IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received June 2, 2020, accepted July 1, 2020, date of publication July 6, 2020, date of current version July 20, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3007291

Approaches to Multi-Objective Feature Selection: A Systematic Literature Review

QASEM AL-TASHI[1,2], SAID JADID ABDULKADIR[1,3], (Senior Member, IEEE), HELMI MD RAIS[1], SEYEDALI MIRJALILI[4], (Senior Member, IEEE), AND HITHAM ALHUSSIAN[1,3]

[1]Department of Computer and Information Sciences, Universiti Teknologi PETRONAS, Seri Iskandar 32160, Malaysia
[2]Faculty of Administrative and Computer Sciences, University of Albaydha, Albaydha CV46+6X, Yemen
[3]Centre for Research in Data Science (CERDAS), Universiti Teknologi PETRONAS, Seri Iskandar 32160, Malaysia
[4]Centre for Artificial Intelligence Research and Optimization, Torrens University Australia, Fortitude Valley, QLD 4006, Australia

Corresponding author: Qasem Al-Tashi (qasemacc22@gmail.com)

This work was supported by the Centre for Research in Data Science (CERDAS), Under Cost Centre (015LC0-119), Universiti Teknologi PETRONAS.

ABSTRACT Feature selection has gained much consideration from scholars working in the domain of machine learning and data mining in recent years. Feature selection is a popular problem in Machine learning with the goal of finding optimal features with increase accuracy. As a result, several studies have been conducted on multi-objective feature selection through numerous multi-objective techniques and algorithms. The objective of this paper is to present a systematic literature review of the challenges and issues of the multi-objective feature selection problem and critically analyses the proposed techniques used to tackle this problem. The conducted review covered all related studies published since 2012 up to 2019. The outcomes of the reviewed of these studies clearly showed that no perfect solution to the multi-objective feature selection problem yet. The authors believed that the conducted review would serve as the main source of the techniques and methods used to resolve the problem of multi-objective feature selection. Furthermore, current challenges and issues are deliberated to find promising research domains for further study.

INDEX TERMS Feature selection, multi-objective optimization, classification, systematic literature review, optimization, benchmark, heuristic.

I. INTRODUCTION

The problems used by machine learning techniques usually have a lot of features. Hence, It is difficult to find an optimal set of feature and omit redundant ones. In any dataset, some of the features are unimportant due to the existence of redundancy and irrelevancy of these features. Hence, considering such features is not beneficial and typically lead to poor classification accuracy. Therefore, feature selection seeks to enhance classification efficiency by selecting from the initial wide range of features only a tiny subset of appropriate features. The removal of irrelevant and redundant features will, therefore, decrease the data dimensionality, accelerate the process of learning by simplifying the model learned as well as performance-boosting [1], [2]. Feature selection is an NP hard problem with 2^n states where n is the number of features. The complexity of the problem is increasing dramatically as N is growing with improvements in data collection methods in many fields. Unlike feature selection, feature extraction approaches such as Principal Component Analysis (PCA) [3] and Linear Discriminant Analysis (LDA) [4], new features are created from the original features to procedure a new reduced search space by merging or transforming the original features utilizing some functional mapping. Hence, in this review our concern on feature selection methods.

Feature selection has two primary conflicting goals, namely, maximizing th perfromance of classification and minimizing features number to overcome the curse of dimensionality. The selection of features can be treated as a multi-objective problem to balance the trade-off between these two conflicting goals. In recent years, research has progressed towards the development of feature selection based multi-objective methods to solve these issues. Given the variety of techniques and methods that now exist, this paper aims to critically analyse the latest related studies to the problem focusing on research outputs from 2012 to 2019. The major databases and leading publishers such as Web of Science, Science Direct, IEEE Xplore, Scopus, Science Direct, etc. which are used to collect the research papers used for this review. Our goal is to provide a comprehensive review of the

The associate editor coordinating the review of this manuscript and approving it for publication was Tao Zhou.

latest work in the domain of multi-objective feature selection and discuss the challenges and current issues for future work. Also, this review aims to attract researchers working on multi-objective optimization paradigms to further study the effective methods to address new challenges in feature selection.

The rest of this systematic review paper is organized into the following sections: Section II presents an overview and essential definitions applied in feature selection. Section III introduces the aims and criteria of this systematic review of the literature on multi-objective feature selection. Data extraction and analysis of selected studies are provided in Section IV, followed by the profile of selected studies. The multi-objective feature selection elaborated in section VI. Lastly, the conclusions and future work are given in the last section.

II. BACKGROUND

This section provides a brief description of feature selection, key factors of feature selection and multi-objective optimization.

A. FEATURE SELECTION

Feature selection is considered as an essential pr-processing step in most machine learning methods. Features which are irrelevant or redundant can adversely affect model performance [5]. With irrelevant data features, the exactness of the models can be decreased and the model learns based on irrelevant features [6]. Therefore, feature selection refers to the method of acquiring the subset from an original feature, which selects the appropriate features in the dataset [2]. Some of the benefits of feature selection are as follows:

I. Reducing overfitting reduces redundant data implies decreasing opportunities for noise-based decision-making.
II. Improving precision and decreasing misleading data implies improving the precision of modelling.
III. Reduce training time, decrease data points, decrease the complexity of algorithms and accelerate algorithm training.

B. KEY FACTORS OF FEATURE SELECTION

As shown in Figure 1, the key factors of feature selection are evaluation measures, search methods, and the number of objectives.

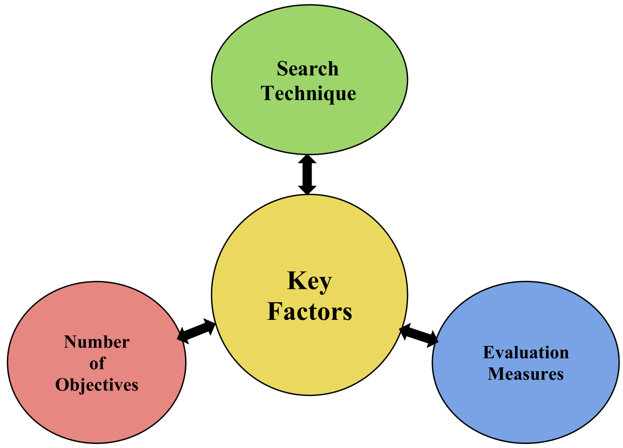

FIGURE 1. Key factors of feature selection.

1) SEARCH TECHNIQUES

Search or optimization methods are required for finding an optimal state for feature selection problem. Figure 2 shows popular search techniques used for this purpose. Exhaustive search is used in a limited number of researches [5], [6]. Several heuristic search approaches have therefore been utilized to the selection of features, including but not limited to greedy search techniques, both sequential forward selection (SFS) [7], and sequential backward selection (SBS) [8].

FIGURE 2. Search techniques of feature selection.

However, in later stages, selected or eliminated features cannot be selected or removed this is due to that the two approaches suffering from the supposed "nesting effect". By utilizing SFS l times, and then SBS r times, these two approaches are compromised and form the "plusl-take-away-r" [9]. This method should reduce the effect of nesting, but it is challenging to decide the correct values between l and r. Two techniques named (sequential backward and forward floating selection) have been proposed in order to prevent this issue [7]. It is stated that both floating search techniques are better than static sequential techniques. Recently, a two- layer cutting plane method to determine the optimum subsets of features were proposed by [8]. In [9] a backtracking method and heuristic search performing an exhaustive search for feature selection based rough set were proposed.

The findings indicate that heuristic search methods have accomplished comparable efficiency to the backtracking algorithm but consume less computational time.

Over the past few years, natural-inspired algorithms such as Genetic algorithms (GA) [10], [11] Particle Swarm Optimization (PSO [12], [13] Ant Colony Optimization (ACO) [14], [15] and Grey Wolf optimizers (GWO) [16]–[19] etc. have been used as efficient methods to solve issues with feature selection. However, these techniques show some limitations such as stagnating in locally optimal solutions and high computational costs. Also, they have been mostly used as a single objective either to reduce the number of features or maximizing classification accuracy. Therefore, a multi-objective idea of these techniques has been adapted to solve the problem of feature selection and shows a great success. Such methodsinclude MOGA [20], [21], MOPSO [22], MOGWO [23], [24] etc. However, these multi-objective techniques not fully investigated.

2) EVALUATION MEASURES

Feature selection approaches normally grouped into a wrapper and filter approaches [1], [2]. These two approaches generally differ in terms of evaluating the subset of features, the wrapper uses the classifiers in the evaluation process, which is the opposite in the filter which usually does not use any classifier in evaluating the subset of features. Wrappers are demanded to be costly in terms of computation and less general than filters. Nevertheless, the literature review shows that wrapper-based feature selectors tend to have a better performance than filters for classification. This is a result of filters disregarding the performance of features selected on a classification algorithm [25], [26]. Besides, other researchers classify the methods of feature selection into three sections: filters, wrappers and embedded methods [25], [26]. Embedded integrates the classifier and feature selected into a single process.

3) SINGLE-OBJECTIVE AND MULTI-OBJECTIVES

A method of aggregating feature numbers and classification accuracy into a single fitness function is considered as a single-objective method. Whereas, in the multi-objective approach, it corresponds to a method aimed in finding a Pareto frontier of the trade-off solution.

The dominance of a solution over other solutions is obtained in the single-objective problem by contrasting the values of their objective function whereas, in multi-objective problem, the dominance is used to find the best solution [27].

In the form of a group of solutions, a non-dominated solution set is a subset of all solutions that are not controlled by any member of the solution set. The Pareto optimum system is a non-dominated system of all feasible decision spaces. The boundary defined by the set of all points mapped by the Pareto Optimum Set is the Pareto Optimal Front [28].

C. MULTI-OBJECTIVE OPTIMIZATION

Identifying the important attributes is an essential factor in formulating an optimized feature selection process as ideal solutions to an optimization problem that increases or reduces a function referent to the standard of the significance of features as a relationship with data classes. The importance of feature subsets can be identified by the optimization concurrency of multiple criteria, taking into account distinct aspects. However, this could be complicated when these goals are conflicting. As a result, Multi-objective Optimization (MO) strategies could be used to overcome this challenge [29]. The problems of multi-objective occur wherever the best decisions essential to be made when a trade-off between two aims that are usullay conflicting to each other. Multi-objective optimization comprises objective functions that minimize or maximize multiple conflicts. Mathematically, the minimization problem with multi-objective functions can be expressed as follows without the loss of generality:

$$minimize\ F(x) = [f_1(x), f_2(x), \ldots, f_k(x)] \tag{1}$$

$$\text{Subject to } g_i(x) \leq 0, \quad i = 1, 2, \ldots m \tag{2}$$

$$h_i(x) = 0, \quad i = 1, 2, \ldots l \tag{3}$$

where $f_k(y)$ is a function of y, and y is decision variables vector, i is the number of objective functions to be reduced, and the constraint functions are $g_k(y)$ and $h_k(y)$. The superiority of multi-objective algorithm solution is clarified by the trade-offs among conflicting aims. For instance, the i-objective minimization problem has two solutions which are c and d. If the following conditions are met, it can be said that $\boldsymbol{c}$ dominates d orc over d:

$$\forall_i : f_i(c) \leq f_i(d) \quad and\ \exists_j : f_i(c) < f_i(d) \tag{4}$$

where i, j ∈ {1, 2, 3…k}

III. SYSTEMATIC LITERATURE REVIEW

In this work, we follow the systematic review used in PRISMA guidelines (www.prisma-statement.org). Based on these guidelines, the following research questions are formulated:

1) What are the search techniques that have been used to select the optimal features?
2) What are the evaluation measures that have been used for evaluating the selected features?
3) Identify how many objectives used in the idea of multi-objective?

A. SEARCH PROCESS

The review began by scanning for pertinent research studies on web sites and in the online library of Universiti Teknologi PETRONAS. The Internet search was led by utilizing search engines to trawl the computerized libraries and databases showed in Figure 3. The search parameter that used was "multi-objective feature selection". We use the Boolean operations AND for example "multi-objective AND Feature AND selection".

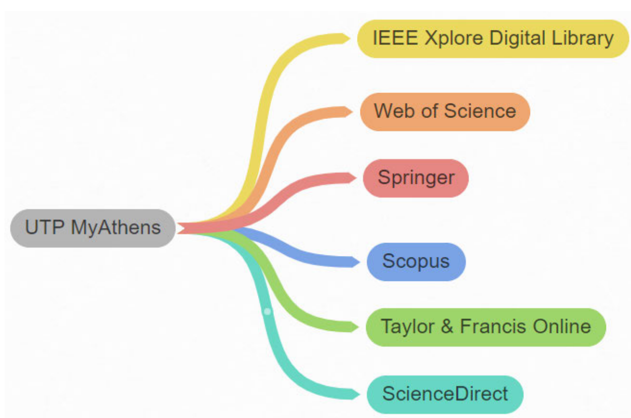

FIGURE 3. Digital libraries and databases searched.

B. RESEARCH SELECTION CRITERIA

There are numerous works on multi-objective feature selection. Thus, to guarantee that the search would be focused, and manageable, inclusion and exclusion criteria were identified to select the papers for analysis as follows:

1) INCLUSION CRITERIA

a) Research studies published from 2012 to 2019 related to multi-objective feature selection.
b) Only the studies that have been published in peer-reviewed journals are included.
c) Choosing the most complete version of the study for inclusion if it had been published in more than one journal.

2) EXCLUSION CRITERIA

a) Studies published before 2012 are excluded.
b) Informal studies (unknown journals); papers irrelevant to the search focus are excluded.
c) Only articles written by the English language are included. Other languages are excluded.

C. DOCUMENT AND BIBLIOGRAPHY MANAGEMENT

Mendeley Desktop was utilized to deal with all the bibliographic subtleties and references. The papers that were selected by the above-defined search procedure were checked by title and abstract as per the inclusion and exclusion criteria. Then, all the studies that were recognized as applicable to our study were formerly downloaded for information extraction and further investigated. Table 1 gives details on the number of research studies that were found by the search of the computerized libraries and databases in Figure 3 above and Figure 4 shows the search procedure as a PRISMA flowchart.

TABLE 1. Data sources and number of papers identified.

Data sources	Number of papers identified: Key terms	Title	Abstract
IEEE explore	54	19	10
Web of science	98	21	15
Springer	450	45	22
Taylor & Francis	85	12	9
Scopus	458	23	11
Science direct	1258	85	33
Total	**2403**	**205**	**100**

IV. EXTRACTION AND ANALYSIS OF DATA

Table 2 summarizes the 38 papers chosen for more detailed study. This table demonstrates the primary details taken from the chosen papers and introduced based on recent papers are published.

A. CHALLENGES IN FEATURE SELECTION

Despite the applicability, achievements and prospects of current multi-objective feature selection techniques, significant problems and difficulties remain and will be discussed in the following subsections.

1) HIGH SCALABILITY OF DATA

The most demanding problem is that the size of the data is getting extremely large because of the trend of big data [30]. In the past years, the selection of features from a dataset with 30 features was considered as large-scale/ high-dimensional feature selection [31]. Yet, the features number in several fields today can simply reach to even millions of features. This expands the computational expense and requires a powerful search mechanism, but these two aspects also have their problems, so the issue cannot only be resolved by maximizing the computing power. Developing new approaches and techniques will be a necessity.

2) THE COST OF COMPUTATION

Multi-objective feature selection methods involve a large number of evaluations which lead to a serious problem of being computationally expensive.

Wrapper methods are claimed to be less effective than filter methods, nonetheless, experiments have demonstrated the opposite in most cases [69]. A rough set theory [9], [70] is a filter type which needs more time than a simple wrapper one. Another example of the filter method is mutual information [71], [72], known to take less time for execution, but the accuracy performance generally takes more than the majority of wrapper methods. Thus, proposing an effective measure to multi-objective feature selection problem is still an open topic to be investigated. To do so, two primary factors, must be regarded to decrease computational costs which are: 1) an effective search technique 2) and a rapid evaluation measure [1].

3) SEARCH MECHANISMS

Selection of features is an NP-hard issue with large complicated solution spaces [73], [74]. Thus, a robust global search algorithm is needed, and recent multi-objective methods still can enhance further. The enhanced search mechanism should be able to search the entire search space and should leverage local areas if necessary. Also, the search mechanisms may comprise local search, hybridization of search mechanisms of different multi-objective techniques, hybridization of conventional approaches and multi-objective [75], surrogate approaches [76], etc.

4) EVALUATION MEASURES

One of the main factors in multi-objective feature selection is the evaluation metrics that shapes the fitness function. It greatly affects the time of computation, the performance of the classification and the search space. A large portion of the execution time is consumed on the assessment procedure of the wrapper strategy and several filter strategies [64]. Although there are some current rapid assessment measures, mutual information is an example [71], [77], [78] they assess features independently instead of groups of features.

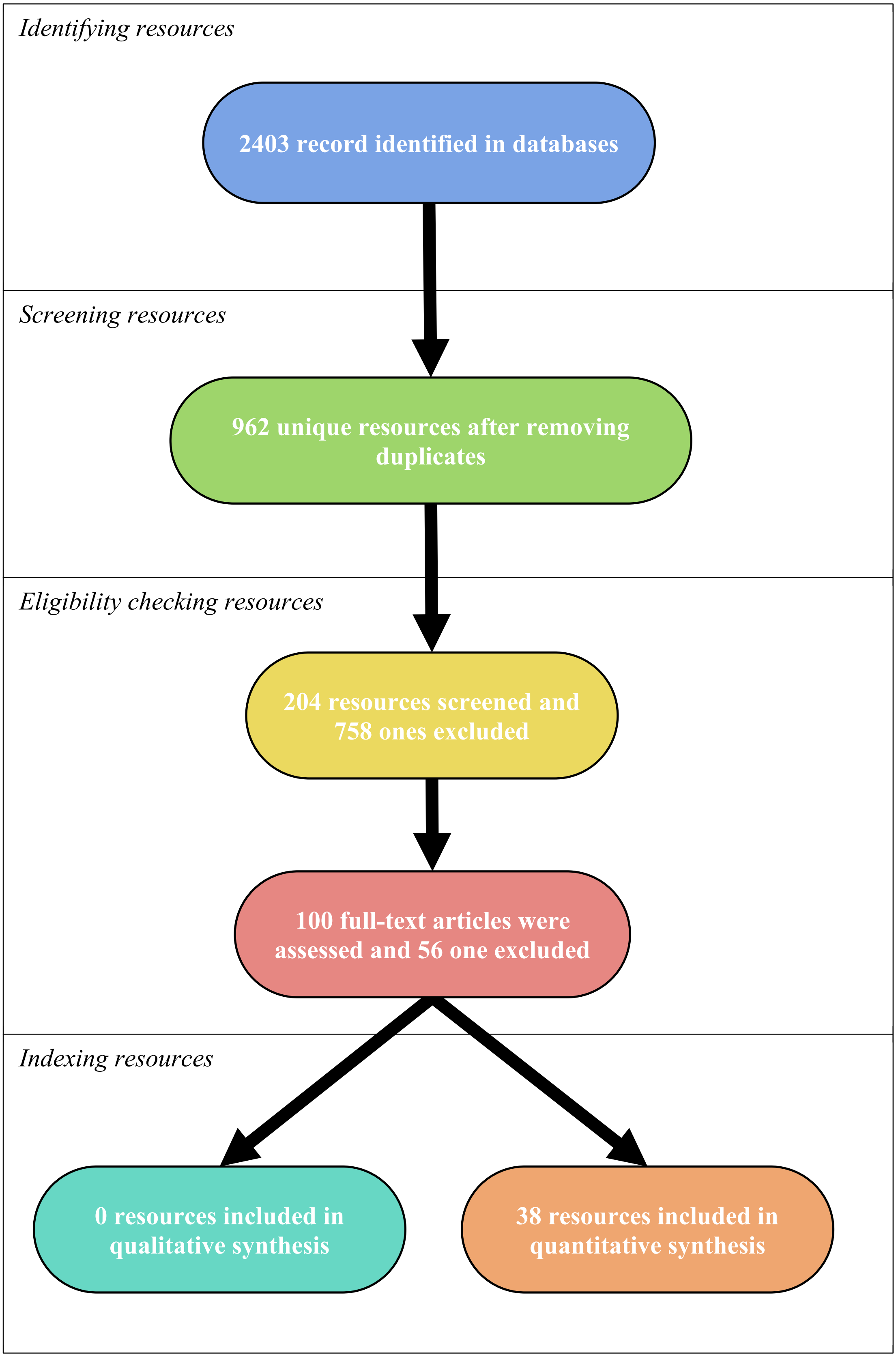

FIGURE 4. PRISMA flowchart adopted from [32].

TABLE 2. Details of methodology and finding of main selected research studies.

Publication	Methodology / Experimental Results / Finding	Key
Binary differential evolution with self-learning for multi-objective feature selection [33]	A binary multi-objective differential evolution based self-learning (MOFS-BDE) was proposed in this paper. To enhance its performance, three operators are proposed: 1) probability difference binary mutation, 2) Purifying Search One-bit operator 3) and crowding distance based non-dominated sorting operator. For evaluation, 20 standard datasets are utilized.	K1
	From the experimental results, the proposed binary mutation with probability difference adequately enhanced the convergence of MOFS-BDE. Dominating the solutions of five existing methods which are multi-objective based on DE, PSO, GA, ABC and MOEA/D. Besides, MOFS-BDE HV values were statistically significant (superior) to the existing methods at the 0.05 level.	
	The findings proved that the proposed MOFS-BDE was able to attain a trade-off among exploitation global exploration. The result compared with benckmarking algorithms was competitive where both the number of attribute and error of classification was reduced	
A multi-objective approach for profit-driven feature selection in credit scoring [34]	This paper proposed a multi-objective wrapper approach based on the NSGA-II with two objective functions: 1) maximinzing the expected profit 2) and reducing features subset. For evaluation, ten retail credit scoring datasets are utilized.	K2
	Experimental results show that, compared with benckmarking algorithms (GA, PSO, SBS, SFS, LASSO), the proposed method maximizes the expected benefit by using fewer features.	
	The proposed methods (NSGA-II boundaries) can yield smaller features on less dimensional datasets and higher differences in the distribution of data between holdout and training samples.	
A new multi-objective wrapper method for feature selection – Accuracy and stability analysis for BCI [35]	A wrapper multi-objective based on NSGA-II using breeding operators was proposed to solve feature selection in motor imagery data. Besides, a feature ranking process was proposed to maintain the stability of the proposed technique. To assess the proposed algorithm performance, four classifiers classification were applied.	K3
	Based on Kappa values, the proposed method was superior to the baseline (No FS method) and filter approaches (FOPT*, SR-LDA, SR-SVC) for the studied datasets irrespective of the choice of the classifier. Further statistical analysis showed that the proposed method yielded results similar (statistically insignificant) to FOPT* but superior (statistically significant) results than the compared methods.	
	From the experimental results, it was observed that the proposed algorithm can minimize and stabilize the features number with high accuracy of classification.	
C-HMOSHSSA: Gene selection for cancer classification using multi-objective meta-heuristic and machine learning methods [36]	This paper proposed a hybrid approach which utilized salp swarm algorithm and multi-objective spotted hyena optimizer to solve the problem of feature selection. This method consists of two stages. Firstly, the filter approach was used to eliminate the irrelevant features. Secondly, the proposed hybrid method used to explore the most optimal features.	K4
	Seven microarray datasets were applied in the experiments to assess the proposed algorithm performance. The experimental results demonsatred that the proposed algorithm attained better results than benchmarking algorithms (MOBBA, MOPSO, NSGA-II, IDGA, PLSVEG). Specifically, the proposed method produced 100% accurate results with lesser feature (4 genes) dataset using SVM and KNN classifier for leukaemia dataset as compared with lesser prediction accuracy by existing methods for studied datasets.	
	The findings shown that the proposed algorithm produced better results than the existing benchmarking algorithms in terms of reducing the error rate of classification and attributes number. That is, the embedded SSA in MOSHO was more efficient in the exploitation and exploration of MOSHO algorithm and consequently improved its performance.	
Novel multi-objective TLBO algorithms for the feature subset selection problem. [37]	A multi-objective teaching learning-based optimization (MOTLBO) has been proposed to solve the problem of feature selection. They proposed three multi-objective TLBO algorithms 1) Scalar transformation (MO-TLBO-ST), 2) Non-dominated Selection (MO-TLBO-NS) and 3) Minimum Distance (MO-TLBO-MD). To evaluate the performance of their proposed method three classifiers namely, LR, ELM, and SVM, have been utilized. Thirteen datasets downloaded from UCI were used in the experiments.	K5
	The results demonstrated that 1) MO-TLBO-ST was the fastest of the three proposed algorithms, nevertheless, it gained a small number of non-dominated solutions. 2) whereas, MO-TLBO-NS resulted in a huge amount of execution time but yielded to more non-dominated solutions. 3) The third method MO-TLBO-MD produced more non-solutions as MO-TLBO-NS with less time as MO-TLBO-ST. For the classifiers, LR was more efficient compared to others in term of execution time, whereas all classifiers attained similar accuracy. Finally, MO-TLBO-MD with LR reported to be the best method, therefore it was compared with benchmarking algorithms.	
	The outcome of the comparison illustrated that the proposed methods attain similar outcomes with NSGA-II, whereas outperformed other experimented methods.	
Optimizing Multi-objective PSO based feature selection method using a feature elitism mechanism. [38]	This paper proposed a PSO-based multi-objective method (RSPSOFS), which ranks features according to the frequency in the set of archives. These levels are then utilized to enhance the set of archives and guide the particles. The proposed algorithm was contrasted to three variants of PSO and one GA multi-objective approaches on nine benchmark data sets.	K6
	The experiment of the proposed approach (RSPSOFS), compared with HMPSOFS and CMDPSOFS, and two simple algorithms, MOPSO and NSGAII. The C-Metric scores of C(RSPSOFS, X) are superior to those of C(X, RFPSOFS) which means that RFPSOFS produced better results than other approaches. Besides, RSPSOFS outperforms both MOPSO and CMDPSO in terms of convergence.	

TABLE 2. *(Continued.)* **Details of methodology and finding of main selected research studies.**

	The findings proved that the proposed (RSPSOFS) can reduce features number with superior performance in the large datasets and acceptable, or similar efficiency in a dataset containing fewer than 100 attributes.	
Pareto front feature selection based on artificial bee colony optimization. [39]	This paper proposed a multi-objective artificial bee colony combined with nondominated sorting genetic operators. A binary version of and the continuous version was developed as two separated forms of the proposed algorithmwhich are Num-MOABC and Bin-MOABC. An experiment was conducted on 12 benchmark datasets. K Nearest Neighbor (KNN) with K = 5, was applied as a classifier to evaluate the feature subset.	K7
	The findings demonstrate that in most cases Bin-MOABC is preferable to other methods, based on the hypervolume ratio of the wilcoxon rank sum. Bin-MOABC performance is similar to NSGAII, Num-MOABC, and MOPSO on datasets with low-dimensional but outperformed NSSABC.	
	The outcomes of the proposed method were compared against two traditional methods namely, greedy stepwise backward and linear forward selections. Also, the proposed method compared with two ABC single-objective methods, as well as three popular multi-objective methods namely, NSGAII, MOPSO and NSSABC. The findings proved that in term of both reductions of dimensionality and classification precision the proposed binary version outperformed the other methods.	
Gene selection for tumour classification using a novel bio-inspired multi-objective approach. [40]	This paper proposed a bio-inspired, multi-objective approach for gene selection and classification of microarray results. The proposed approach (MOBBA) expands the original Bat Algorithm by developing random walks with refined formula, efficient multi-objective operators, and local techniques for research. A hybrid model focused on the Fisher criteria would instead be used to examine essential biomarkers that show the efficacy of the current genomic analysis method on three large-scale microarray cancer datasets to classify the most selective biomarkers. The Fisher score was used to select datasets to render another dataset containing just the maximum measurably valid genes. The consistency requirements of each subset were evaluated using four commonly used classifiers which are: DT, KNN, NBY and SVM.	K8
	The MOBBA established a highly insightful group of three genes that were broadly compatible with the literature's previous studies. The proposed method produced a superior performance with lesser feature dataset using SVM and KNN classifier for leukaemia dataset as compared with lesser prediction accuracy by existing methods for studied datasets.	
	The outcomes of the proposed approach (MOBBA) accomplished high accuracy and a significant reduction of genes, as well as outperforming the state-of-the-art methods against leukaemia dataset in term of accuracy utilizing only three biomarkers.	
Multi-objective simplified swarm optimization with a weighting scheme for gene selection [41].	This paper proposed a multi-objective optimization-based hybrid filter-wrapper technique for gene selection. This technique uses an aggregate filter technique to select finest genes. Moreover, a simplified swarm based multi-objective is suggested and incorporated as a wrapper with SVM to search for the appropriate subset of gene from the chosen genes. The proposed method utilizes a weighting system to lead the search to the suitable areas identified by the preference, unlike most commonly used multi-objective based techniques used to deal with gene selection issues. In this respect, only preferred gene subset are produced and not all Pareto optimal alternatives.	K9
	The experimental findings revealed that AFM-MOSSO had an overall classification accuracy of over 99 percent with fewer than 10 identified genes. This indicates that fewer genes are required to achieve 100 percent accuracy. The Friedman ranks have shown that AFM-MOSSO is the best performing method of comparison with a rank of 1.4.	
	The results confirmed that the proposed technique is a very competitive and good solution to gene selection issues can be regarded. Moreover, both WES and GPS improve the capability of AFM-MOSSO to reduce gene subset and increase accuracy performance. A comparison with six related methods indicates that the suggested algorithm outperformed them.	
MOFSRank: A Multiobjective Evolutionary Algorithm for Feature Selection in Learning to Rank. [42]	This work proposed a feature selection method based on multi-objective optimization named MOFSRank in learning to rank. This method involves three mechanisms. First, a strategy of selection of instances is proposed to select useful examples from the set of ranking, which removes unwanted data and increases training effectiveness. Second, a multi-objective method based adaptive mutation was proposed on the subsets selected, in which good feature subsets are attained through the choice of high-ranking and low redundancies features. Finally, linear SVM was used to produce the ranker which employs these attained feature subsets to generate a set of better features.	K10
	The proposed approach accomplished better accuracy on two statistical points, proving the better performance of the proposed method on the dataset named LETOR and shows the efficiency.	
	The competence of the proposed algorithm has been shown by experimental outcomes on LETOR datasets. They compared their approach in the first portion of the studies with various classical algorithms without the utilization of feature selection.	
Multi-objective grey wolf optimizer for improved cervix lesion classification. [23].	This paper proposed two different methodologies binary GWO based multi-objective: 1) a scalarized way to deal with multi-objective MOGWO 2) a non-dominated sorting NSGWO. Both are utilized for the wrapper method which chooses an ideal subset of textural for enhanced efficiency of cervix lesions. For evaluation, different improved CT-Scan pictures of 62 patients were utilized.	K11
	The results of the proposed two methods are compared against the MOGA and multi-objective firefly technique (MOFA). GWO attains Pareto solutions with better diversification and intensification, that control the solutions acquired by MOFA and MOGA when evaluated on the instances of cervix lesion. Cervix lesions with only five selected features by NSGWO high accuracy of 91%. This proves that NSGWO achieves better result comparing to other approaches in cervix lesion dataset. An additional evaluation was made on large-scale microarray gene expression datasets.	

TABLE 2. *(Continued.)* **Details of methodology and finding of main selected research studies.**

	The outcomes prove the superiority of the proposed algorithm comparing to benchmarking algorithms in term of the features reduction and high accuracy rate.	
Multi-objective differential evolution for feature selection in facial expression recognition systems. [43].	This paper used a multi-objective DE wrapper feature selection for the facial expression recognition systems. linear SVM used as a classifier. Two objectives of the proposed method were the size of feature attributes and the accuracy of classification. The proposed method was validated against three facial recognition databases, namely JAFFE, MMI and CK.	K12
	The proposed method (DEMO) on Cohn Kanade, JAFFE and MNI databases had 98.37%, 92.75%, and 89% mean emotion recognition rate respectively with an 89% reduction in the number of features. Compared to the state-of-the-art, the FER approach produces good efficiency, while at the same time substantially reducing the number of features, which, in effect, minimizes the computational expense of training classifiers.	
	The outcomes of the proposed method demonstrated that the proposed technique is similar to state-of-the-art techniques or even better in the MMI database in terms of precision of recognition, whereas it consists of much simpler processing blocks and fewer features used compared to other techniques.	
Elitism based multi-objective differential evolution for feature selection: A filter approach with an efficient redundancy measure. [44].	This work proposed a filter feature selection method utilizing multi-Objective differential evolution (FAEMODE) based on Elitism and the contribution lies in an objective formula in which linear and nonlinear dependencies between features are considered to remove redundant and undesired features of the processed data set. For evaluation 23 benchmark data sets were used and the selected features tested utilizing 10 cross-validations and 4 recognized classifiers to support the results.	K13
	The proposed method (FAEMODE) outperforms the conventional (MIFS, NMIFS, mRMR, MIFS-U, MIFS-CR, SFS, SBS) with an average AUC score of 89.59% and an average rank of 1.66. Also, the proposed methods had 88.13% average AUC score which is superior to experimented multi-objective FS methods (MECY-FS, MEFS-U, BBA-FS, DEMOFS, MOEA/D-FS)	
	The proposed method has been compared with seven filtering methods and two conventional and three meta heuristic-based packaging methods for verification. The outcomes demonstrate that the technique can be utilized as a good filtering method for feature selection in various fields. This leads to the selection of a subset of features with good and stable data set classification potential.	
An improved NSGA-III algorithm for feature selection used in intrusion detection. [45].	A multi-objective method is proposed in this article for feature selection in IDS, which utilizes two approaches for population evolution, namely a unique method of domination and a predefined multi-targeted search. NSGA-III is used to acquiring an appropriate subset of features good results. Using a new niche preservation method, an enhanced multi-objective algorithm (I-NSGA-III) is further proposed.	K14
	From the results, I-NSGA-III had the higher classification accuracy (99.62%) and selected fewer instances (20) compared with existing methods (NSGA-II (99.2%, 27) and NSGA-III (99.6%, 24)). The proposed method also had a lower computational complexity.	
	Experimental findings indicate that for classes with fewer instances, I-NSGA-III can reduce the imbalance issue with greater classification precision. It can also attain greater classification precision as well as reduced computational complexity.	
Evolutionary multi-objective fault diagnosis of power transformers. [46]	This work utilized a binary variant of PSO based multi-objective MOPSO method to propose a two-step method for diagnosis of fault energy transformer (2-ADOPT). Selection of subset and an ensemble classifier is introduced to enhance the diagnostic accuracy of power transformers for gas dissolved analysis. In the first stage, the proposed technique chooses the best features in the framework of multi-objective and, at the same time, the number of finest features utilized as inputs to train classifiers. The most exact and various classifiers are chosen in the second stage to make a classifier group. Lastly, to determine the real failures of power transformers, the outputs of chosen classifiers were jointly utilizing a (dempster Shafer combination rule).	K15
	The proposed method (2-ADOPT) had an accuracy of 100%, AUC of 1 and F1-score of 1 which outperformed MOECS (94%, 0.97, 0.95), RF(91%, 0.97, 0.91), ORF(97%, 0.97, 0.96). The performance can be attributed to the variants of PSO (FSS-MOPSO and ECS-MOPSO) used feature selection in the proposed diagnostic framework (2-ADOPT)	
	The comparative results were favorable to the proposed approach and showed this approach 's elevated efficiency for energy transformer fault classification.	
Multi-objective evolutionary feature selection for online sales forecasting. [47].	For feature selection in online advertising sales forecasting, this work utilizes a multi-objective algorithm based on the Evolved Non-Dominated Radial Slot Algorithm (ENORA). They suggest a technique for integrating the selection of features into regression, model assessment and decision making to select the most satisfying model based on the posterior.	K16
	The results produced by ENORA are (significantly) superior to that of NSGA-II across the studied datasets.	
	The findings of the proposed method compared with the NSGA-II method and the results demonstrated that the proposed method provides better super-volume values and returns a more precise data set than NSGA-II. Another comparison was conducted against recursive feature elimination (RFE) wrapper approach, which shows that the proposed method performed better in terms of means error and selected subset.	
Multi-objective feature selection for warfarin dose prediction [22].	A multi-objective feature selection method was proposed in this article to predict the dose of warfarin. This study was carried out on 553 patients. NSGA-II and MOPSO were utilized for feature selection. Besides, artificial neural networks are used to evaluate the selected features.	K17
	MOPSO and NSGA-II had superior accuracy and efficacy compared to conventional feature selection methods. MOPSO had a higher precision value than NSGA-II. When only 7 features are selected the for MOPSO obtained the following evalution measures values: MSE, RMSE and MAE were 0.011, 0.1 and 0.109 respectively.	
	The findings of the proposed methods showed better accuracy compared to conventional feature selection techniques. Besides, MOPSO gained more precision and a smaller number of features than NSGA-II. The	

TABLE 2. ***(Continued.)*** **Details of methodology and finding of main selected research studies.**

	findings indicate that both algorithms are more accurate than standard techniques, and the MOPSO has the highest accuracy.	
Robust multiobjective evolutionary feature subset selection algorithm for binary classification using machine learning techniques. [48].	In this paper, the combination of machine learning (ML) and multi-objective algorithm (GA) has been studied with focus on the success feature subset selection (FSS) techniques in binary classification problem (BCP). Two stages have been conducted on the proposed method. FSS using a GA and applying ML techniques for the BCP. GA is firstly used to explore the intelligence of the algorithm because of the difficulty of conducting a comprehensive study of all subset features. In the second stage, the efficiency of the selected subset is identified by the usage of ML techniques.	K18
	The proposed method outperforms the existing methods (PSO, TS, SFS, SBSB, SS-GC, SSS-RGC, PSS) in terms of accuracy and selection of fewer features.	
	Experimentally, the efficiency of the proposed approach was assessed and compared with the state-of-the-art method. The finding proves that the proposed method on most of the datasets performed better than the existing approaches	
A PSO-based multi-objective multi-label feature selection method in classification. [49].	A multi-label feature selection is proposed in this paper utilizing an enhanced MOPSO to search for a set of non-dominated Pareto subsets of features. To enhance the efficiency of the suggested PSO based method, two operators were utilized a mutation and local serach procedure. Furthermore, to discover the Pareto set, the notion of archiving and the crowded distance were introduced to the PSO.	K19
	The proposed method outperformed Relief (RF), Mutual Information (MI) NSGA-II by a 4.2% reduction in its Hamming loss (Hloss) value with fewer features. Also the proposed method recorded the best Set Coverage (SC) metric across all studied data sets which are a sign of good parallel search capability. That is, the proposed method can find a set of optimal solution better than other experimental methods.	
	In the experiment, two traditional techniques (RF-BR and MI-PPT) and NSGA-II method were used for the comparison with the proposed algorithm. The finding illustrated that: (1) High loss value gained due to ability of the proposed method in searching for the extreme solution; (2) The suggested method discovered some tiny Hamming loss subsets of features in one run (3) The proposed method performed better than NSGA-II in term of exploration.	
Feature weighting and selection with a Pareto-optimal trade-off between relevancy and redundancy [50].	In this study, a bi-objective feature selection has been developed. A subset of weighted features was selected as the optimal subset for subsequent data classification. Multi-Objective Evolutionary Algorithm Based on Decomposition (MOEA / D) was proposed based on two data measures, 1) relevancy 2) redundancy. They evaluate the proposed method on 15 standard datasets utilizing k -NN and SVM classifiers.	K20
	From the experimental results, the proposed method presented highly competitive results against other experimented methods (SFSW, RRFSACO, BCFA, FSFOA, FR-FS, FS-FS) with a good prediction accuracy and fewer number of selected features.	
	The finding of the proposed method shows a competitive result comparing to the state-of-the-art in term of accuracy and number of selected features. Also, to prove the efficiency of their proposed framework they used NSGA-II and SVM.	
Feature selection of unreliable data using an improved multi-objective PSO algorithm [51].	Unreliable data feature selection problem was addressed in this paper. First, the multi-objective idea has two goals:1) classification accuracy 2) the reliability. Second, by integrating two new operators, and efficient bare-bones PSO based multi-objective selection method is proposed. The two operators are 1) reinforced memory approach 2) a hybrid mutation.	K21
	The proposed method (BMOPSOFS) produced the smallest set of features, the best hyper-volume (HV) value and had a robust capacity in identifying and eliminating features that are redundant compared to MGC-MOPSO and TV-MOPSO.	
	The finding demonstrated that the proposed method is competitive for unreliable data feature selection problem within the term of classification accuracy, and the reliability.	
Multi-objective feature subset selection using non-dominated sorting genetic algorithm. [52]	In this study, NSGA- II is utilized to select the most optimal features. The classifier ID3 is designed to assess the fitness value of a specific subset of features. The gained test accuracy is then allocated to the fitness value. This method is evaluated against numerous datasets collected from the repository of UCI.	K22
	When the population is 70, the proposed method produces superior results. The results indicate that small datasets require a large number of generation to reach the optimal Pareto front, while large datasets require a small number of generation.	
	The finding of the proposed method NSGA-II confirmed a good result in terms of the performance of classification and reduction of features.	
A multi-objective evolutionary algorithm for feature selection based on mutual information with a new redundancy measure [53].	This paper proposed a multi-objective approach (MECY-FS) based on the correlation metric and the redundancy metric, is proposed. The MECY-FS algorithm uses Pareto optimality to assess candidate feature subsets and find a compact subset of features with maximum correlation and minimum redundancy.	K23
	When evaluating the features through 1NN, C4.5 and RBFNN, MIFS-CR obtained the best mean AUC score. According to the results of NB, MIFS-U obtained the best mean AUC score, while MIFS-CR produced the AUC score near to the best mean. MIFS-CR shows an obvious performance better than other methods when utilizing C4.5 classifier,.	
	The experiment was performed on benchmark datasets to validate the efficiency of the suggested technique, and the findings showed that the suggested technique was capable of generating compact subsets with a high predictive capacity.	
Global mutual information-based feature selection approaches using single-objective and multi-objective optimization [54].	In this paper, the feature selection based on mutual information is converted into a global optimization problem. First, a single-objective (SO-MIFS) that combines redundancy and correlation is proposed. The algorithm has good global search capabilities and high effectiveness in computing. Second, a multi-objective feature selection approach (MO-MIFS) is proposed to enhance the efficiency of feature selection. This strategy can fulfil various demands and reach a compromise between various conflicting objectives. Based on that, the final solution is attained utilizing a hybrid feature selection structure. Both synthetic and real datasets are used for evaluating the proposed method.	K24

TABLE 2. ***(Continued.)*** **Details of methodology and finding of main selected research studies.**

	The performance of the proposed methods is compared to two popular MI-based FS methods (MR and mRMR). The proposed methods had a lower RSME and SMAPE values compared with BE-mRMR and FS-mRMR (variants of mRMR).	
	Results of simulation demonstrate the efficacy and practicality of the technique proposed. To demonstrate the comparative efficiency of the suggested technique, a thorough experiment was conducted. It is observed that in practical applications, the suggested global feature selection technique is superior to other techniques.	
Multi-objective optimization of shared nearest neighbour similarity for feature selection [55].	A feature selection based on SNN-based distance is proposed in this paper. It selects a decreased set of features while maintaining the resemblance of the pairwise sample. In terms of the SNN distance, a feature assessment criterion is developed and optimized in a multi-objective structure concurrently with the feature set cardinality. As an optimization method, they used the NSGA-II to traverse the search space and discover a non-dominated set of features.	K25
	A decreased set of samples selected to maintain sample resemblance reduces the impact of outliers on the selection process while also reducing the complexity of the computation. The proposed approach shows best performances in classification, besides the sample similarity for three datasets COIL20, USPS and ORL.	
	The outcomes showed that the simplified feature subset cannot only sustain the estimate accuracy of the classifier in the decreased feature space, yet, its addition improves the original feature space in some cases. The index of verification demonstrated that the sample similarity additionally stays in the decreased space, and the features selected have less correlation between them. The comparison with related methods verifies the applicability of the proposed approach.	
Parallel alternatives for evolutionary multi-objective optimization in unsupervised feature selection [56].	This paper addresses the large-scale feature selection by using the independent evolution of subgroups that work together after a given number of independent generations, thus contributing to the parallel implementation of MOEA. Besides, feature selection has been resolved as an unsupervised multi-objective clustering problem. Synthetic and BCI benchmarks are used for experimental evaluation.	K26
	the resutls show that MO/SOM beenfit from index superior. The proposed method produced a Kappa index superior to other methods except for BFS but requires much more runtime.	
	The experimental results demonstrated that the proposed method is competitive with other state-of-the-art method and produced a smaller number of features with high classification accuracy.	
Multi-objective unsupervised feature selection algorithm utilizing redundancy measure and negative epsilon-dominance for fault diagnosis. [57]	In this paper, the unsupervised feature selection multi-objective based algorithm (MOUFSA) is proposed. A recent objective that incorporates relationship coefficients and cardinalities for features subsets not only assesses the repetition of the selected features but gives a few objective qualities to every specific size of the feature subset. Besides, a simple chronicling procedure dependent on negative ε-advantage and box-based methodologies are intended to protect promising arrangements regardless of whether they rule. Additionally, three mutation operators with various capacities are proposed to improve the algorithm.	K27
	Compared with five state-of-the-art methods, the proposed methods have higher accuracy or RI value. In most cases, the average results of MOUFSA are also better than MOFSA1, MOFSA2, FMOFSA, WMOFSA and ReliefF. For the feature subsets of specified size, the proposed method can obtain several comparable solutions, while ReliefF, MOFSA1, FMOFSA, WMOFSA, and AEUFS usually only get one solution.	
	For evaluating the proposed method 9 UCI datasets and 5 fault recognition datasets are utilized, and the acquired subsets of features are then utilized for resulting the clustering quality and classification. Test results demonstrate that MOUFSA produced better results than other multi-objective and single-objective approaches.	
Feature selection by multi-objective optimisation: Application to network anomaly detection by hierarchical self-organising maps [58]	This paper proposed a multi-objective feature selection based on the unsupervised clustering process of growth hierarchical self-organizing map (GHSOM), including a new unit marking method and effective determination of the winning unit to solve the intrusion detection problem. In the network anomaly detection problem considered here, this multi-objective method can not only distinguish between normal and abnormal traffic but also distinguish different anomalies. Using the DARPA / NSL-KDD dataset, which includes chosen characteristics and marked attacks obtained from about 2 million links, the effectiveness of the suggested technique has been evaluated. It also involves a multi-objective method for selecting features based on the NSGA-II algorithm to decrease the complexity of GHSOM and enhance the efficiency of classification.	K28
	The outcomes got on the KDD-NSL gives recognition rates up to 99.8% for ordinary traffic, and up to 99.4% for odd traffic with five levels and 225 neurons. The recognition precision accomplished by the proposed strategy is 99.12%, hence improving the outcomes gotten by the considered proposed methodology.	
	The analyses performed with the KDD-NSL dataset demonstrate the relabelling technique proficiency through the relating ROC curves and the rate of classification upgrades.	
A multi-objective evolutionary algorithm-based ensemble optimizer for feature selection and classification with neural network models [59].	This paper proposes an integrated optimizer based on multi-objective method, combined with a neural network model for selecting and classifying features. Specifically, a modified micro-genetic algorithm (MmGA) is utilized to frame a set enhancer. The objective of the MmGA-based set enhancer is twofold, that is, selecting less number of features input for classification and enhancing the performance of classification of the neural network model. To assess the viability of the proposed framework, various benchmark issues were first utilized, and the outcomes were contrasted with the outcomes of other approaches. The applicability of the proposed framework for human movement identification and classification tasks is then assessed. Two UCI benchmark issues and human movement identification and characterization issues have been utilized to assess the adequacy of the proposed MmGA-based set enhancer.	K29
	The proposed method (MmGA) model are higher than benchmarking studies (Decision Tree, Bagged and Boosted Decision Tree based on PCA) for.e. 6.1 and 6.2. there was improvement in the accuracy from 0.2% to 9.5%, and also the feature reduction from 26% to 52%.	
	The results affirmatively showed that the proposed MmGA-based set enhancer can enhance the performance of the classification of neural network models with less subset of features.	

TABLE 2. *(Continued.)* **Details of methodology and finding of main selected research studies.**

Feature selection for face recognition based on multi-objective evolutionary wrappers [60].	A gentic based multi-objective was proposed in this study facial recognition feature selection tasks. The proposed system investigates the space of numerous possible decisions to limit the cardinality of the component subset while boosting its discriminative power. The evaluation was carried out on a familiar face picture dataset	K30
	The feature subset gained by the multi-objective method is obviously smaller, especially when compared with the improved ASM technique, which has better performance of classification. The smallest subset detected by the proposed technique contains only 26 features, and can significantly reduce the classification error to the enhanced ASM with a RER of 35.76%. In addition, the solution found by MOGA provides fewer features, while producing accuracy similar to that of single-target GA.	
	Experimental findings indicate that the suggested technique enables enhanced classification efficiency while decreasing representation sizes compared to other state-of-the-art techniques. Also, the ASM-based representation's dimensionality has been considerably decreased, which also helps to prevent overfitting. The suggested approach, therefore, offers a valid option for selecting appropriate face recognition characteristics.	
An SVM-wrapped multiobjective evolutionary feature selection approach for identifying cancer-microRNA markers. [61].	In this paper, a genetic algorithm-based multi-objective feature selection algorithm is developed, which incorporates an SVM classifier for recognizing miRNA markers from miRNA datasets. The method simultaneously improves diverse execution criteria and builds up the subset of the required features (miRNAs). SVM generalization parameters were created in conjunction with associated subsets of features.	K31
	Compared with other stat-of-the-art algorithms, the proposed multi-objective method affords better classifier performance. A value of 0.958 for the AUC produced by the selected features of MOGA is superior to the AUC value provided by other methods. Therefore, it is clear that the feature selected by MOGA can provide better performance in comparsion to other technologies.	
	The experimental findings indicate that the chosen selected of the proposed technique provide better efficiency measurements for the classifier compared to other algorithms. The proposed method chooses a set of features that are statistically important to provide a performance metric for elevated classification.	
Multi-objective evolutionary algorithms for filter-based feature selection in classification [62].	In this paper, two multi-objective filter-based selection frameworks are created based on NSGAII and strength Pareto developmental algorithm 2 (SPEA2). Then, by implementing mutual information and entropy as two distinct filter assessment criteria in each of the two suggested frameworks, four multi-objective selection techniques are created. The proposed methods are evaluated and contrasted on eight benchmark datasets with state-of-the-art methods. A decision tree is utilized to assess the performance of the classification	K32
	In the proposed multi-objective method, the number of features is always smaller than single-objective method, which demonstrates that they can more efficiently discover the search space to minimize the number of features.	
	Experimental finding demonstrated that the proposed multi-objective algorithms can automatically develop a non-dominated solution that contains reduced features number and attain high classification performance. The proposed methods compared outperformed the state-of-the-art methods in terms of reducing the number of features and increasing the performance of classification.	
A multi-objective particle swarm optimisation for filter-based feature selection in classification problems [63].	A BPSO filter approach and information theory for multi-objective feature selection was proposed in this work with the goal of gaining a non-dominated feature subsets, thereby decreasing the number of feature subsets and achieving greater efficiency in classification. By establishing two multi-objective BPSO schemes and two criteria for data assessment, the objective was efficiently accomplished. Consequently, for the selection of features, four multi-objective algorithms are suggested and then evaluated and contrasted with techniques of related works.	K33
	From the results, the number of features is somewhat larger, while the error rate of classification in NSfsMI-B is less than that in BPSOfsMI. In certain instances, NSfsMI is superior to BPSOfsMI in terms reducing th number of features as well as the error rate of classification. When CMDfsMI compared with BPSOfsMI, the performance of the CMDfsMI out perform BPSOfsMI in terms of reduction of features.	
	Experimental findings indicate that the suggested method can develop a set of alternatives that use a lower amount of feature and attain better efficiency in classification than all features. The proposed method outperformed the related work methods in term of reducing the number of features and classification performance.	
Particle swarm optimization for feature selection in classification: A multi-objective approach [64].	This paper proposed a feature selection method based on MOPSO. Two multi-objective feature selection algorithms based on PSO (CMDPSOFS-B and NSPSOFS-A) are examined. The first algorithm presents the concept of nondominated sorting in PSO to resolve issues with feature selection. The second algorithm refers to PSO the concepts of crowding, mutation and dominance in searching for the front alternatives of Pareto.	K34
	CMDPSOFS-B is superior to NSGAII-B and SPEA2-B, in terms of classification error rate and feature reduction. The NSPSOFS-A performance is somewhat similar to that of NSGAII-A and SPEA2-A when compared. While NSPSOFS-A error rates are higher than NSGAII-A and SPEA2-A error rates, the number of NSPSOFS-A features is typically smaller than NSGAII-A and SPEA2-A. The findings showed that CMDPSOFS is capable of overcoming the NSPSOFS constraint and producing improved performance than NSPSOFS, NSGAII and SPEA2 in terms of classification error rate and feature reduction.	
	Experimental findings indicate that the suggested method can develop a set of alternatives that use a lower amount of feature and attain better efficiency in classification than all features. The proposed method outperformed the related work methods in terms of minimizing classification error rate and features number.	
Safs: A deep feature selection approach for precision medicine [65]	This research implemented stacked auto encoders in higher-level abstraction for representation of features. A feature-learning method was tailored to a different precision medicine issues that focuses on recognizing and organizing risk factors for hypertension in a disadvantaged population subgroup (African-American). The suggested method is to use deep learning to classify major risk factors impacting left ventricular mass indexed to the area of body surface as an measure of risk for heart damages.	K35

TABLE 2. *(Continued.)* **Details of methodology and finding of main selected research studies.**

	The results demonstrated that the proposed approach significantly outperforms state-of-the-art approaches in terms of the ranking and selection of features.	
	The proposed feature learnining method can be a suitable option to assess which patients at greater risk will be recommended for more thorough review and which patients with elevated LVMI benefit more or less from the actual medical issue of diagnosis.	
Deep-FS: A feature selection algorithm for Deep Boltzmann Machines [66]	This paper proposes a deep feature selection that have the ability to eliminate the features that are redundant in large datasets with the aim of minimizing the number of inputs modeled during the learning phase. This method utilized Deep Boltzmann as well as the knowledge learnt throughout the training phase to eliminate features at the start of the learning process.	K36
	The results demonstarted that the proposed approach can perform a selection of features without reducing accuracy. As far as the MNIST dataset is concerned, by more than 45% the proposed approach decreases the number of input features; besides from 0.97% to 0.90% the network errors is decreased; and more than 5.5% reduction on the computional time. Furthermore, comparing to traditional feature selection approaches, the proposed method outerform in terms of accuracy. For the following datasets: GISETTE, MADELON and PANCAN the proposed method reduced the number of input features by 81%, 57% and 77% respectively. Whereas. the computional time reduced by 82%, 70% and 85% respectively.	
	Experiments on diverse datasets containing a large number of features and samples demonstrate that the proposed approach overcomes the key drawbacks of conventional feature selection algorithms. More precisely, most conventional approaches need to maintain a certain number of features as a requirement, but this number is automatically defined in the proposed deep feature selection. Compared to conventional feature selection algorithms, the proposed method performs faster feature selection tasks, making it ideal for deep learning tasks. Furthermore, the proposed deep feature selection is suitable for finding features in large datasets.	
Deep learning-based feature selection for remote sensing scene classification [67]	In this paper, a feature selection method based on deep learning is introduced, which expresses the task of feature selection as the task of feature reconstruction. The common deep learning technique Deep Confidence Network obtains feature abstraction by reducing the reconstruction error of the entire feature set, and the features with smaller reconstruction error will be affected.	K37
	The results demonstarted that the proposed method outperform the baseline method in terms of higher performance which indicate that the proposed method can successfully select discriminant features for the scene classification task.	
	The proposed method chooses the more reconstructible features as the discriminative features. Precisely, an iterative technique was established to acclimate the deep confidence network to yield the requested weights of reconstruction.	
Robust and accurate feature selection for humanoid push recovery and classification: deep learning approach [68]	This paper introduces the human thrust recovery data classification. The features used in this data classification are obtained from the inherent mode function by execution empirical mode decomposition at various angles of the leg joints such as: knee, ankle and hip. The joint angle data of subjects with open eyes and closed eyes were calculated.	K38
	The classification performance was achieved by utilizing deep neural network with an accuracy of 89.28 %. The proposed deep neaural network was evaluated on small, medium, moderately high, high different pushes. Based on 5-flod cross validation the classification accuracy of 88.4 % has been attained. Besides, statistical singnifance have been shown by the analysis of the variance.	
	The appropriate strategies knee, ankle and hip may be used once the push recovery categories small , medium, moderately high, high have been identified accordingly.	

Moreover, ignoring associations within the features leads to redundancy subsets and absence of complimentary features [2], [79], which thusly cannot accomplish ideal classification performance in many research domains. However, it is very difficult to discover complicated interactions within features, and there were only a few studies in this direction [80]. Although, some measures can assess groups of features [8], yet they are normally computationally costly, an example, rough set theory [13], [81]. Besides, numerous investigations demonstrated that filter strategies do not scale well over a huge number of features [30]. Thus, developing new evaluation measures will be a necessity, particularly when dealing with huge scale issues.

5) OBJECTIVES OF MULTIOBJECTIVE IDEA

The vast majority of the current multi-objective methods are intended for continuous issues [82], however, feature selection is a binary problem. Existing multi-objective techniques do not scale well when dealing with large-scale issues [83], [84]. Thus, novel multi-objective algorithms will be a necessity. Furthermore, most of the studies mentioned in Table 3 have two main objectives: minimizing both features number and error rate of the classification, which are not continually conflicting with one another, for example in certain subspaces, reducing features number leads to reduce the error rate of classification as redundant features are eliminated [19], [64], [85], [86]. This makes it precarious to develop a suitable multi-objective technique. Lastly, other goals, such as complexity, time of computation, and scalability, could also be regarded in multi-objective feature selection in addition to the two primary goals.

V. PROFILE OF SELECTED STUDIES

A. RESEARCH FOCUS

The research papers on multi-objective feature selection have attempted to solve the above-mentioned problems associated

TABLE 3. Research questions analysis of main selected research studies.

Studies	Quetion1: Search Technique?	Question 2: Evolutions Measure?	Question 3: Objectives of Multi-objective idea?
K1	MO-BDE algorithm	Wrapper approach, k-Nearest Neighbors (KNN) classifier.	Two minimization objectives: attributes number and error of classification
K2	NSGA-II algorithm	Wrapper approach, logistic regression, extreme gradient boosting and L1-regularized logistic regression	Two conflicting aims: model's comprehensibility and profitability.
K3	NSGA-II algorithm	Wrapper approach, KNN, Linear Discriminant Analysis (LDA)+ KNN, Naïve Bayes Classifier (NBC) and LDA+ NBC	Two minimization objectives: attributes number and error of classification
K4	C-HMOSHSSA algorithm	Wrapper approach, Support Vector Machine (SVM), KNN, NBC, Decision Tree (DT)	Two minimization objectives: attributes number and error of classification
K5	MO-TLBO algorithm	Wrapper approach, three classifiers namely, LR, ELM, and SVM	Two minimization objectives: attributes number and error of classification
K6	MO-PSO algorithm	Filter approach, feature ranking	Two minimization objectives: attributes number and error of classification
K7	MO-ABC algorithm	Wrapper approach, KNN with K = 5.	Two minimization objectives: attributes number and error of classification
K8	MO-Bat algorithm	Wrapper approach, SVM, KNN, NBC, and DT.	Two minimization objectives: attributes number and error of classification
K9	MOSSO algorithm	Hybrid filter/wrapper, emerging aggregate filter & SVM	Two minimization objectives: attributes number and error of classification
K10	MOFSRank	Wrapper approach, linear SVM	Two minimization objectives: attributes number and error of classification
K11	MOGWO algorithm	Wrapper approach, SVM	Two minimization objectives: attributes number and error of classification
K12	MODE algorithm	Wrapper approach, Linear SVM	Two minimization objectives: attributes number and error of classification
K13	FAEMODE algorithm	Filter approach, mRMR, MIFS, NMIFS, MIFS	Two minimization objectives: attributes number and error of classification
K14	I-NSGA-III algorithm	Wrapper approach, GHSOM classifier	Two minimization objectives: attributes number and error of classification
K15	MOPSO algorithm	Wrapper approach, random forests classifier	Two minimization objectives: attributes number and error of classification
K16	ENORA algorithm	Wrapper approach, regression model learner Random Forest.	Two minimization objectives: attributes number and error of classification
K17	NSGA-II algorithm MOPSO algorithm	Wrapper approach, Multi-Layer Perceptron classifier	Two minimization objectives: attributes number and error of classification
K18	MOGA algorithm	Wrapper approach, LR, SVM, ELM, K-means, and AP classifiers.	Two minimization objectives: attributes number and error of classification
K19	MOPSO algorithm	Wrapper approach, ML-KNN classifier	Two objectives: 1) reduce the number of features 2) reduce the Hamming loss value.
K20	MOEA/D algorithm	Wrapper approach, KNN and SVM classifiers	Two minimization objectives: attributes number and error of classification
K21	MOPSO algorithm	Wrapper approach, 1-nearest neighbour method	Two objectives: 1) The reliability 2) classification accuracy.
K22	NSGA-II algorithm	Wrapper approach, ID3 method	Two minimization objectives: attributes number and error of classification
K23	MECY-FS algorithm	Filter approach, MI	Two minimization objectives: error rate of classification and Clustering quality
K24	MO-MIFS algorithm NSGA-II algorithm	Wrapper approach	Two objectives: 1) maximal relevance 2) minimal redundancy
K25	NSGA-II algorithm	Wrapper approach, k-NN, SVM and NB	Two minimization objectives: attributes number and error of classification
K26	MOEA algorithm	Wrapper approach, self-organized map (SOM)	Two minimization objectives: attributes number and error of classification
K27	MOUFSA algorithm	Wrapper approach, k-NN, k-means	Two minimization objectives: attributes number and error of classification
K28	NSGA-II algorithm	Wrapper approach, SOM, Growing Hierarchical SOM (GHSOM)	Two minimization objectives: attributes number and error of classification
K29	MmGA algorithm	Wrapper approach, MLP and RBF classifiers.	Two minimization objectives: attributes number and error of classification
K30	MOGA algorithm	Wrapper approach, KNN classifier	Two minimization objectives: attributes number and error of classification
K31	MOGA (NSGA-II) algorithm	Wrapper approach, SVM classifier	Two objectives:1) reduce the number of features 2) maximize classification accuracy (specificity, and sensitivity).
K32	NSGAII and SPEA2 Algorithms	Filter approach, mutual information and entropy	Two minimization objectives: attributes number and error of classification

TABLE 3. *(Continued.)* **Research questions analysis of main selected research studies.**

K33	NSBPSO and CMDBPSO algorithms	Filter approach, mutual information and entropy	Two minimization objectives: attributes number and error of classification
K34	MOPSO algorithm	Wrapper approach, linear forward selection (LFS) and greedy stepwise backward selection (GSBS)	Two minimization objectives: attributes number and error of classification
K35	Stack auto-encoders: SAFS	Wrapper Approach, Random forests and LASSO	Two minimization objectives: attributes number and feature reconstruction error
K36	Deep Boltzmann Machine	Wrapper Approach, Deep Boltzmann Machine	Two minimization objectives: attributes number and feature reconstruction error
K37	Deep Belief Network	Wrapper Approach, Deep Belief Network	Two minimization objectives: attributes number and feature reconstruction error
K38	Deep Neural Network	Wrapper Approach, Deep Neural Network, feed-forward back-propagation neural network	Two minimization objectives: loss function and objective function

with the previously mentioned three research questions. Hence, few studies concentrated on assessing new methods for multi-objective feature selection. Others try to modify popular multi-objective techniques and compare their outcomes. Some applied the deep learning approaches. Besides, most of the studies utilized wrapper methods and few use filter methods. Table 3 illustrates the 38 selected papers trying to answer the research questions.

B. LANGUAGES AND DATASETS

All the selected papers on multi-objective feature selection are in the English language (K1-K38). Also, most of these works (16 out of 38 use standard benchmark dataset collected from the UCI repository [87]. Some use Medical datasets (K9, K 17, k31), some use Network intrusion datasets (K14, K28). The other selected studies (10 out of 38) utilize a variety of datasets from distinct fields that are publicly accessible, as demonstrated in Table 4.

VI. MULTI-OBJECTIVE FEATURE SELECTION

Multi-objective optimization is a problem with a sort of solution that can be evaluated against two or more unmatched or conflicting goals. Contrary the conventional optimization problems, the outcomes of multi-objective problems are not only one best solution, but a set of solutions, since for each member of the set, no one solution is fully better 'Pareto-set'. Thus, with two competing objectives, increasing accuracy efficiency and reducing the number of attributes, feature selection can be handled as a multi-objective problem.

In this research and based on systematic literature review, two main conflicting objectives of feature selection problem are noticed:

- reducing the required number of relevant attributes.
- reducing the error rate of classification.

Hence, the minimization problem of feature multi-objective selection can be represented mathematically as denoted by equation (5):

$$minimize\ F(x) = [f_1(x), f_2(x)] \tag{5}$$

where $f_1(x)$, denoted by equation (6), refers to minimization of the required number of relevant features as follows:

$$f_1(x) = \frac{\#Features}{AllFeatures} \#Features \in AllFeatures, \in \mathbb{R}^+ \tag{6}$$

where, #Features is the selected features number. #AllFeatures specifies the whole features in the dataset.

$f_2(x)$, denoted by equation (7), refers to the minimization of the classification error rate:

$$f_2(x) = ErrorRate = \frac{FP + FN}{TP + TN + FP + FN} X100, \quad (P + N) \in \mathbb{R}^+ \tag{7}$$

where, TP refers to true positives, TN represents true negatives, FP stands for false positives, and FN indicates false negatives, respectively.

The real problem has multiple objective functions, and different single optimum solutions can be found in each objective function. The optimum solution consistent to different goals is different due to the objective functions usually conflict (competition) with each other. In terms of all objective functions, no single solution can be better than any other solution. Given the non-domination of two goals, the non-dominant solution is when none of the two targets is better than the other. These two goals are equally important. For example, the number of attributes and precision.

By directly observing the evolutionary process, the multi-objective mechanism is more suitable for feature selection tasks than the single target mechanism. The single-objective algorithm maintains only one alternative to guide the search, which is more probable to drop to the optimum local area [60]. The multi-objective mechanism preserves the non-dominated alternatives generated during the search phase that are used as prospective leaders in search techniques guidance and discovering better solutions [60].

VII. MEASURES IN WRAPPER METHODS

The purpose of this section is to explore the typical wrapper measurements used in multi-objective techniques for feature selection, as most of the selected studies in Section 3 use the wrapper approach. Evaluation criteria are the only way to differentiate between the wrapper and filter models. For subset, evaluation wrappers utilize a classifier algorithm, an optimal

TABLE 4. Summary of datasets used for evaluation of multi-objective feature selection methods.

Dataset	Description	Key
Standard benchmark dataset from UCI	Popular datasets collected from UCI Machine Learning Repository.	K1, K5, K6, K7, K13, K18, K20, K21, K22, K23, K25, K27, K29, K32, K33, K34
Microarray cancer datasets	Microarrays are being used to help diagnose diseases, such as cancer.	K4, K8
Other benchmark datasets (Medical datasets)	Standard datasets collected from http://www.gems-system., Warfarin dataset and miRNA expression dataset	K9, K17, K31
LETOR data collections	standard datasets in the problem of learning to rank.	K10
Cervix Lesion dataset	CECT pictures of 62 patients inspected at Eisha Diagnostics Centre, Agra.	K11
Facial Expression Recognition datasets	Datasets such as Cohn Kanade, JAFFE and MMI databases	K12
NSL-KDD datasets	Network intrusion datasets	K14, K28
DGA dataset	Collected from different transformer ranges for Fault diagnosis of transformers	K15
Online Sales dataset	Taken from the Kaggle community.	K2, K16
Mulan: A Java Library for Multi-Label Learning	Datasets used are: bioinformatics,image processing, music emotion.	K19
Friedman dataset	Datasets utilized to test the algorithm of feature selection for selecting appropriate features.	K24
Real datasets	USPS handwritten digit database, ORL, COIL20, Ozone-onehr, Ozone-eighthr contains ground-level ozone data.	K25
BCI (Brain-Computer Interface) benchmarks.	Synthetic benchmark created from the dataset 2D motion provided at EEG motor activity data set	K3, K26
Face images Datasets	Essex Face Database	K30
Cardio-vascular Disease Dataset	Detroit Receiving Hospital	K35
Image Benchmark Datasets	MNIST, MIR-Flickr, GISETTE, MADELO, and PANCAN	K36
Remote Sensing Images	RSSCN7	K37
Real datasets	Robotics (Humanoid) Datasets	K38

subset that best fits the classifier algorithm will be selected. Thus, the wrapper method performance is generally better. The filter selection is independent of the capabilities of any particular classifier. Nevertheless, the main weakness of the filter method is that it completely ignores the influence of the selected subset on the inductive algorithm performance [88]. The optimal feature subset should depend on the specific bias and heuristic of the inductive algorithm. According to this assumption, the wrapper method utilizes a particular classifier to assess the superiority of the subset selected and provides an easy and efficient way to solve the feature selection problem, regardless of which classifier is used [89]. Three stages wrapper method performs for evaluating feature subset:

- Stage 1: searching for feature subset,
- Stage 2: assessing the feature subset selected by the classifier performance,
- Stage 3: redo 1 and 2 until the wanted result is reached.

Classification is the way toward constructing a model that depicts classes or concepts of data and is used to predict object classes whose class labels are unknown [90]. In the two steps of the classification process, the model must first be built using the training data of the class tag, and then the model is evaluated in the test dataset by assigning the class tag to the data object. Consistent with [91], classification is regarded as a supervised learning instance, that is, learning the case of a training set that can obtain a correctly identified observation. The unsupervised learning is recognized as clustering which comprises grouping data into classes based on a certain measure of distance or intrinsic similarity. As shown in Figure 5, classifiers are categorized to the following: 1) generative classifiers, 2) discriminatory classifiers 3) and classifiers based on regression. Table 5 illustrated Most common wrapper approach classifiers.

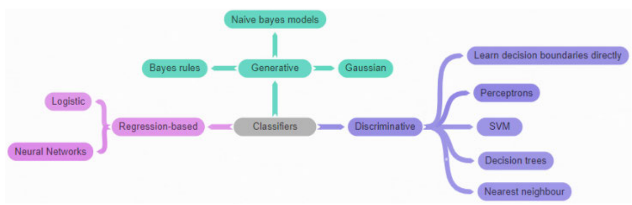

FIGURE 5. Types of classifiers.

The classifier's efficiency depends highly on the characteristics of the information to be categorized. No classifier can fix any problem in the best way. Numerous experimental studies were carried out to compare the classifier's performance and to discover characteristics that show the output of the classifier. Nevertheless, it is still an art, not a science, to find the correct classifier for a specified issue.

VIII. APPLICATIONS

Multi-objective feature selection methods were applied to a multitude of fields. Generally, it is possible to group the main applications into four groups.

1) Benchmark problems: usually a dataset available in UCI repository machine learning as in K1, K5, K6, K7, K13, K18, K20, K21, K22, K23, K25, K27, K29, K32, K33, K34, K35.
2) Image processing: such as brain-computer interface (EEG), face images, CECT images, handwritten digit recognition, remote sensing images, and music emotion as in K3, K11, K12, K19, K26, K30, K36 and K37.

TABLE 5. Measures in wrapper approaches.

Classifier	Description	Key	Performance
SVM	This classifier builds hyperplanes in a large- scale dimensional space and could be utilized for supervised learning (regression or classification) [93]. Naturally, a great partition is accomplished by the hyperplane that has the biggest separation to the closest training-data point of any margin, generally the bigger the margin, the reduced the classifier's generalization error [94].	K4, K5, K8, K9, K10, K11, K12, K20, K25, K31	SVM shows good performance in term of classification accuracy, however, it's computationally expensive.
K-NN	This classifier is one of the finest algorithms of supervised learning [95]. The concept is to scan in function space for the nearest match of the test information.	K1, K3, K4, K7, K8, K20, K21, K25, K30	shows the best performance when dealing with classifying of the selected features classification compared to SVM especially in term of computational cost.
Naive Bayes (NB)	Naive Bayes is a likelihood classifier enlivened by Bayes' hypothesis under a straightforward presumption, the attributes are restrictively independent [96]. Naive Bayes is a basic algorithm that yields great outcomes in most cases.	K3, K4, K8, K25	NB works well on small datasets, however, when classifying large dataset the performance is decreased
Decision Tree (DT)	Based on the procedure of a tree's structure, the decision tree constructs a model of classification or regression [97]. It uses the if-then instruction set, which for classification is mutually exclusive and exhaustive.	K4, K8	DT Rarely used due to its performance is not as good as previously mentioned classifiers, especially for large datasets.
Random Forest (RF)	This algorithm is one of the finest algorithms of classification for machine learning and is capable of highly accurate classification of enormous amounts of data [98].	K15, K16, K35	It is not a cure for small instances datasets. If the instances of the data are small (20), then each of the samples taken with replacement from this data would consist of not more than the twenty-four distinct values.

3) Biomedical issues: such as gene analysis, biomarker detection, and disease diagnosis, as in K4, K8, K9, K17, K31.
4) Financial problems, such as predicting customer churn as in K16.
5) Network intrusion and network security as in K14, K28.
6) Robotics and the embedded system as in K38

All these fields are significant and vital to our daily lives and society. Several areas, such as language learning, oil and gas, and complex engineering tasks, also require multi-objective feature selection, yet these methods have not been thoroughly studied in these fields.

IX. DISCUSSION AND FUTURE WORKS

According to the studied literature, the vast majority of the studies used wrapper method, where only (5 out of 38) used filter method, this is due to the superiority of wrapper methods in terms of high performance compared to filter methods which are experimentally proven to be not true in all cases. Some studies tried to hybridize both filter and wrapper methods to take the benefit of both approaches. Figure 6 demonstrates the utilization of feature selection methods over the years.

Regarding the search algorithms, various types of multi-objective algorithms have been utilized, where (11 out of 38) studies used MOGA and its variants and (7 out of 38) used MOPSO and its variants, where the rest used deep learning, DE, GWO, TLBO, ABC…etc. Table 6 shows the pros and cons of the most popular multi-objective search algorithms.

Moreover, the aim of most of the studies in the field of selection of features based multi-objective algorithms was to reduce both the number of attributes and the error rate of accuracy. Nevertheless, this is a challenging task due to the drawbacks of most of the proposed algorithms such as trapping in local optima and dropping of accuracy. on large scale (datasets) problems. Furthermore, most of the studies considered supervised learning (classification), where only (3 out of 38) studies considered unsupervised learning (clustering). Added to this, there are trends have taken place between 2012 and 2019. Figure 7 demonstrates the different types of fields in terms of the total number of studies from each year. The years showed that the multi-objective feature selection techniques have been utilized in various areas such as image processing, biomedical tasks, economic problem, the intrusion of the network and others.

Overall, previous research has produced many enhancements and changes, and in brief, all distinct research goals

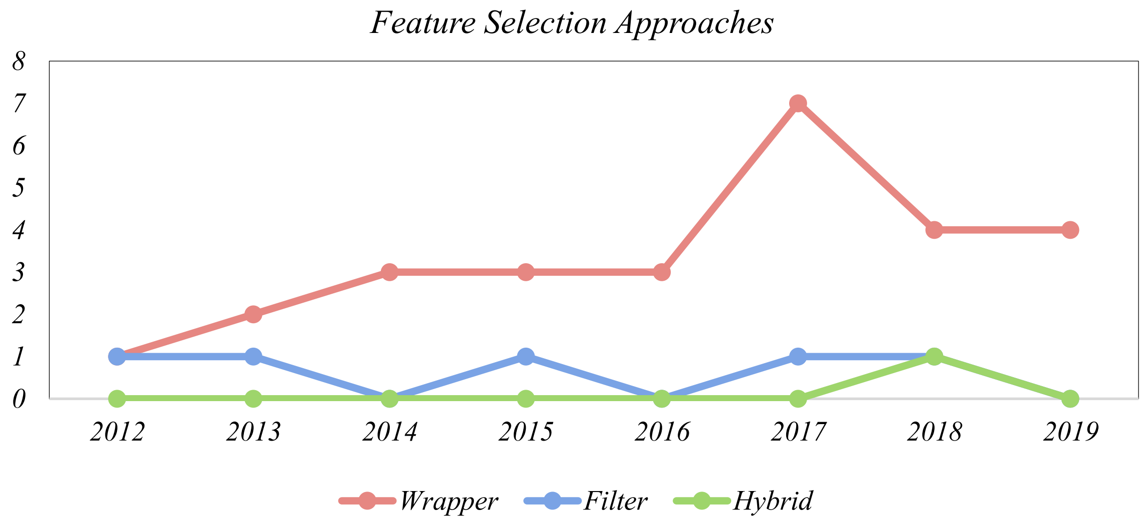

FIGURE 6. Feature selection methods used from the year 2012-2019.

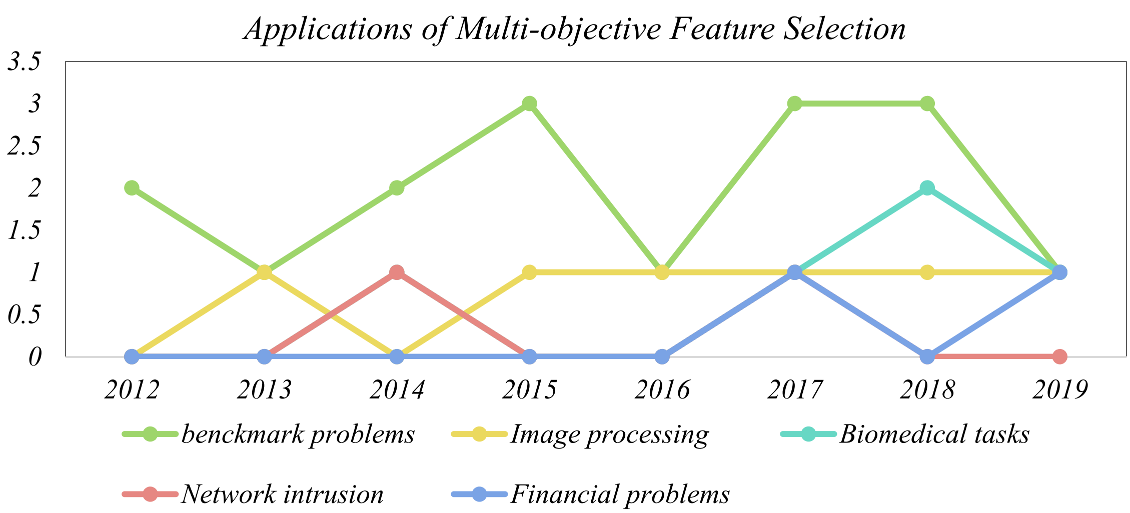

FIGURE 7. Applications used by multi-objective feature selections from the year 2012-2019.

require distinct techniques to accomplish the required outcomes. The solution to every problem cannot be restricted to a single technique and therefore the solution design will grow continually. As multi-objective feature selection is widely used to tackle real-world problems, it is recommended to explore the following issues of multi-objective feature selection:

- The potential of the recently proposed approaches has not been fully investigated, especially in term of scalability, since in many real-world problems both feature attributes and instances are growing, further investigation is recommended.
- Computational cost is one of the primary problems in the multi-objective selection of features, it is recommended to propose an effective measure to decrease computational costs. To do so, two primary factors must be considered: 1) an effective search method 2) rapid assessment measure.
- In addition to the two primary goals of multi-objective feature selection problem, other objectives, such as computational complexity and scalability, can also be considered.
- Few studies have utilized dynamic multi-objective to deal with the problem of feature selection. It's recommended using dynamic multi-objective for the task of feature selection.
- Proposing novel evaluation measures that could soften the fitness landscape will considerably decrease the challenges of the problem and aid in developing an appropriate search technique.
- Applying recent developed multi-objective optimization algorithm to solve the problem of feature selection

TABLE 6. **Multi-objective search techniques pros and cons.**

Technique	Pros	Cons	Key
MOGA + variants	• A simple extension of single-objective GA. • Exchange Information (Crossover or Mutation). • Attained high performance on issues contains hundreds of features. • They usually perform better than traditional and single objective feature selection techniques.	• Slow convergence that prevents from finding the optimum Pareto front. • High computational cost. • With thousands of features, it shows limited success. • The parameters such as mutation can significantly influence the search. • Does not perform well on problems with two or more objectives.	K2, K3, K14, K17, K22, K24, K25, K28, K29, K30, K32.
MOPSO + variants	• MOPSO better than MOGAs in terms of the simplicity of implementation. • Attained high performance on issues contains hundreds of features. • Faster convergence compares with many other Multi-objective techniques • They usually perform better than traditional and single objective feature selection techniques.	• Similar to MOGA shows a limited success with thousands of features. • Converge to global Pareto set in high-dimensional space. • Has a low convergence rate. • Some subsets with good accuracy, but a small number of features could be ignored due to the updating mechanisms of global and personal best of MOPSO.	K6, K15, K19, K21, K33, K34
MODE + variants	• The search can be done randomly. • Has few numbers of parameters. • Attained good performance and appropriate to problems with high-dimensional. • They usually perform better than traditional and single objective feature selection techniques.	• Unbalanced convergence in the last period. • Easy to drop into regional optimum. • Converge to global Pareto set in high-dimensional space. • Run-time complexity.	K1, K12, K13
MOGWO	• One of the recent multi-objective methods, it has a few parameters. • Easy to implement. • Capable of preventing local optima by preserving a balance between a exploitation and exploration. • Needs small size of memory, because it has only a position vector, whereas MOPSO has two vectors velocity and position.	• The low capability to handle multi-modal search landscape. • The three alpha, beta and gamma wolves tend to converge to the same solution. • Alpha, beta, and delta might get trapped in locally optimal solutions.	K11
Deep Learning	• Utilizes a basic auto-encoder to achieve feature selection based on the error of reconstruction. • The basic network is utilized to reconstructing the original data. • non-linear structures modelling of complex systems. • Takes benefits of the structures of deep to model non-linearity	• The basic network may not be able to model large scale data which in turn leads to higher error of reconstruction and redundant or irrelevant feature selection. • They require more data. • They are not used as general-purpose dimensionality reduction algorithms.	K35, K36, K37, K38

such as Harris hawks optimization [98], Dragon algorithm [99], Grasshopper optimization algorithm [100], etc.

X. CONCLUSION

Throughout the years, multi-objective feature selection has gained a high consideration from scholars studying the areas of machine intelligence and information mining. Nevertheless, reliable with (NFL) hypothesis, there was not and will never be an optimization technique to tackle all problems. To help researchers in their endeavours, we attempted a systematic literature review considering the studies published in the 2012-2019 era, to highlight the main multi-objective feature selection challenges and techniques, which secured all the ordinarily utilized multi-objective algorithms and concentrated on the key components, for example, mechanism of search, evaluation measures and the number of objectives of the multi-objective idea as well as the applications. According to this survey, great efforts have been made to improve the efficiency of multi-objective feature selection in terms of accuracy and number of attributes, paving the way for further improvements in the future. Finally, as there is still space for improvements, the multi-objective feature selection could be expanded into numerous hybridizations and modifications depending on the requirements of the problem. Consequently, other potential researches working on multi-objective optimization paradigms could, therefore, use the outcomes of this research to further explore the efficient techniques for addressing recent challenges in multi-objective feature selection.

REFERENCES

[1] M. Dash and H. Liu, "Feature selection for classification," *Intell. Data Anal.*, vol. 1, nos. 1–4, pp. 131–156, 1997.

[2] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, Jan. 2003.
[3] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics Intell. Lab. Syst.*, vol. 2, nos. 1–3, pp. 37–52, 1987.
[4] S. Balakrishnama and A. Ganapathiraju, "Linear discriminant analysis-a brief tutorial," *Inst. Signal Inf. Process.*, vol. 18, pp. 1–8, Mar. 1998.
[5] H. Liu and H. Motoda, *Feature Selection for Knowledge Discovery and Data Mining*, vol. 454. Springer, 2012.
[6] A. L. Blum and P. Langley, "Selection of relevant features and examples in machine learning," *Artif. Intell.*, vol. 97, nos. 1–2, pp. 245–271, Dec. 1997.
[7] P. Pudil, J. Novovi ová, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognit. Lett.*, vol. 15, no. 11, pp. 1119–1125, Nov. 1994.
[8] Q. Mao and I. W.-H. Tsang, "A feature selection method for multivariate performance measures," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 9, pp. 2051–2063, Sep. 2013.
[9] F. Min, Q. Hu, and W. Zhu, "Feature selection with test cost constraint," *Int. J. Approx. Reasoning*, vol. 55, no. 1, pp. 167–179, Jan. 2014.
[10] A. Özçift and A. Gülten, "Genetic algorithm wrapped Bayesian network feature selection applied to differential diagnosis of erythemato-squamous diseases," *Digit. Signal Process.*, vol. 23, no. 1, pp. 230–237, Jan. 2013, doi: 10.1016/j.dsp.2012.07.008.
[11] S. Oreski and G. Oreski, "Genetic algorithm-based heuristic for feature selection in credit risk assessment," *Expert Syst. Appl.*, vol. 41, no. 4, pp. 2052–2064, Mar. 2014, doi: 10.1016/j.eswa.2013.09.004.
[12] H. M. Harb and A. S. Desuky, "Feature selection on classification of medical datasets based on particle swarm optimization," *Int. J. Comput. Appl.*, vol. 104, no. 5, pp. 8887–8975, 2014, doi: 10.5120/18197-9118.
[13] H. H. Inbarani, A. T. Azar, and G. Jothi, "Supervised hybrid feature selection based on PSO and rough sets for medical diagnosis," *Comput. Methods Programs Biomed.*, vol. 113, no. 1, pp. 175–185, Jan. 2014, doi: 10.1016/j.cmpb.2013.10.007.
[14] S. Kashef and H. Nezamabadi-pour, "An advanced ACO algorithm for feature subset selection," *Neurocomputing*, vol. 147, pp. 271–279, Jan. 2015.
[15] H. Rais and T. Mehmood, "Dynamic ant colony system with three level update feature selection for intrusion detection," *IJ Netw. Secur.*, vol. 20, no. 1, pp. 184–192, 2018, doi: 10.6633/IJNS.201801.20(1).20.
[16] E. Emary, H. M. Zawbaa, and A. E. Hassanien, "Binary grey wolf optimization approaches for feature selection," *Neurocomputing*, vol. 172, pp. 371–381, Jan. 2016.
[17] Q. Al-Tashi, H. M. Rais, S. J. Abdulkadir, S. Mirjalili, and H. Alhussian, "A review of grey wolf optimizer-based feature selection methods for classification," in *Evolutionary Machine Learning Techniques*. Singapore: Springer, 2020, pp. 273–286.
[18] J. Too, A. Abdullah, N. M. Saad, N. M. Ali, and W. Tee, "A new competitive binary grey wolf optimizer to solve the feature selection problem in EMG signals classification," *Computers*, vol. 7, no. 4, p. 58, Nov. 2018.
[19] Q. Al-Tashi, S. J. A. Kadir, H. M. Rais, S. Mirjalili, and H. Alhussian, "Binary optimization using hybrid grey wolf optimization for feature selection," *IEEE Access*, vol. 7, pp. 39496–39508, 2019.
[20] D. M. Pierre and N. Zakaria, "Stochastic partially optimized cyclic shift crossover for multi-objective genetic algorithms for the vehicle routing problem with time-windows," *Appl. Soft Comput.*, vol. 52, pp. 863–876, Mar. 2017.
[21] M. H. Zangooei, J. Habibi, and R. Alizadehsani, "Disease diagnosis with a hybrid method SVR using NSGA-II," *Neurocomputing*, vol. 136, pp. 14–29, Jul. 2014, doi: 10.1016/j.neucom.2014.01.042.
[22] M. K. Sohrabi and A. Tajik, "Multi-objective feature selection for Warfarin dose prediction," *Comput. Biol. Chem.*, vol. 69, pp. 126–133, Aug. 2017, doi: 10.1016/j.compbiolchem.2017.06.002.
[23] A. Sahoo and S. Chandra, "Multi-objective grey wolf optimizer for improved cervix lesion classification," *Appl. Soft Comput.*, vol. 52, pp. 64–80, Mar. 2017, doi: 10.1016/j.asoc.2016.12.022.
[24] Q. Al-Tashi, S. J. Abdulkadir, H. M. Rais, S. Mirjalili, H. Alhussian, M. G. Ragab, and A. Alqushaibi, "Binary multi-objective grey wolf optimizer for feature selection in classification," *IEEE Access*, vol. 8, pp. 106247–106263, 2020.
[25] H. Liu and Z. Zhao, "Manipulating data and dimension reduction methods: Feature selection," in *Computational Complexity: Theory, Techniques, and Applications*. New York, NY, USA: Springer, 2012, pp. 1790–1800.
[26] H. Liu, H. Motoda, R. Setiono, and Z. Zhao, "Feature selection: An ever evolving frontier in data mining," in *Proc. Feature Selection Data Mining*, 2010, pp. 4–13.
[27] L. Thiele, K. Miettinen, P. J. Korhonen, and J. Molina, "A preference-based evolutionary algorithm for multi-objective optimization," *Evol. Comput.*, vol. 17, no. 3, pp. 411–436, Sep. 2009.
[28] A. Abraham and L. Jain, "Evolutionary multiobjective optimization," in *Evolutionary Multiobjective Optimization*. London, U.K.: Springer, 2005, pp. 1–6.
[29] L. T. Bui, Ed., *Multi-Objective Optimization in Computational Intelligence: Theory and Practice*. Hershey, PA, USA: IGI Global. 2008.
[30] Y. Zhai, Y.-S. Ong, and I. W. Tsang, "The emerging' big dimensionality,'" *IEEE Comput. Intell. Mag.*, vol. 9, no. 3, pp. 14–26, Aug. 2014.
[31] W. Siedlecki and J. Sklansky, "A note on genetic algorithms for large-scale feature selection," in *Handbook of Pattern Recognition and Computer Vision*. Singapore: World Scientific, 1993, pp. 88–107.
[32] D. Moher, A. Liberati, J. Tetzlaff, D. G. Altman, and G. Prisma, "Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement," *Ann. Internal Med.*, vol. 151, no. 4, pp. 264–269, 2009.
[33] Y. Zhang, D.-W. Gong, X.-Z. Gao, T. Tian, and X.-Y. Sun, "Binary differential evolution with self-learning for multi-objective feature selection," *Inf. Sci.*, vol. 507, pp. 67–85, Jan. 2020.
[34] N. Kozodoi, S. Lessmann, K. Papakonstantinou, Y. Gatsoulis, and B. Baesens, "A multi-objective approach for profit-driven feature selection in credit scoring," *Decis. Support Syst.*, vol. 120, pp. 106–117, May 2019, doi: 10.1016/j.dss.2019.03.011.
[35] J. González, J. Ortega, M. Damas, P. Martín-Smith, and J. Q. Gan, "A new multi-objective wrapper method for feature selection—Accuracy and stability analysis for BCI," *Neurocomputing*, vol. 333, pp. 407–418, Mar. 2019, doi: 10.1016/j.neucom.2019.01.017.
[36] A. Sharma and R. Rani, "C-HMOSHSSA: Gene selection for cancer classification using multi-objective meta-heuristic and machine learning methods," *Comput. Methods Programs Biomed.*, vol. 178, pp. 219–235, Sep. 2019, doi: 10.1016/j.cmpb.2019.06.029.
[37] H. E. Kiziloz, A. Deniz, T. Dokeroglu, and A. Cosar, "Novel multiobjective TLBO algorithms for the feature subset selection problem," *Neurocomputing*, vol. 306, pp. 94–107, Sep. 2018, doi: 10.1016/j.neucom.2018.04.020.
[38] M. Amoozegar and B. Minaei-Bidgoli, "Optimizing multi-objective PSO based feature selection method using a feature elitism mechanism," *Expert Syst. Appl.*, vol. 113, pp. 499–514, Dec. 2018, doi: 10.1016/j.eswa.2018.07.013.
[39] E. Hancer, B. Xue, M. Zhang, D. Karaboga, and B. Akay, "Pareto front feature selection based on artificial bee colony optimization," *Inf. Sci.*, vol. 422, pp. 462–479, Jan. 2018, doi: 10.1016/j.ins.2017.09.028.
[40] M. Dashtban, M. Balafar, and P. Suravajhala, "Gene selection for tumor classification using a novel bio-inspired multi-objective approach," *Genomics*, vol. 110, no. 1, pp. 10–17, Jan. 2018, doi: 10.1016/j.ygeno.2017.07.010.
[41] C.-M. Lai, "Multi-objective simplified swarm optimization with weighting scheme for gene selection," *Appl. Soft Comput.*, vol. 65, pp. 58–68, Apr. 2018, doi: 10.1016/j.asoc.2017.12.049.
[42] F. Cheng, W. Guo, and X. Zhang, "MOFSRank: A multiobjective evolutionary algorithm for feature selection in learning to rank," *Complexity*, vol. 2018, pp. 1–14, Dec. 2018.
[43] U. Mlakar, I. Fister, J. Brest, and B. Poto nik, "Multi-objective differential evolution for feature selection in facial expression recognition systems," *Expert Syst. Appl.*, vol. 89, pp. 129–137, Dec. 2017, doi: 10.1016/j.eswa.2017.07.037.
[44] S. K. Nayak, P. K. Rout, A. K. Jagadev, and T. Swarnkar, "Elitism based multi-objective differential evolution for feature selection: A filter approach with an efficient redundancy measure," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 32, no. 2, pp. 174–187, Feb. 2020, doi: 10.1016/j.jksuci.2017.08.001.
[45] Y. Zhu, J. Liang, J. Chen, and Z. Ming, "An improved NSGA-III algorithm for feature selection used in intrusion detection," *Knowl.-Based Syst.*, vol. 116, pp. 74–85, Jan. 2017, doi: 10.1016/j.knosys.2016.10.030.
[46] A. Peimankar, S. J. Weddell, T. Jalal, and A. C. Lapthorn, "Evolutionary multi-objective fault diagnosis of power transformers," *Swarm Evol. Comput.*, vol. 36, pp. 62–75, Oct. 2017, doi: 10.1016/j.swevo.2017.03.005.

[47] F. Jiménez, G. Sánchez, J. M. García, G. Sciavicco, and L. Miralles, "Multi-objective evolutionary feature selection for online sales forecasting," *Neurocomputing*, vol. 234, pp. 75–92, Apr. 2017, doi: 10.1016/j.neucom.2016.12.045.

[48] A. Deniz, H. E. Kiziloz, T. Dokeroglu, and A. Cosar, "Robust multiobjective evolutionary feature subset selection algorithm for binary classification using machine learning techniques," *Neurocomputing*, vol. 241, pp. 128–146, Jun. 2017, doi: 10.1016/j.neucom.2017.02.033.

[49] Y. Zhang, D.-W. Gong, X.-Y. Sun, and Y.-N. Guo, "A PSO-based multi-objective multi-label feature selection method in classification," *Sci. Rep.*, vol. 7, no. 1, pp. 1–12, Dec. 2017, doi: 10.1038/s41598-017-00416-0.

[50] A. Das and S. Das, "Feature weighting and selection with a Pareto-optimal trade-off between relevancy and redundancy," *Pattern Recognit. Lett.*, vol. 88, pp. 12–19, Mar. 2017.

[51] Z. Yong, G. Dun-wei, and Z. Wan-qiu, "Feature selection of unreliable data using an improved multi-objective PSO algorithm," *Neurocomputing*, vol. 171, pp. 1281–1290, Jan. 2016, doi: 10.1016/j.neucom.2015.07.057.

[52] A. Khan and A. R. Baig, "Multi-objective feature subset selection using non-dominated sorting genetic algorithm," *J. Appl. Res. Technol.*, vol. 13, no. 1, pp. 145–159, 2015, doi: 10.1016/S1665-6423(15)30013-4.

[53] Z. Wang, M. Li, and J. Li, "A multi-objective evolutionary algorithm for feature selection based on mutual information with a new redundancy measure," *Inf. Sci.*, vol. 307, pp. 73–88, Jun. 2015, doi: 10.1016/j.ins.2015.02.031.

[54] M. Han and W. Ren, "Global mutual information-based feature selection approach using single-objective and multi-objective optimization," *Neurocomputing*, vol. 168, pp. 47–54, Nov. 2015, doi: 10.1016/j.neucom.2015.06.016.

[55] P. P. Kundu and S. Mitra, "Multi-objective optimization of shared nearest neighbor similarity for feature selection," *Appl. Soft Comput.*, vol. 37, pp. 751–762, Dec. 2015, doi: 10.1016/j.asoc.2015.08.042.

[56] D. Kimovski, J. Ortega, A. Ortiz, and R. Baños, "Parallel alternatives for evolutionary multi-objective optimization in unsupervised feature selection," *Expert Syst. Appl.*, vol. 42, no. 9, pp. 4239–4252, Jun. 2015, doi: 10.1016/j.eswa.2015.01.061.

[57] H. Xia, J. Zhuang, and D. Yu, "Multi-objective unsupervised feature selection algorithm utilizing redundancy measure and negative epsilon-dominance for fault diagnosis," *Neurocomputing*, vol. 146, pp. 113–124, Dec. 2014, doi: 10.1016/j.neucom.2014.06.075.

[58] E. de la Hoz, E. de la Hoz, A. Ortiz, J. Ortega, and A. Martínez-Álvarez, "Feature selection by multi-objective optimisation: Application to network anomaly detection by hierarchical self-organising maps," *Knowl.-Based Syst.*, vol. 71, pp. 322–338, Nov. 2014, doi: 10.1016/j.knosys.2014.08.013.

[59] C. J. Tan, C. P. Lim, and Y. Cheah, "A multi-objective evolutionary algorithm-based ensemble optimizer for feature selection and classification with neural network models," *Neurocomputing*, vol. 125, pp. 217–228, Feb. 2014.

[60] L. D. Vignolo, D. H. Milone, and J. Scharcanski, "Feature selection for face recognition based on multi-objective evolutionary wrappers," *Expert Syst. Appl.*, vol. 40, no. 13, pp. 5077–5084, Oct. 2013, doi: 10.1016/j.eswa.2013.03.032.

[61] A. Mukhopadhyay and U. Maulik, "An SVM-wrapped multiobjective evolutionary feature selection approach for identifying cancer-MicroRNA markers," *IEEE Trans. Nanobiosci.*, vol. 12, no. 4, pp. 275–281, Dec. 2013.

[62] B. Xue, L. Cervante, L. Shang, W. N. Browne, and M. Zhang, "Multi-objective evolutionary algorithms for filter based feature selection in classification," *Int. J. Artif. Intell. Tools*, vol. 22, no. 4, Aug. 2013, Art. no. 1350024.

[63] B. Xue, L. Cervante, L. Shang, W. N. Browne, and M. Zhang, "A multi-objective particle swarm optimisation for filter-based feature selection in classification problems," *Connection Sci.*, vol. 24, nos. 2–3, pp. 91–116, Sep. 2012.

[64] B. Xue, M. J. Zhang, and W. N. Browne, "Particle swarm optimization for feature selection in classification: A multi-objective approach," *IEEE Trans. Cybern.*, vol. 43, no. 6, pp. 1656–1671, Dec. 2013, doi: 10.1109/TSMCB.2012.2227469.

[65] M. Z. Nezhad, D. Zhu, X. Li, K. Yang, and P. Levy, "SAFS: A deep feature selection approach for precision medicine," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2016, pp. 501–506.

[66] A. Taherkhani, G. Cosma, and T. M. McGinnity, "Deep-FS: A feature selection algorithm for deep Boltzmann machines," *Neurocomputing*, vol. 322, pp. 22–37, Dec. 2018.

[67] Q. Zou, L. Ni, T. Zhang, and Q. Wang, "Deep learning based feature selection for remote sensing scene classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 11, pp. 2321–2325, Nov. 2015.

[68] V. B. Semwal, K. Mondal, and G. C. Nandi, "Robust and accurate feature selection for humanoid push recovery and classification: Deep learning approach," *Neural Comput. Appl.*, vol. 28, no. 3, pp. 565–574, Mar. 2017.

[69] B. Xue, M. Zhang, and W. N. Browne, "Particle swarm optimisation for feature selection in classification: Novel initialisation and updating mechanisms," *Appl. Soft Comput.*, vol. 18, pp. 261–276, May 2014.

[70] L. Cervante, B. Xue, L. Shang, and M. Zhang, "A multi-objective feature selection approach based on binary PSO and rough set theory," in *Proc. Eur. Conf. Evol. Comput. Combinat. Optim.*, 2013, pp. 25–36.

[71] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.

[72] L. Yu and H. Liu, "Efficient feature selection via analysis of relevance and redundancy," *J. Mach. Learn. Res.*, vol. 5, no. 10, pp. 1205–1224, 2004.

[73] E. Amaldi and V. Kann, "On the approximability of minimizing nonzero variables or unsatisfied relations in linear systems," *Theor. Comput. Sci.*, vol. 209, nos. 1–2, pp. 237–260, Dec. 1998.

[74] Q. Al-Tashi, H. Rais, and S. J. Abdulkadir, "Hybrid swarm intelligence algorithms with ensemble machine learning for medical diagnosis," in *Proc. 4th Int. Conf. Comput. Inf. Sci. (ICCOINS)*, Aug. 2018, pp. 1–6.

[75] Y.-S. Jeong, K. S. Shin, and M. K. Jeong, "An evolutionary algorithm with the partial sequential forward floating search mutation for large-scale feature selection problems," *J. Oper. Res. Soc.*, vol. 66, no. 4, pp. 529–538, Apr. 2015.

[76] M. N. Le, Y. S. Ong, S. Menzel, Y. Jin, and B. Sendhoff, "Evolution by adapting surrogates," *Evol. Comput.*, vol. 21, no. 2, pp. 313–340, May 2013.

[77] P. A. Estévez, M. Tesmer, C. A. Perez, and J. M. Zurada, "Normalized mutual information feature selection," *IEEE Trans. Neural Netw.*, vol. 20, no. 2, pp. 189–201, Feb. 2009.

[78] J. R. Vergara and P. A. Estévez, "A review of feature selection methods based on mutual information," *Neural Comput. Appl.*, vol. 24, no. 1, pp. 175–186, Jan. 2014.

[79] A. Jakulin and I. Bratko, "Testing the significance of attribute interactions," in *Proc. 21st Int. Conf. Mach. Learn.*, 2004, p. 52.

[80] Y. Zhai, M. Tan, I. Tsang, and Y. Soon Ong, "Discovering support and affiliated features from very high dimensions," 2012, *arXiv:1206.6477*. [Online]. Available: http://arxiv.org/abs/1206.6477

[81] X. Wang, J. Yang, X. Teng, W. Xia, and R. Jensen, "Feature selection based on rough sets and particle swarm optimization," *Pattern Recognit. Lett.*, vol. 28, no. 4, pp. 459–471, Mar. 2007.

[82] C. Coello, S. De Computación, and C. Zacatenco, "Twenty years of evolutionary multi-objective optimization: A historical view of the field," *IEEE Comput. Intell. Mag.*, vol. 1, no. 1, pp. 28–36, Feb. 2006.

[83] K. Praditwong and X. Yao, "How well do multi-objective evolutionary algorithms scale to large problems," in *Proc. IEEE Congr. Evol. Comput.*, Sep. 2007, pp. 3959–3966.

[84] Z. Yang, K. Tang, and X. Yao, "Scalability of generalized adaptive differential evolution for large-scale continuous optimization," *Soft Comput.*, vol. 15, no. 11, pp. 2141–2155, Nov. 2011.

[85] S. M. Vieira, J. M. C. Sousa, and T. A. Runkler, "Multi-criteria ant feature selection using fuzzy classifiers," in *Swarm Intelligence for Multi-Objective Problems in Data Mining*. New York, NY, USA: Springer, 2009, pp. 19–36.

[86] Q. Al-Tashi, H. Rais, and S. Jadid, "Feature selection method based on grey wolf optimization for coronary artery disease classification," in *Proc. Int. Conf. Reliable Inf. Commun. Technol.*, 2018, pp. 257–266.

[87] C. Blake. (1998). *UCI Repository of Machine Learning Databases*. [Online]. Available: http://www.ics.uci.edu/~mlearn/MLRepository.html

[88] I. Inza, P. Larrañaga, R. Blanco, and A. J. Cerrolaza, "Filter versus wrapper gene selection approaches in DNA microarray domains," *Artif. Intell. Med.*, vol. 31, no. 2, pp. 91–103, Jun. 2004.

[89] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artif. Intell.*, vol. 97, nos. 1–2, pp. 273–324, Dec. 1997.

[90] T. C. Sharma and M. Jain, ''WEKA approach for comparative study of classification algorithm,'' *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 2, no. 4, pp. 1925–1931, 2013.
[91] Y. Kodratoff, *Introduction to Machine Learning*. San Mateo, CA, USA: Morgan Kaufmann, 2014.
[92] S. R. Gunn, ''Support vector machines for classification and regression,'' *ISIS Tech. Rep.*, vol. 14, no. 1, pp. 5–16, 1998.
[93] J. Franklin, ''The elements of statistical learning: Data mining, inference and prediction,'' *Math. Intelligencer*, vol. 27, no. 2, pp. 83–85, Mar. 2005.
[94] N. S. Altman, ''An introduction to kernel and nearest-neighbor nonparametric regression,'' *Amer. Statistician*, vol. 46, no. 3, pp. 175–185, Aug. 1992.
[95] D. D. Lewis, ''Naive (Bayes) at forty: The independence assumption in information retrieval,'' in *Proc. Eur. Conf. Mach. Learn.*, 1998, pp. 4–15.
[96] M. N. Anyanwu and S. G. Shiva, ''Comparative analysis of serial decision tree classification algorithms,'' *Int. J. Comput. Sci. Secur.*, vol. 3, no. 3, pp. 230–240, 2009.
[97] A. Liaw and M. Wiener, ''Classification and regression by randomforest,'' *R News*, vol. 2, no. 3, pp. 18–22, 2002.
[98] A. A. Heidari, S. Mirjalili, H. Faris, I. Aljarah, M. Mafarja, and H. Chen, ''Harris hawks optimization: Algorithm and applications,'' *Future Gener. Comput. Syst.*, vol. 97, pp. 849–872, Aug. 2019.
[99] S. Mirjalili, ''Dragonfly algorithm: A new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems,'' *Neural Comput. Appl.*, vol. 27, no. 4, pp. 1053–1073, 2016.
[100] S. Saremi, S. Mirjalili, and A. Lewis, ''Grasshopper optimisation algorithm: Theory and application,'' *Adv. Eng. Softw.*, vol. 105, pp. 30–47, Mar. 2017.

QASEM AL-TASHI received the B.Sc. degree in software engineering from Universiti Teknologi Malaysia, in 2012, and the M.Sc. degree in software engineering from Universiti Kebangsaan Malaysia, in 2017. He is currently pursuing the Ph.D. degree with Universiti Teknologi PETRONAS. He is also an Academic Staff with Albaydha University, Yemen. His research interests include machine learning, multiobjective optimization, feature selection and classification, and data analytics. He is currently serving as a Journal Reviewer for *Artificial Intelligence Review*, IEEE Access, and *Knowledge-Based Systems*.

SAID JADID ABDULKADIR (Senior Member, IEEE) received the B.Sc. degree in computer science from Moi University, the M.Sc. degree in computer science from Universiti Teknologi Malaysia, and the Ph.D. degree in information technology from Universiti Teknologi PETRONAS. He is currently a Senior Lecturer with the Department of Computer and Information Sciences, Universiti Teknologi PETRONAS. His current research interests include supervised machine learning and predictive and streaming analytics. He is currently serving as a Journal Reviewer for *Artificial Intelligence Review*, IEEE Access, and *Knowledge-Based Systems*.

HELMI MD RAIS received the B.Sc. degree in business administration from Drexel University, USA, the M.Sc. degree in information technology from Griffith University, Australia, and the Ph.D. degree in science and system management from Universiti Kebangsaan Malaysia. He is currently a Senior Lecturer with Universiti Teknologi PETRONAS. His research interests include ant colony algorithms, optimization, swarm intelligence, and database technologies.

SEYEDALI (ALI) MIRJALILI (Senior Member, IEEE) is the Director of the Centre for Artificial Intelligence Research and Optimization, Torrens University Australia at Brisbane. He is internationally recognized for his advances in swarm intelligence and optimization, including the first set of algorithms from a synthetic intelligence standpoint—a radical departure from how natural systems are typically understood—and a systematic design framework to reliably benchmark, evaluate, and propose computationally cheap robust optimization algorithms. He has published over 200 publications with over 18 000 citations and an H-index of 45. As the most cited researcher in robust optimization, he is in the list of 1% highly cited researchers and named as one of the most influential researchers in the world by the Web of Science. He is working on the applications of multi-objective and robust meta-heuristic optimization techniques as well. His research interests include robust optimization, engineering optimization, multi-objective optimization, swarm intelligence, evolutionary algorithms, and artificial neural networks.

Dr. Mirjalili is an Associate Editor of several journals, including *Neurocomputing*, *Applied Soft Computing*, *Advances in Engineering Software*, *Applied Intelligence*, and IEEE Access.

HITHAM ALHUSSIAN received the B.Sc. and M.Sc. degrees in computer science from the School of Mathematical Sciences, Khartoum University, Sudan, and the Ph.D. degree from Universiti Teknologi PETRONAS, Malaysia. He is currently a Senior Lecturer with the Department of Computer and Information Sciences and a Core Research Member of the Centre for Research in Data Science (CERDAS), Universiti Teknologi PETRONAS. His main research interests are in real-time parallel distributed systems, cloud computing, big data mining, machine learning, and secure computer-based management systems.