International Journal of Computers and Informatics

Journal Homepage: https://www.ijci.zu.edu.eg

Int. j. Comp. Info. Vol. 7 (2025) 1–12

Paper Type: Original Article

Metaheuristic Techniques in Feature Selection: A Concise Review

Dina Atef [1,*] and Doaa El-Shahat [1]

[1] Department of Computer Science, Faculty of Computers and Informatics, Zagazig University, Zagazig 44519, Egypt;
Emails: dinaatiff@fci.zu.edu.eg; doaazidan@zu.edu.eg.

Received: 04 Jan 2025 **Revised:** 02 Feb 2025 **Accepted:** 23 Apr 2025 **Published:** 25 Apr 2025

Abstract

Feature selection is an essential process in machine learning, designed to diminish the dimensionality of the feature set while preserving performance accuracy. Since the 1970s, various approaches for feature selection have been proposed, with metaheuristic algorithms being the most effective. This survey analyzes the prominent metaheuristic feature selection algorithms emphasizing their efficacy in exploration/exploitation operators, selection methodologies, transfer functions, fitness evaluations, and parameter optimization strategies. The paper provides a comprehensive literature analysis on addressing feature selection issues with metaheuristic algorithms.Metaheuristic algorithms are categorized into four types according to their behavior, with a compilation of more than one hundred listed. The paper addresses obstacles and issues in acquiring the optimal feature subset through various metaheuristic algorithms and identifies research gaps for researcher.

Keywords: Metaheuristic; Feature Selection; Practical Swarm; Hybrid; Exploration; Explotation.

1 | Introduction

The complexity of data management in real-world issues arises from the extensive volume of data. Feature reduction seeks to decrease the dimensions of source datasets while preserving performance accuracy. This entails feature development and selection, which are essential in machine learning. The feature selection problem is a significant challenge in machine learning, necessitating the identification of the optimal subset of attributes from a set of n qualities. The job of FS can be viewed as a search for "optimal" feature subsets among the rival 2^Noptions given a data set with N features. Depending on the issue at hand, optimality is sometimes subjective. A subset chosen as optimal using one specific assessment tool could not be comparable to that of a subset chosen by another.Feature selection emerges as a compelling research subject due to its many applications across numerous domains, including text mining, image processing, bioinformatics, industrial applications, and computer vision [1].

Diverse approaches, including exhaustive search, greedy search, and random search, have been employed to tackle this problem. Nevertheless, these algorithms frequently encounter premature convergence, unnecessary complexity, and considerable computing costs. Metaheuristic algorithms are regarded as the most efficient and effective techniques for identifying the best subset while preserving model accuracy [2].

Over the past thirty years, numerous metaheuristic algorithms have been developed to address a variety of optimization challenges. This study presents a comprehensive literature review.on metaheuristic algorithms and their application to diverse feature selection situations.Numerous articles have been published by various publishers concerning the development of metaheuristic algorithms and feature selection issues. Earlier, a

Corresponding Author: dinaatiff@fci.zu.edu.eg

Licensee International Journal of Computers and Informatics. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0).

literature survey has been identified on feature selection in which non-evolutionary methods have been considered. In [3], examined feature selection in multimedia applications by means of a thorough literature analysis of seventy related studies spanning 2001 to 2017. In [1], presented a comprehensive study of evolutionary methodologies, primarily emphasizing genetic algorithms, particle swarm optimization, ant colony optimization, and genetic programming in depth. In [4], the feature selection problem, with a focus on medical datasets, and offered a comprehensive assessment of methods inspired by nature. For various algorithms that draw inspiration from nature, they offered a classification system based on binary and chaotic algorithms in [5].

literature review was to identify solutions to difficulties associated with multiclass feature selection by analyzing the various types and descriptions of metaheuristic algorithms utilized for such situations from 2000 to 2022.

In [6], presenting the researchers' methodology to forecast diseases using metaheuristic techniques is the primary purpose of the study.

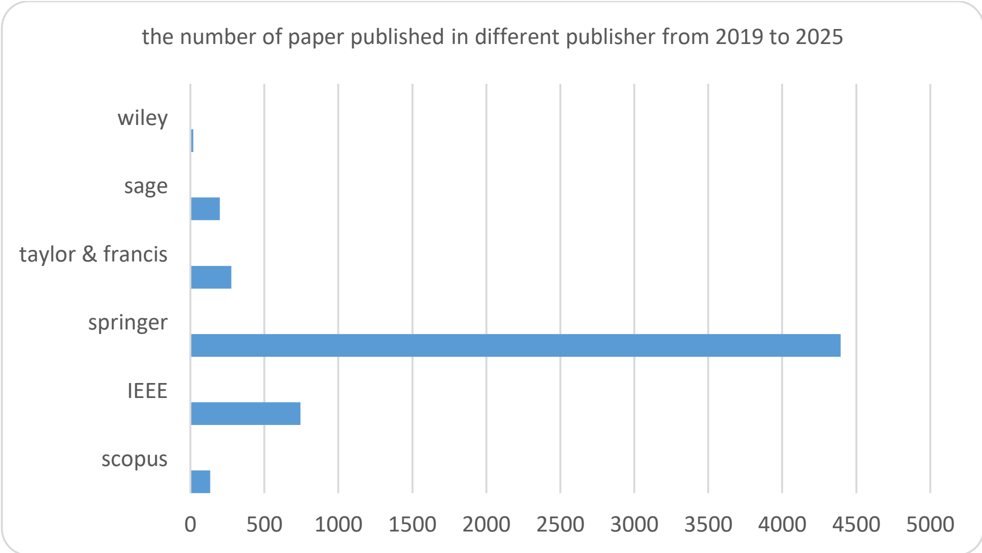

Figure 1. Show the number of paper published in many publisher.

The primary value of this research is outlined as:

- The feature selection problem is defined and methods for solving it are detailed.
- Makes a catalog of metaheuristic algorithms and sorts them into categories.
- Gives a complete bibliography of works pertaining to feature selection problems and binary metaheuristic algorithms.
- Draws attention to important aspects of wrapper feature selection methods, including as the description of the classifier, the names of the datasets, and the assessment metrics.
- Describes difficulties encountered in creating algorithms to solve feature selection problems.

Here is the structure of the paper:Feature selection and metaheuristic algorithms' foundational steps are laid out in Section II.In Section III, we provide the considerable literature on feature selection by means of metaheuristic algorithms.Section IV presents the difficulties and challenges..Section V displays the concludi ngremarks.

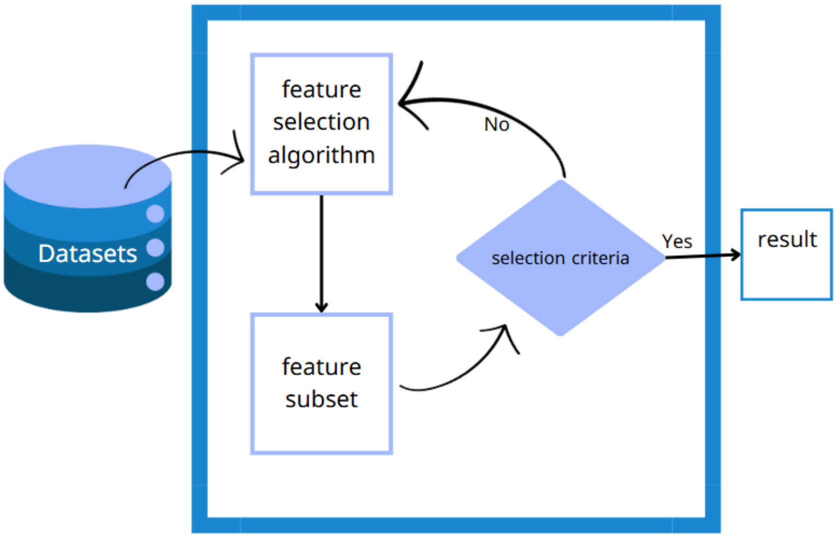

Figure 2. Mechanism of feature selection.

2 | Background

In this part, the mathematical model, as well as the ideas, definitions, and classifications of metaheuristic algorithms, are used to describe the feature selection problem in detail.

2.1 | Feature Selection

Feature selection handles features that aren't needed, irrelevant, or improper. It is a method for gleaning useful information from databases [7].Feature selection has been a major area of research and development in machine learning and data mining for decades, with extensive applications in various domains such as genomic analysis [8], text mining [9], picture retrieval [10], and intrusion detection [11], among others. Recent years have seen the emergence of new applications, which present numerous obstacles necessitating innovative theories and methodologies for managing high-dimensional and complicated data. Stable feature selection, optimal redundancy elimination, and the utilization of auxiliary data and previous knowledge in feature selection are among the most essential and complex issues in the domain of feature selection [12].

A feature selection problem can be mathematically expressed as follows: Assume a dataset S has d features. The operational mechanism of the feature selection problem involves identifying pertinent characteristics from a total of d features. Let the dataset be denoted as S = (f1, f2, f3, ..., fd). The goal is to identify the optimal subsets of features from S. Extract Subset D = (f1, f2, f3, ..., fn) where $n < d$, with f1, f2, f3, ..., fn being the features or attributes of a dataset. Figure 2 illustrates the operational mechanism of the feature selection process. The figure illustrates five primary components of the feature selection process: the original dataset, the selection of the feature subset, the evaluation of the feature subset, the selection criterion, and validation [2].

Numerous feature selection techniques have been devised to identify the optimal subset of features. The approaches are often categorized into three classifications, they split into embedded techniques, filters, and wrappers. Wrappers score subsets of variable by means of the learning machine of interest functioning as a black box. Filters, independent of the selected predictor, choose variable subsets as a pre-processing action. Usually particular to certain learning machines, embedded techniques do variable selection during training [13].

Though they are slower than filter methods, wrapper approaches show better outcomes than filter methods. Wrapper techniques rely on the modelling procedure whereby every subset is produced and then assessed. Different search strategy determines subset creation in wrapper methods.

In [14].classifies search techniques into three categories: exponential, sequential and randomized selection process. The exponential method assesses more characteristics as feature size increases. Although this method generates exact results, its great computational expense renders it not practically deployable. Exhaustive search, branch and bound method are the examples for exponential search methods. Sequential algorithms either add or remove features one after the other. Once a feature is included or removed in the selected subset, it cannot be further changed to generate local optima. Among the linear sequential techniques are best first, floating forward or backward selection, linear forward selection, etc. Randomized algorithms preserve the algorithms from simulated annealing, random generation, metaheuristic algorithms trapping into local optima by means of unpredictability to explore the search space. For example, some people call randomized algorithms population-based methods.

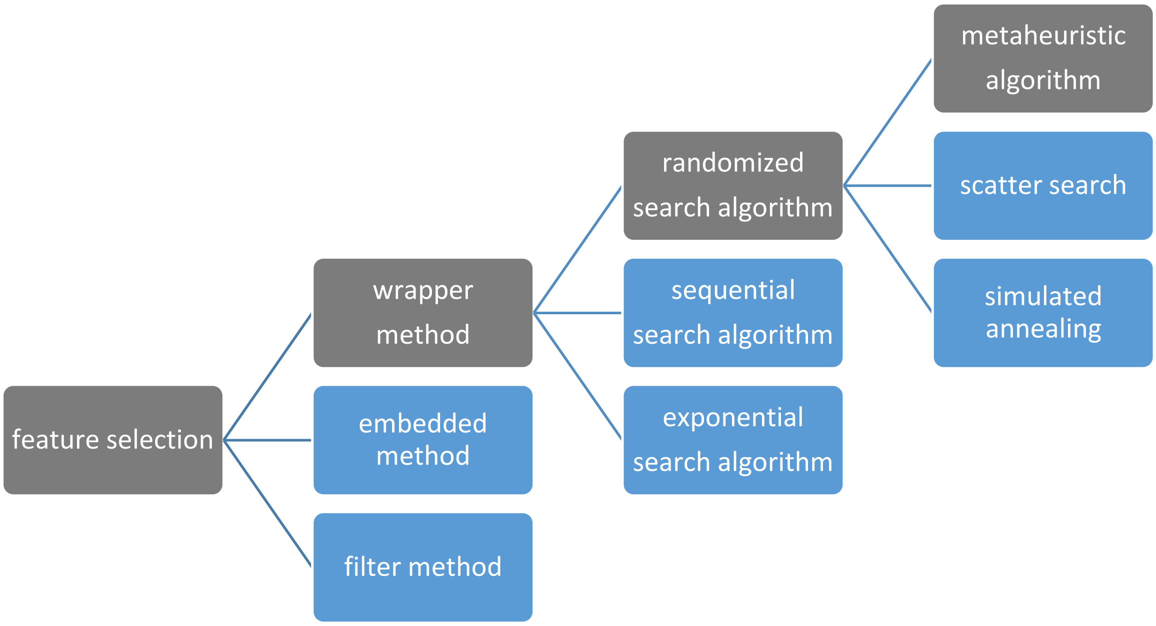

Figure 3. Categorization of feature selection.

Our goal in this work is not to provide a comprehensive overview of feature selection methods.A flow diagr m illustrating the potential solutions to feature selection difficulties is shown in Figure 3. The grey boxes in the methods part of this study show the steps we take to get at metaheuristic algorithms.

2.2 | Metaheuristic Algorithms

When applied to optimization issues, metaheuristic algorithms find the best (or almost best) answer. In addition to being simple, adaptable, and capable of avoiding local optimum, these algorithms are derivative-free. Metaheuristic algorithms begin their optimization process by producing solutions at random, exhibiting stochastic behavior [15]. While gradient search methods do necessitate calculating the derivative of the search space, our approach does not. The simplicity of the idea and the ease of implementation make metaheuristic algorithms both adaptable and easy to understand and use. Adapting the algorithms to a specific problem is a breeze. Metaheuristic algorithms' exceptional ability to avoid premature convergence is their defining feature. The methods operate like a black box, avoiding local optima and efficiently and effectively exploring the search space, all because algorithms exhibit stochastic behavior. Algorithms compromise between exploration and exploitation, two of their most important features [16]. The algorithms do extensive exploration of the promising search space during the exploration phase, and then conduct local searches in the exploitation phase based on the areas that were discovered to be promising. Some examples of the many engineering and scientific fields that make good use of them include electrical engineering, industrial fields, civil engineering , communication , and data mining , prediction, clustering, and other techniques.

There are two primary kinds of metaheuristic algorithms: those that focus on a single solution and those that use a population of solutions. Algorithms that rely on a single solution to begin with update iterations, which could lead to their becoming stuck in a local optimum. In order to avoid local optima and explore the search space more thoroughly, algorithms that are based on populations produce a large number of solutions and update them with generations or iterations. The capacity of population-based algorithms to jump towards promising areas makes them useful for tackling most real-world issues [2].

Because of their unique qualities, metaheuristic algorithms (MH) attract a lot of interest from researchers. Problems of varying kinds have been addressed by a variety of algorithms. In light of The five main types of metaheuristic algorithms are those that are based on evolution (EA), swarm intelligence (SI), physics (PA), mathematics-based algorithms(MA), or human input(HA) as shown in Figure 4 [17].

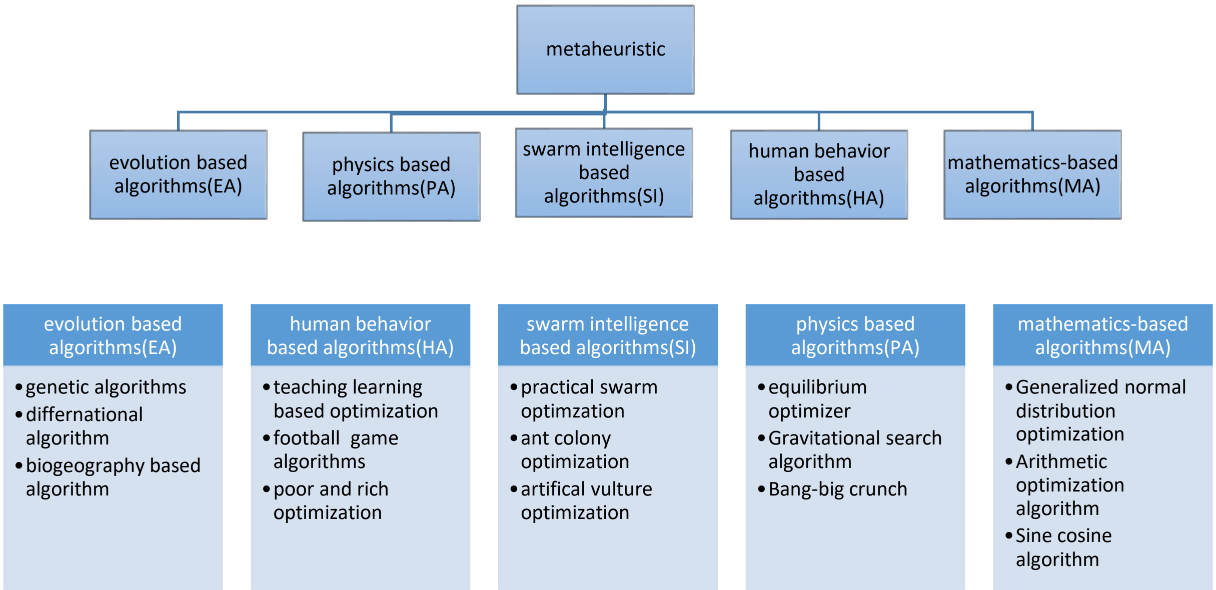

Figure 4. Classification of metaheuristics algorithms.

2.2.1 | Evolution based Algorithms (EA)

EAs draw on methods used in natural selection, including mutation, selection, reproduction, and recombination. An early EA that relied on the three behaviors of crossover, mutation, and selection was the genetic algorithm (GA) [18]. It finds the global optimum by applying these behaviors to a randomly generated population and then using the selection operator. Scheduling, routing, transit network design, feature selection, and traveling salesman are just a few of the optimization challenges that GA is employed for. Its efficacy is highly dependent on the population size, goal functions employed, and rate of crossover and mutation. A strong stochastic optimization technique, differential evolution (DE) uses GA to create two trial vectors by recombining two vectors [19].

BBO is a mathematical framework that models the distributions of biological organisms according to their geographic locations[20]. Quick evolutionary programming, memetic algorithms, cooperative coevolutionary algorithms, invasive tumor growth optimization techniques, and distribution estimation algorithms are among the other EAs.

2.2.2 | Swarm Intelligence based Algorithms (SI)

One kind of metaheuristics (MH) is (SI) algorithm, which attempts to model multi-agent systems after the ways in which creatures interact with one another in social settings. One of the first algorithms in this field, particle swarm optimization (PSO), employs a swarm of particles to find food by mimicking their collective behavior. It finds usage in a wide range of applications, including feature selection, image processing, knapsack issues, and cardiac disorders. Having said that, PSO is extremely sensitive to control parameters and can easily reach local optima. By utilizing three archives—the Lévy battle, mutation, and PSO—researchers have enhanced the convergence of PSO [21].

By combining the actions of scout, worker, and observer bees, the Artificial Bee Colony (ABC) program simulates the smart foraging strategies used by honeybees. On the other hand, ABC isn't without its flaws, like sluggish convergence, poor exploitation, and trouble identifying the best solution from among workable options [22]. While ant colony optimization (ACO) models ant foraging behavior, it is difficult to understand theoretically and nobody knows how long it will take for problems to converge [23].

The cuckoo search mimics the activities of cuckoos that are obligate parasites of their own broods. The main source of inspiration for the marine predator algorithm comes from the way predators in the water use biological interactions with their prey, Brownian motion, and the Lévy battle to find food. The literature also contains proposals for other SI algorithms [24].

2.2.3 | Physics based Algorithms (PA)

Numerous PA, including gravitational search, big bang-big crunch, equilibrium optimizer, and hysteretic optimization, have been developed by researchers by drawing on the laws of physics. These algorithms use Newtonian equations of motion and gravity to simulate the interaction of masses. There are two stages to the theories of the universe that the big bang-big crunch algorithm is based on. To estimate both dynamic and equilibrium states, the equilibrium optimizer employs control volume mass balance models. In magnetism, demagnetization is the source of inspiration for hysteretic optimization. The quantum salp swarm algorithm, the economical frefy algorithm, and quantum multi-verse optimization are only a few examples of the various MHs that feature quantum computing ideas. Charged system search, multi-verse optimization, optimization of gravitational interaction, optimization of thermal exchange, optimization of Henry gas solubility, and optimization of the central force are among the other PAs [25].

2.2.4 | Human Behavior based Algorithms (SI)

Many think humans are the most clever beings on the planet since we're always coming up with novel approaches to old challenges. Sports, social interactions, and politics are just a few examples of the human-centered activities and behaviors that impact MH algorithms [25].

2.2.5 | Mathematics-based Algorithms (MA)

Numerous subfields make up mathematics, such as statistics, probability, calculus, algebra, number theory, and basic arithmetic. Theoretical underpinnings of algorithms have served as sources of inspiration for researchers. An algorithm for sine cosine [26], an algorithm for generalized normal distribution optimization [27] , an algorithm for arithmetic optimization [28] , a new math-inspired algorithm based on the trigonometric sine function [29], and a new MA using basic arithmetic operators and a displacement parameter [30].

3 | Metaheuristic on Feature Selection

The article delves into a metaheuristic approach that use binary vector representations for feature selection. The solution vector of the algorithm is (10101100), where 1 denotes the selection of a feature and 0 denotes its non-selection. This work delves into every possible binary kind of metaheuristic algorithm, including those based on evolution, swarm intelligence, physics, humans, and hybrids. Section one delves into algorithms based on evolution, section two into algorithms based on swarm intelligence, section three into algorithms based on physics, part four into algorithm based on mathmeticas , part five into algorithms connected to humans and part six based on hybrid MHs.

3.1 |EA in Feature Selection

For more information on Genetic Algorithm (GA) and how it is used for feature selection, see [31].Using a binary GA specifically helped with the dimensionality. Enhancing the performance of the classifiers through minimization. From a batch of images, Flavia extracted one hundred (100) attributes. The following features were retrieved: Hu7M, LM, FD, and ZM, or Zertz Moments.

As stated in [32].Feature selection for the diagnosis of breast cancer is the focus of this study. Using a GA-based feature selection and PS-classifier, the current procedure employs a wrapper approach. The experimental results demonstrate that the suggested model performs similarly to the other models when applied to breast cancer in Wisconsin cancer databases.

To handle both continuous and discrete datasets, the differential evolution method presented in [33].integrates filter and wrapper techniques via an enhanced information theoretic local search mechanism grounded in fuzziness. To demonstrate the suggested method's efficacy, it is tested on several benchmarks sourced from various renowned data repositories and contrasted with both older and more contemporary evolutionary feature selection methods.

The difficult processes of detecting cardiac illness and choosing crucial features from the vast collection of accessible features are carried out in [34]. When dealing with classification issues, feature selection is a common pre-processing step. Feature selection and optimization for cardiovascular disease is carried out using a modified differential evolution (DE) algorithm. Compared to other models in use today, the suggested model has a higher accuracy rate of 83%.

In [35], introduces a novel method for selecting optimal feature subsets; it is known as the discrete binary differential evolution (BDE) algorithm. It is on the basis of mutual information that the relativity of qualities is assessed. and ran some tests with support vector machines (SVMs), convolutional neural networks (CNNs), and RBF networks preprocessed with the new feature selection strategy. On certain datasets, the approach significantly improves the accurate classification rate, and the BDE technique proves to be advantageous when selecting feature subsets.

3.2 |HA in Feature Selection

Binary teaching learning based optimization (FS-BTLBO) is a novel wrapper-based feature selection approach that introduced in [36].It requires just common controlling parameters, such as population size and the number of generations, to extract a subset of optimal features from the dataset. In order to measure the efficacy of the suggested system, have computed individual fitness using various classifiers as an objective function. Using the Wisconsin diagnostic breast cancer (WDBC) dataset, the results show that FS-BTLBO achieves better accuracy with less features for classifying benign and malignant tumors.

Using a binary-modified teaching learning-based optimization algorithm (BMTLBO), a new feature selection approach is presented in [37]. In terms of popularity and efficacy, the TLBO algorithm is among the best. This method can converge quickly, but it can also become trapped in a local optimum. A happy medium between exploration and exploitation is what aiming for. Both components make up the suggested approach: To begin, the feature selection problem is addressed using the BMTLBO algorithm, which incorporates the enhanced binary variant of the basic method. This method improves the algorithm's accuracy and convergence rate by increasing the population's variety through the usage of a pool. Secondly, demonstrateing the method's application to a classification issue and assess its performance by utilizing the SLTLBO neural network training algorithm, which is an upgraded version of the TLBO algorithm with a self-learning phase. In terms of classification accuracy and feature count, evaluated the suggested technique on fourteen datasets. The assessment findings indicate that, regarding accuracy, convergence rate, and efficacy in attaining favorable solutions, the suggested algorithm surpasses all other compared optimization methods. The outcomes are really encouraging and nearly optimum.

In [38]. A new hybrid feature selection approach called HBPRO is presented for obtaining the right subset of optimal features. This method is based on the binary poor and rich optimization algorithm. Using two well-known benchmark text corpus datasets, the proposed work evaluates the best feature subset using a Naive Bayes classifier. In comparison to previous feature selection methods, the experimental findings show that the suggested feature selection scheme (HBPRO) achieves better accuracy with less characteristics.

3.3 | SI in Feature Selection

In [39], they suggest a novel feature selection method inspired by cuckoo bird behavior: Binary Cuckoo Search. The suggested method was tested in a setting of power distribution system theft detection using two datasets acquired from a Brazilian electrical power business. The results showed that the method was resilient when compared to other optimization strategies inspired by nature.

In [40] suggests incorporating a new binary variant of the Cuckoo Search method with a pseudobinary mutation neighborhood search procedure. By minimizing the number of selected features and maximising the classification accuracy, the proposed Extended Binary Cuckoo Search method strives to accomplish the feature selection objective. In light of these requirements, we offer a new goal function for feature subset optimization that takes both the total number of features and classification accuracy into account. The Support Vector Machine classifier is employed to evaluate the accuracy of classification while employing a collection of candidate features. Using biomedical datasets, experiments were carried out to test the proposed Extended Binary Cuckoo Search optimization. The outcomes showed that the algorithm outperformed three other algorithms inspired by nature: Binary Ant Colony Optimisation, Binary Genetic Algorithm, and Binary Particle Swarm Optimisation.

3.4 | PA in Feature Selection

In [41], a new model for feature selection is suggested. The first stage is to use the mRMR feature selection method to score the features; the higher the score, the more features are chosen. Using the Improved Equilibrium Optimization (IMEO) method as a foundation, the second stage is to extract the best features using the Wrapper feature selection approach. By managing the algorithm's exploration and exploitation abilities, IMEO improves the speed of the Equilibrium Optimizer (EO) algorithm and employs a novel operator called Entropy-based to prevent it from getting trapped in local optima. In addition to enhancing the EO algorithm's exploration phase, Levy flying is employed to discover novel solutions within the search space. By inserting numerous little and occasionally big jumps into the search space, this technique avoids the local optimum. After making sure the suggested method works, the Binary Improved Equilibrium Optimization (BIMEO) algorithm is constructed using the sigmoid transfer function to address the feature selection issue. The last step is to test the IMEO algorithm on 23 benchmark test functions.

The issue of feature selection is addressed in [42] with the introduction of a novel method based on the Gravitational Search Algorithm (GSA). To give a quick and reliable framework for feature selection, the suggested technique integrates the optimization behavior of GSA with the speed of the Optimum-Path Forest (OPF) classifier. The proposed method is tested against Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), and a Particle Swarm Optimization (PSO)-based algorithm for feature selection using datasets harvested from various applications including vowel recognition, picture classification, and fraud detection in power distribution systems.

3.5 | MA in Feature Selection

Feature selection from medical data is suggested by two binary metaheuristic algorithms in [43]. the S-shaped binary Sine Cosine Algorithm (SBSCA) and the V-shaped binary Sine Cosine Algorithm (VBSCA). The search space is continuous in these methods, and for each solution, two transfer functions, one S-shaped and one V-shaped, are used to construct a binary position vector. We compare the suggested algorithms to four state-of-the-art binary optimization algorithms using five medical datasets stored in the UCI repository. Results from experiments show that when compared to four other algorithms, utilizing both bSCA variants improves classification accuracy on these medical datasets.

In [44], a novel pre-processing technique termed feature selection is introduced, which distinguishes beneficial features from those that may impair machine learning classifier performance owing to irrelevance, redundancy, or insufficient information. The method employs generalized normal distribution optimization (GNDO) and a restarting strategy (RS) to maintain variation among solutions. The strategy combines with GNDO to create an enhanced GNDO (IGNDO), which maintains variety and expedites convergence. The methods are evaluated against seven cutting-edge algorithms over thirteen medical datasets from the UCI library. IGNDO demonstrates superiority in fitness value and classification accuracy, while being competitive on selected criteria. IGNDO is regarded as the most effective method for identifying the optimal subset of features to enhance classification accuracy in the feature selection problem.

3.6 | Hybrid Algorithms in Feature Selection

In [45], introduces an innovative hybrid genetic algorithm for feature selection. Local search operations are integrated into hybrid genetic algorithms to optimize the search process. The procedures are characterized by their fine-tuning capability, and their efficacy and temporal demands are evaluated and contrasted. The hybridization procedure yields two advantageous outcomes: a notable enhancement in final performance and the attainment of subset-size regulation. The hybrid genetic algorithms exhibited superior convergence characteristics relative to the traditional genetic algorithms. A mechanism for conducting thorough timing analysis was established to compare the timing requirements of the standard and proposed methods. Experiments conducted with multiple standard data sets demonstrated that the suggested hybrid genetic algorithm outperforms both a basic genetic algorithm and sequential search techniques.

In [46], provides a hybrid technique that integrates two algorithms, namely Gray Wolf Optimization (GWO) and Particle Swarm Optimization (PSO), which facilitates the identification of key functions while allowing for the exclusion of insignificant ones and the reduction of complexity be eliminated. This facilitates the responsibilities of the machine learning classification by applying training to the classifier using the dataset. A hybrid approach largely relies on metaheuristic swarm intelligence algorithms that emulate the management and hunting behaviors of gray wolves in nature, as well as Particle Swarm Optimization (PSO), where individuals are influenced by their local optimal locations and the global optimal position. This hybridization aims to achieve a balance between exploitation and exploration. We utilized seventeen datasets from the UCI machine learning repository in the experiments and comparative analyses to evaluate the efficacy and quality of the GWOPSO.

In [47], provide a novel hybrid binary variant of the bat algorithm with an upgraded particle swarm optimization technique to address feature selection challenges. The suggested approach is termed Hybrid Binary Bat Enhanced Particle Swarm Optimization approach (HBBEPSO). The proposed HBBEPSO algorithm integrates the bat algorithm, utilizing its echolocation capabilities to navigate the feature space, with an advanced version of particle swarm optimization, which excels at converging to the optimal global solution inside the search space. To evaluate the overall efficacy of the proposed HBBEPSO algorithm, it is compared with both the original optimizers and other previously utilized optimizers for feature selection. A collection of assessment metrics is employed to evaluate and compare various optimizers over 20 standard datasets sourced from the UCI library. The results demonstrate the efficacy of the proposed HBBEPSO algorithm in exploring the feature space for optimal feature combinations.

4 | Issues and Challenges

Notwithstanding the considerable effectiveness of metaheuristic algorithms in addressing feature selection difficulties, certain hurdles and issues arise, which will be explained in the subsequent sections:

4.1 | Stability and Scalability

In practical applications, datasets may encompass thousands or even millions of features. To effectively address the feature selection challenge, the algorithm must exhibit scalability. A robust scalable classifier is imperative for managing extensive datasets. Consequently, scalability is a critical consideration in the development of algorithms for feature selection. Another significant aspect in designing such algorithms is stability. An algorithm is considered stable for feature selection if it consistently identifies the same subset of features across different dataset samples. In the pursuit of optimal classification, feature selection algorithms often demonstrate instability. This instability arises when there is a high correlation among features, leading to their removal in an effort to enhance classification accuracy. Thus, stability is as vital as classification accuracy itself [48].

4.2 | Building Objective Function

A wrapper feature selection method optimizes a specific objective function to identify the optimal feature subset. The formulation of an objective function for feature selection differs based on the classification task. An objective function was previously established that encompasses either the maximizing of classification accuracy or the minimizing of the number of selected characteristics. Furthermore, to reconcile the two opposing purposes, a multi-objective function was devised to address the feature selection problem. The multi-objective function was transformed into a single objective by assigning weights to each objective, followed by the implementation of the learning algorithm. Furthermore, the application of a multi-objective function proved to be highly effective and efficient in optimizing the fitness function and identifying the optimal feature subset from the provided datasets of features [2].

4.3 | Selection of Classifier

The selection of a classifier significantly influences the quality of the result in the design of a wrapper feature selection method. Various classifiers have been employed in addressing the feature selection problem through meta-heuristic algorithms, including K-nearest neighbor (KNN), Support Vector Machine (SVM), Optimum Path Forest (OPF), Naive Bayes (NB), Random Forest (RF), Artificial Neural Network (ANN), Fuzzy Rule-Based (FR), and Kernel Extreme Learning Machine (KLM). KNN is the most often utilized classifier using various datasets from the UCI repository. Conversely, the SVM classifier is commonly employed in intrusion detection systems and medical datasets, including cancer detection and arterial disease [2].

5 | Conclusion

This work provides a thorough examination of metaheuristic algorithms and their binary variations, which have been utilized in the feature selection problem. A comprehensive description and mathematical model of the feature selection challenge are provided to assist researchers in understanding the issue accurately. Furthermore, the methodologies for addressing feature selection issues are delineated. Furthermore, metaheuristic algorithms are employed to address the feature selection problem. Consequently, a fundamental definition, significance, and classification of metaheuristic algorithms are provided. Algorithms related to evolution-based, swarm-based,mathematics based ,and physics-based categories have been developed and employed for feature selection challenges. Nevertheless, metaheuristic algorithms possess some further limitations:

- They experience a sluggish convergence rate attributable to stochastic generation movement.
- They navigate the search space without a defined search direction.
- They may become ensnared in local optima or exhibit premature convergence.
- The parameters employed in the metaheuristic algorithms require calibration, which may also result in premature convergence.

In addition to the limitations of metaheuristic algorithms, improved and expanded versions have been developed and effectively applied to feature selection challenges. A category is provided based on algorithmic behavior: evolution-based, swarm-based, physics-related, mathematics-based, and human behavior-related algorithms.

Acknowledgments

The authors are grateful to the editorial and reviewers, as well as the correspondent author, who offered assistance in the form of advice, assessment, and checking during the study period.

Funding

This research has no funding source.

Conflicts of Interest

The authors declare that there is no conflict of interest in the research.

Ethical Approval

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

References

[1] Xue, B., et al., A survey on evolutionary computation approaches to feature selection. IEEE Transactions on evolutionary computation, 2015. 20(4): p. 606-626.

[2] Agrawal, P., et al., Metaheuristic algorithms on feature selection: A survey of one decade of research (2009-2019). Ieee Access, 2021. 9: p. 26766-26791.

[3] Lee, P.Y., W.P. Loh, and J.F. Chin, Feature selection in multimedia: the state-of-the-art review. Image and vision computing, 2017. 67: p. 29-42.

[4] Sharma, M. and P. Kaur, A comprehensive analysis of nature-inspired meta-heuristic techniques for feature selection problem. Archives of Computational Methods in Engineering, 2021. 28: p. 1103-1127.

[5] Akinola, O.O., et al., Multiclass feature selection with metaheuristic optimization algorithms: a review. Neural Computing and Applications, 2022. 34(22): p. 19751-19790.

[6] Kaur, S., et al., A systematic review on metaheuristic optimization techniques for feature selections in disease diagnosis: open issues and challenges. Archives of Computational Methods in Engineering, 2023. 30(3): p. 1863-1895.

[7] Liu, H. and L. Yu, Toward integrating feature selection algorithms for classification and clustering. IEEE Transactions on knowledge and data engineering, 2005. 17(4): p. 491-502.

[8] Inza, I., et al., Filter versus wrapper gene selection approaches in DNA microarray domains. Artificial intelligence in medicine, 2004. 31(2): p. 91-103.

[9] Forman, G., An extensive empirical study of feature selection metrics for text classification. Journal of machine learning research, 2003. 3(Mar): p. 1289-1305.

[10] Gonzalez, R.C., Digital image processing. 2009: Pearson education india.

[11] Lee, W., S.J. Stolfo, and K.W. Mok, Adaptive intrusion detection: A data mining approach. Artificial Intelligence Review, 2000. 14: p. 533-567.

[12] Zhao, Z., et al., Advancing feature selection research. ASU feature selection repository, 2010: p. 1-28.

[13] Guyon, I. and A. Elisseeff, An introduction to variable and feature selection. Journal of machine learning research, 2003. 3(Mar): p. 1157-1182.

[14] Jović, A., K. Brkić, and N. Bogunović. A review of feature selection methods with applications. in 2015 38th international convention on information and communication technology, electronics and microelectronics (MIPRO). 2015. Ieee.

[15] Olorunda, O. and A.P. Engelbrecht. Measuring exploration/exploitation in particle swarms using swarm diversity. in 2008 IEEE congress on evolutionary computation (IEEE world congress on computational intelligence). 2008. IEEE.

[16] Lin, L. and M. Gen, Auto-tuning strategy for evolutionary algorithms: balancing between exploration and exploitation. Soft Computing, 2009. 13: p. 157-168.

[17] Mohamed, A.W., A.A. Hadi, and A.K. Mohamed, Gaining-sharing knowledge based algorithm for solving optimization problems: a novel nature-inspired algorithm. International Journal of Machine Learning and Cybernetics, 2020. 11(7): p. 1501-1529.

[18] Holland, J.H., Genetic algorithms. Scientific american, 1992. 267(1): p. 66-73.

[19] Storn, R. and K. Price, Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. Journal of global optimization, 1997. 11: p. 341-359.
[20] Simon, D., Biogeography-based optimization. IEEE transactions on evolutionary computation, 2008. 12(6): p. 702-713.
[21] Eberhart, R. and J. Kennedy. Particle swarm optimization. in Proceedings of the IEEE international conference on neural networks. 1995. Citeseer.
[22] Karaboga, D. and B. Basturk. Artificial bee colony (ABC) optimization algorithm for solving constrained optimization problems. in International fuzzy systems association world congress. 2007. Springer.
[23] Dorigo, M., V. Maniezzo, and A. Colorni, Ant system: optimization by a colony of cooperating agents. IEEE transactions on systems, man, and cybernetics, part b (cybernetics), 1996. 26(1): p. 29-41.
[24] Yang, X.-S. and S. Deb. Cuckoo search via Lévy flights. in 2009 World congress on nature & biologically inspired computing (NaBIC). 2009. Ieee.
[25] Abdel-Basset, M., et al., Exponential distribution optimizer (EDO): a novel math-inspired algorithm for global optimization and engineering problems. Artificial Intelligence Review, 2023. 56(9): p. 9329-9400.
[26] Mirjalili, S., SCA: a sine cosine algorithm for solving optimization problems. Knowledge-based systems, 2016. 96: p. 120-133.
[27] Zhang, Y., Z. Jin, and S. Mirjalili, Generalized normal distribution optimization and its applications in parameter extraction of photovoltaic models. Energy Conversion and Management, 2020. 224: p. 113301.
[28] Abualigah, L., et al., The arithmetic optimization algorithm. Computer methods in applied mechanics and engineering, 2021. 376: p. 113609.
[29] Tanyildizi, E. and G. Demir, Golden sine algorithm: a novel math-inspired algorithm. Advances in Electrical & Computer Engineering, 2017. 17(2).
[30] Salem, S.A. BOA: A novel optimization algorithm. in 2012 international conference on engineering and technology (ICET). 2012. IEEE.
[31] Babatunde, O.H., et al., A genetic algorithm-based feature selection. 2014.
[32] Aalaei, S., et al., Feature selection using genetic algorithm for breast cancer diagnosis: experiment on three different datasets. Iranian journal of basic medical sciences, 2016. 19(5): p. 476.
[33] Hancer, E., Differential evolution for feature selection: a fuzzy wrapper–filter approach. Soft Computing, 2019. 23: p. 5233-5248.
[34] Vivekanandan, T. and N.C.S.N. Iyengar, Optimal feature selection using a modified differential evolution algorithm and its effectiveness for prediction of heart disease. Computers in biology and medicine, 2017. 90: p. 125-136.
[35] He, X., et al. Feature selection with discrete binary differential evolution. in 2009 international conference on artificial intelligence and computational intelligence. 2009. IEEE.
[36] Allam, M. and M. Nandhini, Optimal feature selection using binary teaching learning based optimization algorithm. Journal of King Saud University-Computer and Information Sciences, 2022. 34(2): p. 329-341.
[37] Khorashadizade, M. and S. Hosseini, An intelligent feature selection method using binary teaching-learning based optimization algorithm and ANN. Chemometrics and Intelligent Laboratory Systems, 2023. 240: p. 104880.
[38] Thirumoorthy, K. and K. Muneeswaran, Feature selection using hybrid poor and rich optimization algorithm for text classification. Pattern Recognition Letters, 2021. 147: p. 63-70.
[39] Rodrigues, D., et al. BCS: A binary cuckoo search algorithm for feature selection. in 2013 IEEE International symposium on circuits and systems (ISCAS). 2013. IEEE.
[40] Salesi, S. and G. Cosma. A novel extended binary cuckoo search algorithm for feature selection. in 2017 2nd international conference on knowledge engineering and applications (ICKEA). 2017. IEEE.
[41] Varzaneh, Z.A., et al., A new hybrid feature selection based on Improved Equilibrium Optimization. Chemometrics and Intelligent Laboratory Systems, 2022. 228: p. 104618.
[42] Papa, J.P., et al. Feature selection through gravitational search algorithm. in 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2011. IEEE.
[43] Taghian, S. and M.H. Nadimi-Shahraki, Binary sine cosine algorithms for feature selection from medical data. arXiv preprint arXiv:1911.07805, 2019.
[44] Abdel-Basset, M., et al., Medical feature selection approach based on generalized normal distribution algorithm. Computers, Materials and Continua, 2021. 69(3): p. 2883-2901.
[45] Oh, I.-S., J.-S. Lee, and B.-R. Moon, Hybrid genetic algorithms for feature selection. IEEE Transactions on pattern analysis and machine intelligence, 2004. 26(11): p. 1424-1437.
[46] El-Kenawy, E.-S. and M. Eid, Hybrid gray wolf and particle swarm optimization for feature selection. Int. J. Innov. Comput. Inf. Control, 2020. 16(3): p. 831-844.
[47] Tawhid, M.A. and K.B. Dsouza, Hybrid binary bat enhanced particle swarm optimization algorithm for solving feature selection problems. Applied Computing and Informatics, 2018. 16(1/2): p. 117-136.
[48] Bolón-Canedo, V., et al., On the scalability of feature selection methods on high-dimensional data. Knowledge and Information Systems, 2018. 56: p. 395-442.