IMPROVING DETECTION PERFORMANCE OF ARTIFICIAL NEURAL NETWORK BY SHAPLEY VALUE EMBEDDED GENETIC FEATURE SELECTOR

S. Sasikala[*], *S. Appavu alias Balamurugan*[†], *S. Geetha*[‡]

Abstract: This work is motivated by the interest in feature selection that greatly affects the detection accuracy of a classifier. The goals of this paper are (i) identifying optimal feature subset using a novel wrapper based feature selection algorithm called Shapley Value Embedded Genetic Algorithm (SVEGA), (ii) showing the improvement in the detection accuracy of the Artificial Neural Network (ANN) classifier with the optimal features selected, (iii) evaluating the performance of proposed SVEGA-ANN model on the medical datasets. The medical diagnosis system has been built using a wrapper based feature selection algorithm that attempts to maximize the specificity and sensitivity (in turn the accuracy) as well as by employing an ANN for classification. Two memetic operators namely "include" and "remove" features (or genes) are introduced to realize the genetic algorithm (GA) solution. The use of GA for feature selection facilitates quick improvement in the solution through a fine tune search. An extensive experimental evaluation of the proposed SVEGA-ANN method on 26 benchmark datasets from UCI Machine Learning repository and Kent ridge repository, with three conventional classifiers, outperforms state-of-the-art systems in terms of classification accuracy, number of selected features and running time.

Key words: *feature selection, shapley values, genetic algorithm, artificial neural network, medical data mining, classification*

Received: April 25, 2014 **DOI:** 10.14311/NNW.2016.26.010
Revised and accepted: February 25, 2015

1. Introduction

Data mining application in medicine has proved to be a successful strategy in the areas of medical services including prediction of usefulness of surgical procedures, clinical tests, medication procedures , and the discovery of associations among

[*]S. Sasikala – Corresponding author, Research Scholar, Anna university, Tamil Nadu, India, E-mail: nithilannsasikala@yahoo.co.in

[†]S. Appavu alias Balamurugan, K.L.N. College of Information Technology, Tamil Nadu, India, E-mail: app_s@yahoo.com

[‡]S. Geetha, School of Computing Science and Engg., VIT University – Chennai Campus, Tamil Nadu, India, E-mail: geethabaalan@gmail.com

clinical and diagnosis data [21]. The applicability of data mining for healthcare applications is increasingly gaining importance. The availability of diverse-natured medical data for diagnosis and prognosis and of pervasive data mining techniques to process these data offer medical data mining a distinctive place to truly assist and impact patient care.

The unique characteristics of medical databases that pose challenges for data mining are the privacy-sensitive, heterogeneous, and voluminous data. These data may have valuable information which awaits extraction. The required knowledge is found to be encapsulated in/as various regularities and patterns that may not be apparent in the raw data. Extracting such knowledge has proved to be priceless for future medical decision making. Feature selection is crucial for analysing the various dimensional bio-medical data. It is difficult for the biologists or doctors to examine the whole feature-space obtained through clinical laboratories at one time. Machine learning algorithms recommend that only few of these features are significant for the disease diagnosis. These recommended significant features alone are sufficient to help doctors or experts, understand the biomedical mechanism better and deeper along with cause of disease; the faster and accurate diagnosis can help the doctors to provide the best treatment, so that the infected patients recover as early as possible.

Feature selection methods [9, 23] tend to identify the features which are the most relevant for classification and can be broadly categorized as either subset selection methods or ranking methods. The former type returns a subset of the original set of features which are considered to be the most important for classification. Ranking methods sort the features according to their usefulness in the classification task. Most of the classifiers are modeled as per the ranking strategy that selects the final feature subset, based on ad-hoc manner. Feature selection, as a pre-processing step to machine learning, is prominent and effective in dimensionality reduction, by (i) removing irrelevant and redundant data, (ii) increasing learning accuracy, and (iii) improving result comprehensibility.

Feature selection algorithms generally fall into two broad categories, the filter model and the wrapper model. The filter model depends on general characteristics of the training data to select some features without involving any learning algorithm. The filter model assesses the relevance of features from data alone, independent of classifiers, using measures like distance, information, dependency (correlation), and consistency. The filter method is further classified into Feature Subset Selection (FSS) and Feature Ranking (FR) methods. The wrapper model needs one predetermined learning algorithm for classification and uses its performance to evaluate and determine the features to be selected. For each of the generated new subset of features, the wrapper model is supposed to learn the hypothesis of a classifier. It has a propensity to find better features suited to the predetermined learning algorithm resulting in superior learning performance, but it also consumes more computation time and is economically expensive than the filter model. Whenever dealing with the large number of features, the filter model is usually chosen due to its high accuracy and less computational cost [7]. The hybrid model takes the advantages of the two previous models, and uses an independent measure to identify the best subsets for a given cardinality and applies a mining algorithm to select the best subset among all subsets across different cardinalities [15].

In this direction, we have attempted to ensemble of a filter based model with another wrapper based model, i.e., Shapley values embedded into genetic algorithm (GA). The ensemble is brought about in a fashion so as to reduce the number of features as well as enhance the classification accuracy.

The objective of this research work is aimed at selecting more significant and meaningful features from the available raw medical dataset so as to help the physician to arrive at an accurate diagnosis. Aggressive dimensionality reduction is executed with the motive of increasing the prediction accuracy. The features are subjected to a genetic evolution process within which they undergo the memetic operations namely *include* and *remove.* This process is coupled with Shapley Value Analysis that finds out the contribution made by a feature towards the classification process. Eventually, the optimal feature subset that is with minimum cardinality and maximum accuracy is selected.

2. Related work

Numerous works have been carried out in the field of dimensionality reduction for medical diagnosis. The following section summarises them, highlighting the strengths and weaknesses of each method.

John Q. Gan et al. [6] proposed the Filter-Dominating Hybrid Sequential Forward Feature Selection (FDHSFFS) algorithm for high dimensional feature subset selection. This method proved to be fast but demanded huge computational complexity. Another variant of the SFFS method called Improved F-score and Sequential Forward Search (IFSFS) was proposed by Juanying Xie et al [29] for feature selection to diagnose erythemato-squamous disease. This method was designed to improve the F-score of the classifier and measured the discrimination among more than two sets of real numbers instead of measuring between only two sets of real numbers.

Another category of feature selection methods used mutual information score. La Vinh et al. [27] proposed a novel feature selection method based on the normalization of the well-known mutual information measurement to estimate the potential of the features.

An incremental learning algorithm, in which the most informative features are learnt at each step, is proposed by Ruckstieb et al. [19] and is called as Sequential Online Feature Selection (SOFS). Another Scatter Search-based approach coupled with Decision Trees (SS+DT) is proposed by Shih-Wei Lin et al. [13]. The method acquired optimal parameter settings and selected the beneficial subset of features that resulted in better classification. In [12] Irena Koprinska empirically evaluated feature selection methods for classification of Brain-Computer Interface (BCI) data. A new feature selection method based on rough set theory has been proposed by Sushmita Paul et al. [18]. The proposed method identified discriminative and significant genes from high-dimensional microarray gene expression data sets.

Correlation Based Filter [4, 14] is another strategy for feature selection. Ensemble methods have also been proposed. Monirul Kabir et al. [16] presented a new Hybrid Genetic Algorithm (HGA) for Feature Selection (FS), called as HGAFS. It employed a new local search operation that is devised and embedded in HGA to fine-tune the search in feature selection process. Sasikala et al. [21]

proposed a multi-level feature selection process named 'Multi-Filtration Feature Selection (MFFS)' which adjusted the Principal Component Analysis (PCA) parameter named "variance coverage" and recommended the classifier model with the value at which maximum classification accuracy is obtained for 22 benchmark medical datasets.

Hybrid schemes that combine wrapper-based and filter-based approaches are also in the literature. [2, 24] are two such schemes where the features are ranked and then selected so as to offer superior classification accuracy. In the first stage, the features are ranked by the relief algorithm of filter model. In the second stage, the features that exceed the given threshold value are chosen for further processing. They analysed and used shapley values to evaluate the contribution of features to the classification task in the ranked feature subset. J. Sanchez-Monedero et al. [20] investigated the suitability of Extreme Learning Machines (ELM) for resolving bio-informatics and biomedical classification problems.

Genetic Algorithm (GA), one of the universally used contemporary stochastic techniques for global search, has been well known for its ability to generate high quality solutions in a given tractable time even for intricate problems. It has been applied for feature selection process in numerous applications and proved its worth [25]. Contrarily, owing to the intrinsic nature of the GA evolution process, it consumes long time to find the local optimum point of convergence in the solution space and unfortunately sometimes could not locate the optimal solution point with adequate precision. A recommended way of alleviating the local optimal problem is to ensemble GA with few memetic operations (also called as local search operations).

Another upcoming approach to feature selection is the use of game theory. The major benefit of the game theory approach is the facility to calculate a numerical indicator, i.e. a relevance index, which denotes the relevance of each feature under a specific condition. It can be used to analyse performance of the other features under the same condition. Further, the game theory approach developed presents a new characterization for the shapley value, i.e., it is more context sensitive and thus justifies its relevance for ranking/indexing the features.

After reviewing the works on feature selection for medical applications, it is observed that most of the existing methods suffer from the following problems: (1) depending on the complexity of the search method, the iterations of evaluations are too large; (2) they rely on a univariate ranking that does not take into account interaction among the variables already included in the selected subset and the remaining ones. Moreover, a method that produces the maximum accuracy employs more number of features and hence more running time is involved in the construction of the respective classifier. It is also true that a method with the fewest features produces inferior detection accuracy. A holistic and universal method that achieves the maximum classification accuracy with fewest features possible is still an open research problem. This paper makes an attempt to design such a feature selection sequence, called SVEGA that achieves a good trade-off between the number of features selected and detection accuracy.

This paper is organised as follows: Section 3 describes the proposed method and the algorithm. Experimental results and discussions are presented in Section 4. The paper is concluded with a mention on the future scope of this work.

3. System and methodology

3.1 Shapley values

Shapley Value Analysis (SVA) has been proved to be a promising strategy for feature selection process. Shapley Value Analysis [17, 11] is a game theory based technique for causal function localization that addresses the issue in describing and calculating the contributions made by the interactions among the group of elements in a data set with multiple features and their corresponding performance scores.

Consider a set of players denoted as $\mathcal{N}$. Let $\mathcal{N} = |\mathcal{N}|$ be the number of players in this set. Any non-empty set $S \subseteq N$ is referred as a coalition of players. Each coalition has a worth function denoted as $v(S)$, which calculates the total profit produced by the service when all the players in this coalition S are active. The $v(S)$ is given in Eq. (1) as follows:

$$v(S) = \sum_{i \in S} P_i(S), \tag{1}$$

where $P_i(S)$ represent the profit of player i in the coalition S.

Shapley in [3, 5] presented the value as an operator that assigns an expected marginal contribution to each player in the game with respect to a uniform distribution over the set of all permutations on the set of players. Specifically, let $\prod$ be a permutation (or an order) on the set of players, i.e., a mapping exists as one-to-one function from N onto N, and let us imagine the players appearing one by one to collect their payoff according to the order $\prod$. The marginal contribution Δ_i of player i to a coalition S is given in Eq. (2) as follows:

$$\Delta_i(s) = v(s \cup \{i\}) - v(s), \tag{2}$$

where v denotes the function which associates with every non-empty subset S of F, a real number $v(s)$ (the value of S) with $v(\{\varphi\}) = 0$. The unbiased estimator for the shapley value, for a player i is given by the mean of marginal contributions to all possible coalitions of players in N, is given in Eq. (3) as follows:

$$\Phi_i(v) = \frac{1}{n!} \sum_{x \in \prod} \Delta_i(S_i(\pi)), \tag{3}$$

where $\prod$ denotes set of permutations over N and $S_i(\pi)$ is the set of players from π that appear before player i in the permutation.

The feature selection process can be analogously seen as a coalition game where many features cooperate among themselves to achieve optimal performance in a particular task like classification, in our case. Here, the set N represents all the features, n represents the individual features and $v(S)$ stands for the accuracy metric obtained by the classifier using a subset of features S. Evaluation of features using the shapley value involves testing on all possible combinations of subsets of features.

3.2 GA in feature selection

3.2.1 Chromosome encoding

For a GA to search the optimal features in the solution space efficiently, both the chromosome encoding and design of fitness function has to be executed carefully. As far as the bio medical datasets are concerned, a natural encoding prevails over the feature space. Hence a fixed-length binary string encoding in which the value of the i-th gene $\{0, 1\}$ indicates whether or not the *i-th* feature (i = 1 to 'n', where 'n' represents the total number of features) from the overall feature set is included in the specified feature subset. Thus, each individual chromosome in the GA population is of fixed length, i.e., n-bit binary string, representing the respective subset of the given feature set (a gene value '1' denotes that the corresponding feature is selected and '0' denotes that the corresponding feature is omitted. This encoding is preferred and advantageous since it is a standard representation and the basic GA model can be used as such, without any further modification.

3.2.2 Fitness function

Each chromosome of the current GA population denotes a competing feature subset which has to be evaluated by their fitness so as to provide input to the ANN classifier model. This is achieved by calling on the classifier with the particular feature subset and the medical training dataset (which includes only the selected features corresponding to that feature subset). The ANN classifier constructed is then evaluated for its performance on a set of unseen test data. Hence this work is aimed to enhance the detection accuracy of the ANN classifier model that is ultimately achieved by maximizing the sensitivity and specificity of the ANN classifier. Consequently this knowledge is imparted into the model via the GA fitness function. The fitness function is formulated in Eq. (4) as follows:

$$\text{Fitness}(c) = \text{Max}(\text{Obj\_Fun}(SF_c)), \tag{4}$$

where SF_c denotes the Selected Feature subset encoded by a given chromosome c, and the objective function for feature selection Obj_Fun(SF_c) calculates the contribution of the given feature subset SF_c in Eq. (5) as follows:

$$\text{Obj\_Fun}(SF_c) = \alpha(1/\tau) + \beta(\text{Sensitivity}(SF_c)) + \gamma\,(\text{Specificity}(SF_c)), \tag{5}$$

where τ = No. of ones in the SF_c. The sensitivity value attained with the selected feature subset encoded by a given chromosome 'c' is denoted as

$$\text{Sensitivity}(SF_c) = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{6}$$

Also the specificity value attained with the selected feature subset encoded by a given chromosome 'c' is denoted as

$$\text{Specificity}(SF_c) = \frac{\text{TN}}{\text{TN} + \text{FP}}, \tag{7}$$

where TP and TN are the number of instances which are correctly classified in 'healthy' and 'affected' classes respectively. Likewise FP and FN are the number

of instances which are incorrectly classified in 'healthy' and 'affected' classes respectively. We use the number of features and classification accuracy, represented through specificity and sensitivity, as the metrics in our Obj_Fun (SF_c). The former metric has to be minimized and the latter one needs to be minimized. i.e. minimum number of features and maximum accuracy. A higher value of specificity and sensitivity leads to improved detection accuracy. Further, for τ, the number of one's present in the chromosome, a lesser τvalue is preferred over a higher τ. So, the fitness function has been designed to include $(1/\tau)$ as its component, which has to be maximized. Now all the components of fitness function have a common goal, i.e., to be maximized. Weight values are distributed among the three components - number of features, specificity, and sensitivity like $\alpha = 0.4$, $\beta = 0.3$, and $\gamma = 0.3$. If two subsets attain the same performance, while having different number of features, then the chromosome with smaller number of selected features is favored with higher survival priority and is carried over to the next generation. This strategy is preferred in a feature classification problem, where a subset of features with fewer features giving higher classification accuracy is better over a subset of features with more features giving lower or equal classification accuracy.

3.2.3 Genetic operators

The genetic operators like crossover, mutation and selection applied are that of the general simple GA's i.e., ranking selection, restrictive crossover and mutation with elitism. In each of the GA generation, the elite chromosome, i.e., the chromosome having the best fitness value is selected and subjected to shapley value based memetic operators as a part of the Lamarckian learning process. The Lamarckian learning [22] brings improvement in the result by placing the locally improved individual genes back into the population pool so that they acquire the reproductive opportunities. We define two memetic operators in the SVEGA, namely an 'include' operator which includes/adds a feature to the elite chromosome, and a 'remove' operator which removes/omits the existing features from the elite chromosome. The key issue is deciding which features to include and which ones to omit. Preferably, the features to be removed will be the ones which provide the least contribution when considered as a whole set and the ones which provide highest contribution must be included into the solution feature subset. This characteristic has to be brought in the existing GA paradigm. This requirement is fulfilled by the use of shapley value concept. For a given chromosome encoding c of a selected subset, let Q and R be the sets of selected and omitted features encoded in c, respectively. The function of the 'include' operator is to identify and select the feature with maximum shapley score when measured in coalition, from set R and to push it to the set Q. On the other hand, the 'remove' operator serves to identify and select the features with minimum contribution score and deletes from set Q and moves that into the set R. The pseudo code of these memetic operators is outlined in Algorithm 1 and Algorithm 2.

A notable point is that the shapley measure for each feature (i.e., step (1) in Algorithm 1 and Algorithm 2) needs to be calculated only once. Then this feature ranking information is stored for use inside include and remove operators, for fine tuning the entire search of solution space by the GA process.

Algorithm 1 Memetic operator -'include' Algorithm.

BEGIN

Rank the features in $\mathbf{R}$ in decreasing order of their Shapley values. Select a feature $\mathbf{R_i}$ in $\mathbf{R}$ by linear ranking selection in such a way that a feature with larger shapley value of a feature in $\mathbf{R}$ is more likely to be selected. Add $\mathbf{R_i}$ to $\mathbf{Q}$.

END

Algorithm 2 Memetic operator – 'remove' Algorithm.

BEGIN

Rank the features in $\mathbf{Q}$ in decreasing order of Shapley value. Select a feature $\mathbf{Q_i}$ in $\mathbf{Q}$ by linear ranking selection in such a way that a feature with larger Shapley value of a feature in $\mathbf{Q}$ is more likely to be selected. Eliminate all the features in $\mathbf{Q} - \{\mathbf{Q_i}\}$.

END

The computational complexity of these two memetic operators can be quantified according to the search range L, which specifies the upper bound for both 'include' and 'remove'. Therefore, with 'L' possible include operations and 'L' possible remove operations, we get a total of 'L^2' possible combinations of include and remove operations executed on a chromosome. The 'L^2' combinations of 'include' and 'remove' are executed on the candidate chromosome in a random sequence and once an improvement is seen either in the fitness value or reduction is seen in the number of selected features without decline in the fitness value, the procedure is stopped.

The pseudo code of the shapley value embedded memetic operation executed on the elite chromosome of each of the GA generation is outlined in Algorithm 3. After executing the above given Lamarckian learning process over the elite chromosome, the GA population then goes through the typical evolutionary operations like

Algorithm 3 Algorithm for Shapley value based Memetic operation.

BEGIN
Select the elite chromosome c_e to undergo memetic operations.
for $j = 1$ to L^2 **do**
  Generate a unique random pair of values $\{i, r\}$ where $0 \leq i, r \leq L$.
  Apply 'i' times include on the elite chromosome c_e and generate a new chromosome $c_{e'}$.
  Apply 'r' times remove on $c_{e'}$ and generate a new chromosome $c_{e''}$.
  Calculate the fitness of new modified chromosome $c_{e''}$ based on $Obj\_Fun(SF_c)$.
  if $c_{e''}$ is better than c_e either on fitness value or the number of features. **then**
    Replace the genotype c_e with $c_{e''}$ and stop applying the memetic operation.
  end if
end for
END

linear ranking selection, restrictive crossover, and mutation operators with elitism. Since we had a prior knowledge on the optimum number of features for certain datasets, we allowed the integration of such information into our proposed SVEGA by limiting the number of '1' bits in each of the chromosome to a maximum of 'm' ('m' is safely chosen to greater than the optimum number of features) during the evolutionary search process. To facilitate this aspect, we employed restrictive crossover operator and mutation rather than the conventional evolutionary operators of GA, so that the number of '1' bits occurring in each chromosome does not break the constraint imposed by the prior knowledge on 'm' during the search.

3.2.4 Proposed SVEGA algorithm for feature selection

The algorithm of the entire model is given below in Algorithm 4.

Algorithm 4 Shapley value embedded genetic Algorithm (SVEGA) for feature selection.

Input: Encoded n-bit binary string (where n is the number of features), number of generations gencount, population size (PS), crossover probability (P_c), mutation probability (P_m).
Output: A set of selected features, that has lower cardinality and yields higher sensitivity and specificity values.
BEGIN
Randomly generate an initial population, which denoted SF_c of size PS encoded with n-bit binary string. Each gene value can be '0' or '1'. (A gene value of '1' means, the feature at that position is selected and a value of '0' means, the feature at that position is omitted).
Initialize $\alpha = 0.4$, $\beta = 0.3$ and γ =0.3, M (total number of records in the training set), Pc and Pm.
while (not Current_fitness= Previous_fitness <0.0001 or gencount is not reached) **do**
  Apply restrictive cross over and mutation operator to the chromosome at the specified probability Pc and Pm.
  Evaluate the fitness value of all chromosomes in the population according to
  Obj_Fun(SF_c) = $\alpha(1/\tau) + \beta$(Sensitivity(SF_c)) + γ(Specificity(SF_c))
  where τ =No. of ones in the SF_c.
  Select the elite chromosome c_e and subject it to Shapley value based memetic operations.
  Replace the elite chromosome c_e with improved new chromosome $c_{e''}$ by Lamarckian-Learning process.
end while
END

3.3 Classifiers and its performance metric

In order to evaluate the efficiency of the proposed SVEGA method, the selected features by this method are evaluated using three successful classifiers such as Naïve Bayes (NB), J48 and Artificial Neural Network (ANN) [1].The classification models

are evaluated for performance metrics including accuracy, number of features, and CPU running time.

A binary classifier model has two discrete outputs – positive class and negative class. The performance metrics of the classifier is calculated by classification accuracy as

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}.$$

4. Experimental results and discussion

4.1 Biomedical datasets

The proposed approach has been evaluated on 26 biomedical datasets (both synthetic and microarray) from the UCI machine learning repository [8] and Kentridge repository [10]. These datasets are diverse in nature i.e. they cover small, medium and high dimensional dataset. Tab. I summarises the 26 biomedical datasets. The dataset categorisation is done like: 2–30 features as small dimensional datasets, 31–1000 features as medium dimensional datasets and >1000 features as high dimensional datasets. The generalisation ability of the proposed method is tested and proved by executing on all these categories of datasets. The proposed algorithm produces equally good classification accuracy on all these datasets.

4.2 Experimental set-up

The tests are carried out in a high end system with Intel i7, 1TB RAM, DDR3, 500GB hard drive on a Windows XP operating system. The proposed algorithm is implemented in WEKA environment [28]. WEKA is acknowledged as a landmark system in the field of machine learning and data mining. It has attained widespread acceptance among the academia and industry community, and has become a widely used tool for data mining research. Another flavour that is highly encouraging is its "Open Source" nature. The free access given to the source code has enabled us to develop and customize the modules for our work. The stepwise approach is as follows. The input to the system is given in the Attribute-Relation File Format (ARFF). The proposed algorithm is executed and the selected optimal features are obtained as the output. A result is created in WEKA using the name specified in \@relation". The attributes specified under \@attribute" and instances specified under \@data" are retrieved from the ARFF file and then they are added to the created table. 10-fold cross validation is performed for all the classifiers. Fifty runs were done for each classification algorithm on each dataset with features selected by SVEGA method. In each run, the dataset is split in the ratio of 80:20, into training and testing set. Ten runs of genetic algorithm were executed. As a whole, the execution of the ensemble SVEGA and classifier model is an iterative procedure (SVEGA–ANN procedure). Each run results in a complete diagnosis model. After 10 runs, the classifier model with the highest Sensitivity and Specificity and has minimum number of features is identified to be the best classifier model. The following parameter setting is adapted in our SVEGA: Population size (PS): 50, Number of generations *gencount*: 100, Probability of crossover (P_c): 0.6, Probability of mutation (P_m): 0.005.

The memetic operation range 'L' in SVEGA is empirically set to 4 (this value gave the best results on all system constraints). These configurations are consistently maintained in our experiments on all the 26 synthetic and microarray datasets.

4.3 Test results and discussion

The comparison of the proposed SVEGA method against state-of-the-art feature selection systems [3, 26, 19] has been carried out with conventional classifiers using the above specified objectives as the main metrics. The three main objectives such as number of features selected, classifier accuracy on the selected feature subset and the running time has been recorded in Tab. II. It could be observed that SVEGA outperforms other methods. Tab. III shows the precision, specificity, sensitivity and F-measure values obtained using features selected by the proposed system.

Empirical results on few representative datasets are shown in Tab. II and Tab. III while the complete results are shown under Appendix (continuation of Tab. II and Tab. III).

Based on the results obtained in Tab. IV and Tab. V, we conclude that the ANN produces comparably best results than the other conventional methods in terms of accuracy. These results show that by using the proposed SVEGA feature selector, high detection accuracies can be achieved with relatively small number of features. Fig. 1 shows the number of features selected by the proposed SVEGA and other existing systems on $\log_2$ scale. From the graph, it has been observed that SVEGA method outperforms the other existing methods by selecting minimum number of features possible.

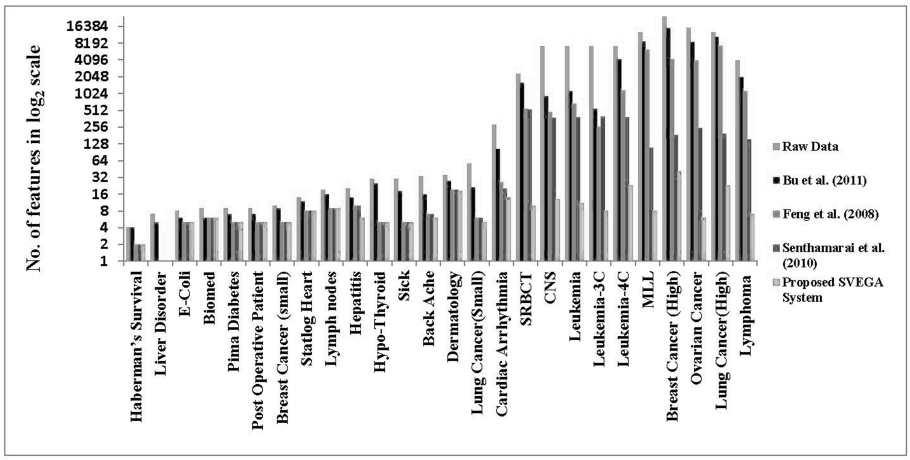

Fig. 1 *Features selected* ($\log_2$ *scale) by the proposed SVEGA method and other existing methods on Bio-Medical dataset.*

An increase of 93.12% from 82.37% is obtained on NB classifier, 92.02% from 78.38% on J48 and 93.88% from 82.42% on ANN models. SVEGA coupled with ANN achieves the maximum accuracy. However this advantage comes at the cost

S.No.	Dataset	Samples	Features	Classes	S.No.	Dataset	Samples	Features	Classes
1	Haberman's Survival	306	4	2	14.	Dermatology	366	35	6
2	Liver Disorder	345	7	2	15.	Lung Cancer	32	57	3
3	E-Coli	336	8	8	16.	Cardiac Arrhythmia	452	280	16
4	Biomed	209	9	2	17.	SRBCT	83	2309	4
5	Pima Diabetes	768	9	2	18.	CNS	60	7130	2
6	Post Operative Patient	90	9	3	19.	Leukemia	72	7130	2
7	Breast Cancer (SD)	286	10	2	20.	Leukemia-3C	72	7130	3
8	Statlog Heart	270	14	2	21.	Leukemia-4C	72	7130	4
9	Lymph nodes	148	19	4	22.	MLL	72	12583	3
10	Hepatitis	155	20	2	23.	Breast Cancer (HD)	97	24482	4
11	Hypo-Thyroid	3772	30	4	24.	Ovarian Cancer	253	15155	2
12	Sick	3772	30	2	25.	Lung Cancer	203	12601	5
13	Back Ache	180	33	2	26.	Lymphoma	66	4027	3

Tab. I *Bio-medical dataset description.*

Datasets (instances, features, classes)	Proposed and Existing Method	Classifier Performance Measures							
		NB			J48			ANN	
		Accuracy	Features	Running Time [s]	Accuracy	Features	Running Time [s]	Accuracy	Features	Running Time [s]
SMALL DIMENSIONAL DATASETS										
Haberman's Survival (306,4,2)	Raw Data	76.14	4	0.04	72.87	4	0.06	66.56	4	1.23
	Bu et al. (2009)[3]	75.16	4	0.08	72.87	4	0.04	71.89	4	2.01
	Feng et al. (2008)[26]	76.12	2	0.08	74.89	2	0.12	76.56	2	3.4
	Senthamarai et al. (2010) [22]	79.86	2	0.06	78.23	2	2.5	81.43	2	5.46
	Proposed SVEGA Method	89.56	2	0.05	83.46	2	0.19	86.56	2	4.1
MEDIUM DIMENSIONAL DATASETS										
Back Ache (180,33,2)	Raw Data	78.33	33	0.42	79.44	33	0.42	85.55	33	7.13
	Bu et al. (2009)[3]	78.45	16	0.36	86.11	16	0.4	86.12	16	6.48
	Feng et al. (2008)[26]	84.44	7	0.58	88.23	7	0.37	88	7	5.46
	Senthamarai et al. (2010) [22]	88.79	7	0.73	91.56	7	0.58	94.56	7	6.14
	Proposed SVEGA Method	90.48	6	0.6	95.68	6	0.45	97.22	6	5.0
HIGH DIMENSIONAL DATASETS										
SRBCT (83,2309,4)	Raw Data	98.79	2309	580	84.33	2309	670	79.56	2309	549
	Bu et al. (2009)[3]	93.13	1568	514	85.79	1568	526	80.40	1568	456
	Feng et al. (2008)[26]	95.48	546	445	89.45	546	456	83.87	546	423
	Senthamarai et al. (2010) [22]	100	526	375	92.56	526	516	85.69	526	345
	Proposed SVEGA Method	100	10	350	93.97	10	462	86.74	10	326

Tab. II *Classifier performance (Accuracy, Features selected, Running Time) on the features selected by the existing systems and proposed SVEGA method.*

Datasets	Proposed and Existing Method	Classifier Performance Measures											
		NB				J48				ANN			
		Prec.	Sens.	Spec.	F-Meas.	Prec.	Sens.	Spec.	F-Meas.	Prec.	Sens.	Spec.	F-Meas.
SMALL DIMENSIONAL DATASETS													
Haberman's	Raw Data	0.738	0.761	0.761	0.719	0.686	0.729	0.729	0.689	0.648	0.623	0.623	0.623
Survival	Bu et al. (2009) [3]	0.721	0.752	0.752	0.706	0.672	0.729	0.729	0.667	0.670	0.719	0.719	0.676
(306,4,2)	Feng et al. (2008)[26]	0.745	0.766	0.766	0.712	0.748	0.789	0.789	0.745	0.754	0.745	0.745	0.765
	Senthamarai et al. (2010) [22]	0.768	0.777	0.777	0.726	0.756	0.792	0.792	0.762	0.789	0.765	0.765	0.745
	Proposed SVEGA Method	0.796	0.778	0.788	0.756	0.879	0.856	0.856	0.832	0.794	0.768	0.768	0.768
MEDIUM DIMENSIONAL DATASETS													
Back Ache	Raw Data	0.807	0.783	0.783	0.794	0.776	0.794	0.794	0.785	0.827	0.856	0.856	0.835
(180,33,2)	Bu et al. (2009) [3]	0.815	0.784	0.783	0.792	0.742	0.861	0.861	0.797	0.835	0.868	0.868	0.842
	Feng et al. (2008)[26]	0.834	0.844	0.844	0.839	0.856	0.879	0.879	0.879	0.847	0.896	0.896	0.896
	Senthamarai et al. (2010) [22]	0.742	0.861	0.858	0.857	0.923	0.912	0.912	0.889	0.889	0.912	0.912	0.912
	Proposed SVEGA Method	0.748	0.889	0.887	0.864	0.968	0.956	0.956	0.912	0.973	0.972	0.972	0.971
HIGH DIMENSIONAL DATASETS													
SRBCT	Raw Data	0.988	0.988	0.988	0.988	0.842	0.843	0.843	0.842	0.745	0.741	0.741	0.741
(83,2309,4)	Bu et al. (2009)[3]	0.843	0.831	0.831	0.832	0.856	0.856	0.861	0.876	0.810	0.804	0.804	0.799
	Feng et al. (2008)[26]	0.826	0.824	0.824	0.831	0.915	0.926	0.926	0.889	0.841	0.839	0.839	0.840
	Senthamarai et al. (2010) [22]	0.856	0.854	0.854	0.845	0.926	0.935	0.935	0.915	0.856	0.842	0.842	0.854
	Proposed SVEGA Method	1.000	1.000	1.000	1.000	0.944	0.940	0.940	0.939	0.872	0.867	0.867	0.868

Tab. III *Classifier performance (Precision, Sensitivity, Specificity, F-Measure) on the features selected by the existing systems and proposed SVEGA method.*

Classifiers	Raw Data	Bu et al. (2009)	Feng et al. (2008)	Senthamarai et al. (2010)	Proposed SVEGA System
Naïve Bayes (NB)	82.37	82.63	85.86	89.22	93.12
J48 (C4.5)	78.38	82.74	84.64	88.58	92.02
Artificial Neural Network (ANN)	82.42	85.7	87.97	91.82	93.88

Tab. IV *Classifier performance on reduced set of features by SVEGA feature selector.*

Classifiers	Raw Data	Bu et al. (2009)	Feng et al. (2008)	Senthamarai et al. (2010)	Proposed SVEGA System
Naïve Bayes (NB)	148.4	147.4	125.49	127.83	108.76
J48 (C4.5)	152.8	139.2	137.51	149.17	126.94
Artificial Neural Network (ANN)	161.2	141.5	137.71	129.36	118

Tab. V *Average Running Time (sec.) of Conventional Classifiers on reduced set of features.*

of higher computational cost. Tab. V displays the average running time obtained on the test dataset over 10 runs of executing SVEGA.It could be observed that ANN consumes a bit higher time than Naïve Bayes.

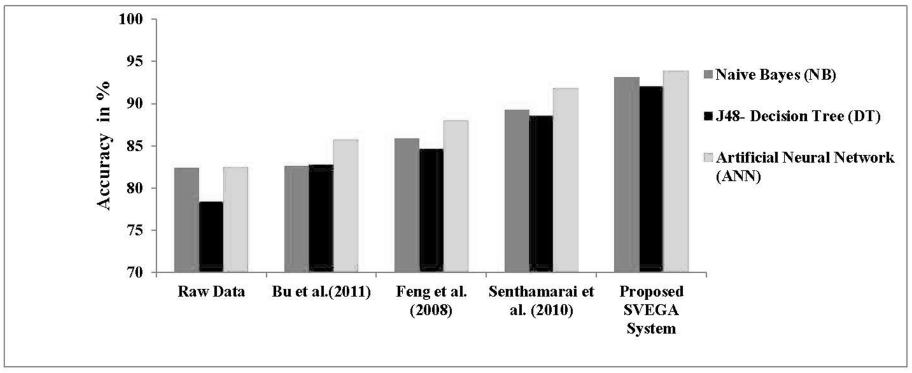

Fig. 2 *Average Accuracy obtained by the Classifier models.*

These results confirm that the proposed methodology is the best fit for improving classification accuracy through the process of feature selection for biomedical datasets.

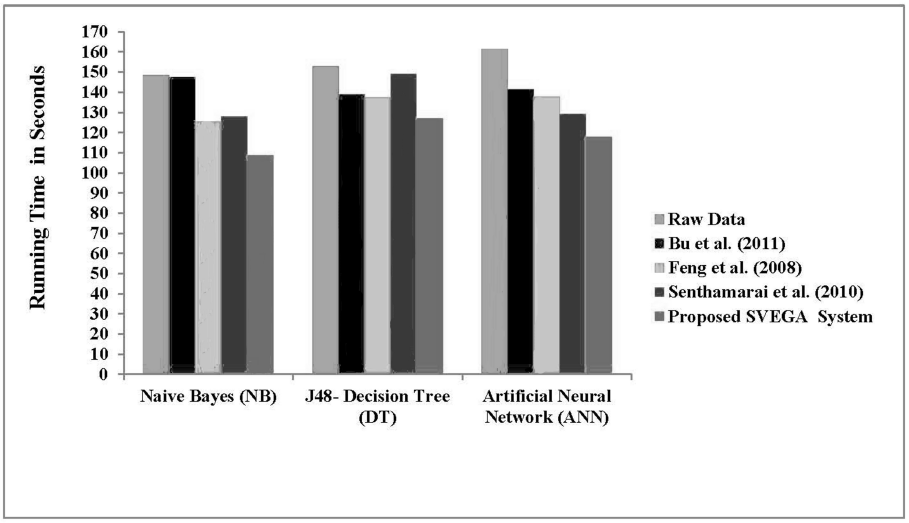

Fig. 3 *Average Running Time consumed by the Classifier models.*

5. Conclusion

This work proposes a feature selection strategy integrating GA and shapley value that enhances classification accuracy of the ANN model. The system is superior to the existing methods in two crucial perspectives such as reduction in the number of features and improvement in classification accuracy, precision, sensitivity, speficity, etc. The proposed FS method SVEGA is evaluated with three already existing systems using three successive classifiers. Experiments conducted on 26 medical datasets summarizes the characteristics of this proposed method with various performance metrics like accuracy, number of features selected and running Time (sec). Precision, Sensitivity, Specificity, F-Measure has been observed that the proposed system performs well even when the dataset has different number of samples, features and classes. This justifies that the proposed features and learning paradigm SVEGA-ANN is a promising strategy to be applied on any data classification problem.

Acknowledgement

This work is supported in part by the University Grant Commission (UGC), New Delhi, INDIA – Major Research Project under grant no. F.No.: 39-899/**(2010)** (SR).

References

[1] BAYRAM B., ACAR U., KOCA H.K., NARIN B., CAVDAROGLU G.C., CELIK L., CUBUK R. An Efficient Algorithm For Automatic Tumor Detection In Contrast Enhanced Breast MRI by Using Artificial Neural Network (NEUBREA). *Neural Network World*. 2013, 5(13), pp. 483–498, doi: 10.14311/NNW.2013.23.031.

[2] BERMEJO P., OSSA L.D.L., GAMEZ J.A., PUERTA J.M. Fast wrapper feature subset selection in high-dimensional datasets by means of filter re-ranking. *Knowledge-Based Systems*. 2012, 25(1), pp. 35–44, doi: 10.1016/j.knosys.2011.01.015.

[3] BU HUALONG, SHANGZHI ZHENG, JING XIA. Genetic algorithm based Semi-feature selection method. In: JOE ZHANG, GUOZHENG LI, JACK Y.YANG.,eds. *Proceedings of International Joint Conference on Bioinformatics, Systems Biology and Intelligent Computing(IJCBS'09),* Shanghai, IEEE, 2009. pp. 521-524, doi: 10.1109/IJCBS.2009.38.

[4] CHEN Y., YU S. Selection of effective features for ECG beat recognition based on nonlinear Correlations. *Artificial Intelligence in Medicine*, 2012, 54(1), pp. 43–52, doi: 10.1016/j.artmed.2011.09.004.

[5] COHEN S.B., RUPPIN E, DROR G. Feature Selection Based on Shapley Value [online]. In: Q. YANG., M. WOOLDRIDGE., eds. *Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI)*, Edinburgh, Scotland. IJCAI Organization, 2005, pp. 665−670 [viewed 2016-04-11]. Available from: http://www.cs.columbia.edu/~scohen/ijcai05features.pdf

[6] GAN J.Q., HASAN B.A.S., TSUI C.S.L. A filter-dominating hybrid sequential forward floating search method for feature subset selection in high-dimensional space. *International Journal of Machine Learning and Cybernetics*. 2012, 3(4), pp. 113–123, doi: 10.1007/s13042-012-0139-z.

[7] HAN Y., YU L. A Variance Reduction Framework for Stable Feature Selection. *Statistical Analysis and Data Mining*. 2012, 5(5), pp. 428−445, doi: 10.1002/sam.11152.

[8] HETTICH S., BLAKE C., MERZ C. UCI repository of machine learning databases. 1998 [accessed 2014-01-06]. Available from: http://www.ics.uci.edu/~mlearn/MLRepository.html

[9] JAZZAR M.M., MUHAMMAD G. Feature Selection Based Verification /Identification System Using Fingerprints and Palm Print. *Arabian Journal for Science and Engineering*. 2013, 38(4), pp. 849–857, doi: 10.1007/s13369-012-0524-7.

[10] JINYAN L., HUIQING L. Kent Ridge bio-medical data set repository. 2002 [accessed 2014-01- 06]. Available from: http://datam.i2r.a-star.edu.sg/datasets/krbd

[11] KEINAN A., SANDBANK B., HILGETAG C.C., ELLISON M.I., RUPPIN E. Fair attribution of functional contribution in artificial and biological networks. *Neural Computation*, 2004, 16(9), pp. 1887–1915, doi: 10.1162/0899766041336387.

[12] KOPRINSKA I. Feature Selection for Brain-Computer Interfaces. In: T. THEERAMUNKONG et al., eds. *Pacific Asia Conference on Knowledge Discovery and Data Mining (PAKDD '09)*, Bangkok, Thailand. Springer, 2009, pp. 106−117. Available from: http://link.springer.com/chapter/10.1007/978-3-642-14640-4_8{#}page-1

[13] LIN S., CHEN S. Parameter determination and feature selection for C4.5 algorithm using scatter search approach. *International Journal of Soft Computing*. 2011, 16(1), pp. 63–75, doi: 10.1007/s00500-011-0734-z.

[14] LU X., PENG X., LIU P., DENG Y., FENG B., LIAO B. A Novel Feature Selection Method Based on CFS in Cancer Recognition. In: CHEN L., ZHANG X., WU L., WANG Y, eds. *Proceedings of the 6th International Conference on Systems Biology (ISB),* Xi'an, China. IEEE, 2012, pp. 226–231, doi: 10.1109/ISB.2012.6314141.

[15] MARCEL J., MARCEL J. jr. GMDH Method with Genetic Selection Algorithm and Cloning. *Neural Network World*. 2013, 5(13), pp. 451–464, doi: 10.1007/978-3-642-01088-05.

[16] MONIRUL KABIR, SHAHJAHAN, KAZUYUKI MURASE. A new local search based hybrid genetic algorithm for feature selection. *International Journal of NeuroComputing*, Elsevier, 2011, 74(17), pp.2914–2928, doi: 10.1016/j.neucom.2011.03.034.

[17] MORETTI S., VAN LEEUWEN D., GMUENDER H., BONASSI S., VAN DELFT J., KLEINJANS J., PATRONE F., MERLO D.F. Combining Shapley value and statistics to the analysis of gene expression data in children exposed to air pollution. *BMC Bioinformatics*, 2008, 9(361), pp. 1−21.

[18] PAUL S., MAJI P. Rough Set Based Gene Selection Algorithm for Microarray Sample Classification. In: KUMAR, VIJAY, RAZA, ZAHID, eds. *Proceedings of the International Conference on Methods and Models in Computer Science (ICM2CS)*, New Delhi, India. IEEE, 2010, pp. 7−13, doi: 10.1109/ICM2CS.2010.5706710.

[19] RUCKSTIE T., OSENDORFER C., SMAGT P.V.D. Minimizing data consumption with sequential online feature selection. *International Journal of Machine Learning and Cybernetics*. 2012, 4(3), pp. 235−243, doi: 10.1007/s13042-012-0092-x.

[20] SANCHEZ-MONEDERO J., CRUZ-RAMÍREZ M., FERNÁNDEZ-NAVARRO F., FERNÁNDEZ J.C., GUTIÉRREZ P.A., HERVÁS-MARTÍNEZ C. On the suitability of Extreme Learning Machine for gene classification using feature Selection. In: A.E. HASSANIEN, A. ABRAHAM, F.,HAGRAS H. MARCELLONI, M. ANTONELLI, T. HONG, eds. *Proceedings of the 10th International Conference on Intelligent Systems Design and Applications (ISDA)*, Cairo, Egypt. IEEE, 2010, pp. 507−512, doi: 10.1109/ISDA.2010.5687215.

[21] SASIKALA S., APPAVU ALIAS BALAMURUGAN S., GEETHA S. Multi Filtration Feature Selection (MFFS) to improve discriminatory ability in clinical data set. *Applied Computing and Informatics*. 2014, pp. 1−20, doi: 10.1016/j.aci.2014.03.002.

[22] SENTHAMARAI KANNAN S., RAMARAJ N. A novel hybrid feature selection via Symmetrical Uncertainty ranking based local memetic search algorithm. *Knowledge-Based Systems*. 2010, 23(6), pp. 580–585, doi: 10.1016/j.knosys.(2010).03.016.

[23] SHEN Q., DIAO R., SU P. Feature Selection Ensemble. In: A.VORONKOV, ed. *EPiC Series in Computing*, 2011, 10(1), pp. 289−306.

[24] SMIALOWSKI P., FRISHMAN D., KRAMER S. Pitfalls of supervised feature selection. *Bioinformatics*. 2010, 26(3), pp. 440−443, doi: 10.1093/bioinformatics/btp621.

[25] STEIN G., CHEN B.,WU A.S., HUA K.A. Decision tree classifier for network intrusion detection with GA-based feature selection. In: M. GUIMARAES, ed. *Proceedings of the 43rd Annual Southeast Regional Conference*, 2(1), Kennesaw, Georgia, USA. ACM, 2005, pp. 136–141, doi: 10.1145/1167253.1167288.

[26] TAN F., FU X., ZHANG Y., BOURGEOIS A.G. A genetic algorithm-based method for feature subset selection. *Soft Computing*. 2008, 11(1), pp. 111–120, doi: 10.1007/s00500-007-0193-8.

[27] VINH L.T., LEE S., PARK Y., AURIOL B.J. A novel feature selection method based on normalized Mutual information. *International Journal of Applied Intelligence*. 2011, 37(1), pp. 100−120, doi: 10.1007/s10489-011-0315-y.

[28] WEKA 3. Weka 3: Data Mining Software in Java v.3.5.7. [software]. 2008 [accessed 2014-01-06]. Available from: http://www.cs.waikato.ac.nz/ml/weka/

[29] XIE J., WANG C. Using Support vector machines with a novel hybrid feature selection method for diagnosis of erythemato-squamous diseases. *Expert Systems with Applications*. 2011, 38(5), pp. 5809–5815, doi: 10.1016/j.eswa.(2010).10.050.

Appendix

Datasets (instances, features, classes)	Proposed and Existing Method	Classifier Performance Measures								
		NB			J48			ANN		
		Accuracy	Features	Running Time (sec.)	Accuracy	Features	Running Time (sec.)	Accuracy	Features	Running Time (sec.)
SMALL DIMENSIONAL DATASETS										
Haberman's Survival (306,4,2)	Raw Data	76.14	4	0.04	72.87	4	0.06	66.56	4	1.23
	Bu et al. (2009)[3]	75.16	4	0.08	72.87	4	0.04	71.89	4	2.01
	Feng et al. (2008)[26]	76.12	2	0.08	74.89	2	0.12	76.56	2	3.4
	Senthamarai et al. (2010) [22]	79.86	2	0.06	78.23	2	2.5	81.43	2	5.46
	Proposed SVEGA Method	89.56	2	0.05	83.46	2	0.19	86.56	2	4.1
Liver Disorder (345,7,2)	Raw Data	55.36	7	0.5	68.69	7	0.5	57.97	7	0.7
	Bu et al. (2009)[3]	55.36	5	0.8	68.72	5	0.2	74.78	5	0.89
	Feng et al. (2008)[26]	57.97	1	0.7	69.79	1	0.8	79.21	1	0.5
	Senthamarai et al. (2010) [22]	64.56	1	1.45	74.56	1	1.2	85.46	1	0.7
	Proposed SVEGA Method	84.89	1	1.2	76.89	1	1.0	87.21	1	0.6
E-Coli (336,8,8)	Raw Data	85.41	8	0.18	84.22	8	0.2	92.55	8	0.18
	Bu et al. (2009)[3]	88.45	6	0.2	84.33	6	0.1	93.56	6	0.25
	Feng et al. (2008)[26]	89.75	5	0.156	85.36	5	0.43	95.45	5	0.36
	Senthamarai et al. (2010) [22]	89.6	5	1.16	89.78	5	0.56	98.21	5	0.48
	Proposed SVEGA Method	96.01	5	0.94	95.63	5	0.49	99.78	5	1.55
Biomed (209,9,2)	Raw Data	90.43	9	0.4	89.47	9	0.2	99.04	9	1.25
	Bu et al. (2009)[3]	90.43	6	0.4	89.95	6	0.1	98.45	6	2.56
	Feng et al. (2008)[26]	91.53	6	0.8	93.56	6	0.36	98.63	6	3.9
	Senthamarai et al. (2010) [22]	94.89	6	0.5	96.53	6	0.56	99.21	6	4.56
	Proposed SVEGA Method	98.43	6	0.6	98.56	6	0.42	100	6	4.1
Pima Diabetes (768,9.2)	Raw Data	76.30	9	0.8	73.82	9	0.7	80.59	9	2.8
	Bu et al. (2009)[3]	76.29	6	0.7	74.86	7	0.6	82.78	6	2.5
	Feng et al. (2008)[26]	77.21	4	1.96	75.86	5	0.23	85.74	4	4.6
	Senthamarai et al. (2010) [22]	79.48	4	1.46	82.56	5	1.45	91.65	4	6.8
	Proposed SVEGA Method	89.78	4	1.23	85.69	5	0.96	94.86	4	5.1
Post Operative Patient (90,9,3)	Raw Data	66.66	9	0.04	70	9	0.05	92.22	9	1.5
	Bu et al. (2009)[3]	68.25	7	0.04	72.53	7	0.04	93.56	7	2.45
	Feng et al. (2008)[26]	71.56	5	0.02	74.56	5	0.12	94.78	5	3.52

Datasets (instances, features, classes)	Proposed and Existing Method	Classifier Performance Measures								
		NB			J48			ANN		
		Accuracy	Features	Running Time (sec.)	Accuracy	Features	Running Time (sec.)	Accuracy	Features	Running Time (sec.)
	Senthamarai et al. (2010) [22]	73.89	5	0.03	78.56	5	0.26	95.36	5	4.56
	Proposed SVEGA Method	75.69	5	0.01	82.56	5	0.18	96.12	5	4.12
Breast Cancer	Raw Data	71.67	9	0.42	75.52	10	1.56	65.23	10	2.3
(286,10,2)	Bu et al. (2009)[3]	71.89	9	0.4	76.07	9	1.23	68.45	9	3.1
	Feng et al. (2008)[26]	72.37	5	0.38	79.56	5	2.13	69.56	5	3.65
	Senthamarai et al. (2010) [22]	80.56	5	0.57	82.79	5	5.63	70.12	5	4.12
	Proposed SVEGA Method	84.56	5	0.45	86.79	5	3.56	71.67	5	3.9
Statlog Heart	Raw Data	83.70	14	0.2	76.66	14	0.35	97.40	14	2.35
(270,14,2)	Bu et al. (2009)[3]	81.11	12	0.4	78.81	12	0.5	97.51	12	3.25
	Feng et al. (2008)[26]	83.49	8	0.75	78.96	8	0.72	97.98	8	4.23
	Senthamarai et al. (2010) [22]	87.89	8	0.63	79.86	8	3.56	98	8	5.36
	Proposed SVEGA Method	88.15	8	0.58	81.56	8	2.13	98.37	8	4.35
Lymph nodes	Raw Data	83.10	19	0.02	77.02	19	0.09	99.32	19	3.56
(148,19,4)	Bu et al. (2009)[3]	84.45	16	0.02	79.05	16	0.10	99.56	16	3.34
	Feng et al. (2008)[26]	84.56	9	0.08	81.56	9	0.25	99.62	9	4.1
	Senthamarai et al. (2010) [22]	85.49	9	0.11	84.63	9	0.86	99.81	9	4.23
	Proposed SVEGA Method	87.18	9	0.09	86.89	9	0.53	100	9	3.65
Hepatitis	Raw Data	84.51	20	0.34	83.87	20	0.2	98.70	20	2.56
(155,20,2)	Bu et al. (2009)[3]	84.51	14	0.32	84.51	14	0.48	98.74	14	3.56
	Feng et al. (2008)[26]	85.78	10	0.28	85.12	10	0.56	98.98	10	1.25
	Senthamarai et al. (2010) [22]	86.89	10	0.32	89.96	10	0.78	98.52	10	3.56
	Proposed SVEGA Method	88.51	6	0.29	94.86	6	0.6	99.81	6	2.46
Hypo-Thyroid	Raw Data	95.28	30	5.18	99.57	30	6.12	95.94	30	6.35
(3772,30,4)	Bu et al. (2009)[3]	94.64	25	5.3	90.81	25	4.86	95.94	25	7.32
	Feng et al. (2008)[26]	96.48	5	4.78	92.53	5	3.45	95.92	5	5.36
	Senthamarai et al. (2010) [22]	97.23	5	5.1	98.56	5	4.96	96	5	7.42
	Proposed SVEGA Method	99.45	5	0.49	100	5	3.96	96.10	5	6.45
Sick	Raw Data	92.60	30	0.03	98.80	30	2.4	95.51	30	6.89
(3772,30,2)	Bu et al. (2009)[3]	90.19	18	0.02	98.87	18	2.0	95.62	18	5.89
	Feng et al. (2008)[26]	94.76	5	0.09	99.12	5	1.56	95.71	5	6.45
	Senthamarai et al. (2010) [22]	96.48	5	0.12	99.26	5	2.86	96	5	5.94
	Proposed SVEGA Method	98.78	5	0.1	99.87	5	2.12	96.02	5	5.21

Datasets (instances, features, classes)	Proposed and Existing Method	Classifier Performance Measures								
		NB			J48			ANN		
		Accuracy	Features	Running Time (sec.)	Accuracy	Features	Running Time (sec.)	Accuracy	Features	Running Time (sec.)
MEDIUM DIMENSIONAL DATASETS										
Back Ache (180,33,2)	Raw Data	78.33	33	0.42	79.44	33	0.42	85.55	33	7.13
	Bu et al. (2009)[3]	78.45	16	0.36	86.11	16	0.4	86.12	16	6.48
	Feng et al. (2008)[26]	84.44	7	0.58	88.23	7	0.37	88	7	5.46
	Senthamarai et al. (2010) [22]	88.79	7	0.73	91.56	7	0.58	94.56	7	6.14
	Proposed SVEGA Method	90.48	6	0.6	95.68	6	0.45	97.22	6	5.0
Dermatology (366,35,6)	Raw Data	97.26	35	0.58	93.98	35	0.89	95.63	35	7.56
	Bu et al. (2009)[3]	97.81	28	0.48	95.35	28	0.78	97.26	28	6.23
	Feng et al. (2008)[26]	97.83	19	0.68	96.12	19	0.96	97.56	19	5.45
	Senthamarai et al. (2010) [22]	98.35	19	0.82	97.56	19	1.56	99	19	6.93
	Proposed SVEGA Method	98.56	18	0.75	98.12	18	1.2	100	18	5.55
Lung Cancer (32,57,3)	Raw Data	78.12	57	0.15	78.125	57	2.1	87.5	57	5.26
	Bu et al. (2009)[3]	78.24	21	0.15	81.25	21	1.98	94.56	21	6.14
	Feng et al. (2008)[26]	80.14	6	0.09	82.23	6	1.63	97.26	6	6.23
	Senthamarai et al. (2010) [22]	83.45	6	0.12	84.37	6	2.75	99.81	6	4.23
	Proposed SVEGA Method	87.5	5	0.1	92.56	5	1.96	100	5	0.01
Cardiac Arrhythmia (452,280,16)	Raw Data	62.38	280	0.06	64.38	280	5.63	67.25	280	3.56
	Bu et al. (2009)[3]	66.15	103	0.03	66.83	103	4.86	79.42	103	4.28
	Feng et al. (2008)[26]	74.45	26	1.4	71.56	26	3.56	81.56	26	3.65
	Senthamarai et al. (2010) [22]	78.45	20	1.9	75.63	20	6.89	84.15	20	4.12
	Proposed SVEGA Method	84	14	1.5	82.49	14	5.62	87.5	14	3.8
HIGH DIMENSIONAL DATASETS										
SRBCT (83,2309,4)	Raw Data	98.79	2309	580	84.33	2309	670	79.56	2309	549
	Bu et al. (2009)[3]	93.13	1568	514	85.79	1568	526	80.40	1568	456
	Feng et al. (2008)[26]	95.48	546	445	89.45	546	456	83.87	546	423
	Senthamarai et al. (2010) [22]	100	526	375	92.56	526	516	85.69	526	345
	Proposed SVEGA Method	100	10	350	93.97	10	462	86.74	10	326
CNS (60,7130,2)	Raw Data	61.66	7130	92	58.33	7130	99.56	75.36	7130	254
	Bu et al. (2009)[3]	65.46	918	95	61.56	918	91.56	78.63	918	124.2
	Feng et al. (2008)[26]	68.75	475	74.5	64.89	475	112.5	81.56	475	256.3
	Senthamarai et al. (2010) [22]	97.78	374	86	79.56	374	124.6	92.63	374	289
	Proposed SVEGA Method	86.66	13	80.2	85	13	115.6	95	13	265
Leukemia	Raw Data	98.61	7130	261.5	83.33	7130	321.3	58.33	7130	356.1

Datasets (instances, features, classes)	Proposed and Existing Method	Classifier Performance Measures								
		NB			J48			ANN		
		Accuracy	Features	Running Time (sec.)	Accuracy	Features	Running Time (sec.)	Accuracy	Features	Running Time (sec.)
(72,7130,2)	Bu et al. (2009)[3]	98.63	1123	287	85.65	1126	256.3	62.48	1126	412.5
	Feng et al. (2008)[26]	98.75	678	118	89.78	678	309.4	68.23	678	389.6
	Senthamarai et al. (2010) [22]	99.56	387	189	92.56	387	456.8	71.23	387	398.6
	Proposed SVEGA Method	100	11	150	95	11	312.6	74.36	11	390.1
Leukemia-3C	Raw Data	94.44	7130	315.5	95.83	7130	356.4	95.63	7130	380
(72,7130,3)	Bu et al. (2009)[3]	95.63	540	380	93.33	540	389.1	97.15	540	315
	Feng et al. (2008)[26]	97.12	262	228	95.12	262	394	98.22	23	272
	Senthamarai et al. (2010) [22]	99.53	394	315	96.56	394	489	99.611	394	255.6
	Proposed SVEGA Method	100	8	255.6	98.12	8	412	100	8	250.2
Leukemia-4C	Raw Data	87.5	7130	296.12	87.5	7130	333.2	85.69	7130	345
(72,7130,4)	Bu et al. (2009)[3]	88.89	4156	345	89.56	4156	304.5	86.74	4156	326
	Feng et al. (2008)[26]	92.56	1146	268	91.23	1146	398.7	89.56	1146	345
	Senthamarai et al. (2010) [22]	98.61	386	350	94.04	386	412.5	92.56	386	324
	Proposed SVEGA Method	97.22	23	272	95.83	23	399.7	98.61	23	333.5
MLL	Raw Data	95.83	12583	395.8	84.72	12583	412.4	82.36	12583	423
(72,12583,3)	Bu et al. (2009)[3]	95.83	8756	368.4	85.79	8756	401.1	85.63	8756	345
	Feng et al. (2008)[26]	97.85	6306	358.9	90.15	6306	385.4	88.79	6306	256.4
	Senthamarai et al. (2010) [22]	100	108	378.56	92.56	108	478.5	94.56	108	312.5
	Proposed SVEGA Method	100	8	360	98.61	8	390.1	95.61	8	288.3
Breast Cancer	Raw Data	54.63	24482	714	60.82	24482	532	64.56	24482	595
(97,24482,4)	Bu et al. (2009)[3]	64.56	15478	654	77.56	15478	524	74.53	15478	586
	Feng et al. (2008)[26]	82.84	4236	631	83.12	4236	496	75.63	4236	573
	Senthamarai et al. (2010) [22]	92.56	183	542	88.56	183	451	89.56	183	499
	Proposed SVEGA Method	88.50	41	436	93.81	41	404	95.86	41	395
Ovarian Cancer	Raw Data	92.49	15155	541	95.65	15155	484	84.56	15155	492
(253,15155,2)	Bu et al. (2009)[3]	85.36	8546	537	94.37	8546	436	89.53	8546	348
	Feng et al. (2008)[26]	93.54	4012	491	85.56	4012	378	91.35	4012	323
	Senthamarai et al. (2010) [22]	100	247	482	95.71	247	363	96.36	247	314
	Proposed SVEGA Method	100	6	414	100	6	301	100	6	256
Lung Cancer	Raw Data	90.53	12601	412	88.56	12601	485	89.54	12601	512
(203,12601,5)	Bu et al. (2009)[3]	89.72	10546	404	91.05	10546	436	89.54	10546	495
	Feng et al. (2008)[26]	91.63	7342	424	91.05	7342	414	91.36	7342	486
	Senthamarai et al. (2010) [22]	94.61	192	397	92.31	192	378	97.35	192	375

Datasets (instances, features, classes)	Proposed and Existing Method	Classifier Performance Measures								
		NB			J48			ANN		
		Accuracy	Features	Running Time (sec.)	Accuracy	Features	Running Time (sec.)	Accuracy	Features	Running Time (sec.)
Lymphoma (66,4027,3)	Proposed SVEGA Method	98.52	23	365	95.56	23	364	99.86	23	352
	Raw Data	89.76	4027	241	88.64	4027	256	88.86	4027	231
	Bu et al. (2009)[3]	89.78	2015	238	85.64	2015	235	89.33	2015	212
	Feng et al. (2008)[26]	95.36	1115	211	91.36	1115	214	91.76	1115	194
	Senthamarai et al. (2010) [22]	98.61	153	194	94.36	153	172	98.76	153	176
	Proposed SVEGA Method	100	7	136	98.53	7	114	100	7	152

Tab. VI *Classifier performance (Accuracy, Features selected, Running Time) on the features selected by the existing systems and proposed SVEGA method.*

Datasets	Proposed and Existing Method	Classifier Performance Measures											
		NB				J48				ANN			
		Prec.	Sens.	Spec.	F-Meas.	Prec.	Sens.	Spec.	F-Meas.	Prec.	Sens.	Spec.	F-Meas.
Haberman's Survival (306,4,2)	Raw Data	0.738	0.761	0.761	0.719	0.686	0.729	0.729	0.689	0.648	0.623	0.623	0.623
	Bu et al. (2009)[3]	0.721	0.752	0.752	0.706	0.672	0.729	0.729	0.667	0.670	0.719	0.719	0.676
	Feng et al. (2008)[26]	0.745	0.766	0.766	0.712	0.748	0.789	0.789	0.745	0.754	0.745	0.745	0.765
	Senthamarai et al. (2010) [22]	0.768	0.777	0.777	0.726	0.756	0.792	0.792	0.762	0.789	0.765	0.765	0.745
	Proposed SVEGA Method	0.796	0.778	0.788	0.756	0.879	0.856	0.856	0.832	0.794	0.768	0.768	0.768
Liver Disorder (345,7,2)	Raw Data	0.609	0.554	0.554	0.544	0.683	0.687	0.687	0.680	0.547	0.580	0.580	0.436
	Bu et al. (2009)[3]	0.331	0.548	0.548	0.479	0.336	0.580	0.580	0.425	0.760	0.748	0.748	0.735
	Feng et al. (2008)[26]	0.336	0.580	0.580	0.425	0.389	0.612	0.612	0.438	0.776	0.759	0.759	0.748
	Senthamarai et al. (2010) [22]	0.489	0.615	0.615	0.489	0.456	0.678	0.678	0.489	0.785	0.768	0.768	0.759
	Proposed SVEGA Method	0.789	0.756	0.756	0.748	0.635	0.789	0.789	0.589	0.896	0.879	0.879	0.879
E-Coli (336,8,8)	Raw Data	0.861	0.854	0.854	0.854	0.832	0.842	0.842	0.836	0.915	0.926	0.926	0.920
	Bu et al. (2009)[3]	0.861	0.860	0.860	0.859	0.820	0.830	0.830	0.824	0.926	0.963	0.963	0.931
	Feng et al. (2008)[26]	0.874	0.862	0.862	0.861	0.831	0.845	0.845	0.836	0.945	0.976	0.976	0.941
	Senthamarai et al. (2010) [22]	0.877	0.863	0.863	0.867	0.879	0.863	0.863	0.845	0.958	0.986	0.986	0.954
	Proposed SVEGA Method	0.961	0.960	0.960	0.912	0.889	0.875	0.873	0.856	0.967	0.994	0.994	0.968

Datasets	Proposed and Existing Method	Classifier Performance Measures											
		NB				J48				ANN			
		Prec.	Sens.	Spec.	F-Meas.	Prec.	Sens.	Spec.	F-Meas.	Prec.	Sens.	Spec.	F-Meas.
Biomed	Raw Data	0.911	0.904	0.904	0.901	0.894	0.895	0.895	0.894	0.991	0.990	0.990	0.990
(209,9,2)	Bu et al. (2009)[3]	0.912	0.901	0.901	0.899	0.899	0.900	0.900	0.899	0.992	0.991	0.991	0.991
	Feng et al. (2008)[26]	0.916	0.903	0.903	0.856	0.912	0.935	0.935	0.923	0.992	0.991	0.991	0.991
	Senthamarai et al. (2010) [22]	0.911	0.903	0.903	0.889	0.963	0.952	0.952	0.946	0.989	0.996	0.996	0.995
	Proposed SVEGA Method	0.909	0.904	0.904	0.902	0.972	0.953	0.963	0.963	1.000	1.000	1.000	1.000
Pima	Raw Data	0.759	0.763	0.763	0.760	0.735	0.738	0.738	0.736	0.819	0.806	0.806	0.809
Diabetes	Bu et al. (2009)[3]	0.746	0.736	0.736	0.762	0.742	0.749	0.749	0.743	0.742	0.749	0.749	0.743
(768,9.2)	Feng et al. (2008)[26]	0.767	0.772	0.772	0.766	0.763	0.752	0.752	0.786	0.897	0.856	0.856	0.863
	Senthamarai et al. (2010) [22]	0.768	0.786	0.786	0.768	0.856	0.862	0.862	0.861	0.924	0.947	0.947	0.947
	Proposed SVEGA Method	0.861	0.849	0.849	0.816	0.864	0.876	0.876	0.854	0.956	0.968	0.968	0.968
Post	Raw Data	0.502	0.667	0.667	0.573	0.503	0.700	0.700	0.586	0.924	0.922	0.922	0.921
Operative	Bu et al. (2009)[3]	0.432	0.589	0.589	0.566	0.500	0.693	0.693	0.589	0.931	0.936	0.936	0.942
Patient	Feng et al. (2008)[26]	0.754	0.746	0.748	0.736	0.578	0.789	0.789	0.689	0.948	0.952	0.952	0.964
(90,9,3)	Senthamarai et al. (2010) [22]	0.548	0.689	0.689	0.678	0.689	0.799	0.799	0.715	0.949	0.958	0.958	0.958
	Proposed SVEGA Method	0.744	0.743	0.732	0.727	0.756	0.845	0.845	0.852	0.951	0.962	0.962	0.962
		SMALL DIMENSIONAL DATASETS											
Breast	Raw Data	0.704	0.717	0.717	0.708	0.752	0.755	0.755	0.713	0.665	0.685	0.685	0.685
Cancer	Bu et al. (2009)[3]	0.789	0.748	0.748	0.739	0.709	0.731	0.731	0.685	0.689	0.672	0.672	0.689
(286,10,2)	Feng et al. (2008)[26]	0.713	0.724	0.724	0.717	0.712	0.756	0.756	0.723	0.701	0.689	0.689	0.695
	Senthamarai et al. (2010) [22]	0.864	0.859	0.859	0.847	0.786	0.792	0.792	0.745	0.701	0.717	0.717	0.705
	Proposed SVEGA Method	0.857	0.867	0.869	0.859	0.880	0.849	0.849	0.856	0.701	0.717	0.717	0.705
Statlog Heart	Raw Data	0.837	0.837	0.837	0.837	0.766	0.767	0.767	0.767	0.804	0.804	0.804	0.862
(270,14,2)	Bu et al. (2009)[3]	0.811	0.811	0.811	0.810	0.748	0.748	0.748	0.746	0.856	0.861	0.861	0.861
	Feng et al. (2008)[26]	0.801	0.801	0.797	0.794	0.756	0.765	0.765	0.772	0.881	0.876	0.876	0.876
	Senthamarai et al. (2010) [22]	0.836	0.824	0.821	0.814	0.789	0.771	0.771	0.776	0.925	0.934	0.934	0.934
	Proposed SVEGA Method	0.843	0.830	0.831	0.829	0.845	0.856	0.856	0.861	0.974	0.974	0.974	0.974
Lymph nodes	Raw Data	0.832	0.831	0.831	0.830	0.776	0.770	0.770	0.772	0.993	0.993	0.993	0.993
(148,19,4)	Bu et al. (2009)[3]	0.848	0.845	0.845	0.846	0.801	0.791	0.791	0.794	0.993	0.993	0.993	0.993
	Feng et al. (2008)[26]	0.824	0.801	0.798	0.764	0.845	0.856	0.856	0.861	0.995	0.998	0.998	0.998
	Senthamarai et al. (2010) [22]	0.811	0.784	0.764	0.724	0.856	0.861	0.861	0.871	0.996	0.998	0.998	0.998
	Proposed SVEGA Method	0.846	0.839	0.839	0.835	0.868	0.896	0.896	0.896	1.000	1.000	1.000	1.000
Hepatitis	Raw Data	0.853	0.845	0.845	0.848	0.825	0.839	0.839	0.825	0.987	0.987	0.987	0.987
(155,20,2)	Bu et al. (2009)[3]	0.836	0.845	0.845	0.839	0.832	0.845	0.845	0.831	0.987	0.987	0.987	0.987
	Feng et al. (2008)[26]	0.810	0.783	0.782	0.714	0.856	0.863	0.863	0.864	0.998	0.998	0.989	0.989

Datasets	Proposed and Existing Method	Classifier Performance Measures											
		NB				J48				ANN			
		Prec.	Sens.	Spec.	F-Meas.	Prec.	Sens.	Spec.	F-Meas.	Prec.	Sens.	Spec.	F-Meas.
	Senthamarai et al. (2010) [22]	0.849	0.837	0.837	0.827	0.896	0.879	0.879	0.889	0.998	0.999	0.999	0.991
	Proposed SVEGA Method	0.886	0.875	0.875	0.839	0.739	0.787	0.787	0.746	0.996	0.998	0.998	0.998
Hypo-	Raw Data	0.946	0.953	0.953	0.945	0.995	0.996	0.996	0.995	0.955	0.959	0.959	0.957
Thyroid	Bu et al. (2009)[3]	0.934	0.946	0.946	0.935	0.971	0.968	0.968	0.969	0.955	0.959	0.959	0.957
(3772,30,4)	Feng et al. (2008)[26]	0.948	0.958	0.958	0.945	0.956	0.948	0.948	0.948	0.956	0.961	0.961	0.959
	Senthamarai et al. (2010) [22]	0.956	0.964	0.964	0.921	0.987	0.986	0.986	0.945	0.957	0.963	0.963	0.963
	Proposed SVEGA Method	0.961	0.967	0.967	0964	1.000	1.000	1.000	1.000	0.959	0.961	0.961	0.960
Sick	Raw Data	0.951	0.926	0.926	0.935	0.988	0.988	0.988	0.988	0.968	0.955	0.955	0.959
(3772,30,2)	Bu et al. (2009)[3]	0.814	0.902	0.902	0.856	0.881	0.939	0.939	0.909	0.969	0.967	0.967	0.967
	Feng et al. (2008)[26]	0.845	0.878	0.878	0.847	0.945	0.936	0.936	0.936	0.972	0.978	0.978	0.978
	Senthamarai et al. (2010) [22]	0.881	0.912	0.913	0.836	0.978	0.969	0.969	0.942	0.983	0.985	0.985	0.985
	Proposed SVEGA Method	0.907	0.923	0.921	0.831	0.981	0.979	0.979	0.959	0.963	0.960	0.960	0.961
		MEDIUM DIMENSIONAL DATASETS											
Back Ache	Raw Data	0.807	0.783	0.783	0.794	0.776	0.794	0.794	0.785	0.827	0.856	0.856	0.835
(180,33,2)	Bu et al. (2009)[3]	0.815	0.784	0.783	0.792	0.742	0.861	0.861	0.797	0.835	0.868	0.868	0.842
	Feng et al. (2008)[26]	0.834	0.844	0.844	0.839	0.856	0.879	0.879	0.879	0.847	0.896	0.896	0.896
	Senthamarai et al. (2010)	0.742	0.861	0.858	0.857	0.923	0.912	0.912	0.889	0.889	0.912	0.912	0.912
	[22]	0.748	0.889	0.887	0.864	0.968	0.956	0.956	0.912	0.973	0.972	0.972	0.971
	Proposed SVEGA Method												
Dermatology	Raw Data	0.974	0.973	0.973	0.973	0.940	0.940	0.940	0.940	0.968	0.978	0.978	0.978
(366,35,6)	Bu et al. (2009)[3]	0.978	0.978	0.978	0.978	0.955	0.954	0.954	0.954	0.973	0.983	0.983	0.983
	Feng et al. (2008)[26]	0.987	0.984	0.983	0.981	0.961	0.976	0.976	0.986	0.986	0.985	0.985	0.985
	Senthamarai et al. (2010)	0.986	0.985	0.982	0.981	0.976	0.989	0.989	0.995	0.998	0.991	0.991	0.991
	[22]	0.984	0.972	0.976	0.984	0.982	0.991	0.991	0.998	1.000	1.000	1.000	1.000
	Proposed SVEGA Method												
Lung Cancer	Raw Data	0.775	0.781	0.781	0.777	0.768	0.781	0.781	0.766	0.874	0.875	0.875	0.870
(32,57,3)	Bu et al. (2009)[3]	0.776	0.794	0.794	0.814	0.789	0.796	0.796	0.851	0.889	0.912	0.912	0.912
	Feng et al. (2008)[26]	0.814	0.824	0.824	0.835	0.795	0.798	0.798	0.894	0.973	0.983	0.983	0.983
	Senthamarai et al. (2010)	0.836	0.834	0.821	0.824	0.843	0.844	0.844	0.833	0.996	0.998	0.998	0.998
	[22]	0.874	0.875	0.875	0.870	0.945	0.916	0.916	0.915	1.000	1.000	1.000	1.000
	Proposed SVEGA Method												
Cardiac	Raw Data	0.627	0.624	0.624	0.623	0.614	0.644	0.644	0.628	0.621	0.673	0.673	0.641
Arrhythmia	Bu et al. (2009)[3]	0.639	0.662	0.662	0.636	0.582	0.628	0.628	0.603	0.820	0.794	0.794	0.755
(452,280,16)	Feng et al. (2008)[26]	0.635	0.658	0.658	0.645	0.789	0.715	0.715	0.726	0.831	0.856	0.856	0.821

Datasets	Proposed and Existing Method	Classifier Performance Measures											
		NB				J48				ANN			
		Prec.	Sens.	Spec.	F-Meas.	Prec.	Sens.	Spec.	F-Meas.	Prec.	Sens.	Spec.	F-Meas.
	Senthamarai et al. (2010)	0.662	0.662	0.636	0.831	0.894	0.876	0.876	0.864	0.874	0.875	0.875	0.870
	[22]	0.621	0.648	0.648	0.634	0.796	0.756	0.756	0.735	0.845	0.861	0.861	0.835
	Proposed SVEGA Method												
		HIGH DIMENSIONAL DATASETS											
SRBCT	Raw Data	0.988	0.988	0.988	0.988	0.842	0.843	0.843	0.842	0.745	0.741	0.741	0.741
(83,2309,4)	Bu et al. (2009)[3]	0.843	0.831	0.831	0.832	0.856	0.856	0.861	0.876	0.810	0.804	0.804	0.799
	Feng et al. (2008)[26]	0.826	0.824	0.824	0.831	0.915	0.926	0.926	0.889	0.841	0.839	0.839	0.840
	Senthamarai et al. (2010) [22]	0.856	0.854	0.854	0.845	0.926	0.935	0.935	0.915	0.856	0.842	0.842	0.854
	Proposed SVEGA Method	1.000	1.000	1.000	1.000	0.944	0.940	0.940	0.939	0.872	0.867	0.867	0.868
CNS	Raw Data	0.630	0.617	0.617	0.622	0.560	0.583	0.583	0.568	0.756	0.763	0.763	0.763
(60,7130,2)	Bu et al. (2009)[3]	0.645	0.625	0.625	0.615	0.626	0.649	0.649	0.649	0.763	0.774	0.774	0.774
	Feng et al. (2008)[26]	0.689	0.684	0.684	0.612	0.636	0.689	0.689	0.652	0.789	0.785	0.785	0.785
	Senthamarai et al. (2010) [22]	0.789	0.778	0.778	0.768	0.896	0.879	0.879	0.886	0.845	0.856	0.856	0.856
	Proposed SVEGA Method	0.871	0.867	0.867	0.868	0.853	0.850	0.850	0.844	0.954	0.950	0.950	0.949
Leukemia	Raw Data	0.986	0.986	0.986	0.986	0.831	0.833	0.833	0.832	0.579	0.583	0.583	0.581
(72,7130,2)	Bu et al. (2009)[3]	0.988	0.987	0.987	0.984	0.846	0.851	0.851	0.864	0.645	0.596	0.596	0.592
	Feng et al. (2008)[26] 0.987	0.985	0.985	0.965	0.904	0.950	0.950	0.962	0.786	0.689	0.689	0.689	
	Senthamarai et al. (2010) [22]	0.999	0.998	0.998	0.998	0.914	0.955	0.955	0.974	0.784	0.789	0.789	0.789
	Proposed SVEGA Method	1.000	1.000	1.000	1.000	0.926	0.965	0.965	0.984	0.786	0.795	0.795	0.795
Leukemia-3C	Raw Data	0.950	0.944	0.944	0.942	0.959	0.958	0.958	0.958	0.958	0.948	0.948	0.956
(72,7130,3)	Bu et al. (2009)[3]	0.958	0.948	0.948	0.956	0.939	0.933	0.933	0.930	0.978	0.956	0.955	0.955
	Feng et al. (2008)[26]	0.961	0.958	0.957	0.957	0.945	0.945	0.961	0.974	0.974	0.972	0.972	0.972
	Senthamarai et al. (2010) [22]	0.978	0.956	0.955	0.955	0.956	0.956	0.974	0.987	0.986	0.986	0.986	0.986
	Proposed SVEGA Method	0.986	0.986	0.986	0.986	0.968	0.967	0.967	0.995	1.000	1.000	1.000	1.000
Leukemia-4C	Raw Data	0.833	0.875	0.875	0.847	0.876	0.875	0.875	0.875	0.856	0.842	0.842	0.854
(72,7130,4)	Bu et al. (2009)[3]	0.845	0.871	0.871	0.846	0.889	0.878	0.878	0.878	0.872	0.867	0.867	0.868
	Feng et al. (2008)[26]	0.858	0.889	0.889	0.856	0.894	0.889	0.889	0.889	0.889	0.872	0.872	0.872
	Senthamarai et al. (2010) [22]	0.862	0.861	0.861	0.858	0.941	0.942	0.942	0.941	0.891	0.878	0.878	0.878
	Proposed SVEGA Method	0.974	0.972	0.972	0.972	0.963	0.958	0.958	0.959	0.946	0.944	0.944	0.944
MLL	Raw Data	0.775	0.781	0.781	0.777	0.848	0.847	0.847	0.848	0.841	0.839	0.839	0.840
(72,12583,3)	Bu et al. (2009)[3]	0.789	0.785	0.698	0.785	0.856	0.865	0.865	0.866	0.856	0.842	0.842	0.854
	Feng et al. (2008)[26]	0.791	0.789	0.789	0.783	0.876	0.875	0.875	0.875	0.861	0.859	0.859	0.859
	Senthamarai et al. (2010) [22]	0.879	0.865	0.865	0.858	0.888	0.876	0.876	0.891	0.875	0.864	0.864	0.864
	Proposed SVEGA Method	1.000	1.000	1.000	1.000	0.987	0.986	0.986	0.986	0.956	0.955	0.955	0.955

Datasets	Proposed and Existing Method	Classifier Performance Measures											
		NB				J48				ANN			
		Prec.	Sens.	Spec.	F-Meas.	Prec.	Sens.	Spec.	F-Meas.	Prec.	Sens.	Spec.	F-Meas.
Breast Cancer (97,24482,4)	Raw Data	0.598	0.546	0.546	0.407	0.645	0.589	0.589	0.572	0.634	0.645	0.645	0.631
	Bu et al. (2009)[3]	0.548	0.689	0.598	0.656	0.723	0.756	0.756	0.743	0.768	0.746	0.746	0.756
	Feng et al. (2008)[26]	0.825	0.817	0.817	0.812	0.798	0.768	0.767	0.790	0.767	0.789	0.789	0.745
	Senthamarai et al. (2010) [22]	0.865	0.865	0.857	0.849	0.832	0.856	0.856	0.879	0.869	0.912	0.912	0.897
	Proposed SVEGA Method	0.885	0.885	0.885	0.885	0.998	0.967	0.964	0.997	0.978	0.896	0.897	0.912
Ovarian Cancer (253,15155,2)	Raw Data	0.934	0.925	0.925	0.926	0.987	0.967	0.965	0.934	0.910	0.897	0.896	0.896
	Bu et al. (2009)[3]	0.867	0.853	0.853	0.817	0.956	0.967	0.967	0.958	0.918	0.897	0.897	0.891
	Feng et al. (2008)[26]	0.934	0.928	0.928	0.916	0.890	0.956	0.956	0.976	0.916	0.890	0.890	0.897
	Senthamarai et al. (2010) [22]	1.000	1.000	1.000	1.000	0.936	0.978	0.978	0.981	0.936	0.924	0.926	0.891
	Proposed SVEGA Method	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Lung Cancer (203,12601,5)	Raw Data	0.912	0.912	0.912	0.912	0.870	0.756	0.756	0.789	0.879	0.912	0.912	0.911
	Bu et al. (2009)[3]	0.897	0.882	0.881	0.881	0.916	0.926	0.926	0.926	0.887	0.812	0.811	0.813
	Feng et al. (2008)[26]	0.923	0.911	0.911	0.910	0.976	0.982	0.982	0.971	0.912	0.891	0.892	0.811
	Senthamarai et al. (2010) [22]	0.978	0.981	0.981	0.978	0.942	0.975	0.975	0.941	0.976	0.962	0.962	0.921
	Proposed SVEGA Method	0.985	0.985	0.985	0.985	0.983	0.981	0.981	0.976	0.989	0.978	0.978	0.967
Lymphoma (66,4027,3)	Raw Data	0.887	0.8952	0.895	0.895	0.885	0.889	0.887	0.887	0.889	0.887	0.887	0.886
	Bu et al. (2009)[3]	0.897	0.878	0.878	0.894	0.879	0.889	0.889	0.889	0.879	0.876	0.876	0.867
	Feng et al. (2008)[26]	0.987	0.956	0.957	0.983	0.913	0.934	0.978	0.978	0.912	0.987	0.987	0.924
	Senthamarai et al. (2010) [22]	0.978	0.989	0.989	0.971	0.946	0.967	0.967	0.979	0.987	0.989	0.989	0.988
	Proposed SVEGA Method	1.000	1.000	1.000	1.000	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Tab. VII *Classifier performance (Precision, Sensitivity, Specificity, F-Measure) on the features selected by the existing systems and proposed SVEGA method.*