A Review on Bio-inspired Optimization Method for Supervised Feature Selection

Montha Petwan[1]

Faculty of Science and Technology
Suratthani Rajabhat University
Khun Taleay, Muang, Surat Thani, 84100, Thailand

Ku Ruhana Ku-Mahamud[2]

School of Computing
Universiti Utara Malaysia, 06010 Sintok, Kedah, Malaysia
Shibaura Institute of Technology, Tokyo, Japan

Abstract—**Feature selection is a technique that is commonly used to prepare particular significant features or produce understandable data for improving the task of classification. Bio-inspired optimization algorithms have been successfully used to perform feature selection techniques. The exploration and exploitation mechanism that is based on the inspiration of living things to find a food source and the biological evolution in nature. Nevertheless, irrelevant, noisy, and redundant features persist from the situation of fall into local optima in case of high dimensionality. Thus, this review is conducted to shed some light on techniques that have been used to overcome the problem. The taxonomy of bio-inspired algorithms is presented, along with its performances and limitations, followed by the techniques used in supervised feature selection in term of data perspectives and applications. This review paper has also included the analysis of supervised feature selection on large dataset which showed that recent studies focus on metaheuristic methods because of their promising results. In addition, a discussion of some open issues is presented for further research.**

Keywords—*Bio-inspired optimization; swarm intelligence; evolutionary algorithm; machine learning*

I. INTRODUCTION

Application of science and engineering such as image classification, machine learning, text mining, image retrieval, intrusion detection, and biology analysis, containing huge number of features that are used for information processing and decision making [1],[2],[3],[4]. These applications must be approached carefully due to the abundance of data dimensions, described as the term of big data [5],[6]. Big data has well-known properties such as velocity, variety, value, volume, and veracity [7]. Memory space and costs will increase with escalated data volume. The variety of data makes data integration difficult due to data structured differently. The veracity is a noisy data and quality fluctuation of data when acquired from multi-sources [8],[9]. In solving this problem, irrelevant and redundant feature elements have to be eliminated from superfluous features leaving only data that represent the actual meaning of all features. This complication can be resolved by utilizing feature selection methods. Feature selection techniques are envisioned in the data-preprocessing stage to reduce the dimensionality by selecting the significant features from the original features in problem domain with higher performance of the task as well as speeding up the algorithm [10],[11],[12].

Filter methods such as the Fisher score [13] rank each feature independently under the Fisher criterion in a supervised model that successfully reduces the feature's size. However, this technique cannot determine the correlation among different features. The permutation of individual features does not necessarily and cannot achieve the desired feature set. Thus, the subset of feature is suboptimal [14]. Linear discriminant analysis is one traditional technique to enhance selected features by maximizing the proportion between the intraclass distribution and interclass distributions. Meanwhile, the inverse matrix calculation within-class distribution of linear discriminant analysis will be tolerated when dealing with a small number of labeled data [15]. Wrapper methods depend on a certain classification algorithm in evaluating the selected feature [16]. Hybrid methods attempt in incorporating the dominant characteristic of filter and wrapper models. The aim of filter phase is to reduce the feature dimensionality. The wrapper stage is then used to select the most optimal feature subset [17]. The author in [18] proposed a filter-wrapper algorithm which applied minimum redundancy maximum relevance algorithm to carry out a local search mechanism. Rough set theory and conditional entropy algorithm as filter method introduced in [19] were proposed in selecting the most significant features from a whole set as the initial population. The wrapper approach, which employs the k-nearest neighbor (KNN) algorithm, was then used as an evaluator of their quality of feature combination. The wrapper method can obtain a higher accuracy rate because they determine the correlations in each feature. In fact, the hybrid method achieves high accuracy with respect to the characteristic of wrapper and filter. However, wrappers are computationally more expensive, have less generalization than the filter and their performances are highly dependent on the particular classifier.

Searching and optimization methods are used to find the best solution of many classification problems. Recently, such techniques exhibit their capacity in dealing with Non-Polynomial (NP) hard problems. Finding the most significant features within reasonable amount of time is also considered as NP-hard problem [20], [21]. For example, there are N features containing in the dataset, 2N features are to be generated and evaluated, which will increase the computational cost, especially when each subset is executed by using the wrapper method. As a consequence finding the possible subsets using the exhaustive search or best-first search technique is not a great choice. Recently, an increasing number of optimization approaches have been focused in handling the issue of both

numerical and combinatorial optimization. Solutions have proposed optimal feature subsets and intensively developing based on a variety of metaheuristic methods. An optimization problem focuses on finding the optimal value which corresponds to the maximizing or minimizing one of its performance criteria, or multiple objectives have been proposed. Metaheuristics search strategy with a population-based approach has shown attractive competency in coping with the different character of optimization problem scenarios that can be used to handle the feature selection tasks [9],[22].

Metaheuristic optimization methods are getting inspiration evidentially from nature. Its mechanism and capabilities are extraordinarily magical and mysterious that researchers have focused on mapping the natural phenomenon onto intelligence algorithms. For example, finding the food source for the ant by using the shortestWE56 path through indirect communication with each other; interaction between organisms to fully matured human being; balancing the ecosystem; hunting movement and echolocation mechanism. Their abilities have been described to solve the complicated problem independently from elementary initial populations and parameters with little or no knowledge of the feature space. Thus, every feature and natural phenomenon used a suitable strategy in getting the best solution. This approach was able to find an optimal solution although simple optimization strategy has been used. One of the dominant categories of metaheuristics optimization methods is bio-inspired optimization. The bio-inspired optimization impersonates the various natural creatures behaviors like fishes, insects, bird swarms, terrestrial animals, reptiles, humans, and other phenomena [5]. The bio-inspired optimization family has emerged and applied in a proposal of feature selection applications, for instance, text mining, information retrieval, robotics, network security, biomedical engineering, power systems, business, agriculture and many more. Their behavior is a random decision that categorizes them as a randomized algorithm. Formulating a bio-inspired optimization algorithm involves modeling a proper problem representation, calculation the obtained solution efficiency through fitness evaluator and identifying operators to generate a new set of solutions [23]. Though, as previously stated, authors have divided a prevalent group in these approaches based on the evolutionary of biological paradigms.

Summarizations of feature selection algorithms have been performed by [11],[24],[25]. These studies have focused on the subset generation techniques in certain application, feature selection by using swarm-based algorithms together with particular classification and clustering tasks. However, a comprehensive overview of supervised feature subset selection based on bio-inspired optimization algorithms in obtaining the most optimal subset association with data perspectives and different applications was not performed. This paper proposes the bio-inspired optimization algorithm taxonomy according to natural biological inspiration as well as the areas in which these algorithms have been employed.

This paper is organized as follows: in Section 2, the description of the taxonomy of the bio-inspired optimization algorithms. The supervised feature selection using bio-inspired optimization algorithms is presented in Section 3. The analysis of technique for large datasets is introduced in Section 4 and discussion of the supervised feature subsets selection using bio-inspired optimization algorithms are provided in Section 5. Finally, in Section 6, the conclusions are presented together with further research directions in supervised feature subset selection.

II. Taxonomy of Bio-inspired Optimization Algorithm

Bio-inspired optimization algorithms have been identified as an excellent approach and play an essential part in finding the best optimum solution in different problem domains. This type of algorithms imitated from the systematic behaviors of natural biological evolutionary such as mutation, selection along with distributed collective of living organisms, including birds, ants and wild animal. These algorithms exhibit high level of diversity, robustness, dynamic, simplicity, and fascinating phenomena as comparison with other existing methods. Studies in computer science area have been broadly used the various bio-inspired optimization algorithms in many kind of literature, like looking for the optimal solutions for hard and complexity problem domains [26],[27],[28]. The popular salient and accomplish classes or directions in bio-inspired optimization algorithms that mimic the biological collective behavior of animals, biophysical environment, and cooperation between species respectively [5],[29],[30]. This review paper aims to form the optimization algorithms category according to the area of such inspiration to perform a widespread view over the domain. This paper attempts to categorize the sources of bio-inspiration into swarm intelligence algorithms, evolutionary-based algorithms, and ecology-based algorithms [5],[31],[32]. Nature inspired algorithm consists of bio-inspired and physics-based algorithms. Swarm intelligence, evolutionary ecology-based algorithms are all bio-inspired algorithms. These algorithms are global optimization metaheuristics that search for solution in stochastic scheme within an appropriately runtime procedure. Generally, they start with initial solution and then generate the best solution in the next iteration. The important mechanism of metaheuristic algorithm is balancing between exploitation and exploration that may produce global optimal solution. Exploitation mechanism (intensification) contributes to the agent convergence in optimality. Exploration mechanism (diversification) prevents the loss of diversity which occurs when algorithm get trapped in local optima [33]. Bio-inspired algorithms help to tackle the global optimization problem for selecting feature subsets in the classification area by extracting and exploiting the collective local and global behavior schemes. Studies have now focused on increasing the performance of the search competency in the problem space and efficiently selecting a minimal and discriminating feature subset.

A. Swarm Intelligence Algorithm

Swarm intelligence is defined as an interactive system with the multi-agent. The system is the emergent of intelligent behavior collaborating with the collecting to complete the particular objective that cannot be completed by a single agent or acting alone [34],[35],[36]. Two expressive behaviors of swarm-based system consist of the self-organization and decentralization mechanism of animal living in nature. Self-organization can be identified as state transition rule and

stability through positive and negative feedback. Decentralization system can be described as the collaboration in groups through the state of environment to collect the communication. The examples of emergent swarm behavior in nature are bat echolocation, birds flocking, fish school, bee mating, mosquitoes host-seeking, cockroaches' infestation, ant foraging, sea creatures, and many others. Swarm intelligence approach has remarkable results in solving a wide range of NP-hard problems and really becomes nearly practical in the different real world problem domains. Furthermore, the number of possible solutions has gone up significantly in the problem that frequently leads to be indefinite. Swarm intelligence is used to solve real-world nonlinear problem applications considering many applications of sciences, engineering, data mining, machine learning, computational intelligence, business and marketing, bioinformatics, and industries. This paper emphasizes different swarm-based algorithms entirely starting from the behavioral and biological creatures perspective, which are specified in the life cycle of the insects, birds and animals (amphibians and mammals) [37].

According to [36],[38] used the term swarm intelligence to describe a system comprised of autonomous robots cooperating to fulfil the task under study. This mechanism only enables to handle with partial and noisy information about their environment, to force with uncertain situations, and to search solutions to complex problems. In this way, many existing theoretical frameworks and algorithms mimic the miracle ability of swarm behavior to deal with different scenario in feature selection problems. Among them, the well-known widespread used are ant colony optimization (ACO) [39], artificial bee colony (ABC) [40], particle swarm optimization (PSO) [41], cuckoo search (CS) [42] and firefly algorithm (FA) [43]. Another popular swarm intelligence technique is the monkey algorithm [44], wolf pack algorithm [45], bee collecting pollen algorithm [46], dolphin partner optimization [47], bat-inspired algorithm (BA) [48], and Hunting Search [49]. Salp swarm algorithm (SSA) has been extensively adapted bio-inspired algorithm on account of its advantage such as: (1) a novelty algorithm, (2) unsophisticated, (3) lesser parameters, and (4) low computation time [50]. Swarm-based approaches are less complex procedure with several parameters to strong solution as compared with evolutionary-based components such as selection, crossover, and mutation.

B. Evolutionary-based Algorithms

Evolutionary algorithm or evolutionary computation is a search method that simulates the biological perspective of generation stands on iterative framework of fittest population selection namely reproduction, mutation, recombination, and selection. This search has taken advantage of the assortment to find the suitable solution by using the historical data that leads to a better new solution. This algorithm simulates Charles Darwin's law of nature evolutionary of "survival of the fittest" in selection process in such environment. Evolutionary algorithms have been designed to search the optimal or near-optimal solution in various optimization frameworks whereas typical statistical techniques may produce ineffective results. The performance of evolutionary algorithms is generally depending on the evolutionary setting. For example, the methods for changing the values from reproduction and

mutation for creating the new populations may yield different optimization results and speeds of convergence. Some of the well-known evolutionary-based approaches are a genetic algorithm (GA), genetic programming (GP), evolution strategy (ES) [51], evolutionary algorithm, and artificial immune system.

C. Ecology-based Algorithm

Ecology-based algorithm has been presented for cooperative stochastic search algorithms. The algorithm mimics an ecosystem balancing on the earth. It relies on the population relationship of individuals in a particular ecosystem. Each population is related to adaptation or optimization strategy in the unit space. This algorithm, like other metaheuristic optimization algorithms such as ACO, ABC, GA, and others, seeks global optimization solutions [52]. The search performance of the individuals in each population are interpreted based on the exploration and exploitation mechanism, and the initial parameters [53]. The ecology-based algorithm is inspired by the ecological concepts of habitation, relationship and interaction, and inheritance ecologically. The examples of well-known ecology-based algorithms in the computer science are biogeography-based optimization, inspiration from the immigration scheme of species or animals to find new environment properly [54],[55]. Flower pollination algorithm (FPA) is presented by Yang [56]. The algorithm is inspired by flowers pollination process with biotic and abiotic pollination forms. Biotic pollination is typically linked with pollinators' livelihood such as birds, bees, bats or insects to transfer the pollen from one to another. Abiotic pollination is a process that depending on wind, rain or water. Biotic and pollination of a different plant are a process of global pollination performing by the Lévy distribution. Abiotic and self-pollination can be described as a local pollination procedure.

III. Supervised Feature Selection Method

The classification task is associated with class labeled and can be classified into three frameworks: supervised [57], unsupervised [58], and semi-supervised [59]. Class labeled participation is based on supervised feature subset selection [60],[61]. On the contrary, unsupervised feature subset selection is a primary challenge due to the class is unlabeled. Meanwhile, semi-supervised feature subset selection approaches utilized both labeled and unlabeled classes [62].

Supervised feature selection has an intention on classification or regression problems. It determines the solution of feature subsets that aim to distinguish the instances from predicted categories or predicting the potential target. After splitting the feature subset selected by supervised feature selection into learning and testing sets, the learning set is learned and evaluated by certain classifiers or regression model. The relevance between feature and classes labels is evaluated via its correlation. Choosing a strategy with a filter method can be independent of the classifier algorithm. In contrast, the wrapper method can take the advantage of the classification or regression performance to evaluate the fitness of selected features from the original set or make utilization the intrinsic predictability of a classification algorithm in embedding feature selection algorithm into their specific

fundamental learning model called embedded method. Finally, the subset features with the unseen data in the test set are employed to label for the result of predefined class or regression target [25]. In the present paper, the proposed supervised feature subset selection methods in improving the classification performance are addressed.

A total of 46 publications on bio-inspired optimization algorithms focusing on supervised feature subset selection from 2018 to 2021 have been reviewed. The papers were obtained from Scopus database in December 2021. These papers might not provide the entire studies, but presents the trend in general. Table I displays the retrieval articles in two broad perspectives: static and dynamic data. The data perspectives are further categorized into two classes: using stand-alone metaheuristic or combination with other metaheuristics (hybridization). The most frequently used algorithms are swarm intelligence group, categorized as insect and reptiles like ACO, ABC and DA have been extremely applied to solve the supervised feature subset selection. In bird group of swarm intelligence, PSO, CS and harris hawk optimization (HHO) are used to combat feature selection problem. In Fig. 1, the overall number of those papers are also illustrated according to the year of publication. Research on swarm intelligence has significantly increased in 2019.

However, greater interest has shown on swarm-based algorithm such as PSO, ACO, ABC and FA as compared to evolutionary-based and ecology-based. This paper reveals that swarm-based algorithms have advantage in controlling their behavior autonomously, self-organization and adaptability [5],[34]. Moreover, the ecology-based algorithm has emerged as a new algorithm to show the performance for the supervised feature selection area. These algorithms with different evaluation methods (filter, wrapper, and hybrid) are described in this following subsection.
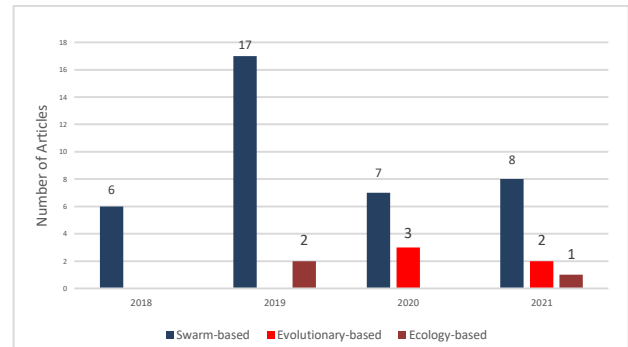

Fig. 1. Usage of the Bio-inspired Algorithm Applied in Supervised Feature Selection.

TABLE I.   SUPERVISED FEATURE SELECTION ALGORITHM WITH DIFFERENT DATA PERSPECTIVES

Inspiration	Data perspective				Total
	Static data		Dynamic data		
	Stand alone	Hybridization	Stand alone	Hybridization	
Swarm intelligence Insects and reptiles	[17], [21], [63], [64], [65], [66], [67], [68], [69], [70]	[19], [71]	[72]	[73]	14
Birds	[3], [10], [26], [74], [75], [76]	[27]	[77], [78], [79]	-	10
Terrestrial animals	[28], [80]	-	-	[81], [82], [83]	5
Sea creatures	[1], [18], [84], [85], [86], [87]	[33]	[88]	[89]	9
Evolutionary	[90], [91]	[92]	[93], [94]	-	5
Ecology-based	[95], [96]	[97]	-	-	3
Total	28	6	7	5	46

A. Filter Methods

Filter methods are generally less computational complexity than wrapper approaches. Filter methods evaluate characteristics of data based on some predefined criteria instead of using capability of certain learning algorithm. In this method, the evaluated features that have lower ranking criteria are filtered out. The filter method can be generally broken down into univariate and multivariate schemes. Individual features are ranked in univariate scheme, while the multivariate scheme ranked each feature simultaneously. Many studies have used different evaluation measure techniques to enhance the accomplishment of feature subset selection [25]. These techniques are relief algorithm [98], feature correlation [99], mutual information [100],[101], Fisher score [102] and principal component analysis (PCA) [103]. These techniques are not guided by a certain learning algorithm, such features could be deteriorated the decision making procedures. Requirement efficient metaheuristic methods can improve the performance when it carries on with large-scale feature space.

In [68], ten chaotic maps have been employed in searching process of the dragonfly algorithm (DA) for choosing the optimal extracted features for achieving convergence speed and efficiency of toxicity drug identification task. The selected features from chaotic DA were then fed into a support vector machine (SVM). The experiment indicated that Gauss chaotic map provide the best performance of DA. In [79] the population initialization in PSO is identified by assigning the Relief scores to distinguish ability of biological features. This technique examined the difference between selected sample as well as its homogeneous and heterogeneous neighbor sample. Then, the threshold selection is used to determine the number of features. The author in [78] carried out streaming features from big datasets through parallel processing with the MapReduce technique, dividing the incoming data into subsection. BA is applied in reduced data dimensionality. Then, the ensemble method with multi-layer perceptron artificial neural network classifier is employed to classify the selected significant features. The result of this work has

shortened the processing time but enhanced the accuracy. The author in [93] fused detection mechanism to scan the existing and incoming of feature drifts. The proposed multi-objective feature selection utilized measurement the quality of the solutions based on mutual information method and GA. The GA is started for evaluating the solutions based on merging population, sorting, and crowding distance mechanism. In addition, [28] proposed reducing the dimension of the binary search space of different scale datasets based on social spider algorithm (SSA), S-shaped and V-shaped transfer function are used to evaluate the binary search space. Each possible best solution is improved to the quality solution through crossover mechanism. The performance of this algorithm named BinSSA4 with crossover operator is superior in terms of fitness values, standard deviation values, number of the selected feature, and the accuracy. In application of network intrusion detection, [66] proposed the performance of ACO by enhancement the exploration process. The proposed algorithm keeps away from falling into a local optimum by designing fitness function, pheromone mitigation and increasing for some special trail.

B. Wrapper Approaches

A wrapper approaches is generally composed of two main steps: 1) explores for a potential feature subset and 2) measure the influential quality of selected features. The model iteratively generates both steps until the predefined stopping condition is met. The possible features are first generated as a feature subset, and then these features are evaluated to measure their quality based on a particular learning algorithm. In other words, the wrapper-based feature selection produces repeatedly until the desired accuracy rate is achieved or the desired minimum number of selected features is acquired and then returned as optimal selected features for a particular problem. However, the well-known impractical condition of wrapper methods is dealing with the high dimensionality of d features (2d) or the large scale of the search space. Additionally, this method will be met high complexity running time compared with the filter methods. However, there is a wealth of literatures on wrapper-based feature selection when handing high dimensionality datasets. In [63] the continuous version of original butterfly optimization algorithm (BOA) is performed by S-shaped and V-shaped activation function. The results shown that the S-shaped is able to boost the capacity of original BOA that can achieved better accuracy rate and number of selected features. The study of [69] proposed the grasshopper optimization algorithm (GOA) to obtain the most optimal solution by more repulsion in unexplored search space. Promising regions were exploited by intensity and length scale of attractive function. The literature supplemented the algorithm by using SVM during iterations to deal with unexplored feature space and duplicated features in the selected subset. In [64], the integration between ABC and gradient boosting decision tree algorithm is established to explore the best final result. The initial problem space is spanned based on gradient boosting decision tree to categorize the sample into positive or negative patterns. The performance of ABC algorithm can remove low correlation into reduced preliminary input of those decision algorithm. The study of [71] used differential evolution to perform the pheromone updating rule mechanism of max-min ant system (MMAS) by disrupting the

pheromone deposition over the features space and raising the behavior of ants in exploring for optimal feature subset for the benefit of classification task. The binary version of DA with dynamic behavior transfer functions, S-shaped and V-shaped are incorporated for beneficial better solution from unexplored regions [67]. Different chaotic map is experimented due to the problem of slow convergence speed and getting stuck in local optimum of SSA algorithm. The five chaotic variables are adapted for salp position. The results of this algorithm is compared with original SSA, GA and PSO that outperformed such algorithms [84]. Binary version of integrated grey wolf optimization algorithm (GWO) and PSO have been proposed in [81] to cope feature subset selection. This combining, the velocity and position have been controlled by weighting function to balance the diversification and intensification of proposed algorithm. KNN algorithm with euclidean distance measurement is employed in the wrapper-based method. Two binary forms of whale optimization algorithm (WOA) algorithm have been integrated with evolutionary operators to perform the exploration and exploitation mechanism in seeking the optimal selected feature for increasing classification targets. In the search process, the Tournament and Roulette Wheel Selection mechanisms are used. Crossover and mutation operators are applied to increase the exploitation mechanism of the WOA algorithm. The results showed that WOA with crossover and mutation outperformed GA, PSO and ant lion optimizer [33]. The study of [70] propose a binary FA with two objectives, accuracy rate and reduction rate to reduce the number of features. In this literature, the new formula is proposed by calculating the distance between two fireflies to enhance the quality of exploration and exploitation of search space. The results of algorithm outperformed the PSO. The discrete cosine transform (DCT) with fixed-size window technique was applied in [77] to exploit the current informative features of data streaming as a baseline. Then, the efficient feature subset produced from the PSO algorithm is fed into the KNN classification algorithm for decision processing. The experimental demonstrated the DCT without and with feature selection. The result show that the automatic feature selection process searches the best feature subset that can give higher performance. The combination metaheuristic approach is proposed by using GWO and WOA to enhance a wrapper-based feature subset selection technique. The hybridization is accomplished by improving the mechanism of both algorithms including immature convergence and stagnation to local optima [83]. As sine-cosine algorithm which can be supplemented the exploration stage, [74] combined this algorithm into HHO in exploration phase, as the result effectiveness in exploitation phase can get the quality information. Additionally, the delta factor is injected in exploitation phase. The results of this proposed outperform sine-cosine algorithm and original HHO algorithm for ten datasets out of sixteen in terms of fitness. In terms of accuracy, the proposed outperform other optimization method on eleven datasets out of sixteen. The author in [75] proposed the opposite point exploration and disruption operation due to problem of struggling in local optimization of CS algorithm, this prevents over random search of features. These enhancements improve the exploration phase and feature selection in complex data. The results show that this algorithm can find the maximizing the classification accuracy rate and

reducing the number of features; however, the computational time still increases. In the same manner, [86] has been adapted opposition-based learning (OBL) technique in slime mould algorithm (SMA) to overcome premature convergence and slow movement. Moreover, [89] presented the concept of OBL. The procedure is the WOA run first and at the same time during the run, population is changed by the OBL. To increase the accuracy and speed convergence, it is used as the initial population of FPA. The best feature set from transformed dataset is proposed in [85]. The continuous values are converted into binary search space using empirical threshold δ. This work shows very excellent results when combined with PCA and independent component analysis. As the results, feature selection method only reduces the unnecessary and unimportant features, not the correlated and higher-order dependencies among features. Due to the problem of CS that is over randomly causes blindly preservation the quality solution, the work of [76] applied chaotic map to enhance the exploration mechanism of CS. The proposed two-population elite preservation strategy can find the attractive one of each generation and preserve it. Levy flight is developed to update the position of a cuckoo, and the proposed uniform mutation strategy avoids the trouble that the search space is too large for the convergence of the algorithm due to Levy flight and improves the algorithm exploration ability. The author in [27] proposed binary PSO with FPA, PSO performs as a global search and FPA conducts a fine-tuned search. The study of [10] performed variant of PSO, competitive swarm optimizer with KNN to handle the large scale optimization problem.

Moreover, an ant colony-based approach has been presented to explore the suitable signals feature subset for application of power quality disturbances classification. S-transform fused with time-time transform is employed for detection and feature extraction. The proposed presented the results of classification with and without feature selection [65]. The author in [90] designed a OBL mechanism to both exploration and exploitation in differential evolution variant have been proposed in the application problem of engineering. The diversity measurement is designed to recognize the convergence behavior of OBL variants. Secondly, the explorative opposition and exploitative opposition are distinguished according to the convergence behavior of OBL variants. Finally, the protective mechanism is introduced to obtain a good ratio between exploration and exploitation for better performance without extra fitness evaluations. This literature carried out experiments on the IEEE Congress of Evolutionary Computation (CEC) 2017, CEC 2011 and CEC 2020 test suites that shows superior performance. The problem of over fitting classification in the swine breed has been solved in [95] by applying feature selection technique to reduce many large original features into the most significant porcine single nucleotide polymorphism. Binary FPA is combined with information gain along with the cut-off-point-finding threshold to identify a 0 or 1 value in feature vector and GA bit-flip mutation operator. The result of this study revealed that the proposed technique outperformed the PSO, CS and FA in terms of classification accuracy. The author in [91] proposed three-dimensional reduction of feature space mechanism under deletion conceptual the unimportant features utilizing the feedback information from evolution algorithm such as DE,

GA, PSO. The assistance of dimension reduction mechanism with evolutionary algorithm is effective way in finding a feature subset with higher classification accuracy and smaller number of features. The WOA with SVM is presented for the task of spam recognition in various languages. The model was proposed to perform automatic detection of arrival spams and gives an insight into the most influential features during the detection process [88]. A artificial fish swam optimization algorithm with a crossover operation have modified for application of text categorization [1]. The method includes the best fish of swam can be brought together to improve the capacity of local search. To reduce the run time complexity of ACO algorithm, the study of [21], the degree based graph representation of ACO algorithm for the field of speech processing domain have been proposed. The proposed method will have benefit over fully connected graph and contributed more flexibility on the problem space compared to binary connected graph representation.

In addition, FPA is used to propose the elimination of irrelevant features in biomedical data analysis [96]. The diversity of the population and search performance of the FPA algorithm have been increased by adopting the absolute balance group strategy and adaptive Gaussian mutation. The classification rate is evaluated by using KNN. The experimental result reveals that the proposed method outperforms other state-of-the-art methods. The author in [97] proposed hybrid version of biogeography-based optimization and GA in breast density classification. The authors of [73] combined ABC and CS for reduction the number of features to utilize the incoming anomalies detection in network. The binary pigeon inspired optimizer based on cosine similarity concept have been proposed to build optimal solution by calculating the velocity of the pigeons [3]. Another swarm algorithm for intrusion detection was proposed by [82]. The proposed algorithm utilized grey wolf optimization algorithm for selection the most optimal features by controlling the balancing between exploration and exploitation mechanism.

C. Hybrid Methods

Hybrid method aggregates the dominant of multiple feature selection approaches such as filter-wrapper. The principal intention is to increase the stability of the solution by combination salient of different available feature subset selection algorithms. For example, dealing with a small dataset with high dimensionality and dealing with a small combination on the training set will result in completely different solutions. By combining multiple selected features based on different approaches, the solutions are more straightforward, so the quality of the selected features can be preserved. In hybrid methods that is stimulated by the way of bio-inspired algorithm, [17] introduced the ranked informative features using filter method. Their quality obtained feature sets are then evaluated based on wrapper method. Improved memory to keep the best ant and normalized pheromone updating mechanism have also been proposed to enhance the feature selection. A logistic map sequence has been used to perform diversity in the problem space of PSO with a spiral-shaped mechanism integrated as an operation of local search around the known optimal solution boundary. The current position and flying velocity are incorporated with two dynamic correction

factors to improve the exploration and exploitation in the feature space [26]. In [87] applied the Pearson's correlation coefficient and correlation distance to adapt the weights of unrelated and consistent features based on filter manner. The random population WOA functions act as a wrapper algorithm. The results achieve the highest classification accuracy in all datasets, but not shortest length of feature set compared with other algorithms. In this manner, chaotic WOA is introduced [18]. Besides, [92] established integration the strengths of GA and PSO to conquer the tradeoff between exploitation and exploration procedure. The wrapper proposed utilizes artificial neural network to evaluate the fitness of feature set. This method applied the data which pull out from smaller datasets to reduce the processing time of subset selection on high volume datasets. Ant lion optimizer is integrated with hill-climbing technique to find the best solution in work of [19].

Furthermore, the feature selection for anomalies detection in the network has been proposed in [72] based on the FA. This literature employs mutation-based in filtering features and FA with wrapper-based methods for evaluating selected features to C4.5 and Bayesian Networks based classifiers. The author in [94] proposed filter-based, information gain metric and the sorting mechanism of evolutionary algorithm which is efficient multiobjective optimization algorithm. The dataset is the World Health Organization Director-General's speeches during the COVID-19 pandemic period and Standford Sentiment Treebank. Due to the abundance of irrelevant and redundant data in microarray datasets, authors in [80] proposed information gain and krill herd algorithm to capture only the important features from the original datasets.

IV. ANALYSIS OF TECHNIQUE FOR HUGE DATASET

The goals of proficient data analysis rely on the providing a large amount of data and these purposes is indispensable to deal with analytics on huge datasets. Working toward data analysis, the facilitation of massive data preservation technologies, the revolution of digital technologies, where huge sizes of data are generated with ever-increasing volumes of transactions over time and diversity. The term "huge dataset" was motivated and officially reported in international conference [104]. One of distinctiveness of huge datasets is that the sample size is generated greater than petabytes level and moves very fast when used to describe a given sample. Recently, the modernizations of technologies and Smartphones have enabled users to use online media to communicate with others in a one-way or two-way manner. Some of examples of sources with huge datasets include Facebook, Twitter, blogs, flick, LinkedIn, Pinterest, sensing devices, and web-based email. Huge collections of these data platforms typically have complex structures which received from multiple sources. The information collected from this online source consists of different data formats such as text, image, videos, log, and so on. It is extremely challenging to obtain important hidden data from these social assemblages in an appropriate manner. The increasing rate of digital adoption has been observable since the initiation of the coronavirus, covid-19 pandemic since 2020. The rate of digital growth has increased dramatically, and more are expected in the second half of 2021 as shown in Fig. 2. It shows that the power of social media continues to drive activities to be connected all over the world, with an

increasing number of social media users worldwide, and an impressive step is rapidly approaching. Also, Facebook and Instagram have attained 5.1 billion users. Twitter generated 8 terabytes of data per day, or 80 million tweets per day [105], and Line and Whatsapp generates approximately 2.5 petabytes of data per hour.

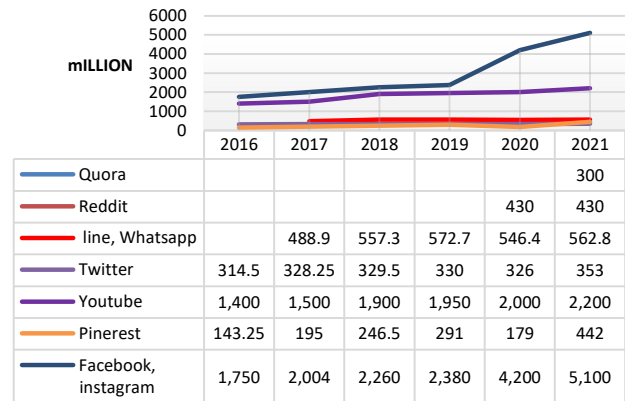| mILLION	2016	2017	2018	2019	2020	2021
Quora						300
Reddit					430	430
line, Whatsapp		488.9	557.3	572.7	546.4	562.8
Twitter	314.5	328.25	329.5	330	326	353
Youtube	1,400	1,500	1,900	1,950	2,000	2,200
Pinerest	143.25	195	246.5	291	179	442
Facebook, instagram	1,750	2,004	2,260	2,380	4,200	5,100

Fig. 2. The Rate of Data Generation in Digital World from 2016 to 2021.

An efficient way in extraction the important information can be measured by the complexity of computing methods and algorithms. Traditional methods and tools are used to operate small and structured data by trail-and-error. This endeavors analysis is not suitable when data sets are large and miscellaneous. This large dataset requires a large amount of memory to store and takes hours or days in the case of using traditional processing methods. At present, a superfluity of different improved techniques are being developed to treat with large amounts of unstructured data sets [106]. The nature of large amounts of data is adequate for learning models to be applied to real-world scenarios efficiently. In addition, the models can be enhanced to digest as much target discrimination as possible. Face forgery detection is an example of an application domain and technique that makes use of a large dataset, with data that is ten times larger than the previous forensic dataset. The increased data source is distorted through face swapping method for robustness the head poses variation due to videos on the internet usually have limitation gesture of head. The feature learning with dimensionality reduction via autoencoder is considered for forensics. The experiment was carried out with as many real-world perturbations as possible. The accuracy results remained low due to the poor quality of the learning set and augmentation method for face diversity [107]. Moreover, the novel chance constrained problem domain is formulated using a huge dataset. A weighted feature reduction operation is proposed to describe a relaxation problem of chance constrained problems. Also, a DE has been adapted and integrated with a pruning technique to force the relaxation concern of chance constrained [108].

In the natural language classification application domain, in [109] employed rule-based classification and Apriori algorithm because it cannot maintain the capability between accuracy and interpretability in the non-big data environment. This problem has been facilitated by proposing the probability integral transform theorem, rule induction and rule selection based on evolutionary optimization. The experimental results show that

fuzzy model-generated models are significantly simpler in terms of classification rate, complexity, and time consumption. A distributed fuzzy decision trees have been proposed due to the problem of time constraints and space requirements. This paper proposed using the MapReduce framework to partition large scale data into binary and multiple decision ways. The relevant features are derived by using information gain method, which will be used in the decision nodes. The author implemented the fuzzy decision tree learning scheme on the Apache Spark framework [110]. Among the given domains that use big data technologies is vehicular ad hoc network to handle their big size data. The author demonstrates a method that entail to detect accidental or irregularity on the way and estimate the distance and time spending on each route in the form of real time system, which allows the user to gain a database having the estimated time spending in all sections, this will serve the vehicles for the reasonable estimated time of attainment consistently throughout their travel and optimize the best route to reach the destination. The experiment reveals that this method effectively warns for crowded vehicles or portions the vehicle overwhelming in all roads, and it can also be used to save road safety [111]. Furthermore, human disease application used huge dataset to prevent the spread of infectious virus to human as the outbreak of covid-19 virus. The epidemic has forced everyone to stay and work at home that has affected on the people's mental health around the world. The global covid-19 pandemic requires a broader overview of data set to analyze the problem of human-to-human transmission of the virus across the country [112]. The covid-19 tracker with HPCC system is used to predict the future trend of the covid-19 outbreak, but this tool is limited in such a way that it cannot explain other factors that may influence trends such as mobility, local weather conditions, and so on. The smoothing filter of the covid-19 tracker is capable of eliminating irregular data transfers, but the system is still incomplete, with only a minor effect on the natural time vector. The automatically ingested data was pulled into the system for data cleaning and then extraction of the important data for subsequent analysis. The system will also be automatically executed when the new information enters to the system [113]. Additionally, human mobilities are extremely restricted affected by the covid-19 situation. Autonomous robot systems are becoming very important and can be conducted to replace human service work such as serving medicine to the infectious patients. The first large-scale elevator panel dataset has been made public in order to challenge the problem of inter-floor navigation. The deep learning based is used to recognize autonomous elevator operation. The performance of that model and dataset is compared to popular network such as ResNet, PSPNet and U-Net. The results of ResNet have the best performance when compared to the remaining networks [114].

V. Discussion

The recent advance in feature selection algorithms have grown exponentially across a wide range of application domains. The most explored areas continue to be in fixed data or static such as bioinformatics and image processing. In addition, social media platforms such as Twitter, Facebook, blogs, and wikis are prevalent in streaming data or dynamic. Among the bio-inspired algorithms, the swarm-based

algorithms have been applied in many areas to produce excellent solutions for problems where the characteristic of the problem is NP-hard, which otherwise produce sub-optimum solutions and consume a lot of processing power. Due to the scalability and simplicity of SI-based algorithms, they have become the first choice in producing the outcomes for an optimization problem. The present researchers expect this review paper to provide other researchers working on different bio-inspired optimization algorithms to effectively and efficiently handle new challenges in supervised feature subset selection.

To this day, many valuable feature selection algorithms have been extensively developed for real-world application and theoretical analysis. However, the present researchers believe that more intelligent behavior from nature can be applied to improve the solutions in this field. There are several contributions and issues related to feature selection method. Firstly, according to the enormous increment in the volume of the data, the recent feature subset selection algorithms may be threatened especially in terms of scalability with online datasets. Secondly, the performance of supervised feature subset selection algorithms is commonly evaluated by the compromised accuracy. As a result, algorithms adaptation should be an essential concern when exploring new searching space and exploitation of the best solution that selected in each iteration to gain most optimal feature. It is determined as the affectability of a feature subset selection algorithm to feature combination during the training phase. Finally, in statistical feature selection algorithms, feature weighting techniques are frequently used to identify the number of selected features. In this method, the number of optimal selected features is discarded. Furthermore, a large number of selected features will jeopardize the learning performance due to the inclusion of noisy, irrelevant, and redundant features. By extension, it should not use a small number of feature subsets because some relevant features are excluded. In practice, many researchers have usually used a metaheuristic way to find and evaluate the candidate feature subset and feed the number of selected features that have the best classification accuracy, but the whole process suffers from computation time. The challenging problem in this domain is determining the optimal number of selected features. Furthermore, the present researchers believe that the selection of the evaluation criteria is also a crucial aspect which requires deeper investigation.

VI. Conclusion

Feature selection techniques aim to provide effective pre-processing of data in eliminating redundant and irrelevant features. It is a fundamental method in preparing the data which is clean and intelligible. It has been an interesting field of research work that has proven to be extremely useful in many application domains including image recognition, machine learning, web and text mining, pattern classification, and medical diagnosis in both offline and online platforms. The past few years the performance of many novel feature selection methods has augmented. This can be observed that, several optimization algorithms have been presented to solve problematic feature selection by optimization the value to gain the best suitable solution. Such algorithms that are inspired by the natural, biological and ecological behavior is to produce

good solutions. However, there are only several bio-inspired algorithms that have been proposed. This study has reviewed works on the enhancement of solving the feature selection problems. In addition, the authors have detailed the algorithms that are practical among which are wolf-based algorithms, salp, and biogeography algorithms. Highlights on each algorithm have been presented, followed by recent advances in the literature and problem domains for each. Nonetheless, it is important to highlight that these algorithms have yet to be demonstrated impressively in large-scale datasets such as streaming data and linked data. This brings up subsequent research application domain such as medical, the environment, and social science.

REFERENCES

[1] Thiyagarajan and N. Shanthi, "A modified multi objective heuristic for effective feature selection in text classification," Cluster Comput., vol. 22, pp. 10625–10635, 2019.

[2] K. Sankar and P. Uma Maheswari, "A dynamic wrapper-based feature selection for improved precision in content-based image retrieval," Concurr. Comput., 2019.

[3] H. Alazzam, A. Sharieh, and K. E. Sabri, "A feature selection algorithm for intrusion detection system based on pigeon inspired optimizer," Expert Syst. Appl., vol. 148, p. 113249, 2020.

[4] L. Ignaczak, G. Goldschmidt, C. A. D. Costa, and R. D. R. Righi, "Text mining in cybersecurity: a systematic literature review," ACM Comput. Surv., vol. 54, no. 7, pp. 1–36, 2021.

[5] M. Sharma and P. Kaur, "A comprehensive analysis of nature-inspired meta-heuristic techniques for feature selection problem," Arch. Comput. Methods Eng., no. 0123456789, 2020.

[6] M. Rong, D. Gong, and X. Gao, "Feature selection and its use in big data: challenges, methods, and trends," IEEE Access, vol. 7, pp. 19709–19725, 2019.

[7] B. H. Nguyen, B. Xue, and M. Zhang, "A survey on swarm intelligence approaches to feature selection in data mining," Swarm Evol. Comput., vol. 54, no. April 2019, p. 100663, 2020.

[8] M. Cherrington, Q. Xu, D. Airehrour, S. Wade, J. Lu, and S. Madanian, "Feature selection methods for linked data: Limitations, capabilities and potentials," in BDCAT 2019 - Proceedings of the 6th IEEE/ACM International Conference on Big Data Computing, Applications and Technologies, 2019, pp. 103–112.

[9] C. Dhaenens and L. Jourdan, "Metaheuristics for data mining: Survey and opportunities for big data," 40R, vol. 17, no. 2, pp. 115–139, 2019.

[10] S. Gu, R. Cheng, and Y. Jin, "Feature selection for high-dimensional classification using a competitive swarm optimizer," Soft Comput., vol. 22, no. 3, pp. 811–822, Feb. 2018.

[11] S. A. K. M. A. M. A. K. Nazish Naheed Muhammad Shaheen, "Importance of features selection, attributes selection, challenges and future directions for medical imaging data: a review," Comput. Model. Eng. \& Sci., vol. 125, no. 1, pp. 315–344, 2020.

[12] S. B. Kotsiantis, "Supervised machine learning: a review of classification techniques," Informatica, vol. 31, pp. 249–268, 2007.

[13] M. Gan and L. Zhang, "Iteratively local fisher score for feature selection," Appl. Intell., vol. 51, no. 8, pp. 6167–6181, 2021.

[14] Hashemi, M. B. Dowlatshahi, and H. Nezamabadi-pour, "Ensemble of feature selection algorithms: a multi-criteria decision-making approach," Int. J. Mach. Learn. Cybern., vol. 13, no. 1, pp. 49–69, 2022.

[15] G. Zhao and Z. Zhou, "Efficient linear feature extraction based on large margin nearest neighbor," IEEE Access, vol. 7, pp. 78616–78624, 2019.

[16] M. Kusy and R. Zajdel, "A weighted wrapper approach to feature selection," Int. J. Appl. Math. Comput. Sci., vol. 31, no. 4, pp. 685–696, 2021.

[17] M. Ghosh, R. Guha, R. Sarkar, and A. Abraham, "A wrapper-filter feature selection technique based on ant colony optimization," Neural Comput. Appl., vol. 3, 2019.

[18] R. Guha, M. Ghosh, S. Mutsuddi, R. Sarkar, and S. Mirjalili, "Embedded chaotic whale survival algorithm for filter–wrapper feature selection," Soft Comput., vol. 24, no. 17, pp. 12821–12843, 2020.

[19] M. M. Mafarja and S. Mirjalili, "Hybrid binary ant lion optimizer with rough set and approximate entropy reducts for feature selection," Soft Comput., vol. 23, no. 15, pp. 6249–6265, 2019.

[20] X. He, M. Ji, C. Zhang, and H. Bao, "A variance minimization criterion to feature selection using laplacian regularization," IEEE Trans. Pattern Anal. Mach. Intell., vol. 33, no. 10, pp. 2013–2025, 2011.

[21] R. R. Rajoo and R. Abdul Salam, "Ant colony optimization based subset feature selection in speech processing: constructing graphs with degree sequences," Int. J. Adv. Sci. Eng. Inf. Technol., vol. 8, no. 4–2, p. 1728, 2018.

[22] J. Swan et al., "Metaheuristics 'In the Large,'" Eur. J. Oper. Res., vol. 297, no. 2, pp. 393–406, 2022.

[23] Jain and D. Zongker, "Feature selection: evaluation, application, and small sample performance," IEEE Trans. Pattern Anal. Mach. Intell., vol. 19, no. 2, pp. 153–158, 1997.

[24] L. Brezočnik, I. Fister, and V. Podgorelec, "Swarm intelligence algorithms for feature selection: a review," Appl. Sci., vol. 8, no. 9, 2018.

[25] Telikani, A. Tahmassebi, W. Banzhaf, and A. H. Gandomi, "Evolutionary machine learning: a survey," ACM Comput. Surv., vol. 54, no. 8, pp. 1–35, 2021.

[26] K. Chen, F. Y. Zhou, and X. F. Yuan, "Hybrid particle swarm optimization with spiral-shaped mechanism for feature selection," Expert Syst. Appl., vol. 128, pp. 140–156, 2019.

[27] M. A. Tawhid and A. M. Ibrahim, "Hybrid binary particle swarm optimization and flower pollination algorithm based on rough set approach for feature selection problem," in Nature-Inspired Computation in Data Mining and Machine Learning, X.-S. Yang and X.-S. He, Eds. Cham: Springer International Publishing, 2020, pp. 249–273.

[28] BAŞ and E. ÜLKER, "An efficient binary social spider algorithm for feature selection problem," Expert Syst. Appl., vol. 146, p. 113185, 2020.

[29] R. Diao and Q. Shen, "Nature inspired feature selection meta-heuristics," Artif. Intell. Rev., vol. 44, no. 3, pp. 311–340, 2015.

[30] B. Xue, M. Zhang, W. N. Browne, and X. Yao, "A survey on Evolutionary Computation approaches to feature selection," IEEE Trans. Evol. Comput., vol. 20, no. 4, pp. 606–626, 2016.

[31] Dhiman and V. Kumar, "Spotted hyena optimizer: A novel bio-inspired based metaheuristic technique for engineering applications," Adv. Eng. Softw., vol. 114, pp. 48–70, 2017.

[32] S. Goel, A. Sharma, and V. K. Panchal, "Performance analysis of bio-inspired techniques," in Proceedings of the Third International Conference on Soft Computing for Problem Solving, 2014, pp. 831–844.

[33] M. Mafarja and S. Mirjalili, "Whale optimization approaches for wrapper feature selection," Appl. Soft Comput., vol. 62, pp. 441–453, 2018.

[34] S. Almufti, "Using swarm intelligence for solving NPhard problems," Acad. J. Nawroz Univ., vol. 6, no. 3, pp. 46–50, 2017.

[35] Y. Li, "Solving TSP by an ACO-and-BOA-based hybrid algorithm," in 2010 International Conference on Computer Application and System Modeling (ICCASM 2010), 2010, vol. 12, pp. V12-189-V12-192.

[36] R. S. Parpinelli and H. S. Lopes, "New inspirations in swarm intelligence: A survey," Int. J. Bio-Inspired Comput., vol. 3, no. 1, pp. 1–16, 2011.

[37] K. Sethi, X. Li, L. Cheng, S. Yadavalli, and L. Zhang, "Swarm intelligence: A review of algorithms," no. March, p. 494, 2017.

[38] Khosravanian, M. Rahmanimanesh, and P. Keshavarzi, "Discrete social spider algorithm for solving traveling salesman problem," Int. J. Comput. Intell. Appl., vol. 20, no. 03, p. 2150020, 2021.

[39] M. Dorigo and G. Di Caro, "Ant colony optimization: a new meta-heuristic," in Proceedings of the 1999 Congress on Evolutionary Computation-CEC99 (Cat. No. 99TH8406), 1999, vol. 2, pp. 1470-1477 Vol. 2.

[40] D. Karaboga and B. Basturk, "A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm," J. Glob. Optim., vol. 39, no. 3, pp. 459–471, 2007.

[41] Kennedy and R. Eberhart, "Particle swarm optimization," in Proceedings of ICNN'95 - International Conference on Neural Networks, 1995, vol. 4, pp. 1942–1948 vol.4.

[42] X. Yang and Suash Deb, "Cuckoo Search via Lévy flights," in 2009 World Congress on Nature Biologically Inspired Computing (NaBIC), 2009, pp. 210–214.

[43] X.-S. Yang, "Firefly Algorithms for Multimodal Optimization," in Stochastic Algorithms: Foundations and Applications, 2009, pp. 169–178.

[44] Mucherino and O. Seref, "Monkey search: A novel metaheuristic search for global optimization," in AIP Conference Proceedings, 2007, vol. 953, no. 1, pp. 162–173.

[45] S. Wu, F. Zhang, and L. Wu, "New swarm intelligence algorithm-wolf pack algorithm," Syst. Eng. Electron., vol. 35, no. 11, pp. 2430–2438, 2013.

[46] X. Lu and Y. Zhou, "A novel global convergence algorithm: bee collecting pollen algorithm," in Advanced Intelligent Computing Theories and Applications. With Aspects of Artificial Intelligence, 2008, pp. 518–525.

[47] Y. Shiqin, J. Jianjun, and Y. Guangxing, "A Dolphin Partner Optimization," in 2009 WRI Global Congress on Intelligent Systems, 2009, vol. 1, pp. 124–128.

[48] X.-S. Yang, "A New Metaheuristic Bat-Inspired Algorithm," in Nature Inspired Cooperative Strategies for Optimization (NICSO 2010), J. R. González, D. A. Pelta, C. Cruz, G. Terrazas, and N. Krasnogor, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 65–74.

[49] R. Oftadeh, M. J. Mahjoob, and M. Shariatpanahi, "A novel meta-heuristic optimization algorithm inspired by group hunting of animals: Hunting search," Comput. Math. with Appl., vol. 60, no. 7, pp. 2087–2098, 2010.

[50] S. Mirjalili, A. H. Gandomi, S. Z. Mirjalili, S. Saremi, H. Faris, and S. M. Mirjalili, "Salp Swarm Algorithm: A bio-inspired optimizer for engineering design problems," Adv. Eng. Softw., vol. 114, pp. 163–191, 2017.

[51] Zhang, M. Chen, X. Xu, and G. G. Yen, "Multi-objective evolution strategy for multimodal multi-objective optimization," Appl. Soft Comput., vol. 101, p. 107004, 2021.

[52] S. H. A. Rahmati and M. Zandieh, "A new biogeography-based optimization (BBO) algorithm for the flexible job shop scheduling problem," Int. J. Adv. Manuf. Technol., vol. 58, no. 9–12, pp. 1115–1129, 2012.

[53] R. Alroomi, F. A. Albasri, and J. H. Talaq, "Essential modifications on biogeograpy-based optimization algorithm," Comput. Sci. Inf. Technol., pp. 141–160, 2013.

[54] D. Simon, "Biogeography-Based Optimization," IEEE Trans. Evol. Comput., vol. 12, no. 6, pp. 702–713, Dec. 2008.

[55] W. Gong, Z. Cai, and C. X. Ling, "DE/BBO: A hybrid differential evolution with biogeography-based optimization for global numerical optimization," Soft Comput., vol. 15, no. 4, pp. 645–665, 2011.

[56] X.-S. Yang, "Flower pollination algorithm for global optimization," in Unconventional Computation and Natural Computation, 2012, pp. 240–249.

[57] S. B. Kotsiantis, "Feature selection for machine learning classification problems: A recent overview," Artif. Intell. Rev., vol. 42, no. 1, pp. 157–176, 2011.

[58] Almazini and K. Ku-Mahamud, "Adaptive technique for feature selection in modified graph clustering-based ant colony optimization," Int. J. Intell. Eng. Syst., vol. 14, p. 2021, Apr. 2021.

[59] R. Sheikhpour, M. A. Sarram, S. Gharaghani, and M. A. Z. Chahooki, "A Survey on semi-supervised feature selection methods," Pattern Recognit., vol. 64, pp. 141–158, 2017.

[60] Z. Zhao, L. Wang, H. Liu, and J. Ye, "On similarity preserving feature selection," IEEE Trans. Knowl. Data Eng., vol. 25, no. 3, pp. 619–632, Mar. 2013.

[61] S. Xiang, F. Nie, G. Meng, C. Pan, and C. Zhang, "Discriminative least squares regression for multiclass classification and feature selection," IEEE Trans. Neural Networks Learn. Syst., vol. 23, no. 11, pp. 1738–1754, Nov. 2012.

[62] Z. Xu, I. King, M. R. Lyu, and R. Jin, "Discriminative semi-supervised feature selection via manifold regularization," IEEE Trans. Neural Networks, vol. 21, no. 7, pp. 1033–1047, Jul. 2010.

[63] S. Arora and P. Anand, "Binary butterfly optimization approaches for feature selection," Expert Syst. Appl., vol. 116, pp. 147–160, 2019.

[64] Rao et al., "Feature selection based on artificial bee colony and gradient boosting decision tree," Appl. Soft Comput., vol. 74, pp. 634–642, 2019.

[65] U. Singh and S. N. Singh, "A new optimal feature selection scheme for classification of power quality disturbances based on ant colony framework," Appl. Soft Comput., vol. 74, pp. 216–225, 2019.

[66] H. Peng, C. Ying, S. Tan, B. Hu, and Z. Sun, "An improved feature selection algorithm based on ant colony optimization," IEEE Access, vol. 6, pp. 69203–69209, 2018.

[67] Mafarja et al., "Binary dragonfly optimization for feature selection using time-varying transfer functions," Knowledge-Based Syst., vol. 161, pp. 185–204, 2018.

[68] G. I. Sayed, A. Tharwat, and A. E. Hassanien, "Chaotic dragonfly algorithm: an improved metaheuristic algorithm for feature selection," Appl. Intell., vol. 49, no. 1, pp. 188–205, 2019.

[69] Zakeri and A. Hokmabadi, "Efficient feature selection method using real-valued grasshopper optimization algorithm," Expert Syst. Appl., vol. 119, pp. 61–72, 2019.

[70] S. Maza and D. Zouache, "Binary firefly algorithm for feature selection in classification," 2019 Int. Conf. Theor. Appl. Asp. Comput. Sci. ICTAACS 2019, 2019.

[71] J. M. Montemayor and R. V Crisostomo, "Feature selection in classification using binary max-min ant system with differential evolution," in 2019 IEEE Congress on Evolutionary Computation (CEC), 2019, pp. 2559–2566.

[72] S. B and M. K, "Firefly algorithm based feature selection for network intrusion detection," Comput. Secur., vol. 81, pp. 148–155, 2019.

[73] H. S. Al-Safi, Z. I. R. Hani, and M. M. Abdul Zahra, "Using a hybrid algorithm and feature selection for network anomaly intrusion detection," J. Mech. Eng. Res. Dev., vol. 44, no. 4, pp. 253–262, 2021.

[74] Hussain, N. Neggaz, W. Zhu, and E. H. Houssein, "An efficient hybrid sine-cosine Harris hawks optimization for low and high-dimensional feature selection," Expert Syst. Appl., vol. 176, no. July 2020, p. 114778, 2021.

[75] kelidari and J. Hamidzadeh, "Feature selection by using chaotic cuckoo optimization algorithm with levy flight, opposition-based learning and disruption operator," Soft Comput., vol. 25, no. 4, pp. 2911–2933, 2021.

[76] Wang, Y. Gao, J. Li, and X. Wang, "A feature selection method by using chaotic cuckoo search optimization algorithm with elitist preservation and uniform mutation for data classification," Discret. Dyn. Nat. Soc., vol. 2021, 2021.

[77] Ö. Aydoğdu and M. Ekinci, "An approach for streaming data feature extraction based on discrete cosine transform and particle swarm optimization," Symmetry (Basel)., vol. 12, no. 2, 2020.

[78] D. Renuka Devi and S. Sasikala, "Online feature selection (OFS) with accelerated bat algorithm (ABA) and ensemble incremental deep multiple layer perceptron (EIDMLP) for big data streams," J. Big Data, vol. 6, no. 1, 2019.

[79] Y. Xue, W. Jia, and A. X. Liu, "A particle swarm optimization with filter-based population initialization for feature selection," 2019 IEEE Congr. Evol. Comput. CEC 2019 - Proc., pp. 1572–1579, 2019.

[80] G. Zhang, J. Hou, J. Wang, C. Yan, and J. Luo, "Feature selection for microarray data classification using hybrid information gain and a

modified binary krill herd algorithm," Interdisc. Sci. Comput. Life Sci., vol. 12, no. 3, pp. 288–301, 2020.

[81] Q. Al-Tashi, S. J. Abdul 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.

[82] H. Almazini and K. Ku-Mahamud, "Grey wolf optimization parameter control for feature selection in anomaly detection," Int. J. Intell. Eng. Syst., vol. 14, p. 2021, Feb. 2021.

[83] Mafarja, A. Qasem, A. A. Heidari, I. Aljarah, H. Faris, and S. Mirjalili, "Efficient hybrid nature-inspired binary optimizers for feature selection," Cognit. Comput., vol. 12, no. 1, pp. 150–175, 2020.

[84] E. Hegazy, M. A. Makhlouf, and G. S. El-Tawel, "Feature selection using chaotic salp swarm algorithm for data classification," Arab. J. Sci. Eng., vol. 44, no. 4, pp. 3801–3816, 2019.

[85] S. S. Shekhawat, H. Sharma, S. Kumar, A. Nayyar, and B. Qureshi, "BSSA: Binary salp swarm algorithm with hybrid data transformation for feature selection," IEEE Access, vol. 9, pp. 14867–14882, 2021.

[86] Y. M. Wazery, E. Saber, E. H. Houssein, A. A. Ali, and E. Amer, "An efficient slime mould algorithm combined with k-nearest neighbor for medical classification tasks," IEEE Access, vol. 9, pp. 113666–113682, 2021.

[87] Y. Zheng et al., "A novel hybrid algorithm for feature selection based on whale optimization algorithm," IEEE Access, vol. 7, pp. 14908–14923, 2019.

[88] M. Al-Zoubi, H. Faris, J. Alqatawna, and M. A. Hassonah, "Evolving support vector machines using whale optimization algorithm for spam profiles detection on online social networks in different lingual contexts," Knowledge-Based Syst., vol. 153, pp. 91–104, 2018.

[89] H. Mohammadzadeh and F. S. Gharehchopogh, "A novel hybrid whale optimization algorithm with flower pollination algorithm for feature selection: Case study Email spam detection," Comput. Intell., vol. 37, no. 1, pp. 176–209, 2021.

[90] J. Li, Y. Gao, K. Wang, and Y. Sun, "A dual opposition-based learning for differential evolution with protective mechanism for engineering optimization problems," Appl. Soft Comput., vol. 113, p. 107942, 2021.

[91] Tan, X. Wang, and Y. Wang, "Dimensionality reduction in evolutionary algorithms-based feature selection for motor imagery brain-computer interface," Swarm Evol. Comput., vol. 52, no. April 2019, p. 100597, 2020.

[92] F. Moslehi and A. Haeri, "A novel hybrid wrapper–filter approach based on genetic algorithm, particle swarm optimization for feature subset selection," J. Ambient Intell. Humaniz. Comput., vol. 11, no. 3, pp. 1105–1127, 2020.

[93] S. Sahmoud and H. R. Topcuoglu, "A general framework based on dynamic multi-objective evolutionary algorithms for handling feature drifts on data streams," Futur. Gener. Comput. Syst., vol. 102, pp. 42–52, 2020.

[94] Deniz, M. Angin, and P. Angin, "Evolutionary multiobjective feature selection for sentiment analysis," IEEE Access, vol. 9, pp. 142982–142996, 2021.

[95] W. Rathasamuth and K. Pasupa, "A modified binary flower pollination algorithm: A fast and effective combination of feature selection techniques for SNP classification," in 2019 11th International Conference on Information Technology and Electrical Engineering (ICITEE), 2019, pp. 1–6.

[96] C. Yan, J. Ma, H. Luo, G. Zhang, and J. Luo, "A novel feature selection method for high-dimensional biomedical data based on an improved binary clonal flower pollination algorithm.," Hum. Hered., vol. 84, no. 1, pp. 34–46, 2019.

[97] R. Hans and H. Kaur, "Hybrid biogeography-based optimization and genetic algorithm for feature selection in mammographic breast density classification," Int. J. Image Graph., p. 2140007, 2021.

[98] Robnik-Šikonja and I. Kononenko, "Theoretical and empirical analysis of ReliefF and RReliefF," Mach. Learn., vol. 53, pp. 23–69, 2003.

[99] Guyon and A. Elisseeff, "An introduction to variable and feature selection," J. Mach. Learn. Res., vol. 39, pp. 1157–1182, 2003.

[100] Shishkin et al., "Efficient high-order interaction-aware feature selection based on conditional mutual information," in International Conference on Neural Information Processing Systems, 2016, pp. 4644–4652.

[101] X. V. Nguyen, J. Chan, S. Romano, and J. Bailey, "Effective global approaches for mutual information based feature selection," Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., pp. 512–521, 2014.

[102] Gu, Z. Li, and J. Han, "Generalized fisher score for feature selection," in Conference on Uncertainty in Artificial Intelligence, 2011, pp. 266–273.

[103] Masaeli, Y. Yan, Y. Cui, G. Fung, and J. G. Dy, "Convex principal feature selection," in Proceedings of the 2010 SIAM International Conference on Data Mining, pp. 619–628.

[104] "VLDB '75: Proceedings of the 1st international conference on very large data bases," 1975.

[105] S. Kemp, "Digital 2021: Global overview report," Datareportal, 2021.

[106] H. Li and P. C. Y. Sheu, A scalable association rule learning heuristic for large datasets, vol. 8, no. 1. Springer International Publishing, 2021.

[107] L. Jiang, R. Li, W. Wu, C. Qian, and C. C. Loy, "Deeperforensics-1.0: A large-scale dataset for real-world face forgery detection," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2020, pp. 2886–2895.

[108] Tagawa, "An approach to chance constrained problems based on huge data sets usingweighted stratified sampling and adaptive differential evolution," Computers, vol. 9, no. 2. 2020.

[109] M. Elkano, J. A. Sanz, E. Barrenechea, H. Bustince, and M. Galar, "CFM-BD: A distributed rule induction algorithm for building compact fuzzy models in big data classification problems," IEEE Trans. Fuzzy Syst., vol. 28, no. 1, pp. 163–177, 2020.

[110] Segatori, F. Marcelloni, and W. Pedrycz, "On distributed fuzzy decision trees for big data," IEEE Trans. Fuzzy Syst., vol. 26, no. 1, pp. 174–192, 2018.

[111] T. Mouad, L. M. Driss, and K. Mustapha, "Big data traffic management in vehicular ad-hoc network," Int. J. Electr. Comput. Eng., vol. 11, no. 4, pp. 3483–3491, 2021.

[112] Sadowski, Z. Galar, R. Walasek, G. Zimon, and P. Engelseth, Big data insight on global mobility during the Covid-19 pandemic lockdown, vol. 8, no. 1. Springer International Publishing, 2021.

[113] F. Villanustre et al., "Modeling and tracking Covid-19 cases using Big Data analytics on HPCC system platformm," J. Big Data, vol. 8, no. 1, 2021.

[114] Liu, Y. Fang, D. Zhu, N. Ma, J. Pan, and M. Q. H. Meng, "A Large-Scale Dataset for Benchmarking Elevator Button Segmentationand Character Recognition," in IEEE International Conference on Robotics and Automation (ICRA), 2021, pp. 14018–14024.