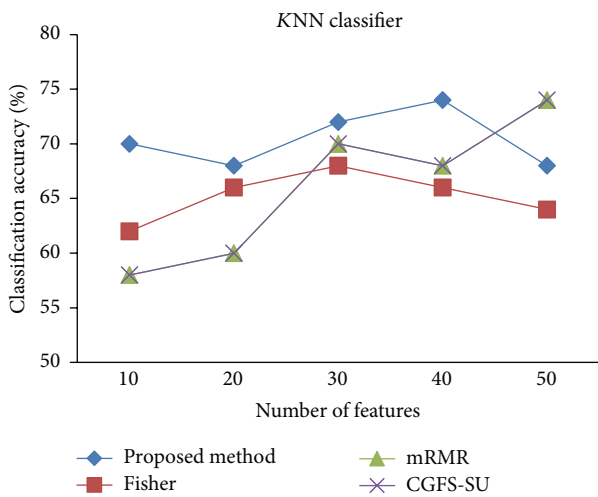


FIGURE 9: KNN classification accuracy for Brain\_Tumor2.

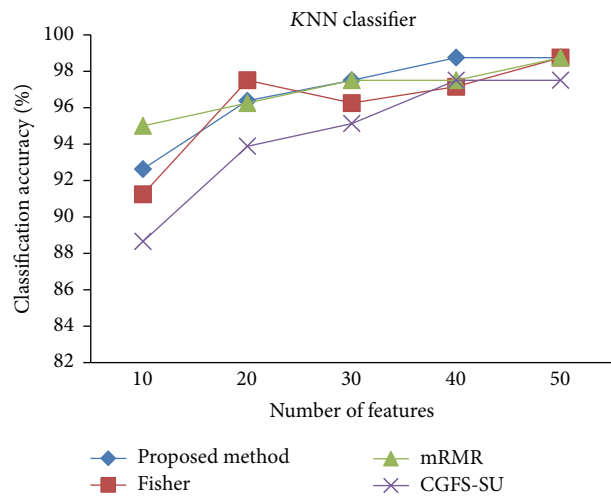


FIGURE 12: KNN classification accuracy for SRBCT.

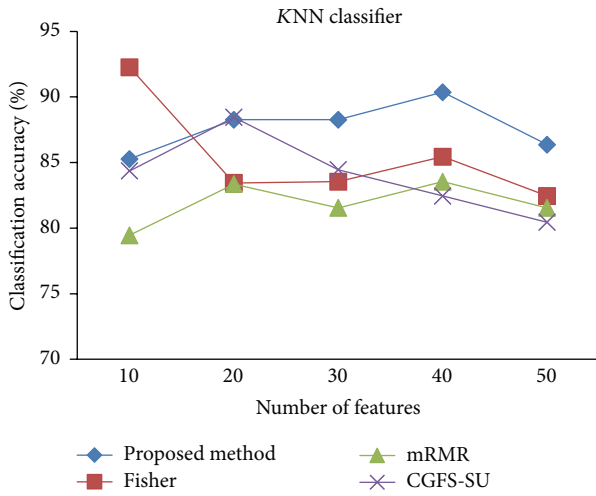


FIGURE 13: KNN classification accuracy for Prostate\_Tumor.

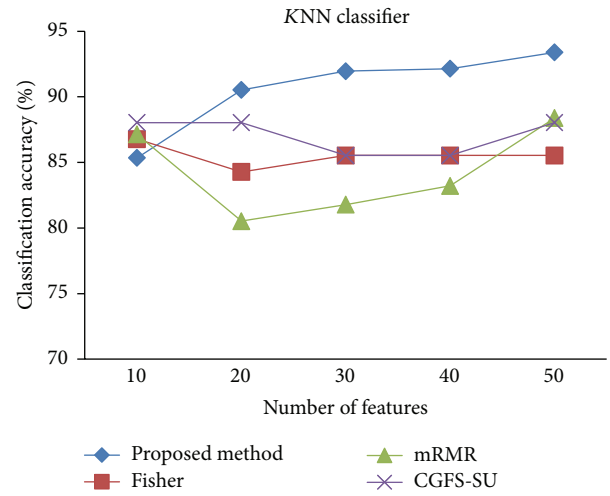


FIGURE 16: KNN classification accuracy for DLBCL.

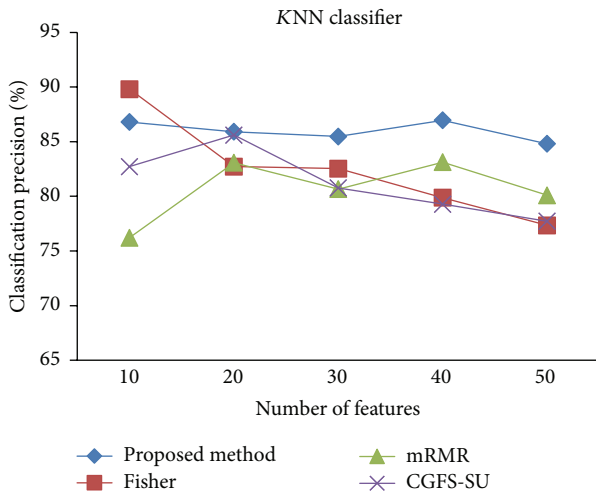


FIGURE 14: KNN classification precision for Prostate\_Tumor.

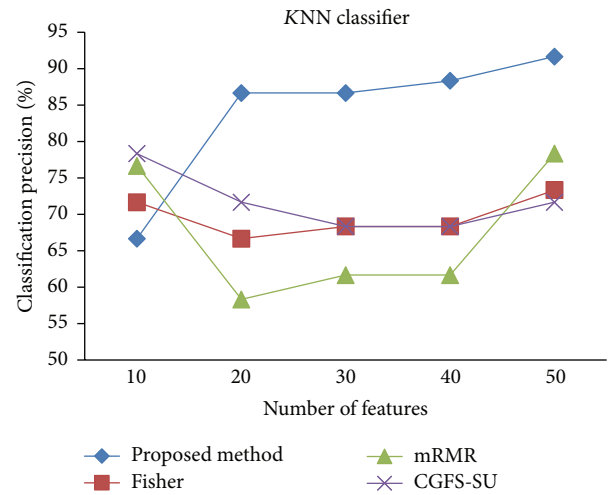


FIGURE 17: KNN classification precision for DLBCL.

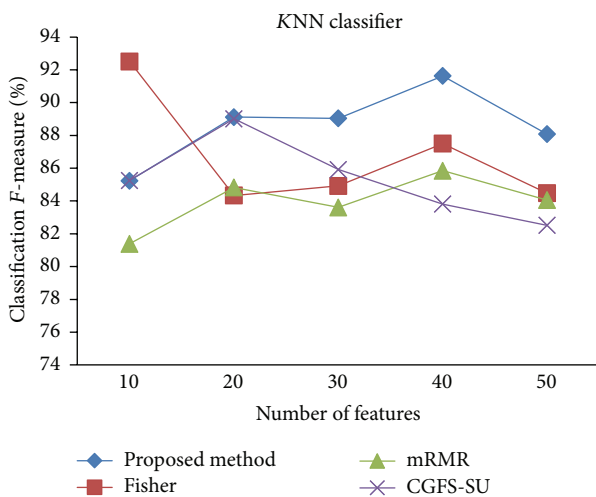


FIGURE 15: KNN classification F-measure for Prostate\_Tumor.

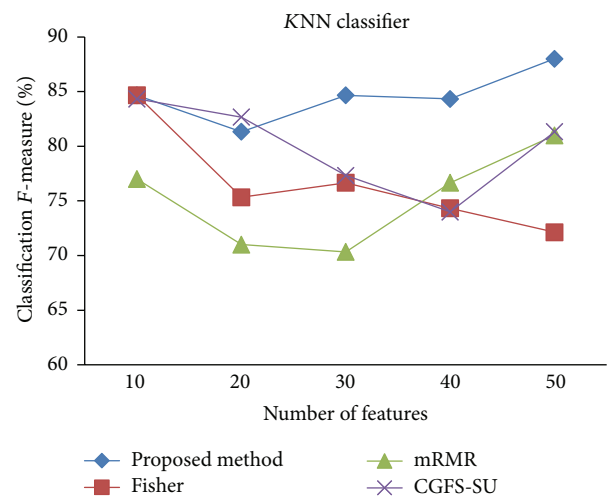


FIGURE 18: KNN classification F-measure for DLBCL.

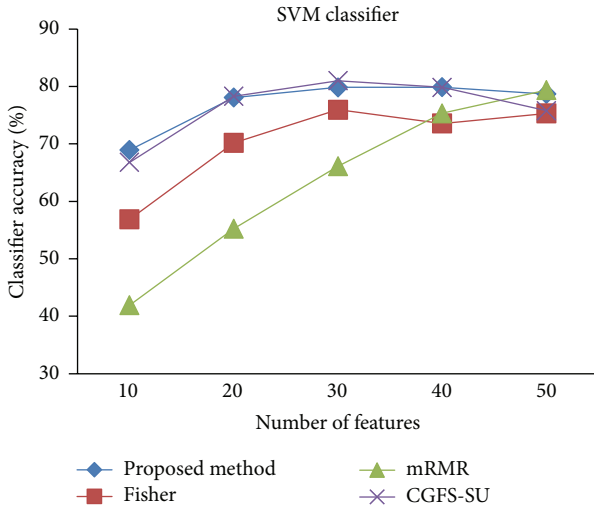


FIGURE 19: SVM classification accuracy for 11\_Tumors.

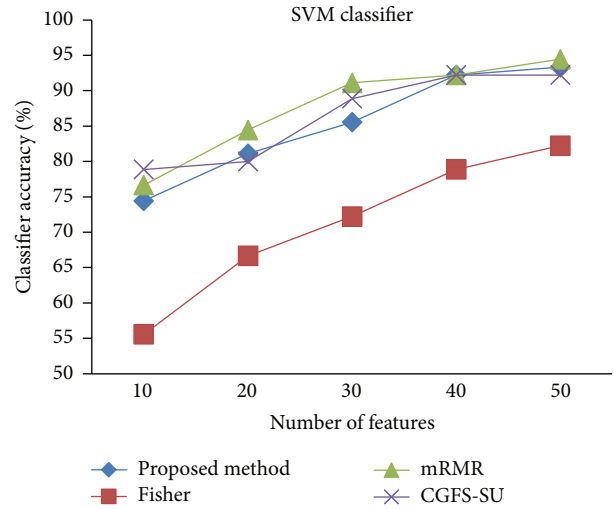


FIGURE 22: SVM classification accuracy for Brain\_Tumor1.

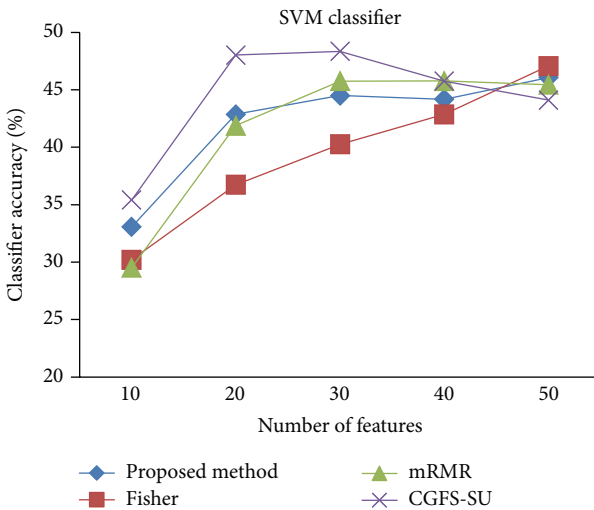


FIGURE 20: SVM classification accuracy for 14\_Tumors.

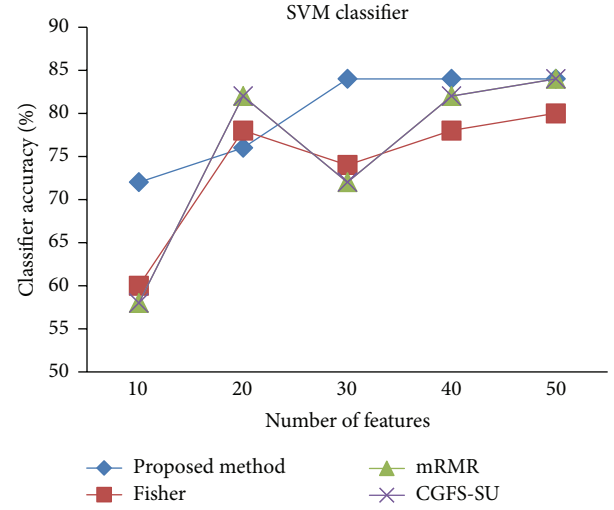


FIGURE 23: SVM classification accuracy for Brain\_Tumor2.

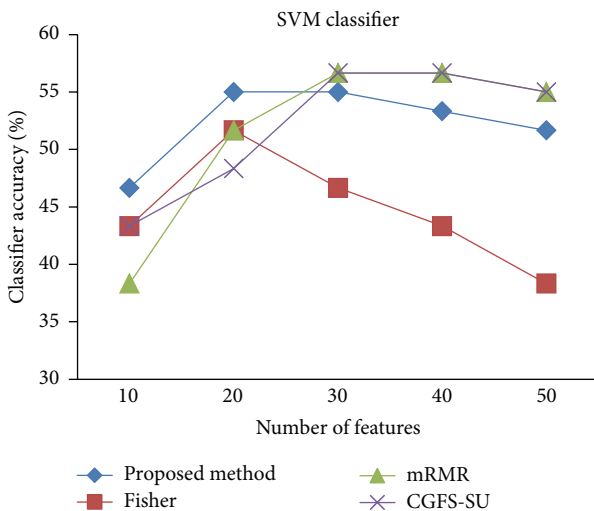


FIGURE 21: SVM classification accuracy for 9\_Tumors.

data sets have two classes; therefore, precision and  $F$ -measure criteria are obtained only for these two data sets and for other data sets only the accuracy criterion is examined. The 300 superior features of 11\_Tumors, 14\_Tumors, Leukemia1, Leukemia2, and DLBCL data sets are selected using mutual information criterion. For other data sets, Fisher ratio is applied for prior feature selection.

For comparison, the classification accuracy is given against the number of features. The  $x$ -axis is the number of features which is considered up to 50 and the  $y$ -axis is the classification accuracy. The proposed method is compared with feature selection using Fisher's ratio, mRMR, and CGFS-SU. The mRMR algorithm is used as information criterion algorithm for comparison. Fisher's ratio method is a univariate filter method for evaluating each feature individually and the CGFS-SU method is proposed in [21] for cooperative game based feature selection. For estimating the performance of classification algorithms, tenfold cross-validation is used.



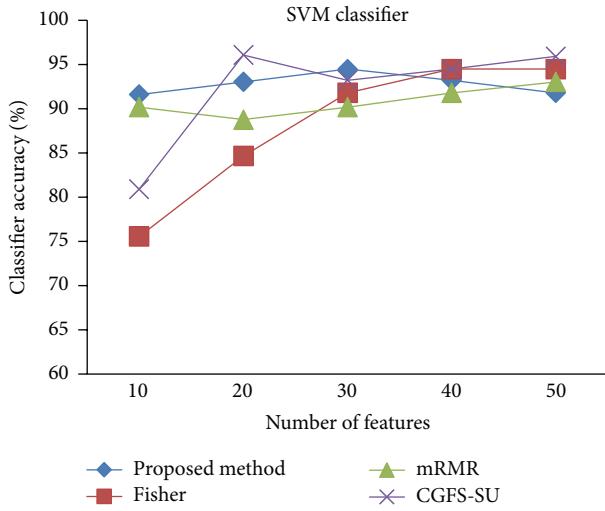


FIGURE 24: SVM classification accuracy for Leukemia1.

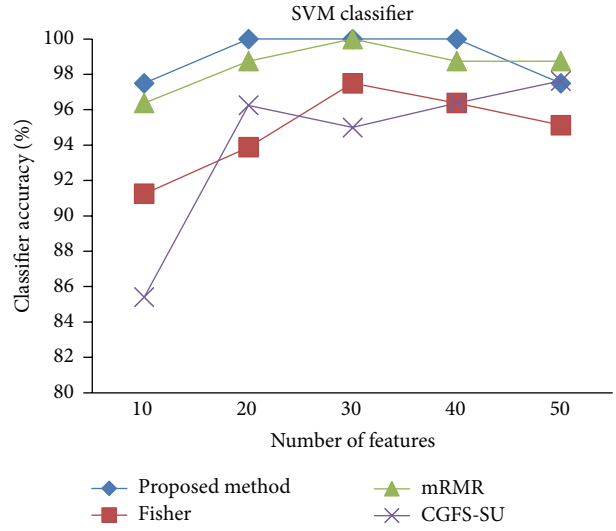


FIGURE 27: SVM classification accuracy for SRBCT.

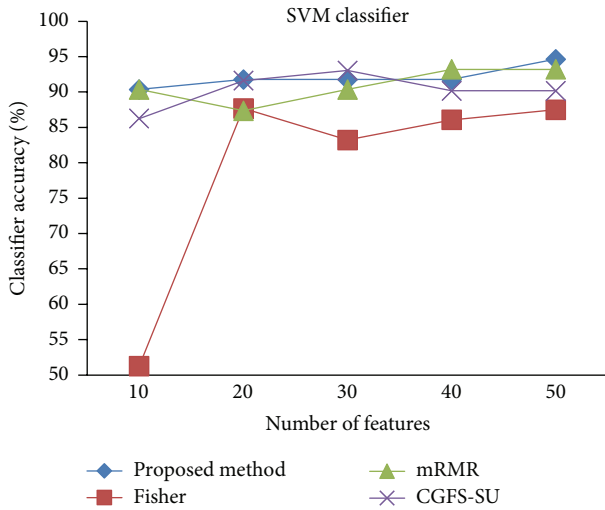


FIGURE 25: SVM classification accuracy for Leukemia2.

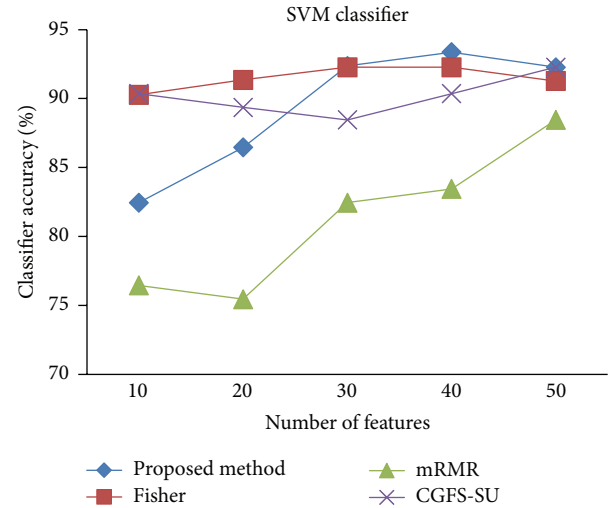


FIGURE 28: SVM classification accuracy for Prostate\_Tumor.

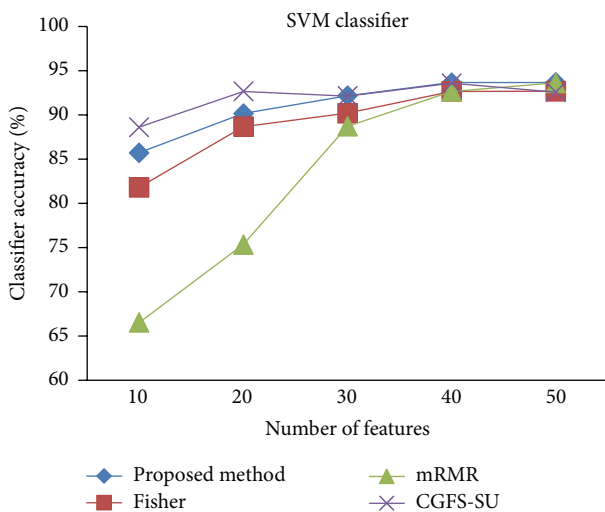


FIGURE 26: SVM classification accuracy for Lung\_Cancer.

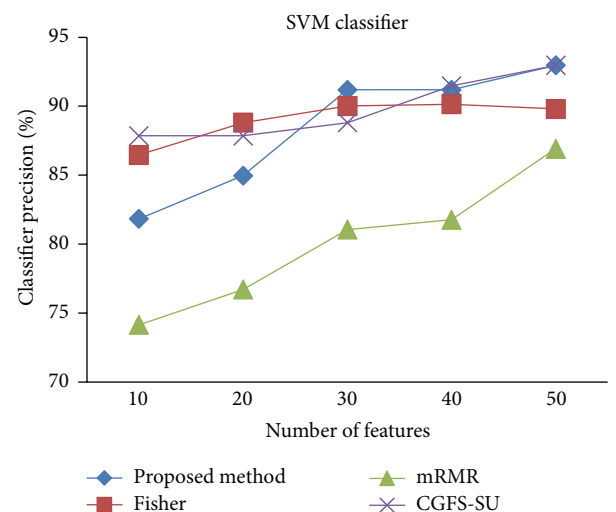


FIGURE 29: SVM classification precision for Prostate\_Tumor.

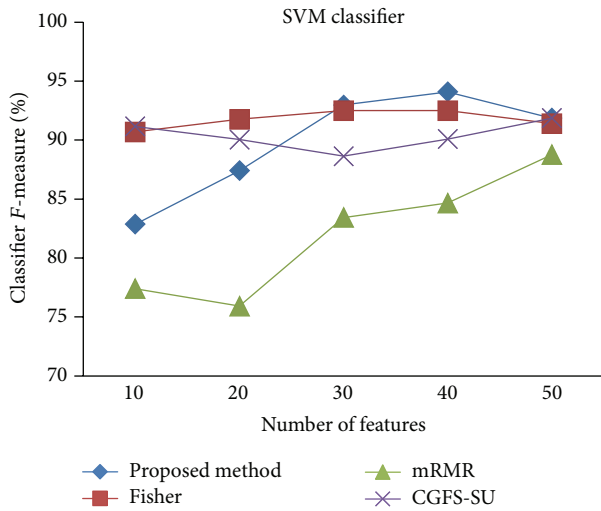


FIGURE 30: SVM classification F-measure for Prostate\_Tumor.

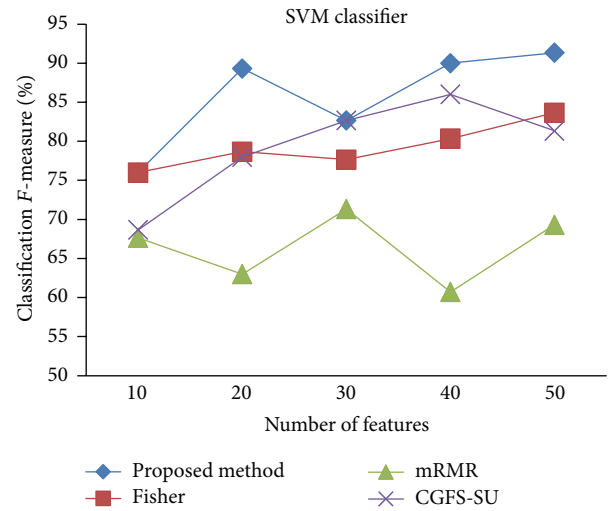


FIGURE 33: SVM classification F-measure for DLBCL.

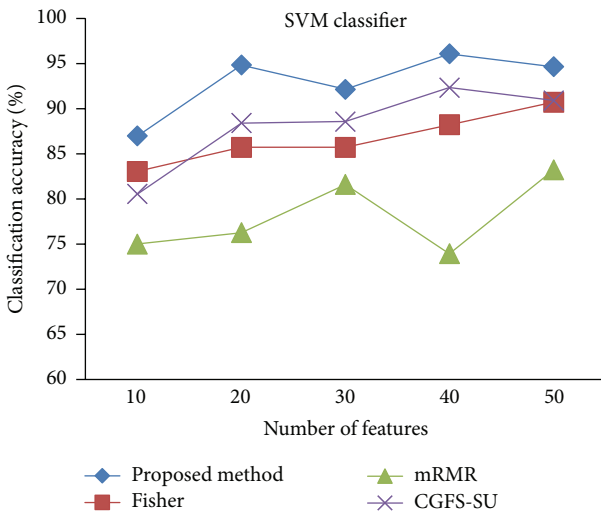


FIGURE 31: SVM classification accuracy for DLBCL.

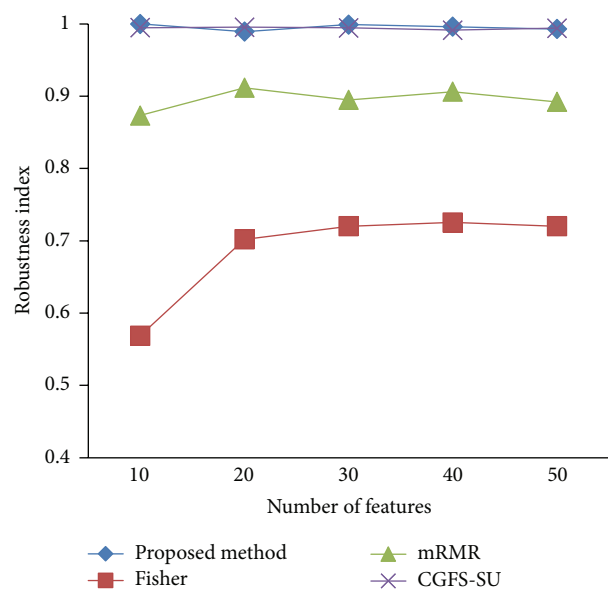


FIGURE 34: Stability results for 11-Tumors.

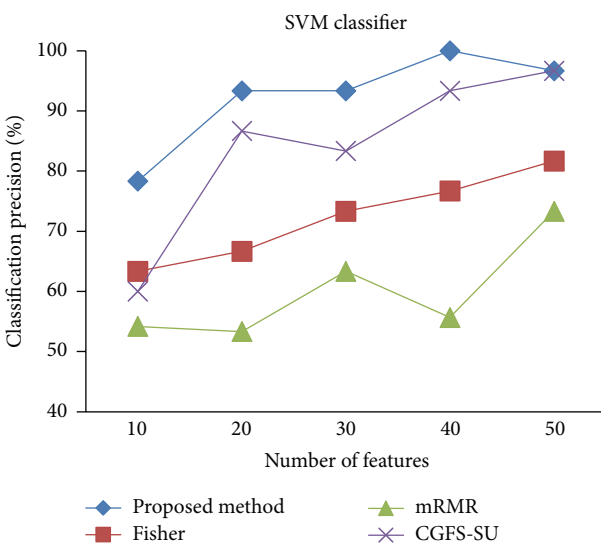


FIGURE 32: SVM classification precision for DLBCL.

Figures 4–18 show the classification results for eleven data sets using KNN classifier.

KNN classifier results show that the proposed method has been improved in most cases compared to the other methods of feature selection. For 11\_Tumors data, the proposed method has achieved the maximum accuracy of 74.73 percent. Also, the result of CGFS-SU method is almost the same as the proposed method and is 74.08 percent, but in 30 first features, its accuracy is less than the proposed method. In 14\_Tumors data, the results are low for all feature selection algorithms. However, the proposed method has the maximum accuracy of 48.34 percent and other methods have much lower accuracy, which is due to the weak correlation between selected features. However, it can be observed that the proposed method achieves higher accuracy in all 50 features as compared to other methods.

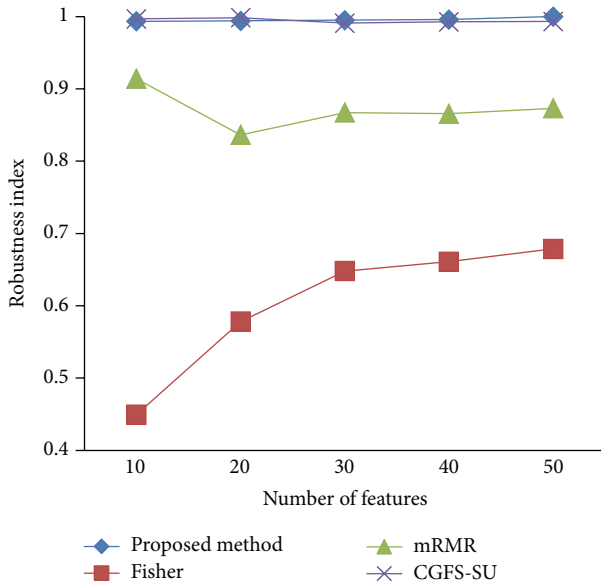


FIGURE 35: Stability results for 14\_Tumors.

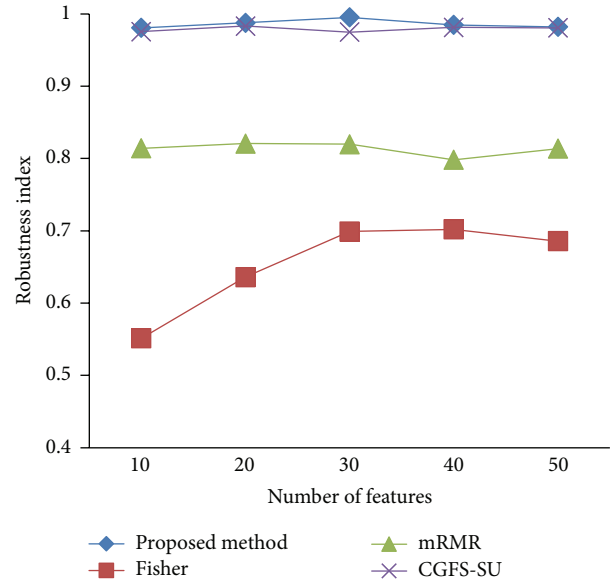


FIGURE 37: Stability results Brain\_Tumor1.

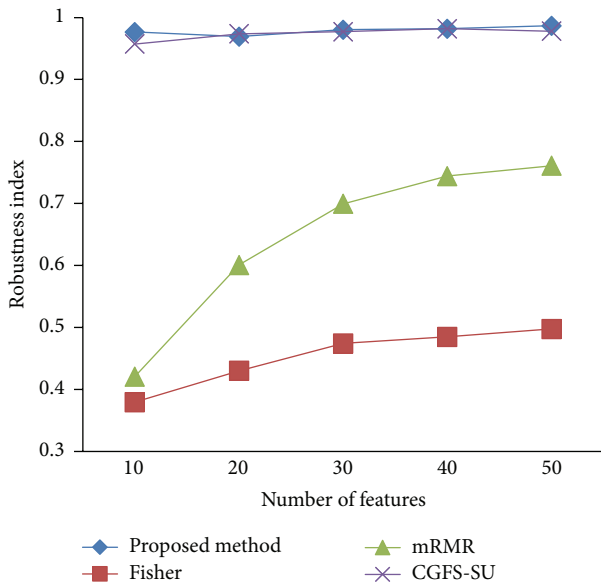


FIGURE 36: Stability results for 9\_Tumors.

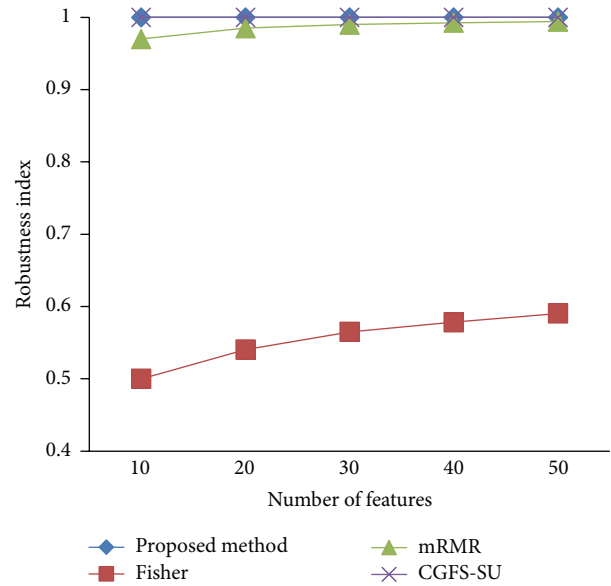


FIGURE 38: Stability results for Brain\_Tumor2.

The accuracy has been increased in SRBCT data, and the proposed method and Fisher and mRMR algorithms reached the maximum accuracy of 98.75 percent. Furthermore, the CGFS-SU method accuracy is 97.56 percent. In Prostate.Tumors data, it has been shown that the proposed method achieved maximum accuracy among all feature selection methods for all three evaluation criteria and for all cases. For DLBCL data, the proposed method achieved the maximum value for most cases for each of the three criteria, and it reached the highest accuracy of 93 percent, precision of 91 percent, and *F*-measure of 88 percent compared to other methods. This shows that the proposed method could reduce interdependency between groups of features and is effective

as compared with other algorithms. Figures 19–33 show the classification results using SVM classifier on the evaluation data sets.

The results of most data sets with SVM classifier are better compared to KNN classifier. In SRBCT data, the proposed method reached the highest rate of accuracy of 100 percent for 10 to 40 features and also the mRMR method. But the CGFS-SU method has lower accuracy and the maximum accuracy is 97.63 percent. Also, Fisher criterion is not much accurate and its maximum accuracy is 97.5 percent. It may be due to the fact that the relationship between the features and target classes is maximum for this data set and the mRMR method and the proposed approach have been able to retain this relationship

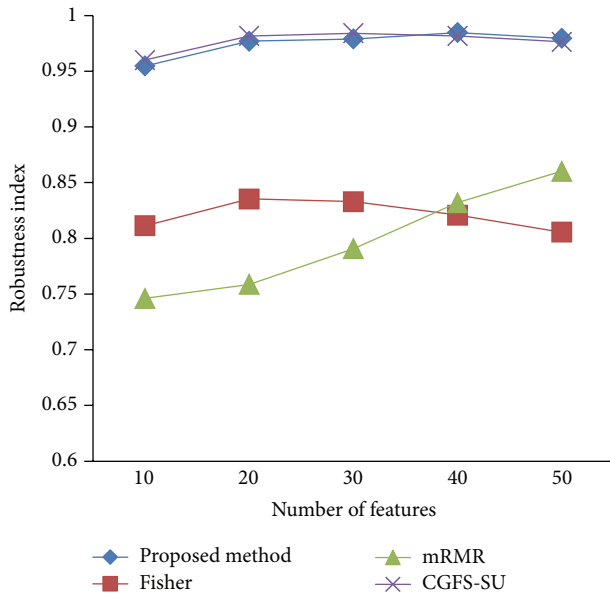


FIGURE 39: Stability results for Leukemia1.

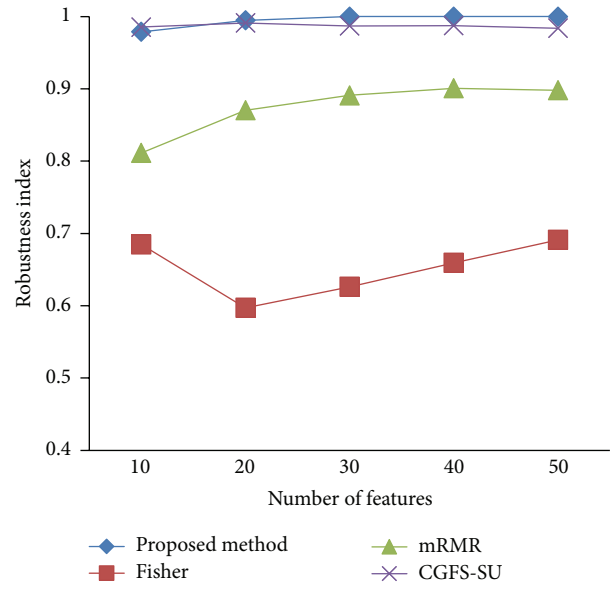


FIGURE 41: Stability results for Lung\_Cancer.

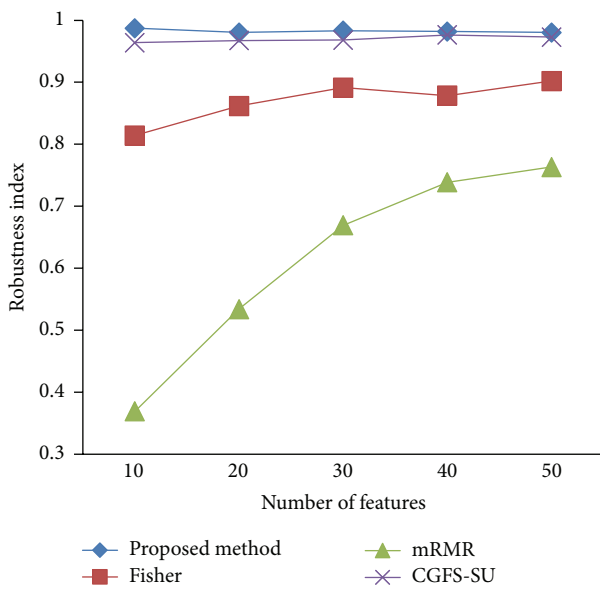


FIGURE 40: Stability results for Leukemia2.

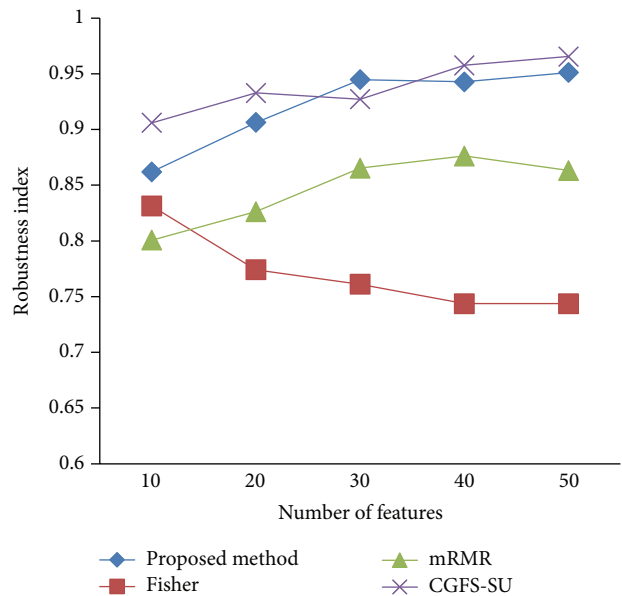


FIGURE 42: Stability results for SRCBT.

between features. Also, in DLBCL data, the proposed method is superior to the other approaches for all three criteria.

**4.4. Stability Results.** In the stability diagram, the  $x$ -axis shows the number of features and the  $y$ -axis shows the stability index. As denoted above, the stability index value is between 0 and 1. For this criterion, features are normalized except for the SRCBT data. Fisher ratio is applied to select the prior 300 best features of all data sets. For estimating the stability, we used tenfold cross-validation. The results of stability are observed in Figures 34–44.

The stability values of the CGFS-SU and the proposed methods are very close to each other and are slightly different.

But we have seen that the proposed method reached the maximum stability in 11\_Tumors, 14\_Tumors, Brain\_Tumor2, and Lung\_Cancer data sets. This shows that the proposed method is much more robust than the other feature selection approaches. However, the CGFS-SU method achieved the maximum stability only for Brain\_Tumor2 data. The stability of other feature selection methods has significant differences with the CGFS-SU and the proposed method; that is, the stability of mRMR method is approximately equal to these methods only for Brain\_Tumor2 data. Table 2 shows the maximum accuracy and average accuracy of feature selection methods on evaluation microarray data using KNN, SVM, NB, and CART classifiers.

TABLE 2: Maximum values of accuracy and average accuracy of feature selection methods on eleven microarray data sets.

| Data set     | Classifier   | Accuracy | Proposed approach | CGFS-SU      | Fisher       | mRMR         |
|--------------|--------------|----------|-------------------|--------------|--------------|--------------|
| 11_Tumors    | KNN          | Max      | <b>74.73</b>      | 74.08        | 71.79        | 74.11        |
|              |              | Average  | <b>71.16</b>      | 70.02        | 65.39        | 56.26        |
|              | SVM          | Max      | 79.86             | <b>81.01</b> | 75.94        | 79.37        |
|              |              | Average  | <b>77.07</b>      | 76.34        | 70.36        | 63.59        |
|              | Naive Bayes  | Max      | <b>84.96</b>      | 82.67        | 77.58        | 83.16        |
|              |              | Average  | 74.75             | <b>77.62</b> | 69.83        | 72.46        |
|              | DecisionTree | Max      | <b>64.37</b>      | 62.05        | 60.35        | 61.07        |
|              |              | Average  | <b>61.64</b>      | 60.20        | 56.05        | 50.33        |
| 14_Tumors    | KNN          | Max      | <b>48.34</b>      | 44.73        | 47.11        | 41.86        |
|              |              | Average  | <b>44.99</b>      | 40.08        | 43.13        | 35.66        |
|              | SVM          | Max      | 46.07             | <b>48.35</b> | 47.06        | 45.76        |
|              |              | Average  | 42.13             | <b>44.33</b> | 39.42        | 41.67        |
|              | Naive Bayes  | Max      | <b>45.22</b>      | 41.18        | 40.92        | 37.23        |
|              |              | Average  | 43.87             | 40.19        | <b>59.8</b>  | 37.04        |
|              | DecisionTree | Max      | <b>46.05</b>      | 43.81        | 41.25        | 37.98        |
|              |              | Average  | <b>41.67</b>      | 39.90        | 37.98        | 33.76        |
| 9_Tumors     | KNN          | Max      | 53.33             | 56.66        | 45           | <b>61.66</b> |
|              |              | Average  | 48.66             | 48.33        | 41.99        | <b>49.46</b> |
|              | SVM          | Max      | 55                | <b>56.66</b> | 51.66        | <b>56.66</b> |
|              |              | Average  | <b>52.33</b>      | 52           | 44.66        | 51.66        |
|              | Naive Bayes  | Max      | 53.33             | <b>56.18</b> | 51.8         | 48.32        |
|              |              | Average  | <b>52.87</b>      | 51.9         | 50.35        | 46.90        |
|              | DecisionTree | Max      | <b>50</b>         | 48.33        | 41.66        | 46.66        |
|              |              | Average  | <b>42</b>         | 40.66        | 37           | <b>42</b>    |
| Brain_Tumor1 | KNN          | Max      | <b>90</b>         | <b>90</b>    | 83.33        | <b>90</b>    |
|              |              | Average  | 88                | 88.22        | 80.44        | <b>88.66</b> |
|              | SVM          | Max      | 93.33             | 92.22        | 82.22        | <b>94.44</b> |
|              |              | Average  | 85.33             | 86.44        | 71.11        | <b>87.77</b> |
|              | Naive Bayes  | Max      | <b>85.55</b>      | 82.22        | <b>85.55</b> | 84.44        |
|              |              | Average  | <b>83.77</b>      | 81.11        | 81.11        | 82.44        |
|              | DecisionTree | Max      | <b>78.44</b>      | 71.11        | 72.44        | 70.22        |
|              |              | Average  | <b>80</b>         | 73.33        | 74.44        | 72.22        |
| Brain_Tumor2 | KNN          | Max      | <b>74</b>         | <b>74</b>    | 68           | <b>74</b>    |
|              |              | Average  | <b>70.4</b>       | 66           | 65.2         | 66           |
|              | SVM          | Max      | <b>84</b>         | <b>84</b>    | 80           | <b>84</b>    |
|              |              | Average  | <b>80</b>         | 75.6         | 74           | 75.6         |
|              | Naive Bayes  | Max      | <b>82</b>         | 80           | 68           | 80           |
|              |              | Average  | <b>73.2</b>       | 68.8         | 63.2         | 68.8         |
|              | DecisionTree | Max      | <b>60</b>         | 56           | <b>60</b>    | 54           |
|              |              | Average  | 50                | 52.40        | <b>53.20</b> | 50.40        |
| Leukemial    | KNN          | Max      | 91.60             | 90.35        | <b>91.78</b> | 91.60        |
|              |              | Average  | <b>89.14</b>      | 86.75        | 87.60        | 88.10        |
|              | SVM          | Max      | 93.21             | <b>96.07</b> | 94.46        | 93.03        |
|              |              | Average  | <b>92.82</b>      | 92.10        | 88.17        | 90.78        |
|              | Naive Bayes  | Max      | <b>97.32</b>      | <b>97.32</b> | 94.64        | 97.14        |
|              |              | Average  | 94.46             | <b>95.89</b> | 92.89        | 95.28        |
|              | DecisionTree | Max      | <b>91.78</b>      | 87.85        | 87.67        | 90.35        |
|              |              | Average  | <b>90.32</b>      | 85           | 81.82        | 88.71        |

TABLE 2: Continued.

| Data set       | Classifier   | Accuracy | Proposed approach | CGFS-SU      | Fisher       | mRMR         |
|----------------|--------------|----------|-------------------|--------------|--------------|--------------|
| Leukemia2      | KNN          | Max      | 93.21             | <b>95.89</b> | 87.50        | 93.21        |
|                |              | Average  | <b>91.25</b>      | 90.82        | 83.10        | 87.53        |
|                | SVM          | Max      | <b>94.64</b>      | 93.03        | 87.67        | 93.21        |
|                |              | Average  | <b>92.07</b>      | 90.25        | 79.14        | 90.99        |
|                | Naive Bayes  | Max      | <b>70.71</b>      | 69.28        | 51.07        | 59.64        |
|                |              | Average  | <b>69.89</b>      | 64.39        | 47.67        | 54.89        |
|                | DecisionTree | Max      | 88.75             | <b>90.17</b> | 85.89        | 87.67        |
|                |              | Average  | <b>86.53</b>      | 86.03        | 77.60        | 84.21        |
| Lung_Cancer    | KNN          | Max      | 89.69             | 89.66        | 85.76        | <b>90.11</b> |
|                |              | Average  | 84.82             | <b>87.84</b> | 84.17        | 84.32        |
|                | SVM          | Max      | <b>93.66</b>      | 93.57        | 92.66        | 93.61        |
|                |              | Average  | 91.06             | <b>91.90</b> | 89.19        | 83.36        |
|                | Naive Bayes  | Max      | 96.09             | <b>96.57</b> | 91.69        | 66.07        |
|                |              | Average  | 92.35             | <b>92.55</b> | 90.57        | 60.12        |
|                | DecisionTree | Max      | <b>88.16</b>      | 86.28        | 83.28        | 87.26        |
|                |              | Average  | <b>85.42</b>      | 85.17        | 81.86        | 79.46        |
| SRCBT          | KNN          | Max      | <b>98.75</b>      | 97.56        | <b>98.75</b> | <b>98.75</b> |
|                |              | Average  | <b>96.80</b>      | 94.13        | 96.25        | 97           |
|                | SVM          | Max      | <b>100</b>        | 97.63        | 97.5         | <b>100</b>   |
|                |              | Average  | <b>99</b>         | 94.13        | 94.88        | 98.52        |
|                | Naive Bayes  | Max      | 98.75             | 98.75        | <b>100</b>   | 98.75        |
|                |              | Average  | <b>97.27</b>      | 95.33        | 94.88        | 95.75        |
|                | DecisionTree | Max      | 87.91             | 85.41        | <b>89.16</b> | 88.05        |
|                |              | Average  | 82.55             | 83.55        | <b>85.08</b> | 80.08        |
| Prostate_Tumor | KNN          | Max      | 90.36             | 84.45        | <b>92.27</b> | 83.54        |
|                |              | Average  | <b>87.70</b>      | 84.03        | 85.43        | 81.89        |
|                | SVM          | Max      | <b>93.36</b>      | 92.27        | 92.27        | 88.45        |
|                |              | Average  | 89.38             | 90.14        | <b>91.49</b> | 81.25        |
|                | Naive Bayes  | Max      | <b>91.27</b>      | 85.45        | 82.45        | 79.45        |
|                |              | Average  | <b>90.12</b>      | 82.03        | 80.85        | 77.45        |
|                | DecisionTree | Max      | <b>91.27</b>      | 80.36        | 82.36        | 76.27        |
|                |              | Average  | <b>87.07</b>      | 78.36        | 78.92        | 73.41        |
| DLBCL          | KNN          | Max      | <b>93.39</b>      | 88.03        | 86.78        | 88.39        |
|                |              | Average  | <b>90.67</b>      | 86.03        | 85.53        | 84.21        |
|                | SVM          | Max      | <b>96.07</b>      | 92.32        | 90.71        | 83.21        |
|                |              | Average  | <b>92.92</b>      | 88.14        | 86.67        | 78           |
|                | Naive Bayes  | Max      | 91.96             | <b>93.57</b> | 84.46        | 83.03        |
|                |              | Average  | 87.10             | <b>90.64</b> | 83.21        | 79.71        |
|                | DecisionTree | Max      | <b>83.21</b>      | 81.78        | 82.58        | 81.78        |
|                |              | Average  | <b>81.39</b>      | 78.60        | 80.78        | 78.53        |

According to Table 2, it can be seen that the proposed method achieved the highest accuracy among all the algorithms in most cases. Also, results of the average classification accuracy show that the proposed method has improved the average accuracy between 1 and 5 percent as compared with the CGFS-SU method, also between 2 and 14 percent as compared with the Fisher ratio method, and between 1 and 14 percent as compared with the minimum redundancy maximum relevance approach. In Table 3, average stability

$\pm$  standard deviation of four feature selection methods is shown.

In these experiments, if the average stability is high and the standard deviation is low, the approach is more robust. It is observed that, among the eleven microarray data sets, the stability of the proposed method is more than the other feature selection methods on eight data sets and the stability is less than the CGFS-SU method only on three data sets. Furthermore, the results of these experiments show that

TABLE 3: Average stability  $\pm$  standard deviation stability of the evaluated feature selection approaches.

| Data set       | Proposed method                       | CGFS-SU                               | Fisher              | mRMR                |
|----------------|---------------------------------------|---------------------------------------|---------------------|---------------------|
| 11_Tumors      | <b>0.9954 <math>\pm</math> 0.0044</b> | 0.9939 $\pm$ 0.0015                   | 0.6872 $\pm$ 0.0668 | 0.8956 $\pm$ 0.0149 |
| 14_Tumors      | <b>0.9956 <math>\pm</math> 0.0026</b> | 0.9944 $\pm$ 0.0030                   | 0.6029 $\pm$ 0.0940 | 0.8712 $\pm$ 0.0279 |
| 9_Tumors       | <b>0.9790 <math>\pm</math> 0.0065</b> | 0.9736 $\pm$ 0.0096                   | 0.4557 $\pm$ 0.0472 | 0.6451 $\pm$ 0.1400 |
| Brain_Tumor1   | <b>0.9873 <math>\pm</math> 0.0053</b> | 0.9790 $\pm$ 0.0037                   | 0.6548 $\pm$ 0.0636 | 0.8132 $\pm$ 0.0090 |
| Brain_Tumor2   | <b>1 <math>\pm</math> 0</b>           | <b>1 <math>\pm</math> 0</b>           | 0.5548 $\pm$ 0.0358 | 0.9863 $\pm$ 0.0097 |
| Leukemia1      | 0.9750 $\pm$ 0.0117                   | <b>0.9768 <math>\pm</math> 0.0098</b> | 0.8212 $\pm$ 0.0130 | 0.7975 $\pm$ 0.0483 |
| Leukemia2      | <b>0.9826 <math>\pm</math> 0.0029</b> | 0.9697 $\pm$ 0.0048                   | 0.8695 $\pm$ 0.0345 | 0.6203 $\pm$ 0.1538 |
| Lung_Cancer    | <b>0.9947 <math>\pm</math> 0.0092</b> | 0.9869 $\pm$ 0.0027                   | 0.6517 $\pm$ 0.0398 | 0.8743 $\pm$ 0.0372 |
| SRCBT          | 0.9214 $\pm$ 0.0375                   | <b>0.9378 <math>\pm</math> 0.0240</b> | 0.7707 $\pm$ 0.0362 | 0.8463 $\pm$ 0.0317 |
| Prostate_Tumor | 0.9235 $\pm$ 0.0250                   | <b>0.9551 <math>\pm</math> 0.0066</b> | 0.7876 $\pm$ 0.0165 | 0.7462 $\pm$ 0.0655 |
| DLBCL          | <b>0.9766 <math>\pm</math> 0.0094</b> | 0.9569 $\pm$ 0.0161                   | 0.6719 $\pm$ 0.0700 | 0.7748 $\pm$ 0.0543 |

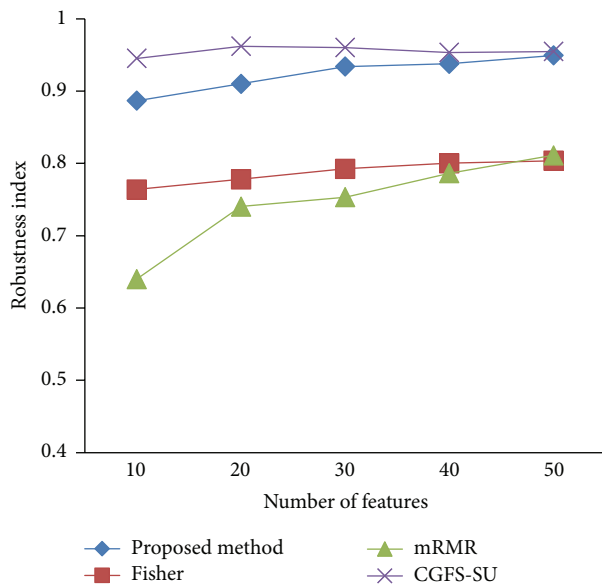


FIGURE 43: Stability results for Prostate\_Tumor.

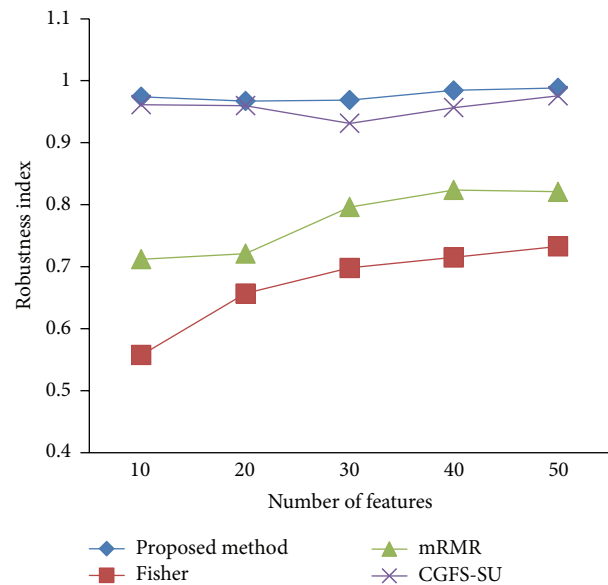


FIGURE 44: Stability results for DLBCL.

the proposed method has improved the average stability between 0.001 and 0.01 as compared with the CGFS-SU method, between 0.1 and 0.5 as compared with the Fisher ratio method, and between 0.01 and 0.3 as compared with the minimum redundancy maximum relevance method.

## 5. Conclusion and Future Work

This paper proposed a feature selection approach for robust gene selection of microarray data. In the proposed method, a cooperative game theory method is introduced to evaluate the weight of each feature considering the complex and inherent relationships among features and Qualitative Mutual Information (QMI) measure is used for more robust feature selection. The idea used for the stability is to use Fisher ratio as a utility in calculating QMI. The results on eleven evaluation microarray data sets show that the proposed method is an effective and stable method for reducing the dimensions of data and is able to reach relative improvement as compared

to the other feature selection methods. As future works, we propose to calculate the weight of each feature using fuzzy Shapley value or fuzzy Banzhaf power index.

## Competing Interests

The authors declare that they have no competing interests.

## References

- [1] M. Yassi and M. H. Moattar, "Robust and stable feature selection by integrating ranking methods and wrapper technique in genetic data classification," *Biochemical and Biophysical Research Communications*, vol. 446, no. 4, pp. 850–856, 2014.
- [2] F. Yang and K. Z. Mao, "Robust feature selection for microarray data based on multicriterion fusion," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 8, no. 4, pp. 1080–1092, 2011.

- [3] X. Sun, Y. Liu, J. Li, J. Zhu, H. Chen, and X. Liu, "Feature evaluation and selection with cooperative game theory," *Pattern Recognition*, vol. 45, no. 8, pp. 2992–3002, 2012.
- [4] V. Bolón-Canedo, N. Sánchez-Marroño, A. Alonso-Betanzos, J. M. Benítez, and F. Herrera, "A review of microarray datasets and applied feature selection methods," *Information Sciences*, vol. 282, pp. 111–135, 2014.
- [5] K. Kira and L. Rendell, "The feature selection problem: traditional methods and a new algorithm," in *Proceedings of the 10th National Conference on Artificial Intelligence (AAAI '92)*, pp. 129–134, AAAI Press, 1992.
- [6] I. Kononenko, "Estimating attributes: analysis and extensions of relief," in *Machine Learning: ECML-94*, F. Bergadano and L. De Raedt, Eds., vol. 784 of *Lecture Notes in Computer Science*, pp. 171–182, Springer, 1994.
- [7] K. Thangavel and A. Pethalakshmi, "Dimensionality reduction based on rough set theory: a review," *Applied Soft Computing Journal*, vol. 9, no. 1, pp. 1–12, 2009.
- [8] G. Brown, "A new perspective for information theoretic feature selection," in *Proceedings of the 12th International Conference on Artificial Intelligence and Statistics (AISTATS '09)*, pp. 49–56, Clearwater Beach, Fla, USA, 2009.
- [9] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.
- [10] J. Wang, L. Wu, J. Kong, Y. Li, and B. Zhang, "Maximum weight and minimum redundancy: a novel framework for feature subset selection," *Pattern Recognition*, vol. 46, no. 6, pp. 1616–1627, 2013.
- [11] T. Nguyen, A. Khosravi, D. Creighton, and S. Nahavandi, "A novel aggregate gene selection method for microarray data classification," *Pattern Recognition Letters*, vol. 60–61, pp. 16–23, 2015.
- [12] I. Inza, B. Sierra, R. Blanco, and P. Larrañaga, "Gene selection by sequential search wrapper approaches in microarray cancer class prediction," *Journal of Intelligent & Fuzzy Systems*, vol. 12, no. 1, pp. 25–33, 2002.
- [13] S. Tabakhi, A. Najafi, R. Ranjbar, and P. Moradi, "Gene selection for microarray data classification using a novel ant colony optimization," *Neurocomputing*, vol. 168, pp. 1024–1036, 2015.
- [14] H. M. Alshamlan, G. H. Badr, and Y. A. Alohal, "Genetic bee colony (GBC) algorithm: a new gene selection method for microarray cancer classification," *Computational Biology and Chemistry*, vol. 56, pp. 49–60, 2015.
- [15] R. Ruiz, J. C. Riquelme, and J. S. Aguilar-Ruiz, "Incremental wrapper-based gene selection from microarray data for cancer classification," *Pattern Recognition*, vol. 39, no. 12, pp. 2383–2392, 2006.
- [16] A. Sharma, S. Imoto, and S. Miyano, "A top-r feature selection algorithm for microarray gene expression data," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 9, no. 3, pp. 754–764, 2012.
- [17] M. Wanderley, V. Gardeux, R. Natowicz, and A. Braga, "GA-KDE-BAYES: an evolutionary wrapper method based on non-parametric density estimation applied to bioinformatics problems," in *Proceedings of the 21st European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN '13)*, pp. 155–160, 2013.
- [18] J. Apolloni, G. Leguizamón, and E. Alba, "Two hybrid wrapper-filter feature selection algorithms applied to high-dimensional microarray experiments," *Applied Soft Computing*, vol. 38, pp. 922–932, 2016.
- [19] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine Learning*, vol. 46, no. 1–3, pp. 389–422, 2002.
- [20] J. Canul-Reich, L. O. Hall, D. Goldgof, J. N. Korecki, and S. Eschrich, "Iterative feature perturbation as a gene selector for microarray data," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 26, no. 5, Article ID 1260003, 25 pages, 2012.
- [21] X. Sun, Y. Liu, J. Li, J. Zhu, X. Liu, and H. Chen, "Using cooperative game theory to optimize the feature selection problem," *Neurocomputing*, vol. 97, pp. 86–93, 2012.
- [22] K. Zeng, K. She, and X. Niu, "Feature selection with neighborhood entropy-based cooperative game theory," *Computational Intelligence and Neuroscience*, vol. 2014, Article ID 479289, 10 pages, 2014.
- [23] Y. Lv, S. Liao, H. Shi, Y. Qian, and S. Ji, "QMIQPN: an enhanced QPN based on qualitative mutual information for reducing ambiguity," *Knowledge-Based Systems*, vol. 71, pp. 114–125, 2014.
- [24] H. Luan, F. Qi, and D. Shen, "Multi-modal image registration by quantitative-qualitative measure of mutual information (QMI)," in *Computer Vision for Biomedical Image Applications*, vol. 3765 of *Lecture Notes in Computer Science*, pp. 378–387, Springer, Berlin, Germany, 2005.
- [25] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artificial Intelligence*, vol. 97, no. 1–2, pp. 273–324, 1997.
- [26] L. Yu and H. Liu, "Efficient feature selection via analysis of relevance and redundancy," *Journal of Machine Learning Research*, vol. 5, pp. 1205–1224, 2004.
- [27] A. K. Jain, R. P. W. Duin, and J. C. Mao, "Statistical pattern recognition: a review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 4–37, 2000.
- [28] S. Cohen, G. Dror, and E. Ruppin, "Feature selection via coalitional game theory," *Neural Computation*, vol. 19, no. 7, pp. 1939–1961, 2007.
- [29] L. S. Shapley, "A value for n-person games," in *Contributions to the Theory of Games*, vol. 2, no. 28 of *Annals of Mathematics Studies*, pp. 307–317, Princeton University Press, Princeton, NJ, USA, 1953.
- [30] P. E. Meyer, C. Schretter, and G. Bontempi, "Information-theoretic feature selection in microarray data using variable complementarity," *IEEE Journal on Selected Topics in Signal Processing*, vol. 2, no. 3, pp. 261–274, 2008.
- [31] Microarray Cancers, Plymouth University, January 2014, [http://www.tech.plym.ac.uk/spmc/links/bioinformatics/microarray/microarray\\_cancers.html](http://www.tech.plym.ac.uk/spmc/links/bioinformatics/microarray/microarray_cancers.html).