An Ensemble of Intelligent Water Drop Algorithm for Feature Selection Optimization Problem

Article *in* Applied Soft Computing · February 2018

DOI: 10.1016/j.asoc.2018.02.003

5 authors, including:

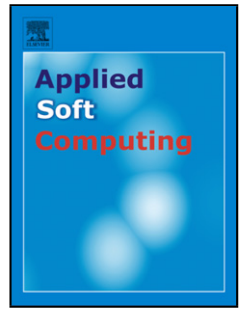
Basem Alijla
Islamic University of Gaza

Li-Pei Wong
Universiti Sains Malaysia

Ahamad Tajudin Khader
Universiti Sains Malaysia

Mohammed Azmi Al-Betar
Al-Balqa' Applied University

Accepted Manuscript



Title: An Ensemble of Intelligent Water Drop Algorithm for Feature Selection Optimization Problem

Author: Basem O. Alijla Chee Peng Lim Li-Pei Wong Ahamad Tajudin Khader Mohammed Azmi Al-Betar

PII: S1568-4946(18)30052-8
DOI: https://doi.org/doi:10.1016/j.asoc.2018.02.003
Reference: ASOC 4693

To appear in: *Applied Soft Computing*

Received date: 24-10-2016
Revised date: 26-1-2018
Accepted date: 3-2-2018

The manuscript highlights:

- FS-MRMC-IWD is introduced as an ensemble of the IWD for feature selection.
- The potential of the MRMC-IWD using real-world optimization is assessed.
- The FS-MRMC-IWD is evaluated using benchmark optimization tasks.
- The FS-MRMC-IWD is evaluated using two real-world problems
- Comparative studies are conducted to the classification accuracy and feature reduction.

An Ensemble of Intelligent Water Drop Algorithm for Feature Selection Optimization Problem

Basem O. Alijla[a,*], Chee Peng Lim[b], Li-Pei Wong[c], Ahamad Tajudin Khader[c], Mohammed Azmi Al-Betar[d]

[a]*Faculty of Information Technology, Islamic University of Gaza, Palestine*
[b]*Centre for Intelligent Systems Research, Deakin University, Australia*
[c]*School of Computer Sciences, Universiti Sains Malaysia, Malaysia*
[d]*Department of Information Technology, Al-Huson University College, Al-Balqa Applied Univer sity, P.O. Box 50, Al-Huson, Irbid, Jordan*

Abstract

Master River Multiple Creeks Intelligent Water Drops (MRMC-IWD) is an ensemble model of the intelligent water Drop, whereby a divide-and-conquer strategy is utilized to improve the search process. In this paper, the potential of the MRMC-IWD using real-world optimization problems related to feature selection and classification tasks is assessed. An experimental study on a number of publicly available benchmark data sets and two real-world problems, namely human motion detection and motor fault detection, are conducted. Comparative studies pertaining to the features reduction and classification accuracies using different evaluation techniques (consistency-based, CFS, and FRFS) and classifiers (i.e., C4.5, VQNN, and SVM) are conducted. The results ascertain the effectiveness of the MRMC-IWD in improving the performance of the original IWD algorithm as well as undertaking real-world optimization problems.

Keywords: Intelligent water drops, optimization, swarm intelligence, feature selection, motion detection, motor fault detection

*Corresponding author
Email address: balijla@iugaza.edu.ps (Basem O. Alijla)

Preprint submitted to Journal of Applied Soft Computing *February 5, 2018*

1. Introduction

Feature selection (FS) is a fundamental pre-processing step of a classification system. It refers to the process of selecting the most informative features, which represent the original set of features [1]. The importance of FS comes from the problems of the high dimensionality of the data set, e.g. text mining applications[2], gene expression array analysis [3]. FS is crucial in pattern recognition applications and widely used in the literature [4, 5, 6, 7]. FS techniques reduce the number of features by removing noisy, irrelevant, and redundant features. It enhances the performance of classification systems either in the terms of prediction accuracy or computation time.

Figure 1 depicts the structure of FS methods, which contains the following fundamental components: Subset generation, subset evaluation, stopping features. Subset generation is a search technique which explores the problem space for the optimal subset of features. Subset evaluation is an evaluation function which is used to score the goodness of the generated subset. Stopping criteria is a condition that terminates the search process. FS process starts with the subset generation, which utilizes a certain search algorithm to generate the candidate subsets. Evaluation function is used to evaluate the fitness of generated subset. This process is iterated until a stopping criterion is met. The outcome is a subset that optimizes the fitness value. The selected subset can be validated suing a classifier to ensure the classification accuracy.

Figure 1: structure of FS methods.

FS techniques can be broadly classified into two categories: wrapper-based and filter-based [5]. Wrapper-based methods are often used in conjunction with machine learning or data mining algorithms, which are used as a black box to score the subset of features. Filter-based methods are usually used as a pre-possessing step, and are independent of any learning (predicator) algorithms, which can be subsequently applied to evaluate classification accuracy with the selected subset of features.

Various evaluation techniques have been reported in the literature to evaluate the quality of the discovered feature subsets, e.g. rough sets [8, 6, 9], fuzzy rough sets [10, 11, 6, 12], probabilistic consistence [13, 14], correlation analysis [15, 16],information gain, mutual Information [17]. In [5] a good analysis on FS techniques is provided. The optimality of the subset of features discovered by filter-based methods is relative to the evaluation techniques. Filter-based methods are computationally less intensive, as compared with wrapper-based methods. However, wrapper-based methods can be more efficient as compared with filter-based methods. A wrapper-based method selects the feature subset and utilizes it directly with a learning algorithm [5]. The optimality of the selected subset is relative to the employed learning algorithm [18].

The simplest solution for feature selection is to generate all possible combinations, and choose the one with the minimum cardinality and maximum evaluation score [19]. Obviously, this requires an exhaustive search, and it is impractical for large data sets, where the number of alternatives grows exponentially with the data set size. Consider a given data set with N features, then 2^N possible feature subsets have to be searched [20]. To manage the complexity of the search process, several optimization methods, e.g., HC with forward selection and backward elimination [21], GA [21], PSO [22], ACO [23, 24], great deluge and non-linear great deluge [25, 26] have been used. A detailed taxonomy and the associated algorithms of FS can be found in[5].

Intelligent Water Drops (IWD) algorithm is a swarm based nature-inspired optimization introduced by Shah-Hosseini [27]. IWD is a constructive-based algorithm constructs an optimal solution through cooperation among a group of agents called water drops. The algorithm imitates the phenomena of a swarm of water drops flowing with soil along a river bed. Procedurally, each water drop incrementally constructs a solution through a series of iterative transitions from one node to the next until a complete solution is obtained. Water drops communicate with each other through an attribute called soil, which is associated with the path between any two points. The soil value is used to determine the direction of movement from the current node to the next, whereby a path with a lower amount of soil is likely to be followed. A detailed description for the IWD can be found in [28].

The IWD algorithm has been successfully employed to solve numerous combinatorial, and continuous optimization problems from different application fields [29]. It has been adopted to solve optimization problems such as function optimization, travelling salesman, multiple knapsack, n-queen puzzle problems, feature selection, parallel processor scheduling [30, 31, 28, 32, 33, 34, 35, 36]. IWD has been successfully used to solve multi-objectives optimization problem [37, 38, 39]. Some efforts have been made by researchers in investigating the fundamental algorithmic aspects of IWD in order to enhance the search capability [40, 41, 42, 43, 44].

This paper, investigates the applicability of the Master River Multiple Creeks Intelligent Water Drops (MRMC-IWD) model to real-world optimization problems related to feature selection and classification problems. To assess the performance of MRMC-IWD and to facilitate a performance comparison study with other state-of-the-art methods, benchmark and real-world optimization problems have been used. The problems include UCI (University of California Irvine machine learning repository) benchmark data sets [45] and two real-world problems, namely human motion detection and motor fault detection.

The rest of this paper is organized as follows. In Section 2, the MRMC-IWD model is briefly explained. Section 3 describes the experimental study of MRMC-IWD using the UCI benchmark data sets and the two real-world problems. Conclusion is presented in Section 4 .

2. The Master-River, Multiple-Creek IWD model (MRMC-IWD)

MRMC-IWD is an ensemble of the IWD algorithm proposed in [41]. It is inspired by the natural phenomena pertaining to a main river with multiple independent creeks flowing down the stream. MRMC-IWD utilizes divide-and-conquer strategy to enhance the search capability of the IWD algorithm. The rationale is based on dividing a complex problem into a number of sub-problems, (i.e., divide-and-conquer). Figure 2 depicts the structure of MRMC-IWD and its communication scheme.

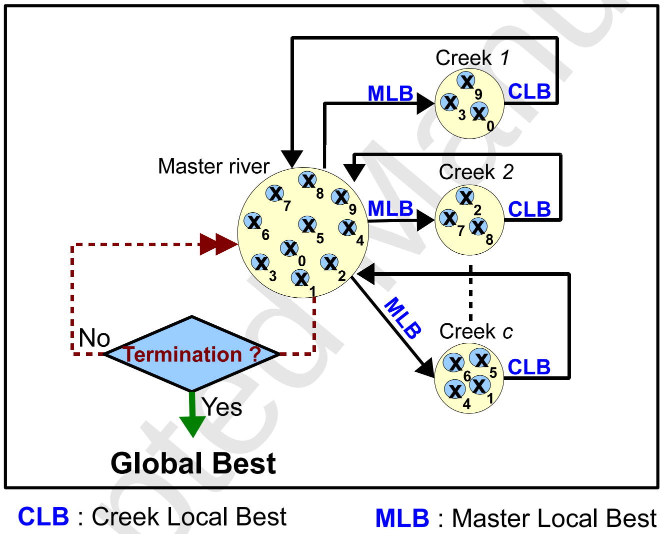


Figure 2: The structure of the MRMC-IWD and its communication model.

MRMC-IWD comprises a master river and multiple independent creeks. Firstly, a suitable decomposition technique (e.g. clustering algorithm) can be used to decompose the entire problem into a number of sub-problems, e.g., the k-means algorithm is used to cluster the entire problem into several sub-problems. The master river handles the entire problem, while each creek handles a sub-problem. In other words, the master river constructs a complete solution for the problem, while each creek contributes a partial solution. Both the master river and independent creeks maintain their parameters, (i.e. IWD static and dynamic parameters). A bilateral cooperative scheme between the master river and multiple creeks is introduced, in order to enable exchange of partial solutions between the master river and each creek, as shown in Figure 2. The partial solutions are known as the creek local best (CLB) water drops. The complete solution is known as the master local best (MLB) water drop. A sequential optimization process is adopted. Algorithm 1 depicts a pseudo-code of MRMC-IWD. A detailed description to the MRMC-IWD is found in [41].

Algorithm 1 Pseudo-code of the MRMC-IWD model.

1: Initialize the Master River and the C numbers of Creeks.
2: **while** Termination condition not reached **do**
3:     **for** each creek $i = 1...C$ **do**
4:         The master applies the IWD to construct the complete solution.
5:         The master river passes its MLB water drop to creek i.
6:         Creek i applies the IWD algorithm to construct its solutions
7:         Creek i passes its CLB water drop back to the master river
8:     **end for**
9: **end while**

3. MRMC-IWD for features subset selection

FS is a fundamental process in any data mining techniques. It is used to discover (i.e., select) a high quality feature subset that represents information of the original set of features [1]. The subset quality is evaluated in two aspects, namely the subset evaluation score and the subset size. FS is the search process for the subset that has the maximum evaluation score [4]. The objective function shown in Eq. (1), in [6], can be used to represent FS as a maximization problem.

$$\max_{x} \left\{ \gamma_{\xi}(x)) \times \frac{|D| - S(x)}{|D|} \right\} \tag{1}$$

where $\gamma_{\xi}(x))$ is the evaluation score of feature subset x, $|D|$ is the dimension of the complete set of features, and $S(x)$ is the dimension of feature subset x.

In this paper, a filter-based FS method is built by coupling the MRMC-IWD (i.e, search method) with different subset evaluation techniques, which are available in WEKA (Waikato Environment for Knowledge Analysis). WEKA is open source software contains a number of state-of-the-art knowledge analysis and data mining models. It has been widely used in both academic and industrial domains [46]. The aim of building MRMC-IWD with WEKA is to leverage on its built-in models (i.e., subset evaluation techniques, and classifiers) for evaluating the effectiveness of MRMC-IWD in tackling FS tasks. Three evaluation techniques, namely Fuzzy Rough Set Feature Selection (FRFS) [6], correlation-based Feature Selection (CFS) [15], and probabilistic consistence [13] are employed in MRMC-IWD. They cover three different principle for assisting the informative representation of data set (i.e., fuzzy rough theory, probabilistic data consistence, and statistical valuables correlation). The potential of MRMC-IWD is evaluated using UCI benchmark data sets [45] and two real-world problems related to FS and classification, namely human motion detection and motor fault detection. Detailed descriptions are as follows.

3.1. MRMC-IWD for FS using benchmark data sets

The applicability of MRMC-IWD is investigated in tackling FS problems using UCI benchmark data set. The main characteristics of the UCI data sets are described in Section 3.1.1 The experiments results and the comparative studies are discussed in Section 3.1.2.

3.1.1. Data sets

A total of seven real-valued UCI FS and classification benchmark data sets are used in the experimental study. Table 1 summarizes the main characteristics of the data sets. They have different degrees of complexity, i.e., varying number of features, number of data samples, and number of target outputs. The numbers of features and target outputs vary from 35 and 2 in Ionosphere to 280 and 16 in Arrhythmia, respectively, while the number of data samples varies from 230 in Ionosphere to 5000 in Waveform. All data sets have real-valued features; therefore discretizing is applied to the features.

Table 1: The main properties of the real UCI data sets.

Data-sets	No. of features	No. of instances	Decisions
Ionosphere	35	230	2
Water	39	390	3
Waveform	41	5000	3
Sonar	61	208	2
Ozone	73	2534	2
Libras	91	360	15
Arrhythmia	280	452	16

3.1.2. Experimental study

The stratified 10-fold cross validation (10-FCV) [47] schemes is adopted to confirm the reliability and validity of the results. As such, each data set is divided into ten subsets, nine of them are used for training, where MRMC-IWD is used to select the feature subsets. The remaining subset is used for testing the classifier. This process is repeated ten times. The advantage of 10-FCV over random sub-sampling is that all data samples are used for both training and testing, and each data sample is used for testing only once per fold. Data stratification prior to its division into different folds ensures that each class label has as equal representation in all folds as possible, therefore mitigating the bias/variance problems [48]. The parameters setting of MRMC-IWD is shown in Table 2.

Using the real-valued UCI benchmark data sets, as shown in Table 1, a series of experiments was carried out with the 10-FCV method. For each data set, a total of 100 (10 runs $\times$ *10-FCV*) experimental outcomes were obtained. The results were compared with those from other state-of-the-art methods in the literature. The performance indicators used were the feature subset size and its evaluation score. Four standard classifiers in WEKA were employed in the experimental study, namely C4.5 [49], Naive Bayes (NB)[50], Vaguely Quantified Nearest Neighbor (VQNN)[11], and Support Vector Machine (SVM) [51]. Classifiers were selected to cover different the commonly used classifica-

Table 2: Parameters setting for the hybrid MRMC-IWD for feature selection.

Parameter type	Parameters	Values
Static	IWD	30
	a_v, b_v, c_v	1, 0.01, 1
	a_s, b_s, c_s	1, 0.01, 1
	$initSoil$	1000
	Max_iter	1000
	ε_s ,ρ_{IWD}, ρ_n	0.01,0.9,0.9
Dynamic	$V^k_{visited}$	Empty
	$intiVel^k$	4
	$soil^k$	0
Number of creeks	C	3

tion methods, i.e., decision tree, uncertainty modeling, machine learning and probabilistic models [52].

The bootstrap method [53] was used to compute the 95% confidence intervals of the feature subset size and evaluation score of MRMC-IWD. Bootstrap is a statistical method that does not rely on the assumption that the samples must be drawn from normal distribution, and can be used with small sample sizes. The results were compared with those from other state-of-the-art methods, namely Harmony Search (HS). Furthermore, a comparative study pertaining to the classification accuracies using the features selected from different evaluation techniques (consistency-based, CFS, and FRFS) was conducted.

3.1.3. Comparing the performance of MRMC-IWD against other methods

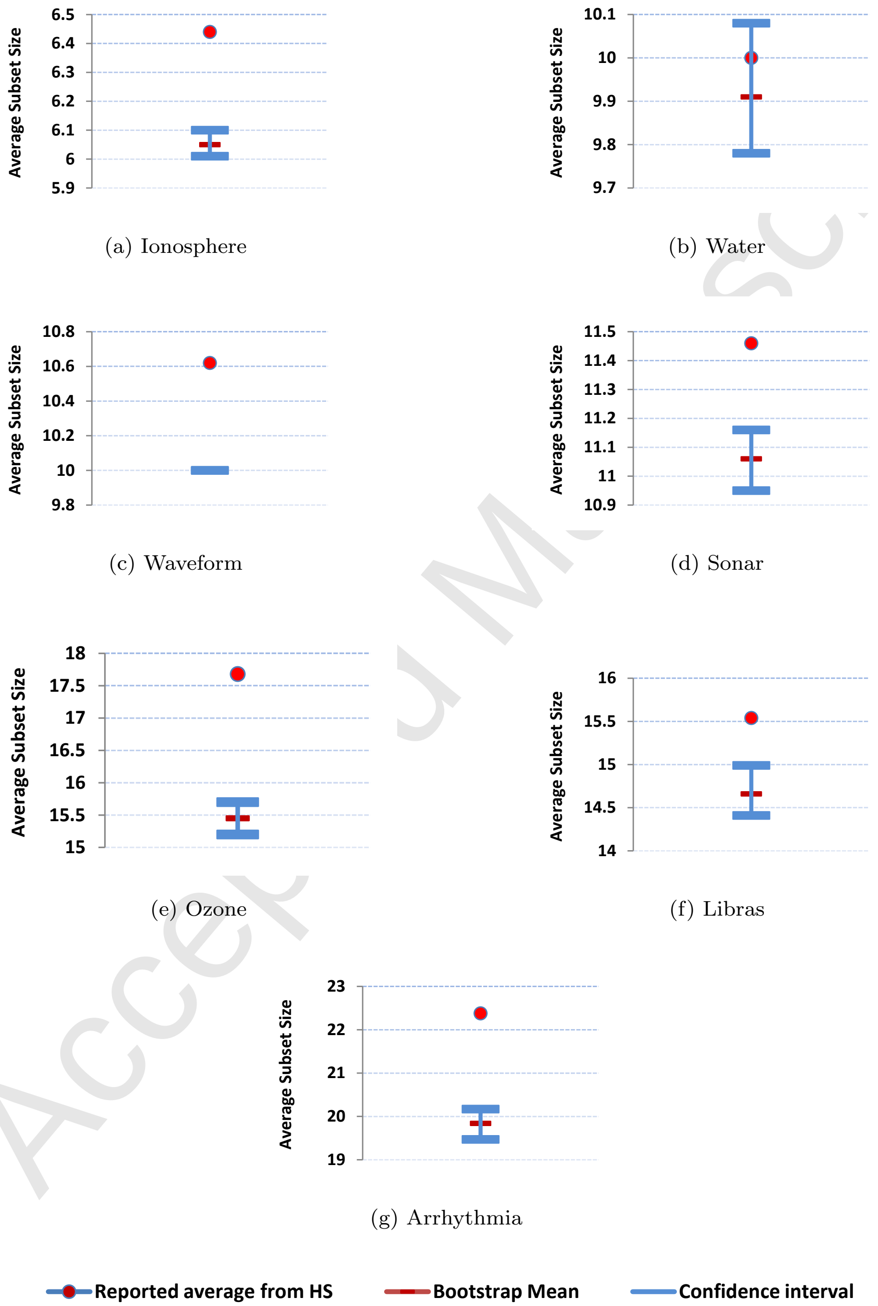
In this section, the experimental results of MRMC-IWD using the three subset evaluation techniques(i.e., CFS, Consistency-based, and FRFS), were analyzed and discussed. It was reported in [4], that HS performed better than other optimization methods. In this study, the results of the MRMC-IWD is compared with the those that obtained by HS [4].

i. **Consistency-based evaluation technique**.

The performance of MRMC-IWD with 95% confidence intervals is compared with the average subset sizes obtained by HS [4]. As can be seen in Figures 3 (a) to (g), the upper 95% confidence intervals of the subset sizes from MRMC-IWD are lower than those from HS for all the data sets, except one (i.e. Water) in Figure 3 (b).

Table 3 shows the average subset sizes and evaluation scores between MRMC-IWD and four state-of-the-art methods (i.e., HS, GS, PSO, and HC). Symbols v, $-$, and $*$, respectively, denote that the bootstrapped results (i.e., subset size or the evaluation score) yielded by MRMC-IWD is not significantly better, has no statistical difference, or worse than those provided by other methods. Comparing the feature subset sizes, MRMC-IWD significantly outperformed HS as well as the other methods (i.e., GA, PSO, and HC) for all data sets except one (i.e., Water), which showed no significant difference comparing with HS. The results show that GA and PSO can optimize the evaluation score but unable to reduce the feature subset size further. In terms of the evaluation score, no significant differences between the results of MRMC-IWD and those from HS [4].

In general, global optimization methods (i.e. MRMC-IWD, PSO, HS, and GA) discovered features subsets with equally good evaluation scores. Local-based methods had the tendency to be stuck in local optima, e.g. HC stuck in a local solution in three (i.e., Sonar, Ozone, and Libras) out of seven data sets. Overall, MRMC-IWD optimized both the subset size and evaluation score for the all seven data sets. This is owing to the exploration-exploitation balance that enabled it to perform well in the both global and local optimization conditions.



(a) Ionosphere (b) Water

(c) Waveform (d) Sonar

(e) Ozone (f) Libras

(g) Arrhythmia

Figure 3: Comparing the bootstrapped means and 95% confidence intervals of the subset size from MRMC-IWD and average subset sizes from HS [4] using the consistency-based evaluation technique.

Table 3: Comparison between MRMC-IWD and four state-of-the-art methods (i.e., HS,GA,PSO, and HC) published in [4], using the consistency-based evaluation technique in terms of the average subset size and evaluation score. Symbols v, -, and * respectively, denote that the result is significantly better, no statistical difference, and worse than those provided by HS.

		MRMC-IWD			HS			GA		PSO		HC	
Data sets	Full	Size	Ind	Eval.	Ind	Size	Eval.	Size	Eval.	Size	Eval.	Size	Eval.
Ionosphere	35	6.05	v	0.998	-	6.44	0.997	10.08	0.997	8.16	0.997	7.20	0.996
Water	39	9.91	-	0.995	-	10.00	0.995	10.24	0.995	12.68	0.995	10.90	0.995
Waveform	41	10.00	v	1.000	-	10.62	1	11.38	1	12.54	0.999	11.70	1
Sonar	61	11.06	v	0.992	-	11.46	0.99	12.06	0.99	18.92	0.99	12.40	0.986
Ozone	73	15.45	v	0.999	-	17.68	0.999	19.62	0.999	25.60	0.999	20.10	0.929
Libras	91	14.66	v	0.972	-	15.54	0.972	16.52	0.972	26.58	0.972	15.80	0.97
Arrhythmia	280	19.84	v	0.988	-	22.38	0.988	59.42	0.988	111.38	0.988	22.30	0.988

ii. **CFS evaluation technique**.

Table 4 shows the results obtained using CFS as the evolution technique. The bootstrapped results from MRMC-IWD were compared with those from HS in regards to the subset size and evaluation score. Figures 4 (a) to (g) summarize the results presented in Table 4. They depict the difference between the average subset sizes from HS and the bootstrapped results (mean and 95% confidence interval) from MRMC-IWD using CFS for seven data sets.

In terms of the feature subset size, MRMC-IWD provided significantly better results for four data sets (i.e., Water, Waveform, Ozone, and Arrhythmia) as compared with those provided by HS. As can be seen in 4(b), (c), (e) and (g), the upper bounds 95% confidence intervals of the bootstrapped subset size from MRMC-IWD are lower than those from HS. For the other three data sets (i.e., Ionosphere, Sonar, and Libras), no statistically significant difference between the results from both MRMC-IWD and from HS. As shown in Figures 4(a),(d), and (f), the average subset sizes of the results obtained from HS are within the 95% confidence intervals. In terms of the evaluation score, MRMC-IWD yielded equal results for two data sets (i.e., Libras, and Arrhythmia), better results for four data sets (i.e. Ionosphere, Waveform, Sonar, and Ozone), and an inferior result for one (i.e., Water) data set.

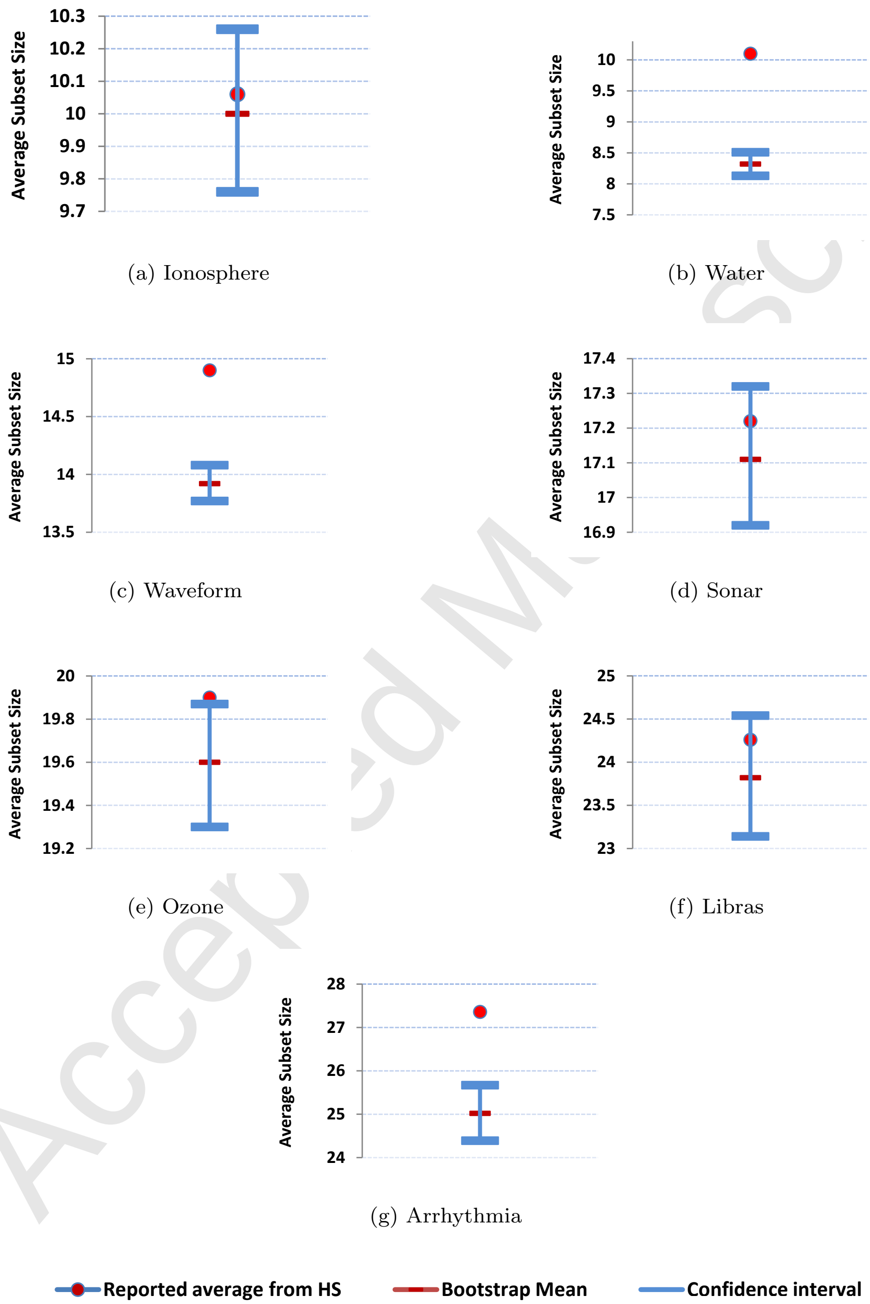


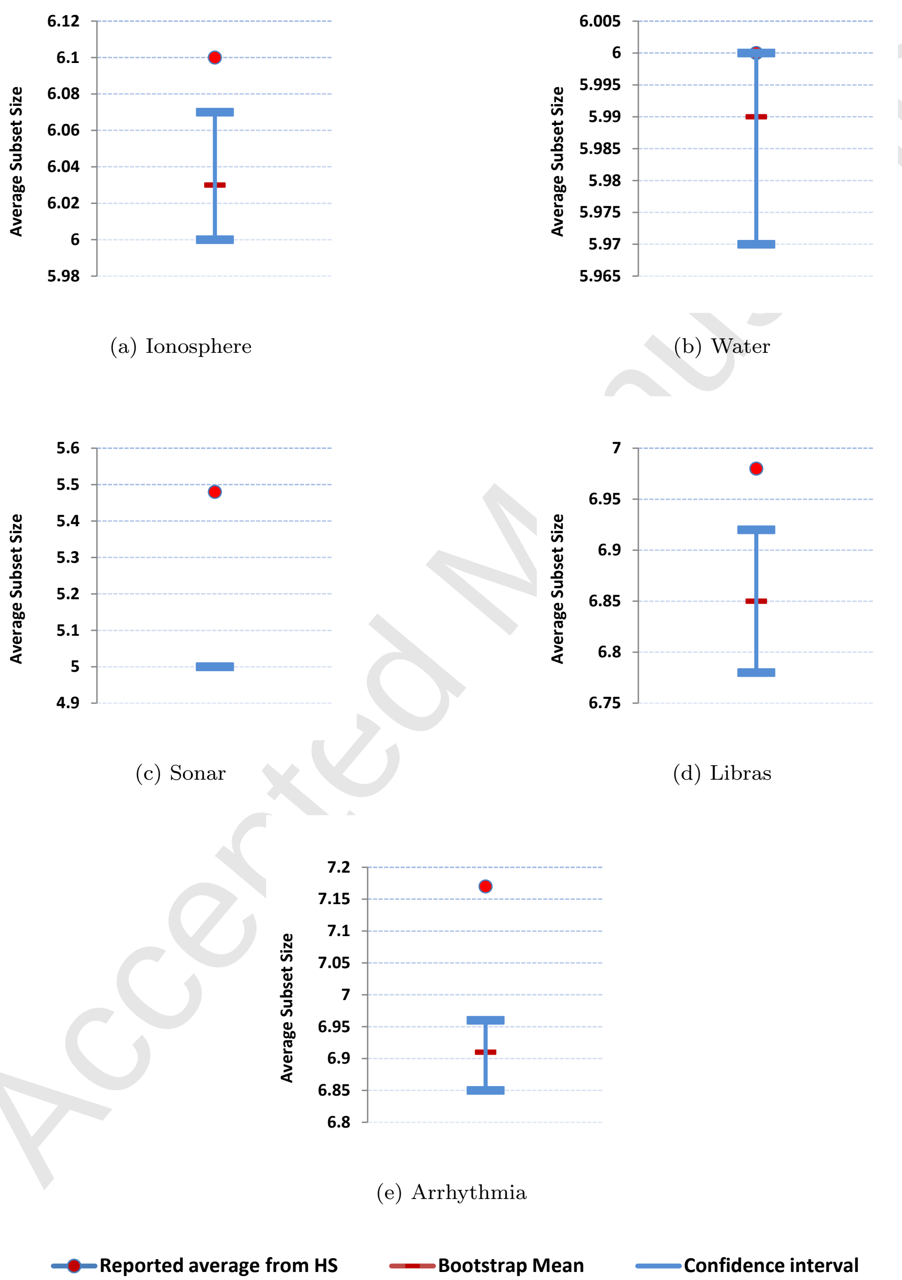
Figure 4: Comparing the average subset sizes of HS, published in [4], with the bootstrap results (i.e., means and 95% confidence intervals) of MRMC-IWD using the CFS evaluation technique.

Table 4: Comparison between MRMC-IWD and four state-of-the-art methods (i.e., HS,GA,PSO, and HC published in [4], (2012), using CFS in terms of the average subset size and evaluation score. Symbols v, -, and *, respectively, denote that the result is significantly better, no statistical difference, and worse than those provided by HS.

		MRMC-IWD				HS		GA		PSO		HC	
Data sets	Full	Size	Ind	Eval.	Ind	Size	Eval.	Size	Eval.	Size	Eval.	Size	Eval.
Ionosphere	35	10.00	-	0.544	-	10.06	0.542	10.38	0.542	11.06	0.537	10.00	0.542
Water	39	8.32	v	0.359	*	10.10	0.427	10.38	0.427	8.68	0.419	10.20	0.427
Waveform	41	13.92	v	0.384	-	14.90	0.384	14.90	0.384	12.41	0.369	14.90	0.384
Sonar	61	17.11	-	0.361	-	17.22	0.359	18.02	0.359	12.02	0.325	18.00	0.359
Ozone	73	19.60	v	0.115	-	19.90	0.114	21.58	0.114	25.34	0.11	21.30	0.114
Libras	91	23.82	-	0.610	v	24.26	0.607	30.82	0.607	26.36	0.594	24.00	0.607
Arrhythmia	280	25.02	v	0.467	v	27.36	0.441	65.38	0.189	13.48	0.279	24.30	0.437

iii. **FRFS evaluation technique**.

Table 5 shows the results from MRMC-IWD and those from HS, GA, PSO, and HC [4] using FRFS as the evolution technique. Only five data sets were used in the experimental study because of the high computational load of RSFS. Figures 5(a) to (e) summarize the results presented in Table 5 Comparing the results in terms of the evaluation score, all optimization methods including MRMC-IWD were able to discover feature subsets with the best evaluation score (i.e., eval = 1) for all five data sets. In terms of the feature subset size, MRMC-IWD significantly outperformed HS for all data sets except for one (i.e., Water) that showed no significant difference, as shown in Figure 5 (a). This is owing the property of the search space landscape of the *Water* data set, where the local optimum solution happened to be the global optima[4]. This can be observed from the results obtained from the local search method (i.e., HC), which performed better than the global-based methods (i.e., HS, GA, PSO) for this data set . However, MRMC-IWD performed well in this case due to its exploration and exploitation capability.



(a) Ionosphere

(b) Water

(c) Sonar

(d) Libras

(e) Arrhythmia

Figure 5: Comparing the average subset size of HS published in [4] with the bootstrap results (i.e., mean and 95% confidence interval) of MRMC-IWD using the FRFS evaluation technique..

Table 5: Comparison between MRMC-IWD and four state-of-the-art methods (i.e., HS,GA,PSO, and HC published in [4], using the FRFS evaluation technique in terms of the average subset size and evaluation score. Symbols v, -, and *, respectively denote that the result is significantly better, no statistical difference, and worse than those provided by HS.

		MRMC-IWD			HS			GA		PSO		HC	
Data sets	Full	Size	Ind	Eval.	Ind	Size	Eval.	Size	Eval.	Size	Eval.	Size	Eval.
Ionosphere	35	6.03	v	1	-	6.10	1	9.94	1	7.30	1	7.00	1
Water	39	5.99	-	1	-	6.00	1	8.34	1	6.68	0.998	6.00	1
Sonar	61	5.00	v	1	-	5.48	1	8.66	1	6.00	0.997	5.00	1
Libras	91	6.85	v	1	-	6.98	1	16.26	1	8.24	0.999	7.60	1
Arrhythmia	280	6.91	v	1	-	7.17	1	28.28	1	9.62	1	7.10	1

3.1.4. Comparison of classification accuracies.

Based on the features selected using different evaluation techniques, the impact of FS on classification accuracies was evaluated using four standard classifiers (i.e., C4.5, NB, VQNN, and SVM) in WEKA. The evaluation compared classification accuracies between the selected features and the full features.

i. **C4.5**.

Tables 6, 7, and 8 respectively, show the average classification accuracy rates of C4.5 based on the features selected using the consistency-based, CFS, and FRFS evaluation techniques. The results were compared with those using the original set of features (i.e., full features). A statistical test, namely the *t*-test, was carried out to indicate the reliability of the result statistically. Specifically, the paired *t*-test with p= 0.05 was conducted.

The results show that the classification accuracy rates after FS varied as compared against the use of full features. In some cases, the classification accuracy rates were better with more attributes, while in other cases the classification results were enhanced by selecting a subset of features. Symbols v, -, and * , respectively, denote that classification accuracy using the selected subsets by MRMC-IWD is better than, no statistical difference, or worse than those using the full features.

Table 6 show the accuracy rates of C4.5 based on the feature subsets discovered by different optimization methods using the consistency-based evaluator technique. When the C4.5 used the subset that discovered by MRMC-IWD, better classification accuracies for four data sets (i.e., Ionosphere, Sonar, Ozone, and Arrhythmia) were obtained. The classification accuracy for one data set (i.e., Libras) using the full features was better than that using the discovered data set from MRMC-IWD. There was no significant difference between the classification accuracies using the full and selected features for the other two data sets (i.e., Water, and Waveform).

Tables 7 and 8, which show the results of C4.5 using the subsets discovered using CFS and FRFS, respectively. A number of observations can be made:

Table 6: Classification accuracy rates (measured in %) of C4.5 using features selected by the consistency-based evaluation technique and from different optimization methods published in [4].

		MRMC-IWD		HS	GA	PSO	HC
Data sets	Full	Acc.	Ind	Acc.	Acc.	Acc.	Acc.
Ionosphere	86.44	88.22	v	87.04	86.52	84.61	85.22
Water	82.89	82.94	-	83.38	82.77	82.26	82.25
Waveform	75.29	75.31	-	74.3	74.75	75.29	76.88
Sonar	72.68	75.33	v	73.8	71.79	70.99	70.5
Ozone	92.69	93.25	v	93.05	93.22	93.17	93.21
Libras	68.77	65.86	*	65.33	65.22	67.39	65.83
Arrhythmia	66.12	66.96	v	66.12	66.66	65.99	66.38

a). Referring to Table 7 the classification accuracies for two the all data sets except for one data set (i.e., Libras) are significantly better using the discovered subset. There is no significant different in the classification accuracy using the full feature and the discovered subset for the case of the Libaras data set.

b). Referring to Table 8, which shows the FRFS results, two data sets (i.e., Ionosphere, and Water) show better classification accuracies, one (i.e., Sonar) shows no significant difference, and two (i.e., Libras, and Arrhythmia) show inferior results using the discovered subsets as compared with those from the full feature set.

ii. **Different classifiers using the feature subsets discovered by MRMC-IWD**.

To demonstrate the generality, Table 9 shows the results of three standard classifiers (i.e., VQNN, NB, and SVM) using the feature subsets discovered by MRMC-IWD. The results are compared with those using the full features. Overall, CFS performed the best in terms of preserving and improving classification accuracy. Out of the 21 cases (i.e., 7 data sets $\times$ 3 classifiers), 4 cases indicated reduction, and 17 cases showed improvement

Table 7: C4.5 classification accuracy rates (measured in %) using the feature subsets selected by CFS and different optimization methods published in [4].

		MRMC-IWD		HS	GA	PSO	HC
Data sets	Full	Acc.	Ind	Acc.	Acc.	Acc.	Acc.
Ionosphere	86.44	88.75	v	85.3	85.22	85.57	85.21
Water	82.89	83.40	v	82.46	82.36	81.18	82.56
Waveform	75.29	77.59	v	77.23	77.22	77.19	77.22
Sonar	72.68	76.56	v	72.95	72.48	72.74	73.14
Ozone	92.69	93.26	v	93.28	93.31	93.47	93.49
Libras	68.77	69.23	-	69.33	71.33	67.5	70.83
Arrhythmia	66.12	68.99	v	67.27	66.74	63.19	66.81

Table 8: C4.5 classification accuracy (measured in %) using feature subsets selected by FRFS and different optimization methods published in [4].

		MRMC-IWD		HS	GA	PSO	HC
Data sets	Full	Acc.	Ind	Acc.	Acc.	Acc.	Acc.
Ionosphere	86.44	87.27	v	86.96	86	87.04	86.52
Water	82.89	85.04	v	79.03	80.21	78.15	80.51
Sonar	72.68	72.42	-	70.54	70.1	70.19	76.51
Libras	68.77	58.53	*	60.39	64.06	56.67	61.11
Arrhythmia	66.12	54.66	*	62.27	64.33	63.44	66

in classification performance. On the other hand, the results obtained using the consistency-based and FRFS evaluation techniques reflect loss of accuracy in most cases. Note that for the Ozone data set, VQNN and SVM obtained identical cross-validated accuracy of 93.69%. Furthermore, the results from the reduced feature subsets discovered by both consistency-based and CFS were in agreement in terms of classification accuracy. Overall, the experimental results demonstrate that CFS works very well in preserving the information of the original features in the reduced feature subsets. FRFS can produce smaller features subsets, but compromise classification performance slightly. The consistency-based evaluation technique compromises the classification performance and the feature subset size. The results conform to the common understanding that evaluation techniques used in filter-based FS and the actual classification performance are independent. In other words, it is not necessary for a feature subset that has the highest evaluation score to yield the highest classification performance.

Table 9: Comparing the classification accuracies between full and reduced features using different evaluation techniques. Highlight (bold) denote that result is significantly better.

	Nave Bayes				VQNN				SVM			
Data sets	Full	CON	CFS	FRFS	Full	CON	CFS	FRFS	Full	CON	CFS	FRFS
Ionosphere	83.73	80.15	**86.85**	86.05	88.08	**88.68**	88.45	88.29	**92.39**	89.70	92.08	90.43
Water	69.74	84.99	**88.27**	85.76	85.23	87.82	**88.92**	85.63	**87.74**	80	80	86.84
Waveform	79.99	79.75	**80.48**		81.74	79.36	**84.53**		33.84	38.79	**33.94**	
Sonar	68.17	67.38	68.15	**69.76**	79.46	80.39	**81.83**	77.32	**87.06**	81.46	83.25	76.86
Ozone	67.63	**76.08**	74.28		93.68	93.68	**93.68**		93.68	93.68	**93.68**	
Libras	**63.8**	60.03	62.82	47.53	**71.16**	66.86	67.38	59.76	86.22	86.56	**87.65**	75.06
Arrhythmia	61.61	68.69	**70.00**	58.86	60.57	64.56	**67.05**	57.23	54.2	54.2	54.2	**61.75**

3.2. MRMC-IWD for real-world problems

In this section, MRMC-IWD is used to tackle FS on two real-world problems, namely human motion detection and motor fault detection. The main objective of this study is to investigate the impact of FS on classification accuracy for both problems. Figure 6 shows the main components of a pattern recognition system, viz., (i) data acquisition (ii) features extraction (iii) features subset selection (iv) classification. The details are as follows.

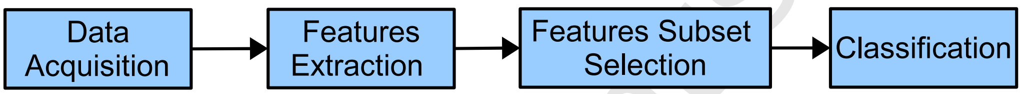


Figure 6: The main components of a pattern recognition system.

- **Data acquisition**

  Data acquisition is a process of collecting data samples that represent the physical condition and behavior of a real-world problem. As an example, sensors can be used to record signal samples, which represent the physical condition of a system. The recorded signals can be converted to numerical values, which can then be manipulated by using a computing model [54].

- **Features extraction**

  Extracting the relevant information that represents the characteristics of the underlying problem is an important task. Two main types of features (i.e., time-domain and frequency-domain) can be derived from the collected data samples. Time-domain features comprise set of statistical information (e.g., mean, median, variance, skewness, kurtosis) of the data samples [55]. Frequency-domain features describe the periodical properties of a signals. Fast Fourier transform (FFT) is an efficient approach used to extract periodicity of signals [55].

  The data samples can be transformed from the time domain to the frequency domain. The output of FFT typically gives a set of basis coefficients that represents the amplitudes of the frequency components of the signal and the distribution of the signal energy. Different attributes can then be used to characterize the spectral distribution from these coefficients. Tables 10, and 11, respectively, show the description of 15 time-domain and frequency-domain features used in this study.

Table 10: Time-domain features.

Features	Equation
Mean	$\mu_x = \frac{1}{K}\sum_{k=1}^{K} x(k)$
Standard Deviation (std)	$std_x = \sqrt{\frac{\sum_{k=1}^{K}(x(k)-\mu_x)^2}{K-1}}$
Root Mean Square (RMS)	$RMS_x = \sqrt{\frac{\sum_{k=1}^{K}(x(k))^2}{K}}$
Maximum Amplitude	$MAX_x = max(x(k))$
Minimum Amplitude	$MIN_x = min(x(k))$
Skewness	$Skew_x = \frac{\sum_{k=1}^{K}(x(k)-\mu_x)^3}{(K-1)\times std(x)^3}$
Kurtosis	$Kurto_x = \frac{\sum_{k=1}^{K}(x(k)-\mu_x)^4}{(K-1)\times std(x)^4}$
Clearance Factor	$CLF = \frac{MAX_x}{\left(\frac{1}{K}\sum_{k=1}^{K}\sqrt{\lvert x(k)\rvert}\right)^2}$
Shape Factor	$SF = \frac{MAX_x}{\frac{1}{K}\sum_{k=1}^{K}\lvert x(k)\rvert}$
Crest Factor	$CF = \frac{MAX_x}{RMS_x}$

- **Features subset selection**

Not all features are important to the learner classifiers. Selecting the significant features by removing irrelevant, noisy, and redundant features can enhance the classification performance. In this study MRMC-IWD with three evaluation techniques (i.e., CFS, FRFS, and probabilistic consistence) was used. FS was repeated 10 times, each time the 10-FCV scheme

Table 11: Frequency-domain features.

Features	Equation
Maximum Power Spectrum	$MAXF_x = max(power(n))$
Maximum Envelope	$MAXF_env = max(Env)$
Frequency Center	$F_c = \frac{\sum_{n=1}^{N} FS(n)}{\sum_{n=1}^{N} S(n)}$
RMS Frequency	$RMSF_x = \sqrt{\frac{\sum_{n=1}^{N} F^2 \times S(n)}{\sum_{n=1}^{N} S(n)}}$
Standard Deviation Frequency	$stdF_x = \sqrt{\frac{(F-F_c)^2 \times S(n)}{\sum_{n=1}^{N} S(n)}}$

was performed to indicate the reliability and validity of the results.

- **Classification**

  The outcomes of FS (i.e., selected feature subsets) was used by the learner classifiers. In this study, four standard classifiers in WEKA (i.e., VQNN, NB, SVM, C4.5) were employed. The 10 × *10-FCV* FCV scheme was employed in the classification process.

3.3. Human motion detection

Human motion detection is an important and challenging research area with many different applications, e.g. safety surveillance, fraud detection, clinical management, and healthcare [56]. For an instance, in the healthcare area, it is beneficial to identify the energy consumption rate during human activities [57]. The wearable based sensors are an efficient data acquisition unit to acquire human motion activities (i.e. data acquisition unit). Wearable sensors are small size mobile sensors designed to be worn by humans [58]. They can be used to record humans' physiological states such as location changes, moving directions, and speed. Most of the smartphones are equipped with sensors such as accelerometers nowadays [56]. In the following sub-sections, the procedure of human motion detection and the experimental results and discussion are presented.

3.3.1. Data acquisition

In this case study, a binary classification problem was formulated. A data set was collected using smartphones with the built-in tri-axial accelerometer. A total of 57 subjects including children, adults, and students of both gender (i.e., male and female) participated in the data collection process. Three smartphones with built-in accelerometers were placed in different positions (i.e., belt pocket, shirt pocket, and front pants pocket). Each subjects performed two types of activities (i.e., walking, and running) with 100 steps. The tri-axial accelerometer measured the subjects' acceleration along three axes (i.e., x-axis, y-axis and z-axis). Data pre-processing was conducted to remove noise from the motion waveforms. Table 12 shows the details of the data samples.

Table 12: Number of samples collected for human motion data set.

Pocket Position	Walking Samples	Running Samples	Total Samples
Belt Pocket	76	57	133
Front Pants Pocket	49	39	88
Shirt Pocket	67	55	122
Overall	192	151	343

3.3.2. Feature extraction

As shown in Tables 10 and 11 , a total of 15 features (10 time-domain and 5 frequency-domain) were extracted from the motion waveforms. An augmented feature vector of 45 components (i.e., 15 features $\times$ 3 axes) was formed for each data sample. Given the data set of 343 instances, each instance has 45 features, the problem was to identify the type of human motion, i.e. walking or running.

3.3.3. Features selection

Table 13 show the results obtained by applying the FS procedure. The average subset sizes and average evaluation scores of 10×10-FCV for three different evaluation techniques are summarized. As shown in Table 13 the number of

Table 13: The results of feature selection for human motion detection using MRMC-IWD.

Evaluation technique	Average subsets Size	Average evaluation value
FRFS	6.21	1
CFS	6.09	0.626
Consistency-based	8.96	0.996

features has been drastically reduced from 45 to an average of 6 to 9 features, based on the evaluation techniques.

3.3.4. Classification

Table 14 shows the classification results of 10×10-FCV. Three performance indicators, namely classification accuracy, specificity, and sensitivity using four classifiers are summarized. MRMC-IWD with FRFS reduced the subset size to 6.21, and it produced the best evaluation score of 1. However, it was unable to preserve important features, as indicated by the classification performance using the selected features. Improved classification results using the feature subsets selected by both CFS, and consistency-based evaluation techniques were obtained. Overall, CFS performed the best in terms of preserving important original features and the subset size. It yielded the smallest subset size (6.09 features) and the best classification performance for different standard classifiers.

Table 14: Comparing the classification performance accuracies (measured in %) between full and selected features for human motion detection. Highlight (bold) denote that result is significantly better.

	Average Size	Average Evaluation Value	Nave Bayes			VQNN			SVM			C4.5		
			Acc	Spec	Senc	Acc	Spec	Senc	Acc	Spec	Senc	Acc	Spec	Senc
Full	45		86.623	0.849	0.892	92.581	0.944	0.895	92.111	0.935	0.898	91.991	0.930	0.902
FRFS	6.21	1	84.438	0.855	0.825	89.222	0.931	0.830	92.529	0.944	0.894	90.575	0.911	0.896
CFS	6.09	0.626	92.454	0.899	0.964	93.664	0.942	0.928	93.686	0.939	0.932	93.582	0.947	0.917
Cons	8.96	0.996	90.237	0.876	0.943	92.956	0.942	0.909	93.167	0.941	0.916	92.861	0.942	0.907

3.4. Fault Motor fault detection

Induction motors are widely used in different industrial areas, include manufacturing machines and pumps. Fault detection and diagnosis is important issue that can reduce the maintenance and downtime costs in the manufacturing domain. In this study, the motor current signature analysis (MCSA), which is a condition monitoring technique, was used for fault detection of induction motors [59, 60]. In the following sub-sections, the procedure of fault detection induction motors and the experimental results and discussion are presented.

3.4.1. Data acquisition

Data set comprising three-phases stator currents (A, B, and C) from induction motors was collected. The task was to identify the motor conditions, either normal or faulty (with broken rotor bars).The data samples comprised current spectrum from normal and faulty (one or two broken rotor bars) motors in two load conditions (i.e., 50%, and 100%) are considered. Table 15 shows the details of data samples used for experimentation.

Table 15: Data samples for fault detection of induction motors.

	Load condition		Total Samples
	50 %	100 %	
Broken rotor bar #1	10	10	20
Broken rotor bar #1	10	10	20
Healthy motor	10	10	20
Overall	30	30	60

3.4.2. Feature extraction

To extract the relevant features, the three-phases current signals were preprocessed by dividing each signal into its perspective cycle of the sine waveform. Each data sample was represented by the 15 features, which are shown in Tables 10,and 10. As such, a data set of 60 instances, each with 45 features (15 features

$\times$ 3 current phases) was established to determine the motor condition (i.e., normal or faulty with broken rotor bars).

3.4.3. Features selection

Table 16 shows the results obtained by using MRMC-IWD for the three different evaluation techniques. As shown in Table 16,the number of features was drastically reduced from 45 features to an average range of 2 to 6 features, based on the respective evaluation techniques.

Table 16: The results of feature subset selected using MRMC-IWD for motor fault detection.

Evaluation technique	Average subsets Size	Average evaluation value
FRFS	1.86	1
CFS	5.76	0.87062
Consistency-based	1.83	1

3.4.4. Classification

Table 17 shows the classification results of $10 \times 10 - FCV$. In case of FRFS and Consistency based evaluation techniques, the average subset size was significantly reduced; however, they failed to preserve the important original features. This was reflected by the inferior classification accuracy rates from the classifiers using the selected feature subset as compared with those from the full set of features. CFS was able to compromise between preserving the important features and the subset size, since the accuracy rate is better using the reduced feature set as compared with those of the full features.

Table 17: Comparing the classification accuracies performance (measured in %) between full and reduced feature sets for motor fault detection.

	Average Size	Average Evaluation Value	Nave Bayes			VQNN			SVM			C4.5		
			Acc	Spec	Senc	Acc	Spec	Senc	Acc	Spec	Senc	Acc	Spec	Senc
Full	45		99.222	0.977	1.000	100.000	1.000	1.000	100.000	1.000	1.000	96.555	0.947	0.975
FRFS	1.86	1	82.456	0.694	0.89	94.211	0.837	0.995	90.331	0.746	0.982	94.585	0.903	0.967
CFS	5.76	0.87062	99.898	0.999	0.999	99.987	1	1	99.978	1	1	96.569	0.947	0.975
Cons	1.83	1	82.361	0.687	0.892	94.421	0.843	0.995	90.183	0.741	0.982	94.508	0.903	0.966

4. Conclusion and Future Study

This paper, investigated the potential of the MRMC-IWD model pertaining to real-world FS problems was investigated. Firstly, benchmark data sets from UCI was evaluated using three evaluation techniques (i.e., CFS, Consistency-based, and FRFS) for FS. The comparative results indicate the superiority of MRMC-IWD to other state-of-the-art (i.e., HS, PSO, GA, and HC) methods. Furthermore, the impact of FS on the classification performance of different classifiers (i.e., NB, VQNN, SVM, and C4.5) was investigated. Secondly, the applicability of MRMC-IWD to real-world human motion detection and motor fault detection was evaluated. The results indicate the usefulness of MRMC-IWD in tackling real-world FS problems.

References

[1] M. Dash, H. Liu, Feature selection for classification, Intelligent data analysis 1 (1-4) (1997) 131–156.

[2] M. H. Aghdam, N. Ghasem-Aghaee, M. E. Basiri, Text feature selection using ant colony optimization, Expert Systems with Applications 36 (3) (2009) 6843–6853.

[3] P. Maji, S. Paul, Rough set based maximum relevance-maximum significance criterion and gene selection from microarray data, International Journal of Approximate Reasoning 52 (3) (2011) 408–426.

[4] R. Diao, Q. Shen, Feature selection with harmony search, Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 42 (6) (2012) 1509–1523.

[5] R. Jensen, Q. Shen, Computational intelligence and feature selection: rough and fuzzy approaches, Vol. 8, Wiley-IEEE Press, 2008.

[6] R. Jensen, Q. Shen, New approaches to fuzzy-rough feature selection, Fuzzy Systems, IEEE Transactions on 17 (4) (2009) 824–838.

[7] R. Jensen, A. Tuson, Q. Shen, Finding rough and fuzzy-rough set reducts with sat, Information Sciences 255 (2014) 100–120.

[8] R. W. Swiniarski, A. Skowron, Rough set methods in feature selection and recognition, Pattern recognition letters 24 (6) (2003) 833–849.

[9] N. Parthalain, Q. Shen, R. Jensen, A distance measure approach to exploring the rough set boundary region for attribute reduction, Knowledge and Data Engineering, IEEE Transactions on 22 (3) (2010) 305–317.

[10] R. B. Bhatt, M. Gopal, On fuzzy-rough sets approach to feature selection, Pattern recognition letters 26 (7) (2005) 965–975.

[11] Jensen, R. Chris Cornelis, A new approach to fuzzy-rough nearest neighbour classification, in: Proceedings 6th International Conference on Rough Sets and Current Trends in Computing: (RSCTC), Vol. 5306, Springer, Akron, OH, USA, 2008, p. 310.

[12] E. C. C. Tsang, D. Chen, D. S. Yeung, X.-Z. Wang, J. W. T. Lee, Attributes reduction using fuzzy rough sets, Fuzzy Systems, IEEE Transactions on 16 (5) (2008) 1130–1141.

[13] M. Dash, H. Liu, Consistency-based search in feature selection, Artificial Intelligence 151 (1) (2003) 155–176.

[14] H. Liu, L. Yu, Toward integrating feature selection algorithms for classification and clustering, Knowledge and Data Engineering, IEEE Transactions on 17 (4) (2005) 491–502.

[15] M. A. Hall, L. A. Smith, Feature subset selection: a correlation based filter approach.

[16] M. A. Hall, Correlation-based feature selection for machine learning, Ph.D. thesis, The University of Waikato (1999).

[17] H. Peng, F. Long, C. Ding, Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy, Pattern

Analysis and Machine Intelligence, IEEE Transactions on 27 (8) (2005) 1226–1238.

[18] R. Kohavi, G. H. John, Wrappers for feature subset selection, Artificial intelligence 97 (1) (1997) 273–324.

[19] J. G. Bazan, H. S. Nguyen, S. H. Nguyen, P. Synak, J. Wrblewski, Rough set algorithms in classification problem, in: Rough set methods and applications, Springer, 2000, pp. 49–88.

[20] L. Yu, H. Liu, Efficient feature selection via analysis of relevance and redundancy, The Journal of Machine Learning Research 5 (2004) 1205–1224.

[21] H. Vafaie, I. F. Imam, Feature selection methods: genetic algorithms vs. greedy-like search, in: Proceedings of International Conference on Fuzzy and Intelligent Control Systems, 1994.

[22] C. Bae, W.-C. Yeh, Y. Y. Chung, S.-L. Liu, Feature selection with intelligent dynamic swarm and rough set, Expert Systems with Applications 37 (10) (2010) 7026–7032.

[23] R. Jensen, Q. Shen, Finding rough set reducts with ant colony optimization, in: Proceedings of the workshop on computational intelligence, Vol. 1, UK, 2003.

[24] M. F. Zaiyadi, B. Baharudin, A proposed hybrid approach for feature selection in text document categorization, World Academy of Science, Engineering and Technology 48 (2010) 111 – 116.

[25] S. Abdullah, N. S. Jaddi, Great deluge algorithm for rough set attribute reduction, Database Theory and Application, Bio-Science and Bio-Technology 118 (2010) 189–197.

[26] M. Mafarja, S. Abdullah, Modified great deluge for attribute reduction in rough set theory, in: Eighth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), Vol. 3, IEEE, 2011, pp. 1464–1469.

[27] H. Shah-Hosseini, Problem solving by intelligent water drops, in: Proceedings of the IEEE Congress on Evolutionary Computation, IEEE, Singapore, 2007, pp. 3226–3231.

[28] H. Shah-Hosseini, Intelligent water drops algorithm: A new optimization method for solving the multiple knapsack problem, International Journal of Intelligent Computing and Cybernetics 1 (2) (2008) 193–212.

[29] N. Siddique, H. Adeli, Water drop algorithms, International Journal on Artificial Intelligence Tools 23 (06).

[30] B. O. Alijla, L. C. Peng, A. T. Khader, M. A. Al-Betar, Intelligent water drops algorithm for rough set feature selection, in: Intelligent Information and Database Systems, Springer, 2013, pp. 356–365.

[31] B. O. Alijla, L. C. Peng, A. T. A.-B. M. A. Khader, W. L. Pei, Fuzzy rough set approach for selecting the most significant texture features in mammogram images, in: Palestinian International Conference on Information and Communication Technology, 2013.(PICICT 2013). Proceedings. IEEE International Conference on, 2013.

[32] H. Shah-Hosseini, Intelligent water drops algorithm for automatic multilevel thresholding of grey–level images using a modified otsu's criterion, International Journal of Modelling, Identification and Control 15 (4) (2012) 241–249.

[33] H. Shah-Hosseini, An approach to continuous optimization by the intelligent water drops algorithm, Procedia-Social and Behavioral Sciences 32 (2012) 224–229.

[34] H. Mokhtari, A nature inspired intelligent water drops evolutionary algorithm for parallel processor scheduling with rejection, Applied Soft Computing 26 (2015) 166–179.

[35] N. Acharya, S. Singh, An iwd-based feature selection method for intrusion detection system, Soft Computing (2017) 1–10.

[36] V. Kayvanfar, M. Zandieh, E. Teymourian, An intelligent water drop algorithm to identical parallel machine scheduling with controllable processing times: a just-in-time approach, Computational and Applied Mathematics 36 (1) (2017) 159–184.

[37] S. H. Niu, S. K. Ong, A. Y. C. Nee, An improved intelligent water drops algorithm for solving multi-objective job shop scheduling, Engineering Applications of Artificial Intelligence 26 (10) (2013) 2431–2442.

[38] C. Sur, S. Sharma, A. Shukla, Multi-objective adaptive intelligent water drops algorithm for optimization & vehicle guidance in road graph network, in: International Conference on Informatics, Electronics & Vision (ICIEV), IEEE, 2013, pp. 1–6.

[39] L. A. Moncayo-Martínez, E. Mastrocinque, A multi-objective intelligent water drop algorithm to minimise cost of goods sold and time to market in logistics networks, Expert Systems with Applications 64 (2016) 455–466.

[40] B. O. Alijla, L.-P. Wong, C. P. Lim, A. T. Khader, M. A. Al-Betar, A modified intelligent water drops algorithm and its application to optimization problems, Expert Systems with Applications 41 (15) (2014) 6555–6569.

[41] B. O. Alijla, L.-P. Wong, C. P. Lim, A. T. Khader, M. A. Al-Betar, An ensemble of intelligent water drop algorithms and its application to optimization problems, Information Sciences 325 (2015) 175–189.

[42] E. Teymourian, V. Kayvanfar, G. M. Komaki, M. Zandieh, Enhanced intelligent water drops and cuckoo search algorithms for solving the capacitated vehicle routing problem, Information Sciences 334 (2016) 354–378.

[43] E. Teymourian, V. Kayvanfar, G. M. Komaki, M. Khodarahmi, An enhanced intelligent water drops algorithm for scheduling of an agile manufacturing system, International Journal of Information Technology & Decision Making 15 (02) (2016) 239–266.

[44] G. Liu, S. Guo, H. Zhao, F. Wang, Intelligent water drops based joint subcarrier pairing and power allocation with fairness in cooperative relay networks, Journal of Communications and Networks 19 (1) (2017) 10–22.

[45] K. Bache, M. Lichman, Uci machine learning repository. irvine, ca: University of california, school of information and computer science, See¡ http://archive. ics. uci. edu/ml.

[46] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I. H. Witten, The weka data mining software: an update, ACM SIGKDD explorations newsletter 11 (1) (2009) 10–18.

[47] M. Markatou, H. Tian, S. Biswas, G. M. Hripcsak, Analysis of variance of cross-validation estimators of the generalization error, Journal of Machine Learning Research 6 (2005) 1127–1168.

[48] Y. Bengio, Y. Grandvalet, Bias in estimating the variance of k-fold cross-validation, in: Statistical modeling and analysis for complex data problems, Springer, 2005, pp. 75–95.

[49] J. Wrblewskii, Ensembles of classifiers based on approximate reducts, Fundamenta Informaticae 47 (3) (2001) 351–360.

[50] G. H. John, P. Langley, Estimating continuous distributions in bayesian classifiers, in: Proceedings of the Eleventh conference on Uncertainty in artificial intelligence, Morgan Kaufmann Publishers Inc., 1995, pp. 338–345.

[51] S. S. Keerthi, S. K. Shevade, C. Bhattacharyya, K. R. K. Murthy, Improvements to platt's smo algorithm for svm classifier design, Neural Computation 13 (3) (2001) 637–649.

[52] Z. Chen, H. Strange, E. Denton, A. Oliver, C. Boggis, R. Zwiggelaar, Topological modelling and classification of mammographic microcalcification clusters, IEEE Transactions on Biomedical Engineering In press.

[53] B. Efron, Bootstrap methods: another look at the jackknife, The annals of Statistics (1979) 1–26.

[54] O. Maimon, L. Rokach, Data mining and knowledge discovery handbook, Vol. 2, Springer, 2005.

[55] A. M. Khan, Human activity recognition using a single tri-axial accelerometer, Ph.D. thesis, Kyung Hee University Seoul, Korea (2011).

[56] A. Bayat, M. Pomplun, D. A. Tran, A study on human activity recognition using accelerometer data from smartphones, Procedia Computer Science 34 (2014) 450–457.

[57] J. Mantyjarvi, J. Himberg, T. Seppanen, Recognizing human motion with multiple acceleration sensors, in: Systems, Man, and Cybernetics, 2001 IEEE International Conference on, Vol. 2, IEEE, 2001, pp. 747–752.

[58] X. Su, H. Tong, P. Ji, Activity recognition with smartphone sensors, Tsinghua Science and Technology 19 (3) (2014) 235–249.

[59] A. K. Mahamad, T. Hiyama, Fault classification based artificial intelligent methods of induction motor bearing, International Journal of innovative computing, information and control 7 (9).

[60] M. Seera, C. P. Lim, D. Ishak, H. Singh, Fault detection and diagnosis of induction motors using motor current signature analysis and a hybrid fmm-cart model, Neural Networks and Learning Systems, IEEE Transactions on 23 (1) (2012) 97–108.