A Novel Two-Stage Selection of Feature Subsets in Machine Learning

F. Rosita Kamala
Department of Computer Science,
Bharathiar University,
Coimbatore, India
rositakamala@gmail.com

P. Ranjit Jeba Thangaiah
Department of Information Technology,
Karunya Institute of Technology and Sciences,
Coimbatore, India
ranjit@karunya.edu

Abstract—**In feature subset selection the variable selection procedure selects a subset of the most relevant features. Filter and wrapper methods are categories of variable selection methods. Feature subsets are similar to data pre-processing and are applied to reduce feature dimensions in a very large dataset. In this paper, in order to deal with this kind of problems, the selection of feature subset methods depending on the fitness evaluation of the classifier is introduced to alleviate the classification task and to progress the classification performance. To curtail the dimensions of the feature space, a novel approach for selecting optimal features on two-stage selection of feature subsets (TSFS) method is done, both theoretically and experimentally. The results of this method include improvements in the performance measures like efficiency, accuracy, and scalability of machine learning algorithms. Comparison of the proposed method is made with known relevant methods using benchmark databases. The proposed method performs better than the earlier hybrid feature selection methodologies discussed in relevant works, regarding classifiers' accuracy and error.**

Keywords-dimensionality reduction; feature subset selection; filter method; hybrid method; variable selection; wrapper method

I. INTRODUCTION

An extensive, high dimensionality problem is caused by the numerous volumes of feature dimensions [1]. To find a solution, feature subsets are applied to alleviate the features dimensions and to give better outcomes in performance by reducing the less significant features. These methods boost the classification accuracy and reduce the training time of the learning techniques. The methods which select features, namely filters and wrappers, are distinguished on the basis of classifier evaluation [2]. The filter method finds the weight of the attributes on the basis of relevance that is computed by using different measures like information, distance, consistency, and correlation. It is fast and simple computationally, not depending on any learning algorithm, and scalable to huge-dimensional datasets. The feature subset can undergo an evaluation by the classifier subsequently, once the feature selection (FS) is performed. The outcome of feature selected results in the worst classification performance because the dependency of features is mostly overlooked. Wrapper methods depend on classifiers and work together between feature subset searches and model selection dependent on feature selection methods. The impediments in this method are the high risk of overfitting and that it is computationally intensive [3], particularly if the classifier is built. The embedded technique chooses features on the basis of filter methods and performs evaluation on the basis of the classifier of the wrapper method within the model. This method is less computationally intensive than wrappers. The selection of feature subsets is considered as an initial processing technique on the basis of an evaluation criterion for high dimensional data-sets. The four step process of selecting features stated in [1] is generating subsets, evaluating subsets, setting criteria to stop the process, and validating results. The first step produces a subset of features and performs an evaluation of features. The process goes on until the stopping criterion is reached. The feature subsets selected in the previous step are validated for performance analysis by the classification algorithm. A feature space of n features causes the subsets of 2^n to exist generally. The most important drawbacks are the point to start the search and the direction of search. The subset of features begins with a null set, and the features are added in the onward direction which satisfies the evaluation criteria, otherwise they are eliminated.

A. Objectives

Feature subset methods depend on parameters. The number of features in the final set is identified by the inputs and the threshold values. The issues of the above-mentioned filter-wrapper techniques claim the improvement of innovative algorithms. The overall objectives of this paper are:

- To eliminate irrelevant and redundant features.
- To select the optimal features from a huge set.
- To build up a multiple objective, filter-wrapper hybrid framework for continuous, categorical and hybrid data.
- To conduct an extensive experiment with the hybrid framework for evaluating the proposed methodology.
- To improve the classification accuracy and to minimize the error rate.

A novel algorithm is proposed and compared with the state-of-the-art methods. The results of the proposed method, using a subset of features, are similar or superior with the ones of the existing methods.

Corresponding author: F. Rosita Kamala

II. Relevant work

A lot of attribute selection methodologies have been proposed in recent studies for classification by machine learning. In high-dimensional feature sets, the choice of relevant features is indispensable, as a result of the large space of search of 2^n for n dimension variables. It is a challenging task to perform a comprehensive search to enhance the significant measures of the learning system [2]. A lot of filters, wrappers, and hybrid methodologies for feature selection have been adopted to accomplish a smaller set of features with relevance and significance [3]. To deal with this issue, a collection of techniques to perform a search in every feasible way for a solution, and explore an algorithm that is guaranteed to uncover a solution. The greedy algorithm always makes the choice that seems to be the best at that moment. Heuristic search and arbitrary search have been adopted in [4]. Particle swarm optimization (PSO) is proposed in [5]. In comparison with different evolution algorithms such as genetic algorithms (GAs) and genetic programming, PSO is cost effective, and more swift convergence is possible. A combination of PSO, and ACO (ant colony optimization) hybrid method was initiated for classification in [6]. The demerit of PSO was the requirement of conversion of nominal data to binary overcoming the need for preprocessing. A hybrid method of the filter - wrapper methods on the basis of PSO-GA which aimed to incorporate the merits of filter and wrapper techniques resulting in a smaller number of optimum features with better efficiency was proposed in [7]. A novel Gini-Index filter was proposed in [8]. It was formulated and adapted by the theory of Gini-Index for text classification in the selection of features, and produced better performance than the other methods.

Authors in [9] proposed a hybrid FS algorithm for gene data by combining mutual information maximization and adaptive GA (MIMAGA) to enhance the competence of MIMAGA algorithm. Authors in [10] employed a hybrid technique for GA by considering the merits of filters to improve the crossover and mutation operators. Hybrid FS approaches were created by subsets with features of different sizes and importance. Authors in [11] proposed a hybrid method using filters and wrappers based on instance learning. In the wrapper approach, a classification algorithm is adopted in a cooperative subset search in [11]. Authors in [3] proposed a hybrid FS method by combining the information gain ratio (IGR) filter and backward elimination (BE) wrapper in the first phase and PSO in the second phase. The method performs better for continuous features than for categorical features [3]. The relevant study leads to a framework of hybrid methodology of feature selection.

III. Materials and methods

A. The Proposed Methodology

In this paper, feature subsets are formulated from Chi square, Gini index, and PSO algorithms to solve FS problems in machine learning.

1) Chi Square Test Statistic (CHI)

The characteristics of a categorical data were studied in various perspectives like data size, number of features, possible values of attributes, and values of frequency distribution for the attributes of a dataset. For categorical data the similarity measure CHI was acknowledged as a goodness of fit test [12], with an estimated CHI distribution. Comparison of several classes can be assessed with the help of this test. It is considered as a test for independence and homogeneity [13, 14].

Authors in [15] accounted a comparison study of these methods and found that information gain and CHI are the best methods in feature selection. Chi-square distance is computed between every attribute and class. It is a measure of an attribute weighting task, as a result of its capability in attributes ranking [16]. This test is applicable for categorical datasets, and it does not perform well for data of quantitative nature. Frequency or count of data is needed for calculations with chi square test. No relationship of associativity between attributes is stated as null hypothesis. It creates a model by distributing data in different categories with an assumption that it follows the null hypothesis. Thus, this test compares the given data values of distribution with the expected data values. The frequency outcomes observed for the cell C_{ij} are Oij. The frequency outcomes expected for the cell C_{ij} are Eij.

$$\chi 2 = \mathrm{Sum}\,[(Oij\; - \;Eij)2\;/\;Eij, \{\mathrm{i},1,\mathrm{r}\}, \{\mathrm{j},1,\mathrm{c}\}] \quad (1)$$

where $Eij{=}n_i{*}n_j/n$.

Chi square value is contributed by rows, having different actual and expected values. The maximum value indicates the related features.

2) Gini Index(GINI)

GINI is the most suitable to classify attributes which have distributions clump together. During the evaluation process, GINI uses the combination of feature condition probability with its previous probability to evade the unbalanced class effects [17]. The inequality of a distribution is measured by the Gini coefficient [17]. The GINI is explained as the inequality percent within a specified population. Gini index is a correlation-based criterion. It approximates the features and discriminates among classes. It was first proposed as a rule for splitting in the generation of a decision tree [18]. It reveals the reduction of impurity, if features are chosen. The feature is represented by Y.

$$\mathrm{GINI}(Y) = \mathrm{Sum}\big[P\big(Y_j\big), j\big]\mathrm{Sum}\big[P\big(X_c\big|Y_j\big)^2, c\big] - \mathrm{Sum}\big[P(X_c)^2, c\big] \quad (2)$$

where $P(Y_j)$ is the previous probability of the feature Y has value Y_j, $P(X_c|Y_j)$ is the probability of a random sample from the dataset, that feature Y has value Y_j , appertain to class X_c, and $P(\,X_c\,)$ the previous probability of a random sample appertain to class X_c. Gini index, as a measure of inequality has some advantages. Gini index has very low computation requirements in high-dimensional data analysis. Analysis of increasing or decreasing inequality is possible. It can be used to compare feature distributions across different populations. It is sufficiently simple and easily interpreted. The weakness of Gini index is that it is not able to detect redundant features, with inter-feature relationships.

3) PSO

The wrapper method PSO [4] is selected due to advantages such as simple implementation, sufficiently less parameters to fine-tune, it is more robust, it has fast convergence, less computing time, and finds global optima with high probability and efficiency. PSO is inspired by the societal way of behaving birds in flocks. PSO's fundamental characteristic is the social interaction in the population that optimizes information. Every result can be described as a particle in the swarm. Every particle's position in the search space is described by a vector $S_i=(S_{i1}, S_{i2}, \ldots, S_{iD})$, where D is the dimensionality of the space for search. In order to achieve optimal solutions, the particles move in the space for search. As a result, every particle's velocity is described as $C_i=(C_{i1}, C_{i2}, \ldots, C_{iD})$. Every particle's position and velocity can be revised with respect to the movement of its neighbors. *pbest* is the best preceding location and is depicted as the personal best of the particle, and *gbest* is the best location in the population. Derived from *pbest* and *gbest* the optimal solution is searched by improving the velocity C and every particle's position in space for search S in relation to the subsequent equations [5]:

$$C_i(t+1) = \omega C_i(t) + a_1 r_1 \big(pbest(i,t) - S_i(t)\big) + a_2 r_2 \big(gbest(t) - S_i(t)\big) \quad (3)$$

$$\mathrm{S_i}\,(\mathrm{t}+1) = \mathrm{S_i}\,(\mathrm{t}) + \mathrm{C_i}\,(\mathrm{t}+1) \quad (4)$$

where t describes the number of repetitions in the evolution process, ω performs inertia weight, which controls the responses of the earlier velocities, a_1 and a_2 are acceleration constants, r_1 and r_2 are random numbers following a uniform distribution in [0, 1]. The algorithm stops after a predefined condition is attained, which might be the best suitable assessment or a specified count of iterations.

B. The Proposed Algorithm

The methods of the filter and wrapper undergo a combination in the proposed two staged method. The pseudocode of the proposed framework, named two-stage selection of feature subsets (TSFS), is presented in Figure 1. Figure 2 shows a flowchart of the TSFS algorithm. In Figure 1, lines 1-5 account for the preprocessing step for the selection of relevant features of the algorithm TSFS. It is an introductory step to be performed for the datasets downloaded from UCI (University of California, Irvine). In the first phase of the algorithm, filter techniques, namely Chi square and Gini index, are applied separately to the datasets for the elimination of the superfluous or inappropriate features. The CHI filter selects the feature subset f_1 features and the GINI filter selects the feature subset f_2 features. The selected subsets of features f_1 and f_2 are calculated as the most relevant features associated with the class label. The methods which are described above, weigh up the significance of the features by computing the weight for each and every feature of the dataset, with the class label of the dataset. The outputs from the filter techniques undergo a combination of two feature subsets f_1 and f_2 in line 6 for better performance outcomes by removing the subset of features available in both subsets resulting in a feature subset with fewer features.

```
Input: Features set $D=\{f_i, i=1.....n\}$ $C$: class labels
Output: The subset $P$ of $D$ features
1. (Initialize) Let $D \leftarrow$ "Original set having d dimensions"; $P \leftarrow \{\ \}$
2. Evaluate the significance of CHI with the resultant class $C$. For each $f_d \in D$ find CHI($C$;$f_d$) applying (1)
3. (possibility of the first dimensions) Find a dimension $f_1$ that improves CHI($C$,$f_d$); Let $D \leftarrow D \setminus \{f_1\}$; Set $P \leftarrow \{f_1\}$
4. Evaluate the significance of GINI with the output class $C$. For each $f_d \in D$ find GINI($C$;$f_d$) applying (2).
5. (possibility of the next dimensions) Find a dimension $f_2$ that improves GINI($C$,$f_d$); Let $D \leftarrow D \setminus \{f_2\}$
6. (optimal output) Output the set $P \leftarrow P \cup \{f_2\}$
7. Split the feature set $P$ into $s$ for training dataset and testing dataset. For $P$ generate particles $P$.
8. For each particle from 1 to $N$
9. Initialize particle
9.1 Initialization of particle's position in the search space.
9.2. Initialization of pbest and gbest.
9.3. Initialization of velocity.
10. Repetition until the termination condition is met.
10.1 Update particle's velocity using (3).
10.2 Update particle's position using (4).
10.3 $t \leftarrow t+1$
11. Best found solution for the output gbest.
12. Fitness evaluation through CV.
13. Split the data into $k$ equal sized folds.
14. for $s = 1.....k$
14.1 Training a model with basis features on $s^{th}$ fold's training set.
14.2 Computation of testing error on this corresponding fold.
15. Return values over all $k$ folds with the lowest average of testing error.
16. Learning results and significance of a predicted outcome for $P$ as subset of selected features.
```

Fig. 1. The pseudocode of the suggested TSFS algorithm

Lines 7-11 correspond to the wrapper approach using PSO to select the subset of optimal features in the feature space. The PSO algorithm computes the two best values for each and every particle in each iteration. The selection for an optimal subset is achieved to diminish the dimension of features. Lines 12-16 correspond to the resultant optimal subset by the wrapper approach PSO. Training and testing sets in the second phase undergo tenfold cross-validation (CV) for improvement in learning efficiency. This is the tuning step to remove redundancy for optimal subsets.

C. K-Nearest Neighbour Classification (kNN)

To classify data, their nature is most important. It can be either parametric or non-parametric. K-nearest neighbor classification is non-parametric. The instances of the data which are together in nearby proximity are self-reliant and distributed separately. Those instances have similar classification results [19].

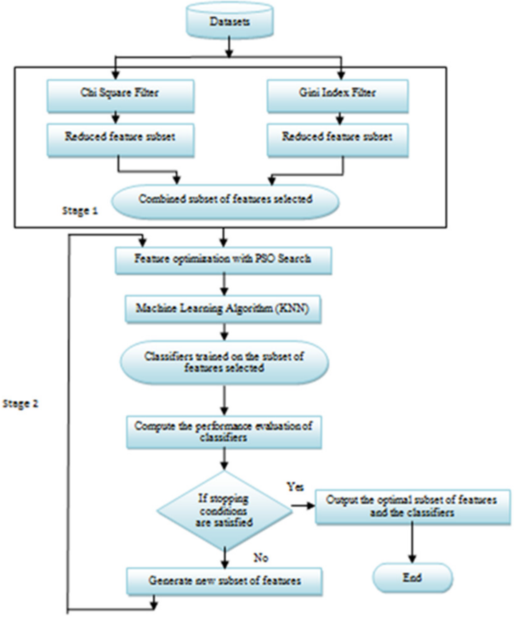
Datasets
Chi Square Filter
Gini Index Filter
Reduced feature subset
Reduced feature subset
Combined subset of features selected
Stage 1
Feature optimization with PSO Search
Machine Learning Algorithm (KNN)
Classifiers trained on the subset of features selected
Stage 2
Compute the performance evaluation of classifiers
If stopping conditions are satisfied
Yes
Output the optimal subset of features and the classifiers
No
Generate new subset of features
End

Fig. 2. The flowchart of the suggested algorithm TSFS

IV. RESULTS AND DISCUSSION

A. Datasets

The experimentation in java environment using RapidMiner confirms the efficiency of the suggested algorithm. Eight UCI [20] datasets are shown in Table I. The PSO parameters *pbest* and *gbest* are assigned to 1. True value is set for dynamic inertia weight which makes the enhancement of inertia during its run. The upper limit for the generations to be performed is 30. The size of population is 100. The model is evaluated by the conduction of experiments on all sample instances by performing 10 CVs using kNN classifier for better performance outcomes.

TABLE I. EXPERIMENTAL DATASET INFORMATION

Datasets	#Samples	#Features				#Classes
		Total	Numeric	Nominal	Type	
Anneal	898	38	6	32	Discrete+ Continuous	6
Vowel	990	13	10	3	Discrete+ Continuous	11
Hypothyroid	3772	29	6	23	Discrete+ Continuous	4
Sick	3772	29	7	22	Discrete+ Continuous	2
Splice	3190	61	0	61	Discrete	3
Chess	3196	36	0	36	Discrete	2
Sonar	208	60	60	0	Continuous	2
Waveform	5000	40	40	0	Continuous	3

B. Performance Metrics

The most common measures for the evaluation of feature subset are Precision and Recall. Precision is the ratio of instances of relevance among the instances of retrieved while Recall is the ratio of instances of relevance that have been retrieved over the total number of instances of relevance. Precision and recall are the estimates for finding relevance in terms of true positives, false positives, false negatives and true negatives [21]. Figures 3 and 4 show the outcomes of Precision and Recall fitness functions on the efficiency of various variable selection techniques and the suggested TSFS.

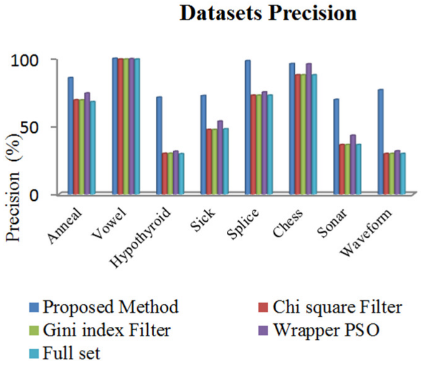

Fig. 3. Comparative study of the significance measure Precision of the TSFS with other existing methods

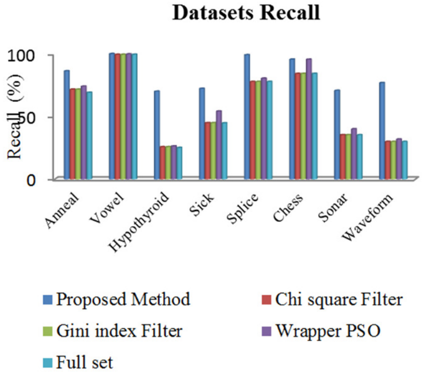

Fig. 4. Comparative study of the significance measure Recall of the TSFS with other existing methods

C. Statistical Analysis

The proposed TSFS algorithm's evaluation has been carried out by the kNN classifier. The main objectives, i. e. the number of selected features, the selected feature subsets' classification accuracy, and the computing time with the measures of performance like Precision, Sensitivity, and Kappa have been documented. In Figure 5 the accuracy of the suggested algorithm is compared with the traditional existing methods of features subsets in the literature. Depending on the results obtained in Figure 5, the proposed method TSFS is considered to be superior in terms of accuracy than the traditional methods. It can be seen that TSFS performs better for continuous datasets like sonar and waveform than for discrete datasets like splice and chess and hybrid datasets like sick, anneal, and hypothyroid.

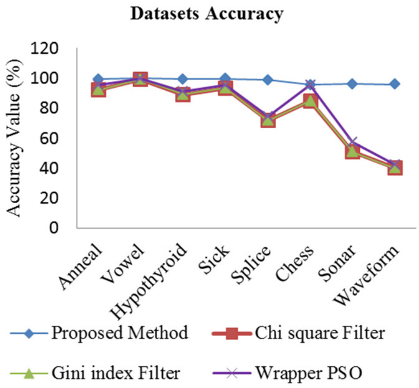

Fig. 5. Accuracy comparison for each feature selection algorithm with the proposed algorithm.

D. Kappa Statistic as Performance Measure

The kappa statistic measures whether the instances are closely classified by the learning algorithm with the matching data label, controlling the accuracy of random classification. Classifiers constructed and calculated with datasets of various distributions of learning can undergo comparison by kappa in association with expected accuracy. This shows a better indication of how the classification of instances occurred, because accuracy could be skewed, provided that class distribution is skewed. There is no identical representation of kappa. Authors in [22] represent 0-0.20 as minor, 0.21-0.40 as light, 0.41-0.60 as reasonable, 0.61-0.80 as significant, and 0.81-1 as almost ideal, while author in [23] considers kappa>0.75 as outstanding, 0.40-0.75 as fine, and kappa<0.40 as deprived. It is commonly considered as a more vigorous estimate. The proposed TSFS's kappa analysis is outlined in Table II.

TABLE II. KAPPA COEFFICIENT OF THE FRAMEWORK TSFS

Dataset	Anneal	Vowel	Hypothyroid	Sick
Kappa	0.982	1.00	0.961	0.972
Dataset	**Splice**	**Chess**	**Waveform**	**Sonar**
Kappa	0.980	0.914	0.937	0.928

E. MAE and RMSE Analysis

Mean absolute error (MAE) [18] is the average of the discrepancy between anticipated and actual data values, and is given by:

$$\mathrm{MAE} = \frac{1}{n}\sum_{j=1}^{n}|y_j - \hat{y}_j| \quad (5)$$

MAE ranges from zero to infinity and a perfect fit is obtained when MAE=0. Root Mean Squared Error (RMSE) [21] is one of the most commonly used measures for calculating the average amount of error in numerical prediction. It is the square root of the average of squared discrepancies between prediction and actual observation. Its value is computed by:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2} \quad (6)$$

The smaller the RMSE value, the better the model performance. Table III presents the results of MAE and RMSE.

TABLE III. MAE AND RMSE

Dataset	Anneal	Vowel	Hypothyroid	Sick
MAE	0.010	0.000	0.054	0.052
RMSE	0.085	0.000	0.075	0.056
Dataset	**Splice**	**Chess**	**Waveform**	**Sonar**
MAE	0.045	0.082	0.042	0.073
RMSE	0.112	0.207	0.205	0.196

TABLE IV. COMPARATIVE ANALYSIS OF THE FRAMEWORK TSFS TO RELEVANT METHODS

Datasets	TSFS %	Results obtained from the literature					
		Methodology & Accuracy		Methodology & Accuracy		Methodology & Accuracy	
Anneal	99.28	[24] RCRF	99.63	[24] CRF	99.71	[24]RF	99.68
Vowel	**100**	[25] IEM	99.29	[25] EM-1NN	98.88	[25]IGA	98.18
Hypothyroid	99.44	[24] RCRF	**99.73**	[24] BA-CDT	99.59	[24] BA-C4.5	99.62
Sick	**99.68**	[24] BA-C4.5	98.97	[24] RCRF	98.64	[24]CRF	98.59
Splice	**98.75**	[24] RF	95.88	[24] RCRF	96.48	[24]CRF	96.31
Chess	95.71	[25] BPNN	**99.28**	[25] LR	97.43	[25]C4.5	97.90
Sonar	96.17	[3] HFSM	**97.13**	[26] WOASA	95.67	[26]MEGWO	94.81
Waveform	**95.78**	[24] RCRF	85.01	[24] CRF	85.15	[24] RF	85.2

Best results are shown in bold

TABLE V. TSFS COMPARISON ON COMPUTING TIME WITH MOST RELEVANT METHODS

Datasets	TSFS (s)	Results obtained from the literature (in seconds)					
		Methodology & run time		Methodology & run time		Methodology & run time	
Anneal	**59**	[27] A4	162.06	[27] EFSH	59.39	[28] OM	90.80
Vowel	**1**	[27] A4	41.82	[27] EFSH	4.66	[28] OM	34.40
Hypothyroid	732	[27] A4	1446.28	[27] EFSH	**148.09**	[28] OM	1362.80
Sick	793	[27] A4	1516.71	[27] EFSH	**310.06**	[28] OM(Wrapper)	540.70
Splice	690	[29] FSSMC	25.7	[29] Relief	164	[29] Relief-RS	**18.4**
Chess	542	[30] MA+ SVM$^{\text{optimized}}$	**176.3**	[31] BF-MLP	302.13 ± 1.05	[31] BF-RBF	4347.9 ± 21.17
Sonar	**4**	[32] ISSA	61.43	[32] ALO	58.48	[33] PSOPG2	25.2
Waveform	1201	[32] PSO	290.37	[32] ISSA	293.45	[32] GA	**288.50**

Best results are shown in bold

F. Result Analysis Comparison of Classification Accuracy and Computing Time with Existing Methods

In Tables IV and V the accuracy and run time of TSFS is compared with other methods, available in the literature. Regarding accuracy, the TSFS method shows appreciable improvement for all datasets except chess and sonar, when compared with the other existing methods. TSFS achieves the best accuracy values for vowel, sick, splice, and waveform datasets and the best run time for anneal, vowel and sonar datasets.

V. CONCLUSION AND FUTURE WORK

Feature selection is an imperative method for dimensional reduction in machine learning. In this paper, a proficient methodology for selecting instructive features is being proposed for massive hybrid datasets. Experimentation confirms that the methodology is effective and efficient, especially considering classification performance. It was shown that the proposed methodology provided better accuracy rates than the existing methods. Additionally, the use of the proposed method decreased error rates regarding two error measures, MAE and RMSE. These results are important and show that the use of feature selection can provide better performance and higher efficiency for hybrid systems. The weakness of this work is that it fails to produce better computation time even though the accuracy outperforms other methodologies, under experimentation with UCI repository datasets. It would be notable to use mutual information [34] for estimating and finding features that are minimal redundancy-maximal relevance (mRMR) in future research. These concepts could be integrated with the proposed method in order to get a quicker and improved method, which may be applied effectively to huge datasets.

REFERENCES

[1] M. Dash, H. Liu, "Feature selection for classification", Intelligent Data Analysis, Vol. 1, No. 1-4, pp. 131–156, 1997

[2] R. Kohavi, G. H. John, "Wrappers for feature subset selection", Artificial Intelligence, Vol. 97, No. 1-2, pp. 273–324, 1997

[3] F. R. Kamala, P. R. J. Thangaiah, "A proposed two phase hybrid feature selection method using backward Elimination and PSO", International Journal of Applied Engineering Research, Vol. 11, No. 1, pp. 77–83, 2016

[4] M. Dash, H. Liu. "Consistency-based search in feature selection", Artificial Intelligence, Vol. 151, No. 1-2, pp. 155–176, 2003

[5] J. Kennedy, R. C. Eberhart, "Particle swarm optimization", IEEE International Conference on Neural Networks, Perth, Australia, November27-December 1, 1995

[6] N. Holden, A. A. Freitas, "A Hybrid PSO/ACO Algorithm for Discovering Classification Rules in Data Mining", Journal of Artificial Evolution and Applications, Vol. 2008, ArticleID 316145, pp. 1-11, 2008

[7] Indriyani, W. Gunawan, A. Rakhmadi, "Filter-Wrapper Approach to Feature Selection Using PSO-GA for Arabic Document Classification with Naive Bayes Multinomial", IOSR Journal of Computer Engineering, Vol. 17, No. 6, pp. 45-51, 2015

[8] W. Shang, H. Huang, H. Zhu, Y. Lin, Y. Qu, Z. Wang, "A Novel feature selection algorithm for text categorization", Expert Systems with Applications, Vol. 33, No. 1, pp. 1-5, 2007

[9] H. Lu, J. Chen, K. Yan, Q. Jin, Z. Gao, "A hybrid feature selection algorithm for gene expression data classification", Neurocomputing, Vol. 256, pp. 56-62, 2017

[10] A. S. Ghareb, A. A. Bakar, A. R. Hamdan, "Hybrid feature selection based on enhanced genetic algorithm for text categorization", Expert Systems with Applications, Vol. 49, pp. 31-47, 2016

[11] A. B. Brahim, M. Limam, "A hybrid feature selection method based on instance learning and cooperative subset search", Pattern Recognition Letters, Vol. 69, pp. 28-34, 2016

[12] K. Pearson, "On a criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling", The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, Vol. 50, No. 302, pp. 157–175, 1900

[13] P. E. Greenwood, M. S. Nikulin, A Guide to Chi-squared Testing, John Wiley & Sons, 1996

[14] H. O. Lancaster, The Chi-squared Distribution, John Wiley & Sons, 1969

[15] Y. Yang, J. O. Pedersen, "A comparative study on feature selection in text categorization", in: Proceedings of the 14th international conference on machine learning(ICML) San Francisco, USA, pp. 412–420, Morgan Kaufmann Publishers, 1997

[16] I. H. Witten, E. Frank, Data Mining Practical Machine Learning Tools and Techniques with Java Implementations, Morgan Kaufmann, 2005

[17] A. Jacobson, A. D. Milman, D. M. Kammen, "Letting the (energy) Gini out of the bottle: Lorenz curves of cumulative electricity consumption and Gini coefficients as metrics of energy distribution and equity", Energy Policy, Vol. 33, No. 14, pp. 1825-1832, 2005

[18] L. Breiman, J. Friedman, R. Olshen, C. Stone, Classification and Regression Trees, Chapman and Hall, 1984

[19] D. Wettschereck, D. Aha, W. T. Mohri, "A review and empirical evaluation of feature weighting methods for a class of lazy learning algorithms", Artificial Intelligence Review, Vol. 11, No. 1-5, pp. 273-314, 1997

[20] UCI Machine Learning Repository, University of California, available at: https://archive.ics.uci.edu/ml/index.php

[21] V. S. Stehman, "Selecting and interpreting measures of thematic classification accuracy", Remote Sensing of Environment, Vol. 62, No. 1, pp. 77–89, 1997

[22] J. R. Landis, G. G. Koch, "The measurement of observer agreement for categorical data", Biometrics, Vol. 33, No. 1, pp. 159–174, 1977

[23] J. L. Fleiss, Statistical Methods for Rates and Proportions, John Wiley, 1981

[24] J. Abellan, C. J. Mantas, J. G. Castellano, S. Moral-Garcia, "Increasing diversity in random forest learning algorithm via imprecise probabilities", Expert Systems With Applications, Vol. 97, pp. 228–243, 2018

[25] K. J. Wang, A. M. Adrian, K. H. Chen, K. M. Wang, "An improved electromagnetism-like mechanism algorithm and its application to the prediction of diabetes mellitus", Journal of Biomedical Informatics, Vol. 54, pp. 220–229, 2015

[26] Q. Tu, X. Chen, X. Liu, "Multi-strategy ensemble grey wolf optimizer and its application to feature selection", Applied Soft Computing, Vol. 76, pp. 16–30, 2019

[27] F. Wang, J. Liang, "An efficient feature selection algorithm for hybrid data", Neurocomputing, Vol. 193, pp. 33–41, 2016

[28] C. Li, H. Li. "A Survey of Distance Metrics for Nominal Attributes", Journal of Software, Vol. 5, No. 11, pp. 1262-1269, 2010

[29] Y. Huang, P. J. McCullagh, N. D. Black, "An optimization of ReliefF for classification in large datasets", Data & Knowledge Engineering, Vol. 68, No. 11, pp. 1348-1356, 2009

[30] M. Nekkaa, D. Boughaci, "A memetic algorithm with support vector machine for feature selection and classification", Memetic Computing. Vol. 7, No. 1, pp. 59–73, 2015

[31] L. T. Kueti, N. Tsopze, C. Mbiethieu, E. Mephu-Nguifo, L. P. Fotso, "Using Boolean factors for the construction of an artificial neural network", International Journal of General Systems, Vol. 47, No. 8, pp. 849-868, 2018

[32] A. E. Hegazy, M. A. Makhlouf, G. S. El-Tawel, "Improved salp swarm algorithm for feature selection", Journal of King Saud University – Computer and Information Sciences, 2018

[33] B. Xue, Particle Swarm Optimisation for Feature Selection in Classification, PhD Thesis, Victoria University of Wellington, 2014

[34] C. Pascoal, M. R. Oliveira, A. Pacheco, R. Valadas, "Theoretical evaluation of feature selection methods based on mutual information", Neurocomputing, Vol. 226, pp. 168–181, 2017

AUTHORS PROFILE

F. Rosita Kamala got her Masters in Computer Application (MCA) from the Bharathidasan University in 1999 and her Master of Engineering in Systems Engineering and Operations Research from Anna University, Chennai in 2008. She is currently an Assistant Professor in an Affiliated College under Visvesvaraya Technological University, Karnataka, and a doctoral student at Bharathiar University, Tamil Nadu. Her current research interests include data mining, data analytics, machine learning, and intelligent information processing.

P. Ranjit Jeba Thangaiah did his BSc (Physics) at the P.S.G. College of Arts and Science, Coimbatore, M.C.A. at Karunya Institute of Technology, Coimbatore, MPhil at Manaonmanian Sundaranar University, Tirunelveli, He did his full time PhD at the Bharathiar University, Coimbatore. He is working in the Karunya Institute of Technology and Sciences, Coimbatore in the last 15 Years. He has published 25 papers on International Journals and Conferences. His areas of research include Networking, Data Mining and Machine Learning.