The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 2206 - 2225

An Exhaustive Wrapper Method for Feature Selection in Large Dimensional Datasets (WFS)

Damodar Patel[1], Amit Kumar Saxena[1,] *, Suman Laha[2], Rajeshwar Prasad[1], Utpal Roy[2]

[1]Department of CSIT, Guru Ghasidas Vishwavidyalaya Bilaspur, Bilaspur, India, damodarpatel7497@gmail.com, amitsaxena65@rediffmail.com, rp4464867@gmail.com

[2] Department of Computer and System Sciences Visva-Bharati University Santiniketan, Santiniketan, India, mrlaha@gmail.com, utpal.roy@visva-bharati.ac.in

Abstract: In this paper, a novel algorithm for randomly selecting a small subset of features from a dataset is presented. Using different combinations of features across a number of trials, the algorithm discovers the best subsets of features. When these subsets of features are obtained, the classification accuracies produced by three classifiers (K-Nearest Neighbor, Support Vector Machines, and Random Forest) are considered to evaluate the performance criterion of the proposed wrapper-based method. Further, to improve the classification accuracy and reduce the cardinality of the selected feature sets, an exhaustive feature selection method (the wrapper method) is used. The proposed algorithm is simulated on eighteen datasets, and the results are compared with those reported using nine comparable algorithms using three classifiers to justify the performance of the proposed algorithm. The average classification accuracies of eighteen datasets achieved are 88.66% in K-NN, 89.88% in SVM, and 89.14% in RF classifier with at most 10 features. The proposed algorithm archives better CA compared to nine comparable algorithms and the results of the experiments prove the proposed algorithm's performance is better in selecting the most effective features compared to other algorithms.

Keywords: data mining, dimensionality reduction, machine learning, feature section, wrapper method.

1. Introduction

We need huge storage media today since so many daily tasks in our culture are automated. Mostly, the formatted datasets that have a well-formatted structure (like a relational database) nowadays often include a significant number of patterns and a limited number of classes in computer-based applications. Every pattern has a set of features that help to represent it, and every pattern belongs to one of the classes. Data mining [1] is the process of extracting relevant information from a large database. A core phase in the data mining process is classification. The study of features is essential to classification. Feature selection (FS) and feature extraction are two important parts of feature analysis. FS [2], [3], is the process of selecting a subset of features from a dataset. On the other hand, feature extraction may combine or recalculate existing features to produce new ones. There may be redundant or noisy features in a dataset. These unnecessary features make the classifier more difficult and expensive while also creating confusion. Sometimes a classifier with the optimal number of features may yield more accurate results than one with additional features. The method is referred to as supervised if the FS uses data (such as the class of a pattern) that was provided before the process was applied. An unsupervised algorithm [4] is one in which the patterns are classified without any previous information (such as class) being provided. Many supervised FS [5] techniques make use of neural networks [6], fuzzy logic [7], and K-NN [8] search algorithms.

To select the most important features from a dataset with a large number of features, wrapper methods could be used. Starting with a random FS for each fold, the scheme decides on the best fold based on classification accuracy (CA). Then, to increase CA and reduce the number of selected features, the exhaustive wrapper-based FS method is used. K-NN (K-Nearest Neighbor) [8], SVM

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 2206 - 2225

(Support Vector Machines) [9], and RF (Random Forest) [10] classifiers are used in the experiments to evaluate the CA of prediction by the algorithm.

In this study we propose, A novel wrapper method for FS in large dimensional datasets (WFS) and then determine the usefulness of the WFS algorithm by comparing with a recently published paper by Zhao et al. [11], in which eight FS algorithms namely as, FSRRW (relevant-redundant weight-based feature criterion) [11], MIFS (Mutual information feature selector) [12], JMI (join mutual information) [13], mRMR (minimum-redundancy maximum-relevance) [14], CIFE (class-relevant redundancy) [15], MRI (max-relevance and max-independence) [16], DCSF (Dynamic Change of Selected Feature with the class) [17], & CWJR (conditional weight-based joint relevance) [18] are used and these algorithms could obtain results with at most 30 features. The objective is to achieve better CA with a reduced number of selected features compared to using eight algorithms.

The paper is organized as follows: Some feature selection techniques are given in Section 2. Section 3 explains the proposed method. A dataset description is presented in Section 4. Section 5 presents the experiments. Section 6 covers simulation experiments and outcome analyses. The last part of the essay presents conclusions and potential areas for further study.

2. Some Feature Selection Techniques

Feature selection (FS) is one of the essential machine learning pre-processing steps that eliminates redundant and irrelevant data with the goal of improving prediction accuracy and minimizing computing complexity. According to Saxena [5], there are mainly two basic types of FS methods: filter methods and wrapper methods. According to various evaluation factors, filter methods assign a score to each of the features. The observable contributions are further arranged in decreasing order. So, until the essential number of features is not obtained or the unique threshold (or CA, for example) is not reached, the significant features are sequentially selected. A feature's individual property is used to determine whether or not it is a significant feature. Using filter approaches, the majority of work has been done. In [19], Chernbumroong proposed the MRMC algorithm, based on neural networks for FS, for FS in [20], the ABACO algorithm is used; this method is a modified version of ant colony optimization [21]. In [22], FS using GNMF (Graph Regularized Non-Matrix Factorization) algorithms is proposed. A target function is created that locates a subspace where all samples are very separated from one another. The repeated optimization of this target function resulted in unsupervised FS [23]. In order to select features for multi-label datasets, a mutual information-based multi-label feature selection approach using the information interaction method is proposed. This algorithm evaluates feature dependence [24]. For email identification, a brand-new combined, document frequency and term frequency feature selection approach (DTFS) is proposed in [25]. To avoid premature convergence and provide more accuracy, a memetic feature selection technique for multi-label classification has been developed in [26]. Based on forward approximation, fuzzy-rough [27] feature selection is developed and used for large dimensional datasets in [28]. A FS technique based on the Fast Fourier Transform [29] is proposed for mechanical systems. RTBFS (Robust Twin Boosting Feature Selection), a novel ensemble method, is developed for FS in [30].

in [31], implementation of the genetic algorithm as a filter model-based feature subset selection technique is used. Two kinds of weights (input-hidden and hidden-output) are obtained from trained neural networks. After this, since each node's general formula is based only on inputs, a genetic algorithm is utilized to improve this formula [32]. For classifying text, the ant colony optimization technique is used to select features in [33]. To get the best feature subset, a hierarchical search

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 2206 - 2225

framework and the Tabu search approach are combined in [34]. A technique for selecting features based on bits is proposed. It consists of two phases: the first phase creates a bitmap indexing matrix from a given dataset, and the second phase selects a collection of relevant features for the classification process and evaluates them using domain expertise [35]. For FS, a hybrid approach based on artificial neural networks (ANNs) and ant colony optimization is applied in [36]. The group method of data handling (GMDH) algorithm is developed in [37] and features are ranked based on the predictive power of those rankings using a learning algorithm. Zhu et al. [38] proposed a measurement known as the relative importance factor (RIF) to obtain data on less significant features. Higher accuracy and shorter calculation times are obtained by removing these less important features from the dataset.

The wrapper technique to FS generates a subset of features and uses a classifier to assess the subset's usefulness. The Feature Subset Selection by Estimation of Bayesian Network Algorithm (FSS-EBNA) method is an evolutionary, population-based, randomized search technique. As feature subset evaluations, Naive-Bayes [39] and ID3 [40] learning algorithms are used. It employs the EDA (Estimation of Distribution Algorithm) paradigm and ignores using crossover and mutation operators, like in genetic algorithms, to generate the populations. The factorization of the probability distribution of the best solutions is obtained throughout a generation of the search [41]. Unsupervised FS is utilized to reduce the dimension of the datasets while still keeping the structure of the high dimensional dataset using a genetic algorithm using Sammon's stress function as the fitness function in [42]. By Dy and Brodley in [43], a wrapper framework for FS, clustering, and order identification concurrently. Furthermore, they compared the scatter separability feature selection threshold.

Several techniques [44–50] for unsupervised FS are proposed. In [44], a technique is presented that divides the original feature set into multiple subsets or clusters, resulting in features that are extremely similar in one cluster. The last step is to select one feature from each cluster to produce a reduced feature set. The suggested technique for FS is dependent on the method used to estimate univariate data, although multiple techniques based on maximum entropy and maximum likelihood criteria are discussed in [45]. An unsupervised neuro-fuzzy feature ranking algorithm has been presented by Pal et al. [46]. In [47], a novel correlation-based FS technique is proposed. CFS based its search for a suitable subset of features on the predicted performances and inter-correlations of the features. CFS may significantly reduce the dimensionality of datasets while preserving or boosting the effectiveness of learning algorithms, according to tests on both continuous and discrete class datasets. A concept of redundancy between two random variables, X and Y, is described by Heydorn [48], and this concept is also used to construct a redundancy test. By using this test, redundant features may be eliminated without affecting classifier performance. Linear approaches cannot classify patterns using features that are linearly dependent on other variables. A measure of linear dependency is proposed in [49] as a FS tool to help identify linearly dependent features. Yan [50], proposed an efficient unsupervised feature selection method through feature clustering to address the redundancy among features [EUFSFC] and determine the size of the final feature subset. This paper used twelve high dimensional datasets. In addition, Xie, Wang [51], and Peng [52] proposed the hybrid filter and wrapper approach. Saxena et al. proposed hybrid feature selection [53], and in [54], Dubey et al. proposed a feature selection technique based on mutual information and cosine similarity.

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 2206 - 2225

3. Proposed Methodology

The WFS starts with a set of features randomly selected from the set of entire feature set of the original dataset. After repeated sets of experiments, the final subset of features is obtained. Following is a description of the whole scheme:

1. Let F be the set of all features.
2. Divide the F features into folds of q features each, q=10 in present method. Let N=F/q be the number of folds in the method. If mod(F/q) ~=0 then (N+1) th fold will be containing (F-N*q) features. Thus, in this case N folds will contain q features each, and the last (N+1) th fold will have remaining features as mentioned before. In case F is completely divisible by q, then all N folds will contain exactly q features each. For simplicity in understanding, we will mention N folds which will mean N or N+1 as the case may be.
3. Select randomly q features for all N folds (for last fold, number of features may be <q as stated above).
4. Use classifier Ci (i=1,2,3 for K-NN, SVM and RF respectively) to determine each fold's CAi.
5. Compare all N folds' CAi using Ci and note the features of the fold with highest CAi .
6. Apply exhaustive FS on Ci by taking all possible combinations of features obtained from best feature set. Find CA for each possible subset. Let EFSi(t), t=1 in first trial, be the feature subset which produces highest CA after exhaustive search.
7. Repeat steps 2 to 6 for T (T=100) trials and save EFSi(t), t=1,…,T.
8. Determine EFi= EFSi(tj) for highest CAi achieved by classifier Ci for the jth trial.
9. Determine EFi for i=1,2,3 classifiers.

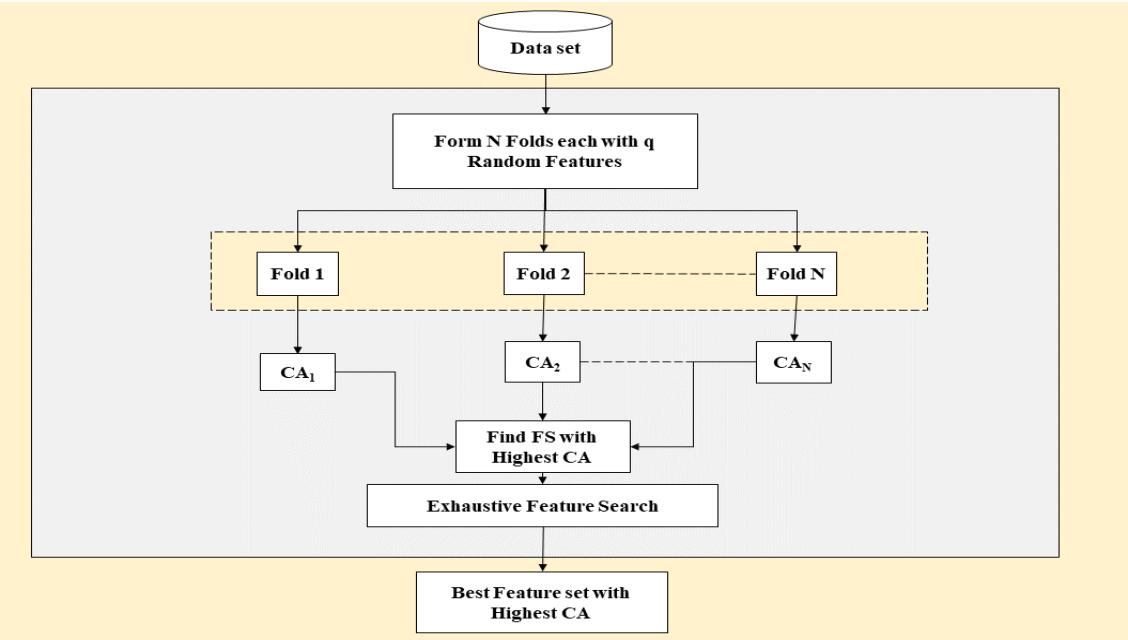

Figure 1: Process used in Proposed Work

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 2206 - 2225

4. Experiments

4.1. Experiment Setting

Using Python 3.9.7, the WFS algorithm was tested on a Windows 10 system with an Intel (R) Core i3, 4 GB of memory, a 2.30 GHz CPU, and a 500 GB SSD card. Ten independent runs of each experiment are performed in Python. For analysis, the average CA values over 10 times are taken in the Tables.

In the experiment, we have used a standard scaler for normalization in all datasets, and three classification models, K-NN (K = 3), SVM (kernel = rbf), and RF, are used to evaluate the CA of the proposed scheme with the existing label present at the data source. In this study, 10-fold cross-validation [55] was performed. Each experiment is run 10 times, and the average CAs of 10 times runs are shown in the Tables.

4.2. Dataset Description

Eighteen benchmark datasets from ASU [56] and UCI's machine learning repository [57] are taken in order to evaluate the effectiveness of the proposed algorithm. Table 1 provides the essential details of the datasets. Out of these datasets, which range in size from 17 to 11340 features; most are high-dimensional datasets. Twelve multiclass and six binary-class datasets are included in the 18 datasets.

Table 1: Datasets Descriptions

Data sets	Features	Records	class
ALLAML	7129	72	2
CBSMR(ConnectionistBench_Sonar_Mines_Rocks)	60	208	2
CLL_SUB_1111	11340	111	3
FTM (ForestTypeMapping)	27	326	4
GLIOMA	4434	50	4
Ionosphere	34	351	2
Lung	3312	203	5
lung-discrete	325	73	7
Lymphoma	4026	96	9
ORL	1024	400	40
orlraws10P	10304	100	10
ProsteGE	5966	102	2
QSAR_B (QSAR_Biodegradation)	41	1055	2
SCADI	206	70	7
ThoracisSurgery	17	470	2

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 2206 - 2225

UrbanLandCover	148	168	9
warpAR10P	2400	130	10
warpPIE10P	2420	210	10

4.3. Models Used

4.3.1. K-NN

The K-Nearest Neighbor classifier (K-NN) [8] uses a distance measure like Euclidean distance to classify test data based on the known labels of the k closest neighbors. The majority of the class labels predicted by training patterns determine the class of the test patterns.

4.3.2. SVM

Support Vector Machines (SVM) [9] is operated by locating the maximal margin hyperplane, or the linear separator, that is as far away from the positive and negative training data as possible. To make the linear separator very non-linear in the input space, kernel functions may be implemented to project the data into a high-dimensional space.

4.3.3. RF

The supervised learning method includes the well-known machine learning algorithm Random Forest (RF) [10]. It is used for ML problems involving both classification and regression. The idea of ensemble learning provides its basis. The outcomes of many decision trees are merged to get a single conclusion. Its popularity is increasing as a result of its adaptability and simplicity. The ultimate result is predicted by the random forest using predictions from each tree, and the majority votes for those predictions.

5. Results and Discussion

Tables 2–4 present the experimental results of the proposed WFS algorithm tested on eighteen datasets using K-NN, SVM, and RF classifiers, respectively. The first column presents the names of the datasets; the second to ninth columns present CA obtained using FS algorithms, namely WFS, FSRRW, MIFS, JMI, mRMR, CIFE, MRI, DCSF, and CWJR, and the tenth column presents the Number of Selected Features (NSF) using the proposed algorithm.

For comparison of the proposed algorithm WFS, eight renowned FS algorithms (FSRRW, MIFS, JMI, mRMR, CIFE, MRI, DCSF, and CWJR) are taken, which are termed as "other algorithms" now onwards. The W/L in the last row of each table represents win/loss scores. Win indicates that the proposed algorithm WFS performs "better or equal to" other algorithms, and loss indicates that the proposed algorithm WFS performs "lesser than" the other algorithms. The highest CA in each row is written in bold letters.

As per Table 2, WFS achieves better CA on fourteen datasets compared to those obtained by other FS algorithms. For CBSMR, lung, lung-discrete, lymphoma, and ORL datasets, the WFS algorithm achieves slightly less CA. On CBSMR datasets, the MRI algorithm achieves better CA compared to other algorithms. On lung, lung-discrete, lymphoma, and ORL datasets, the FSRRW method achieves better CA compared to other algorithms. In the WFS algorithm, the ALLAML

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 2206 - 2225

dataset achieves a maximum CA of 98.75% compared to eighteen datasets and a minimum CA of 76.25% on ORL datasets compared to eighteen datasets. Overall, the WFS algorithm achieves an average CA of 88.66% across all datasets with an average of seven selected features, which is better compared to other algorithms. With respect to W/L, WFS directs maximum win, i.e., seventeen in comparison with the CIFE algorithm, and minimum win, i.e., thirteen in comparison with the FSRRW algorithm. This refers to the hardest competitor of WFS is FSRRW.

According to Table 3, we archived the better CA on twelve datasets using the proposed algorithm WFS when compared to other algorithms. The WFS algorithm produces slightly less CA for the lung, lung-discrete, lymphoma, ORL, warpAR10P, and warpPIE10P datasets. In the lung datasets, the CWJR algorithm outperforms other algorithms in terms of CA. In comparison to other algorithms, the FSRRW algorithm produces better CA in the lung-discrete, lymphoma, ORL, warpAR10P, and warpPIE10P datasets. When compared to eighteen datasets, the ALLAML dataset in the WFS algorithm achieves a maximum CA of 100%, while the warpAR10P dataset obtains a minimum CA of 83%. The WFS algorithm achieves an average CA of 89.88% across all datasets with only an average of 7 features, and the average CA of the WFS algorithm is better compared to other algorithms. W/L, WFS directs maximum win, i.e., eighteen in comparison with the CIFE algorithm, and minimum win, i.e., twelve in comparison with the FSRRW algorithm.

Table 2: Averages CA (in %) Evaluated by the K-NN

Datasets	WFS	FSRRW	MIFS	JMI	mRMR	CIFE	MRI	DCSF	CWJR	NSF
ALLAML	**98.75**	98.39	95.03	97.42	95.95	75.53	91.58	89.02	96.78	3
CBSMR	80.88	83.15	77.42	82.67	78.38	81.39	**84.48**	82.9	80.73	3
CLL_SUB_1111	**84.77**	84.77	70.7	80.6	74.54	52.99	78.92	74.99	83.41	5
FTM	**87.66**	82.62	81.45	77.76	81.54	73.76	81.36	81.42	82.15	6
GLIOMA	**90**	87.59	76.75	86.97	76.05	58.34	73.72	70.67	85.7	4
Ionosphere	**94.29**	89.92	89.83	88.92	89.91	87.82	89.29	87.34	89.45	4
Lung	95.57	**96.04**	89.17	95.56	93.74	73.13	94.93	94.37	95.36	7
lung-discrete	86.52	**90.84**	84.42	88.37	80.46	66.28	87.87	87.3	90.08	9
lymphoma	85.11	**94.33**	89.21	89.9	89.68	58.56	92.01	92.62	92.3	9
ORL	76.25	**83.78**	81.03	81.12	82.92	52.69	80.63	82.43	83.54	9
orlraws10P	**98**	94.6	89.5	92.6	93.93	63.17	87.6	85.37	94.7	8
Prostate_GE	**97**	95.04	89.01	93.64	92.9	86.99	93.33	93.17	94.31	7
QSAR_B	**85.01**	78.46	75.67	75.14	77.94	75.96	76.86	77.79	76.21	8
SCADI	**88.57**	78.69	70.61	71.43	71.12	60.67	78.67	75.1	78.29	5
ThoracisSurgery	**85.95**	81.84	81.01	80.75	80.57	80.94	81.49	81.35	81.16	5
UrbanLandCover	**85.77**	79.27	72.74	64.48	70.06	44.51	80.03	78.36	62.74	7
warpAR10P	**80**	76.52	77.72	78.2	78.06	36.87	64.62	66.2	79.77	5
warpPIE10P	**95.76**	95.58	82.63	93.85	94.38	80.41	93.14	94.59	91.12	10
Average	**88.66**	87.3	81.88	84.41	83.45	67.22	83.92	83.06	85.43	7
W/L		13/5	16/2	14/4	16/2	17/1	14/4	14/4	15/3	

In Table 5, we compare the proposed algorithm WFS to other algorithms and archive all eighteen datasets for better CA. In the WFS algorithm, ALLAML and ORLraws10P datasets achieve a maximum CA of 100% compared to eighteen datasets and a minimum CA of 75.75% on ORL datasets compared to eighteen datasets. It has been observed that WFS achieves an average CA of up to 89.14% across eighteen datasets, and WFS average CA is better than compared to other algorithms. With respect to W/L, WFS directs eighteen wins and zero losses compared to other algorithms.

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 2206 - 2225

Figures 2-4 effectively summarize the statistical win/loss statistics of WFS in comparison to other algorithms. Most of the proposed algorithm consists of the win frequencies on the K-NN, SVM, and RF classifiers. This research demonstrates that in the WFS algorithm, selected features provide more significant information, which can mostly increase the classification effectiveness, and that the WFS algorithm is more effective than other algorithms.

Table 5, represents the summative comparison of WFS and other algorithms: the first block presents the best CA out of all classifiers in other algorithms, the second column represents the best CA out of all classifiers in the WFS algorithm; the third block represents the number of selected features in the WFS algorithm; and the last block represents the index of selected feature sets in WFS. Figure 5 represents the Summative comparison of WFS and other algorithms.

Table 3: Averages CA (in %) Evaluated by the SVM

Datasets	WFS	FSRRW	MIFS	JMI	mRMR	CIFE	MRI	DCSF	CWJR	NFS
ALLAML	**100**	99.02	94.69	97.44	94.92	82.73	93.89	93.81	93.64	6
CBSMR	**80.73**	74.65	75.44	78.61	75.54	75.54	76.15	76.64	75.69	7
CLL_SUB_1111	**86.59**	77.36	70.98	73.31	67.93	48.77	74.83	75.16	79.7	8
FTM	**88.94**	87.19	84.9	79.62	76.5	76.5	81.96	83.81	84.16	5
GLIOMA	**92**	86.72	76.82	82.24	70.7	59.53	71	70.01	84.01	7
Ionosphere	**95.72**	87.54	88.26	84.37	83.19	83.19	83	82.75	88.24	8
Lung	95.5	95.14	90.79	94.36	92.04	77.61	93.46	93.34	**95.75**	9
lung-discrete	86.25	**91.96**	88.29	88.77	81.98	69.5	88.22	88.29	90.17	8
lymphoma	87.22	**96.28**	90.58	94.45	90.09	56.98	94.37	93.56	95.73	2
ORL	83.5	**86.02**	82.03	83.38	82.88	58.55	83.64	85.22	84.86	9
orlraws10P	**99**	95.93	92.5	95.73	95.43	72.97	92.9	91.23	97.07	8
Prostate_GE	**97.09**	93.2	89.83	91.38	89.81	85.02	90.88	90.76	90.08	6
QSAR_B	**86.24**	81.48	80.1	78.48	77.76	77.76	79.73	80.83	79.72	9
SCADI	**88.57**	84.79	84.7	75.59	72.67	72.67	86.03	84.33	84.52	4
ThoracisSurgery	**85.96**	85.11	85.11	85.11	85.11	85.11	85.11	85.11	85.11	3
UrbanLandCover	**86.83**	80.56	78.82	67.5	66.57	58.24	81.2	80.46	66.57	7
warpAR10P	83	**91.03**	86.6	78.6	86.41	58.8	81.72	81.11	82.22	7
warpPIE10P	95.23	**96.52**	91.88	94.71	94.71	90.14	93.86	95.36	91.89	8
Average	**89.88**	88.36	85.13	84.65	82.46	71.65	85.11	85.1	86.06	7
W/L		12/6	15/3	15/3	16/2	18/0	15/3	14/4	16/4	

Table 5 and Figure 5 clearly show that the WFS algorithm selected feature set is more relevant compared to other algorithms, and the WFS algorithm selected features are mostly non-redundant feature. The WFS algorithm selects at most 10 features in each dataset, whereas other algorithms

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 2206 - 2225

select at most 30 features. In the WFS algorithm, 90.16% of the average CA is better compared to other algorithms 89.97% of the average CA. Overall, results show that the WFS algorithm has the best ability to select relevant features and significantly improves the classifier's efficiency.

Table 6 and Figure 6 shows the comparison between the CA of the WFS algorithm (in selected feature set) and the CA of the PCA algorithm (with the same number of selected features and classifier as in Table 5).in Table 6, the first block presents the best CA out of all classifiers in the WFS algorithm; the second column represents the CA of the PCA algorithm with the same number of selected feature sets and classifiers as in the WFS algorithm. Figure 6 represents the comparison of the WFS algorithm with the PCA algorithm.

In Table 6, the WFS algorithm archives better CA on the seventeen-dataset compared to CA of PCA algorithm with same number of selected features. For the ORL dataset WFS algorithm archives 83.5% of CA and PCA algorithm archives 84.25% of CA. in the ORL dataset WFS algorithm archives slightly low accuracy compare to PCA algorithm. The WFS algorithm, archives average CA of eighteen datasets is 90.13% and for the PCA algorithm archives average CA of eighteen datasets is 81.14%.

Table 4: Averages CA (in %) Evaluated by the RF

Datasets	WFS	FSRRW	MIFS	JMI	mRMR	CIFE	MRI	DCSF	CWJR	NFS
ALLAML	**100**	96.58	94.25	95.39	94.37	82.04	92.47	91.47	95.31	6
CBSMR	**81.19**	79.45	74.43	79.03	77.1	78.33	80.12	80.87	78.38	5
CLL_SUB_1111	**85.75**	81.9	70.3	76.53	71.23	53.14	76.43	73.68	80.06	8
FTM	**87.7**	83.26	81.9	78.23	82.48	74.89	81.82	82.3	83.12	4
GLIOMA	**92**	82.44	80.2	79.96	65.19	48.72	62.57	62.66	79.88	5
Ionosphere	**94.89**	92.61	91.29	92.51	91.54	91.03	91.56	92.16	92	8
Lung	**95.54**	92.5	88.11	91.91	89.84	77.79	90.83	89.37	93.03	6
lung-discrete	**87.85**	78.5	69.5	76.43	66.89	58.45	75.73	76.12	77.85	8
lymphoma	**87.33**	86.73	77.69	84.26	78.1	55.06	84.47	84.09	86.44	7
ORL	**75.75**	69.79	61.95	67.46	67.19	39.18	62.64	63.41	69.78	10
orlraws10P	**100**	93.67	82.2	92.83	87.1	53.9	84.03	78.57	95.27	6
Prostate_GE	**95.09**	91.76	88.5	91.23	89.64	85.48	92.62	90.53	91.49	6
QSAR_B	**86.25**	81.22	80.02	78.58	81.24	77.9	79.08	80.86	80.77	5
SCADI	**88.57**	84.35	77.86	76.24	80.35	68.68	84.01	83.56	85.28	7
ThoracisSurgery	**85.74**	82.09	83.54	81.71	82.63	81.49	81.57	82.07	83.47	4
UrbanLandCover	**87.53**	79.68	75.96	68.37	70.79	51.5	79.05	78.19	66.04	5
warpAR10P	**80**	79.02	71.86	76.63	77.08	38.8	68.95	64.97	78.37	7
warpPIE10P	**93.33**	88.66	86.56	86.8	86.78	74.22	88.31	88.54	87.32	8
Average	**89.14**	84.68	79.78	81.89	79.97	66.14	80.9	80.19	83.55	7
W/L		18/0	18/0	18/0	18/0	18/0	18/0	18/0	18/0	

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 2206 - 2225

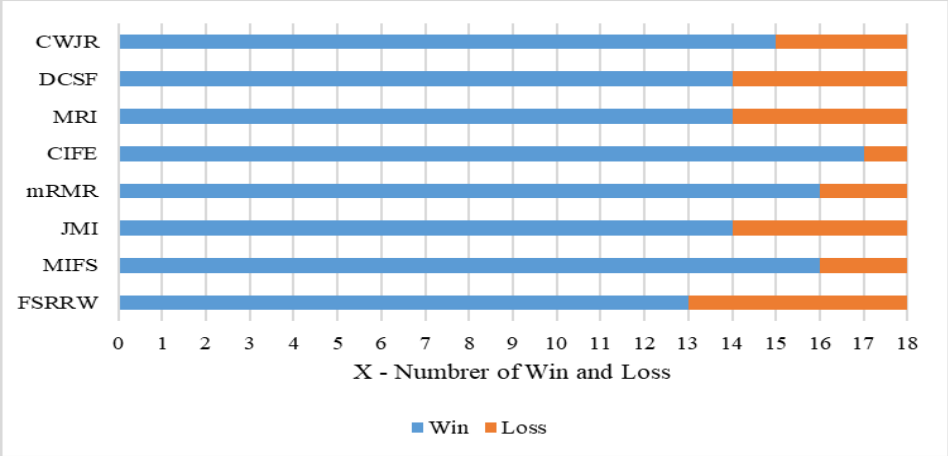

Figure 2: WFS Algorithm Performance with the K-NN Classifier

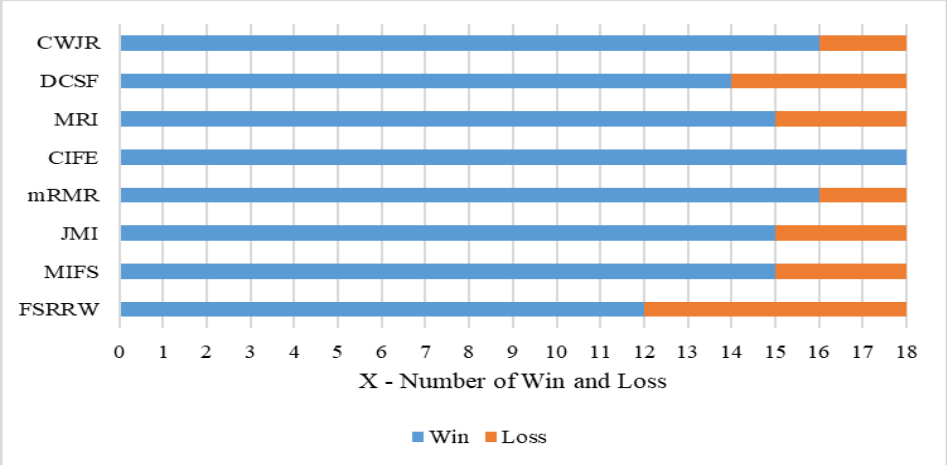

Figure 3: WFS Algorithm Performance with the SVM Classifier

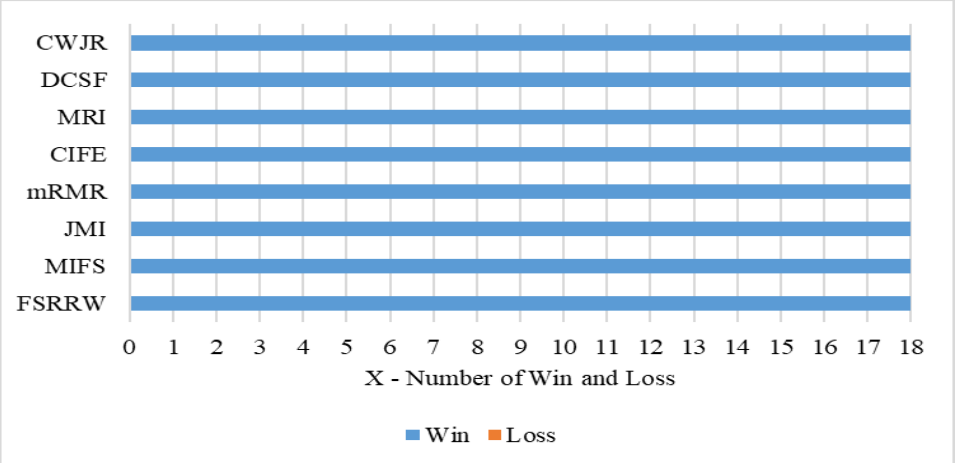

Figure 4: WFS Algorithm Performance with the RF Classifiers

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 2206 - 2225

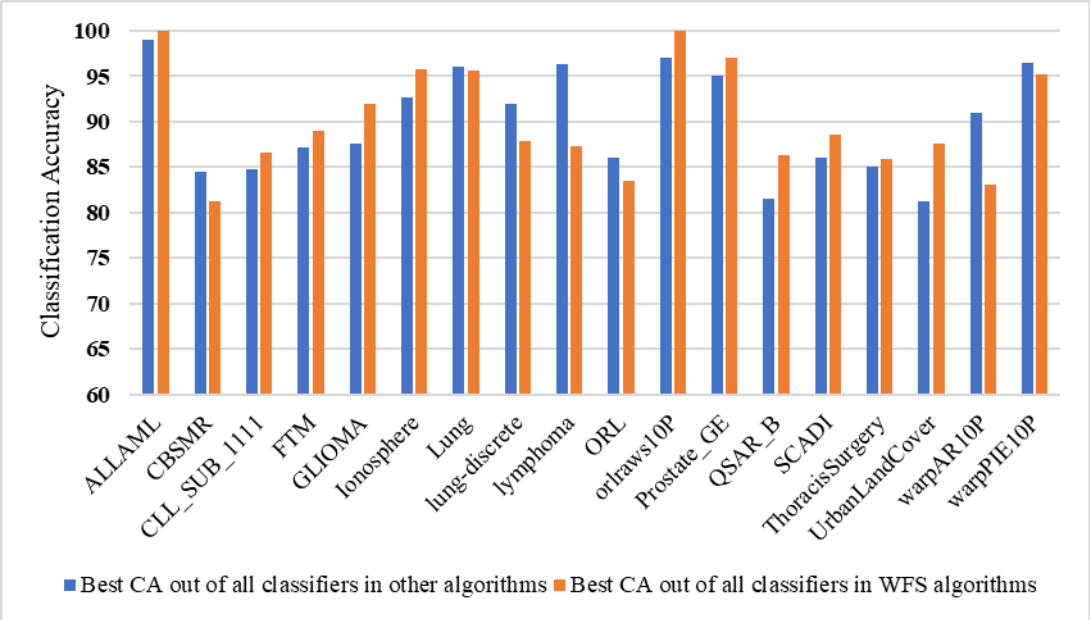

Figure 5: Summative comparison of WFS and other algorithms.

Table 5: Summative comparison of WFS and other algorithms.

Datasets	Best CA out of all classifiers in other algorithms		Best CA out of all classifiers in WFS algorithms		NSF	Index of selected feature sets
	CA	Classifier	CA	classifier		
ALLAML	99.02	SVM	**100**	SVM	6	4231, 6540, 4380, 6184, 6544, 1833
CBSMR	84.48	K-NN	81.19	RF	5	20, 59, 44, 10, 36
CLL_SUB_1111	84.77	K-NN	**86.59**	SVM	8	10295, 7335, 8880, 4564, 7147, 10612, 6285, 2631
FTM	87.19	SVM	**88.94**	SVM	5	0, 1, 7, 8, 24
GLIOMA	87.59	K-NN	**92**	SVM	7	1029, 884, 3837, 2045, 2002, 2759, 1156
Ionosphere	92.61	RF	**95.72**	SVM	8	21, 7, 33, 27, 4, 9, 14, 8
Lung	**96.04**	K-NN	95.57	K-NN	7	2239, 566, 1492, 2341, 3242, 1182, 2009
lung-discrete	**91.96**	SVM	87.85	RF	8	8, 120, 70, 55, 130, 222, 18, 202
lymphoma	**96.28**	SVM	87.33	RF	7	2287, 236, 2746, 3552, 3656, 2295, 1189
ORL	**86.02**	SVM	83.5	SVM	9	173, 348, 800, 28, 980, 904, 101, 151, 228, 108
orlraws10P	97.07	SVM	**100**	RF	6	944, 9621, 95, 10195, 6854, 781
Prostate_GE	95.04	K-NN	**97.09**	SVM	6	4183, 3685, 4646, 1978, 2585,

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 2206 - 2225

						3079
QSAR_B	81.48	SVM	**86.25**	RF	5	17, 35, 9, 8, 15, 5, 26
SCADI	86.03	SVM	**88.57**	SVM	4	40, 104, 83, 75
ThoracisSurgery	85.11	SVM	**85.96**	SVM	3	8,6,4
UrbanLandCover	81.2	RF	**87.53**	RF	5	4, 69, 39, 18, 92
warpAR10P	**91.03**	SVM	83.07	SVM	7	843,2342,1866,1140,2144,2145,2174
warpPIE10P	**96.52**	SVM	95.23	K-NN	8	1664, 1925, 697, 107, 2419, 97, 730, 1235
Average	89.97		**90.13**		7	

Figure 7 to 24 represents a correlation matrix heat map of selected features in eighteen datasets. A correlation matrix is a table that displays the correlation coefficients for different features. The matrix displays the correlation between each group of numbers in a table. A correlation heatmap uses colored cells to display the data on a typically monochrome scale while providing a 2D correlation coefficient between two discrete dimensions. The values of the second dimension are represented as columns in the table, while the values of the first dimension are shown as rows. The color of the cell indicates the percentage of measurements that match the dimensional value. The range of correlation is +1 to -1, with 1 representing highly redundant features in the positive direction and -1 representing highly redundant features in the negative direction. Hence, a value of 0 denotes that the features are non-redundant. Correlation heatmaps are perfect for data analysis and redundancy checking. Figure 7 to 24 clearly show that the WFS algorithm selected feature are mostly non-redundant features.

Table 6: Comparison of the WFS algorithm with the PCA algorithm

Datasets	classifier	NFS	Best CA out of all classifiers in WFS algorithm	CA of PCA algorithm with same number of selected feature set and classifier respected to WFS algorithm
ALLAML	SVM	6	**100**	91.61
CBSMR	RF	5	**81.19**	71.62
CLL_SUB_1111	SVM	8	**86.59**	64.85
FTM	SVM	5	**88.94**	82.77
GLIOMA	SVM	7	**92**	80
Ionosphere	SVM	8	**95.72**	92.05
Lung	K-NN	7	**95.57**	95.05
lung-discrete	RF	8	**87.85**	80.77
lymphoma	RF	7	**87.33**	82.54
ORL	SVM	9	83.5	**84.25**
orlraws10P	RF	6	**100**	95.2

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 2206 - 2225

Prostate_GE	SVM	6	**97.09**	84.18
QSAR_B	RF	5	**86.25**	82.73
SCADI	SVM	4	**88.57**	84.29
ThoracisSurgery	SVM	3	**85.96**	85.11
UrbanLandCover	RF	5	**87.53**	73.78
warpAR10P	SVM	7	**83.07**	46.92
warpPIE10P	K-NN	8	**95.23**	82.86
Average			**90.13**	81.14

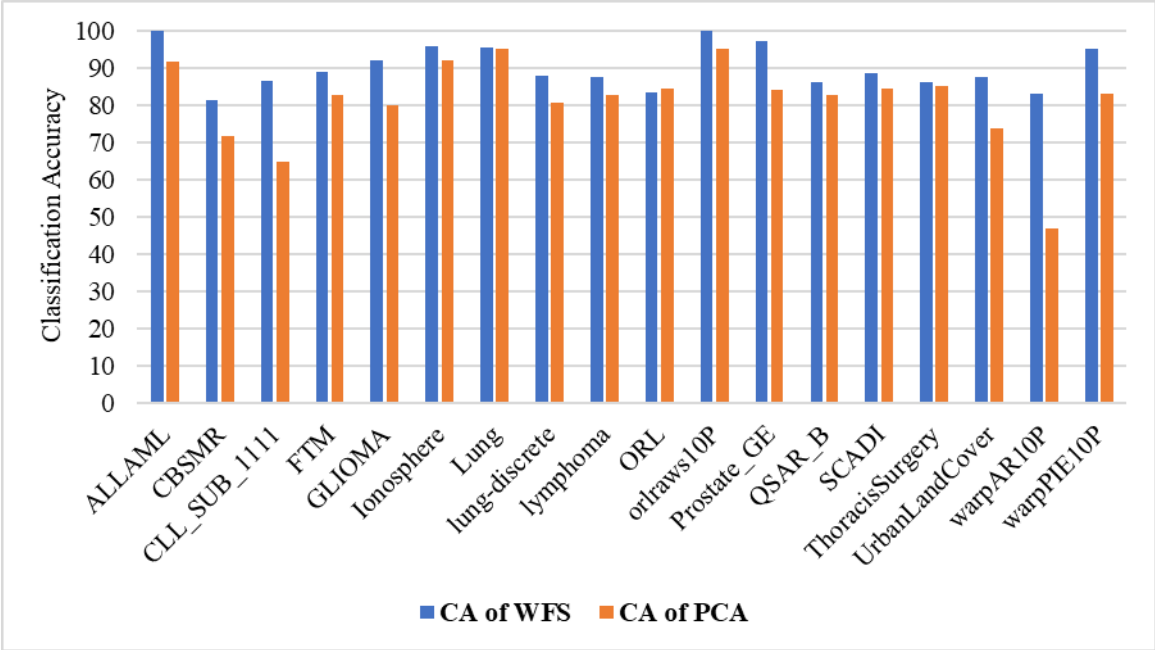

Figure 6: Comparison of the WFS algorithm with the PCA algorithm

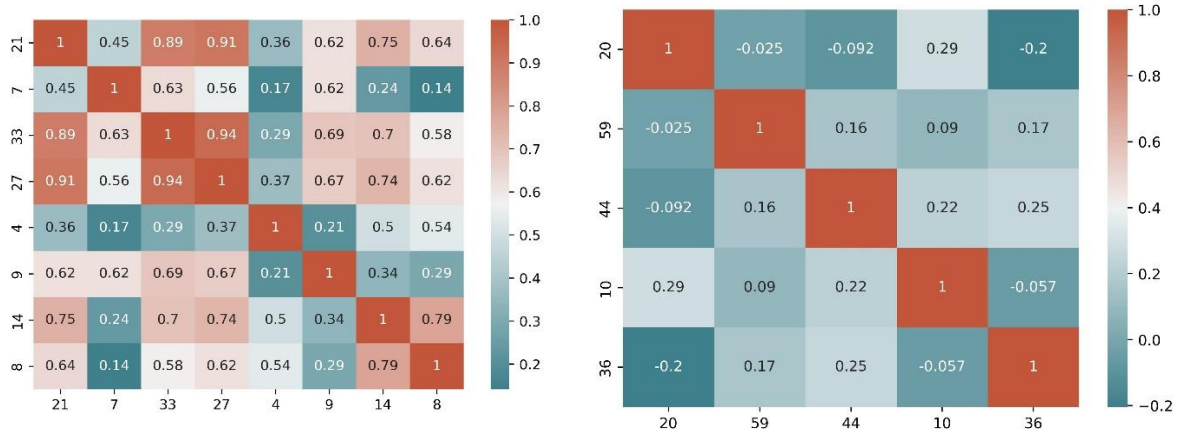

Figure 8: Correlation Matrix using Heat Map Representation on Selected Feature sets of CBSMR Dataset.

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 2206 - 2225

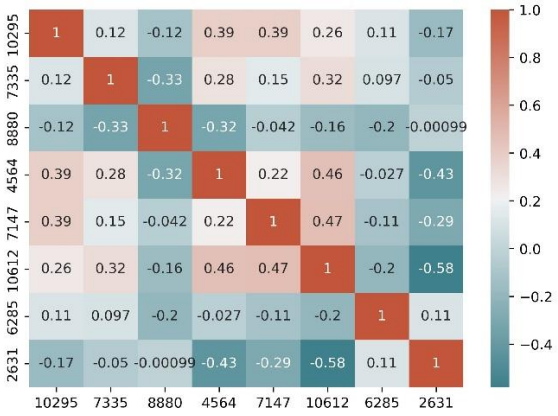

Figure 9: Correlation Matrix using Heat Map Representation on Selected Feature sets of CLL_SUB_1111 Dataset.

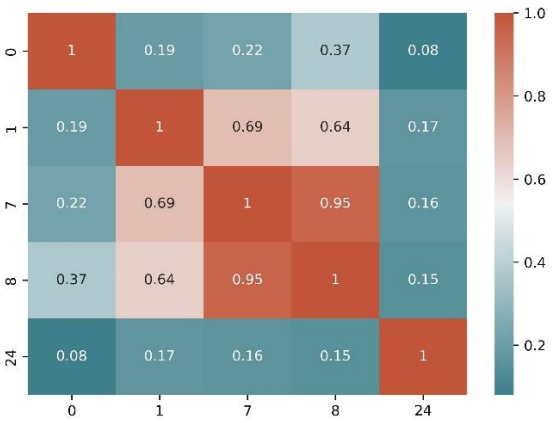

Figure 10: Correlation Matrix using Heat Map Representation on Selected Feature sets of FTM Dataset.

Figure 11: Correlation Matrix using Heat Map Representation on Selected Feature sets of GLIOMA Dataset.

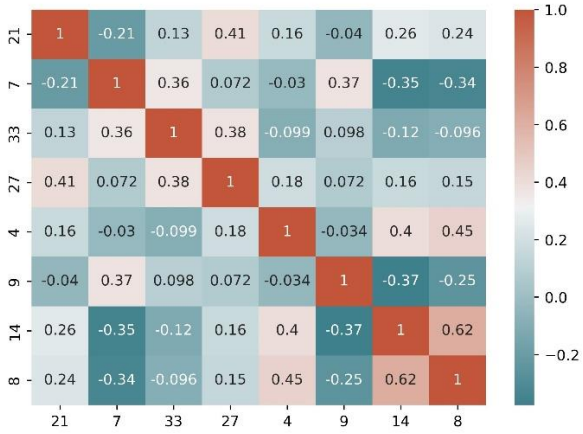

Figure 12: Correlation Matrix using Heat Map Representation on Selected Feature sets of Ionosphere Dataset.

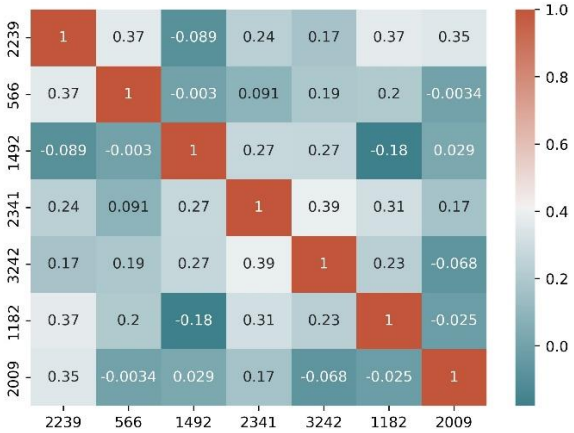

Figure 13: Correlation Matrix using Heat Map Representation on Selected Feature sets of Lung Dataset.

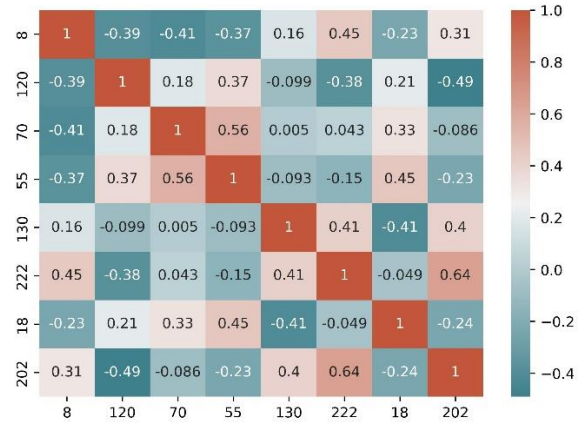

Figure 14: Correlation Matrix using Heat Map Representation on Selected Feature sets of lung-discrete Dataset.

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 2206 - 2225

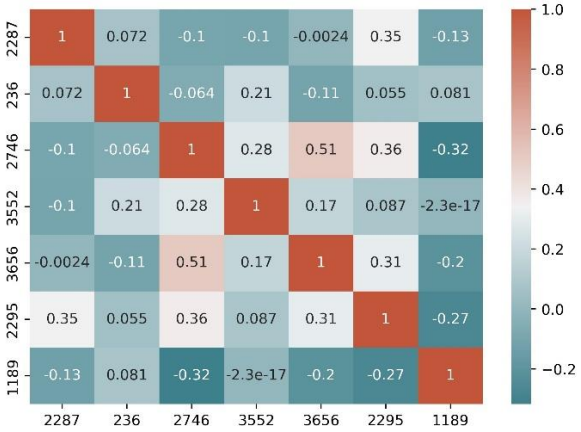

Figure 15: Correlation Matrix using Heat Map Representation on Selected Feature sets of lymphoma Dataset.

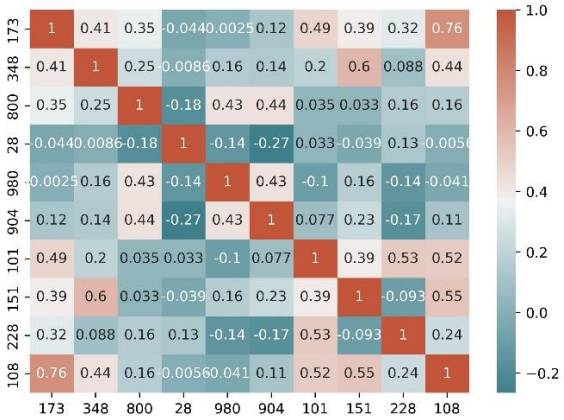

Figure 16: Correlation Matrix using Heat Map Representation on Selected Feature sets of ORL Dataset.

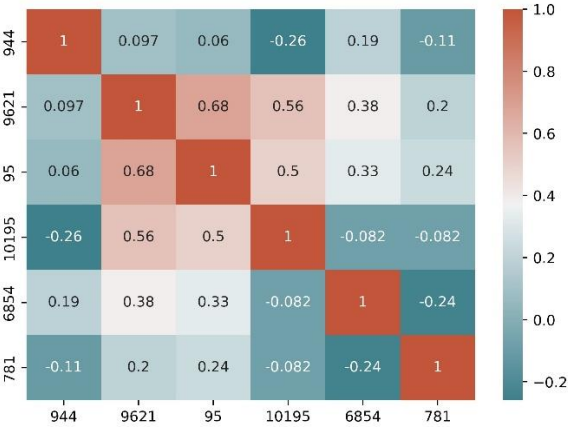

Figure 17: Correlation Matrix using Heat Map Representation on Selected Feature sets of orlraws10P Dataset.

Figure 18: Correlation Matrix using Heat Map Representation on Selected Feature sets of Prostate_GE Dataset.

Figure 19: Correlation Matrix using Heat Map Representation on Selected Feature sets of QSAR_B Dataset.

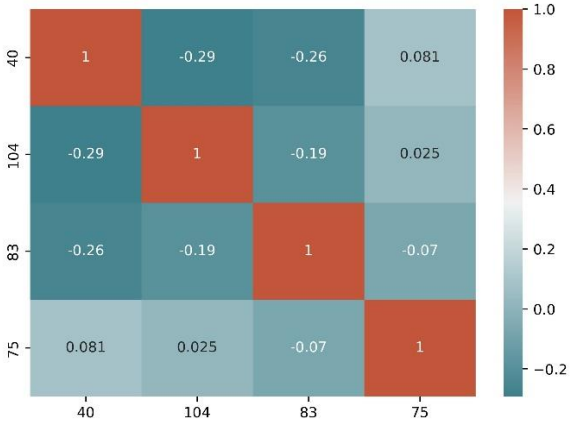

Figure 20: Correlation Matrix using Heat Map Representation on Selected Feature sets of SCADI Dataset.

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 2206 - 2225

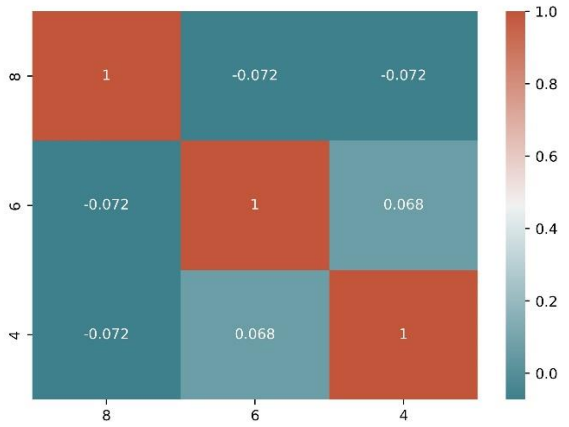

Figure 21: Correlation Matrix using Heat Map Representation on Selected Feature sets of ThoracisSurgery Dataset.

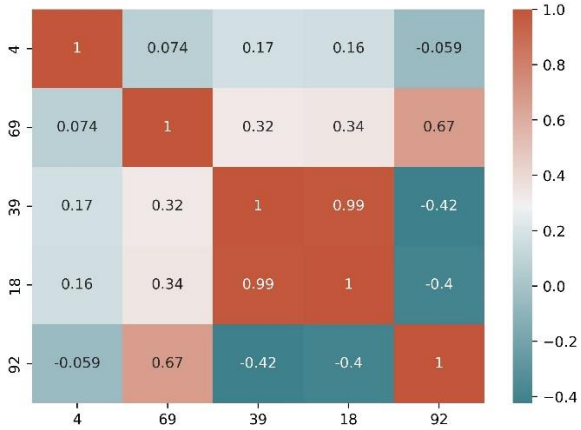

Figure 22: Correlation Matrix using Heat Map Representation on Selected Feature sets of UrbanLandCover Dataset.

Figure 23: Correlation Matrix using Heat Map Representation on Selected Feature sets of warpAR10P Dataset.

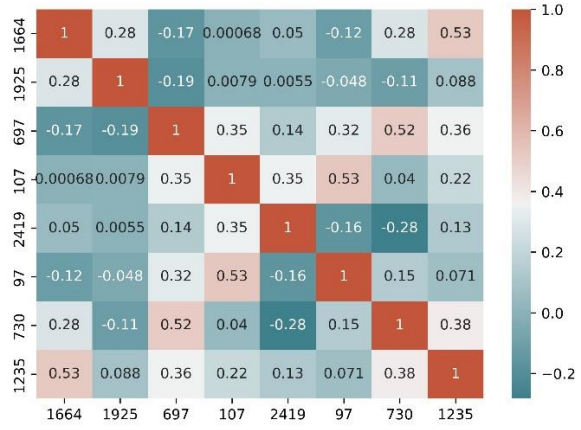

Figure 24: Correlation Matrix using Heat Map Representation on Selected Feature sets of warpPIE10P Dataset.

6. Conclusion and Future Work

The wrapper methods can be used to select the most relevant features from a dataset with a huge number of features. The scheme begins by randomly selecting features for each fold and selecting the best fold based on classification accuracy (CA). To improve CA and reduce the number of selected features, the exhaustive feature selection method (Wrapper method) is used.

The size of the selected features is much smaller than that of the original dataset of higher dimensionality. On several datasets with reduced features (selected features), the CAs achieved using K-NN, SVM, and RF classifiers are better than or almost similar to those obtained by nine comparable algorithms. In maximum datasets, WFS achieved better CA compared to the nine comparable algorithms. The simulation of the scheme makes use of eighteen datasets. In the selected feature sets of the proposed algorithms, the average CA of eighteen datasets is 88.66% in K-NN, 89.88% in SVM, and 89.14% in RF classifiers with at most 10 selected features. The challenges and future scope include testing the performance of the model with thousands or millions of features. Further, it will be interesting to calculate classification accuracy, using various machine learning and deep learning algorithms with a carefully selected set of parameter values.

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 2206 - 2225

Compliance with ethical standards

Conflict of Interest: The authors declare that there is no conflict of interest regarding the publication of this paper.

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent: Informed consent was obtained from all individual participants included in the study.

Data availability statements: Data is available from the authors upon reasonable request.

Funding: Not Applicable

References

[1] Kamber, M., Han, J., Pei, J.: Data Mining Concepts and Techniques, 2nd ed. Morgan Kaufmann. San Francisco (2006).

[2] Iguyon, I., Elisseeff, A.: An introduction to variable and feature selection. Journal of Machine Learning Research, 3, 1157-1182 (2003).

[3] Chakraborty, D., Pal, N. R.: Selecting useful groups of features in a connectionist framework. IEEE Transactions on Neural Networks 19(3), 381-396 (2008).

[4] Basak, J., De, R. K., Pal, S. K.: Unsupervised feature selection using a neuro-fuzzy approach. Pattern Recognition Letters 19(11), 997-1006 (1998).

[5] Saxena, A. K., Dubey, V. K.: A Survey on feature selection algorithms. International Journal on Recent and Innovation Trends in Computing and Communication 3(4), 1895–1899 (2015).

[6] Setiono, R., Liu, H.: Neural-network feature selector. IEEE Transactions on Neural Networks 8(3), 654–662 (1997).

[7] Zadeh, L. A.: Fuzzy Logic. Computer 21(4), 83–93 (1988).

[8] Yang, S., Jian, H., Ding, Z., Hongyuan, Z., Giles, C. L.: IK-NN: Informative K-nearest neighbor pattern classification. , in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), (2007).

[9] Wang, Q.: Support Vector Machine Algorithm in Machine Learning. , in: 2022 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), pp 750–756. IEEE, (2022)

[10] Genuer, R., Poggi, J. M., Tuleau-Malot, C.: Variable selection using random forests. Pattern Recognition Letters 31(14), 2225–2236 (2010).

[11] Zhao, S., Wang, M., Ma, S., Cui, Q.: A feature selection method via relevant-redundant weight. Expert Systems with Applications 207, 117923 (2022).

[12] Kwak, N., Choi, C. H.: Input feature selection for classification problems. IEEE Transactions on Neural Networks 13(1), 143–159 (2002).

[13] Yang, H., John, M.: Data visualization and feature selection: New algorithms for nongaussian data. Advances in neural information processing systems 12 (1999).

[14] Peng, H., Long, F., Ding, C.: Feature selection based on mutual information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(8), 1226–1238 (2005).

[15] Lin, D., Tang, X.: Conditional infomax learning: An integrated framework for feature extraction and fusion. European conference on computer vision , 68–82 (2006).

[16] Wang, J., Wei, J. M., Yang, Z., Wang, S. Q.: Feature selection by maximizing independent classification information. IEEE Transactions on Knowledge and Data Engineering 29(4), 828–841 (2017).
[17] Gao, W., Hu, L., Zhang, P., Wang, F.: Feature selection by integrating two groups of feature evaluation criteria. Expert Systems with Applications 110, 11–19 (2018).
[18] Zhang, P., Gao, W., Hu, J., Li, Y.: A conditional-weight joint relevance metric for feature relevancy term. Engineering Applications of Artificial Intelligence 106, 104481 (2021).
[19] Chernbumroong, S., Cang, S., Yu, H.: Maximum relevancy maximum complementary feature selection for multi-sensor activity recognition. Expert Systems with Applications 42(1), 573–583 (2015).
[20] Kashef, S., Nezamabadi-pour, H.: An advanced ACO algorithm for feature subset selection. Neurocomputing 147(1), 271–279 (2015).
[21] Dorigo, M., Gambardella, L. M.: Ant colony system: A cooperative learning approach to the traveling salesman problem. IEEE Transactions on Evolutionary Computation 1(1), 53–66 (1997).
[22] Zhu, P., Zuo, W., Zhang, L., Hu, Q., Shiu, S. C. K.: Unsupervised feature selection by regularized self-representation. Pattern Recognition 48(2), 438–446 (2015).
[23] Yao, J., Mao, Q., Goodison, S., Mai, V., Sun, Y.: Feature selection for unsupervised learning through local learning. Pattern Recognition Letters 53, 100–107 (2015).
[24] Lee, J., Kim, D. W.: Mutual Information-based multi-label feature selection using interaction information. Expert Systems with Applications 42(4), 2013–2025 (2015).
[25] Wang, Y., Liu, Y., Feng, L., Zhu, X.: Novel feature selection method based on harmony search for email classification. Knowledge-Based Systems 73, 311–323 (2015).
[26] Lee, J., Kim, D.-W.: Memetic feature selection algorithm for multi-label classification. Information Sciences 293, 80–96 (2015).
[27] Hu, Q., Xie, Z., Yu, D.: Hybrid attribute reduction based on a novel fuzzy-rough model and information granulation. Pattern Recognition 40(12), 3509–3521 (2007).
[28] Qian, Y., Wang, Q., Cheng, H., Liang, J., Dang, C.: Fuzzy-rough feature selection accelerator. Fuzzy Sets and Systems 258, 61–78 (2015).
[29] Duhamel, P., Vetterli, M.: Fast fourier transforms: A tutorial review and a state of the art. Signal Processing 19(4), 259–299 (1990).
[30] He, S., Chen, H., Zhu, Z., Ward, D. G., Cooper, H. J., Viant, M. R., Heath, J. K., Yao, X.: Robust twin boosting for feature selection from high-dimensional omics data with label noise. Information Sciences 291, 1–18 (2015).
[31] Mitchell, M.: An introduction to genetic algorithms. MIT press, 1998.
[32] ElAlami, M. E.: A filter model for feature subset selection based on genetic algorithm. Knowledge-Based Systems 22(5), 356–362 (2009).
[33] Aghdam, M. H., Ghasem-Aghaee, N., Basiri, M. E.: Text feature selection using ant colony optimization. Expert Systems with Applications 36(3 PART 2), 6843–6853 (2009).
[34] Oduntan, I. O., Toulouse, M., Baumgartner, R., Bowman, C., Somorjai, R., Crainic, T. G.: A multilevel tabu search algorithm for the feature selection problem in biomedical data. Computers and Mathematics with Applications 55(5), 1019–1033 (2008).
[35] Chen, W. C., Tseng, S. S., Hong, T. P.: An efficient bit-based feature selection method. Expert Systems with Applications 34(4), 2858–2869 (2008).
[36] Sivagaminathan, R. K., Ramakrishnan, S.: A hybrid approach for feature subset selection using neural networks and ant colony optimization. Expert Systems with Applications 33(1), 49–60 (2007).

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 2206 - 2225

[37] Abdel-Aal, R. E.: GMDH-based feature ranking and selection for improved classification of medical data. Journal of Biomedical Informatics 38(6), 456–468 (2005).
[38] Zhu, F., Guan, S.: Feature selection for modular GA-based classification. Applied Soft Computing Journal 4(4), 381–393 (2004).
[39] FRIEDMAN, N., GEIGER, D., GOLDSZMIDT, M.: Bayesian network classifiers. Machine Learning, 131–163 (1997).
[40] Quinlan, J. R.: Induction of decision trees. Machine Learning 1(1), 81–106 (1986).
[41] Inza, I., Larrañaga, P., Etxeberria, R., Sierra, B.: Feature Subset Selection by Bayesian network-based optimization. Artificial Intelligence 123(1–2), 157–184 (2000).
[42] Saxena, A., Pal, N. R., Vora, M.: Evolutionary Methods for Unsupervised Feature Selection Using Sammon's Stress Function. Fuzzy Information and Engineering 2(3), 229–247 (2010).
[43] Dy, J. G., Brodley, C. E.: Feature Subset Selection and Order Identification for Unsupervised Learning., in: Proceedings of 17th International Conference on Machine Learning, 2000.
[44] Mitra, P., Murthy, C. A., Pal, S. K.: Unsupervised feature selection using feature similarity. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(3), 301–312 (2002).
[45] Basu, S., Micchelli, C. A., Olsen, P.: Maximum entropy and maximum likelihood criteria for feature selection from multivariate data, in: IEEE International Symposium on Circuits and Systems (ISCAS), vol 3, pp 267-270 IEEE, Geneva, Switzerland (2000)
[46] Pal, S. K., De, R. K., Basak, J.: Unsupervised feature evaluation: A neuro-fuzzy approach. IEEE Transactions on Neural Networks 11(2), 366–376 (2000).
[47] Hall, A.: Correlation-based feature selection for discrete and numeric class machine learning. (2000).
[48] Heydorn, R. P.: Redundancy in Feature Extraction. IEEE Transactions on Computers C–20(9), 1051–1054 (1971).
[49] Das, S. K.: Feature selection with a linear dependence measure. IEEE transactions on Computers 100(9), 1106–1109 (1971).
[50] Yan, X., Nazmi, S., Erol, B. A., Homaifar, A., Gebru, B., Tunstel, E.: An efficient unsupervised feature selection procedure through feature clustering. Pattern Recognition Letters 131, 277–284 (2020).
[51] 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 38(5), 5809–5815 (2011).
[52] Peng, Y., Wu, Z., Jiang, J.: A novel feature selection approach for biomedical data classification. Journal of Biomedical Informatics 43(1), 15–23 (2010).
[53] Saxena, A. K., Dubey, V. K., Wang, J.: Hybrid feature selection methods for high-dimensional multi-class datasets. International Journal of Data Mining, Modelling and Management 9(4), 315–339 (2017).
[54] Dubey, V. K., Saxena, A. K.: A cosine-similarity mutual-information approach for feature selection on high dimensional datasets. Journal of Information Technology Research 10(1), 15–28 (2017).
[55] Browne, M. W.: Cross-validation methods. Journal of Mathematical Psychology 44(1), 108–132 (2000).
[56] Arizona State University Library, https://lib.asu.edu/, last accessed 2023/03/09.
[57] UCI Machine Learning Repository, http://archive.ics.uci.edu/ml, last accessed 2023/03/09.
[58] Dhablia, D., & Timande, S. (n.d.). Ensuring Data Integrity and Security in Cloud Storage.
[59] Dhabalia, D. (2019). A Brief Study of Windopower Renewable Energy Sources its Importance, Reviews, Benefits and Drwabacks. Journal of Innovative Research and Practice, 1(1), 01–05.

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 2206 - 2225

[60] Mr. Dharmesh Dhabliya, M. A. P. (2019). Threats, Solution and Benefits of Secure Shell. International Journal of Control and Automation, 12(6s), 30–35.

[61] Verma, M. K., & Dhabliya, M. D. (2015). Design of Hand Motion Assist Robot for Rehabilitation Physiotherapy. International Journal of New Practices in Management and Engineering, 4(04), 07–11.