International Journal of Computer Information Systems and Industrial Management Applications.
ISSN 2150-7988 Volume 14 (2022) pp. 285-296


Submitted: 27 December, 2021; Accepted: 5 April, 2022; Publish: 12 May, 2022

# Multi-Objective Evolutionary Algorithms and Metaheuristics for Feature Selection: a Review

**Duarte Coelho[1,2], Ana Madureira[2], Ivo Pereira[1,2,3] and Ramiro Gonçalves[4]**

[1]E-goi, Av. Menéres 840,
4450-190 Matosinhos, Portugal
{*dcoelho, ipereira*}*@e-goi.com*

[2]Interdisciplinary Studies Research Center,
Rua Dr. António Bernardino de Almeida 431,
4200-072 Porto, Portugal
*amd@isep.ipp.pt*

[3]University Fernando Pessoa, Praça 9 de Abril 349
4249-004 Porto, Portugal
*ivopereira@ufp.edu.pt*

[4]Universidade de Trás-os-Montes e Alto Douro (UTAD)
Vila Real, Portugal
*ramiro@utad.pt*

***Abstract***: **In the areas of machine learning / big data, when collecting data, sometimes too many features may be stored. Some of them may be redundant or irrelevant for the problem to be solved, adding noise to the dataset. Feature selection allows to create a subset from the original feature set, according to certain criteria. By creating a smaller subset of relevant features, it is possible to improve the learning accuracy while reducing the amount of data. This means means better results obtained in a shorter learning time. However, feature selection is normally regarded as a very important problem to be solved, as it directly impacts both data analysis and model creation. The problem of optimizing the selected features of a given dataset is not always trivial but, throughout the years, different ways to counter this optimization problem have been presented. This work presents how feature selection fits in the larger context of multi-objective problems as well as a review of how both multi-objective evolutionary algorithms and metaheuristics are being used in order to solve feature selection problems.**

***Keywords***: Big Data, Feature Selection, Multi-objective, Evolutionary Algorithms, Machine Learning

## I. Introduction

Data is at the core of machine learning. However, when collecting data, there are times when too many features may end up being stored in order to solve a given problem [11]. Features may be redundant or even irrelevant for the problem, meaning they only add more noise to the dataset [52].
In order to deal with this problem, feature selection (also known as dimensional reduction) is used to find the best subset of features that transmit most of the important information contained in the initial dataset. This step is a regular occurrence in any machine/deep learning pipeline (as seen in Figure 1). By removing data related errors/anomalies, various benefits can be ripped, such as better model interpretability, shorter training times, and reduced overfitting risk [64]. Additionally, depending on the problem, these algorithms may be able to reduce the operational and risk costs (i.e. in clinical trials) [52].
The literature for feature selection proposes many methods, each having advantages and downfalls. Generally two main types can be referred, namely, wrapper and filter techniques [52]. In the filter approach, features are selected based on a performance metric independently of the classifier being used. In wrapper approaches, a classifier is used to test each feature subset, which means they are classifier-dependent [52]. They distinguish themselves from one another via their speed and efficiency: *"filter methods are usually faster than wrapper methods since they have lower computational cost, however, wrapper methods have usually better performance than the filter methods since they select more representative features"* [64].
The main problem tackled by the different methods included in these two major approaches is that searching all possible subset spaces for a considerable feature number becomes an impossible operation, since it is too costly and restrictive. Instead of the best one, the accepted techniques tend to find an acceptable sub-optimum feature subset. To find these sub-optimal solutions both heuristic and random search methods can be applied [52, 64].

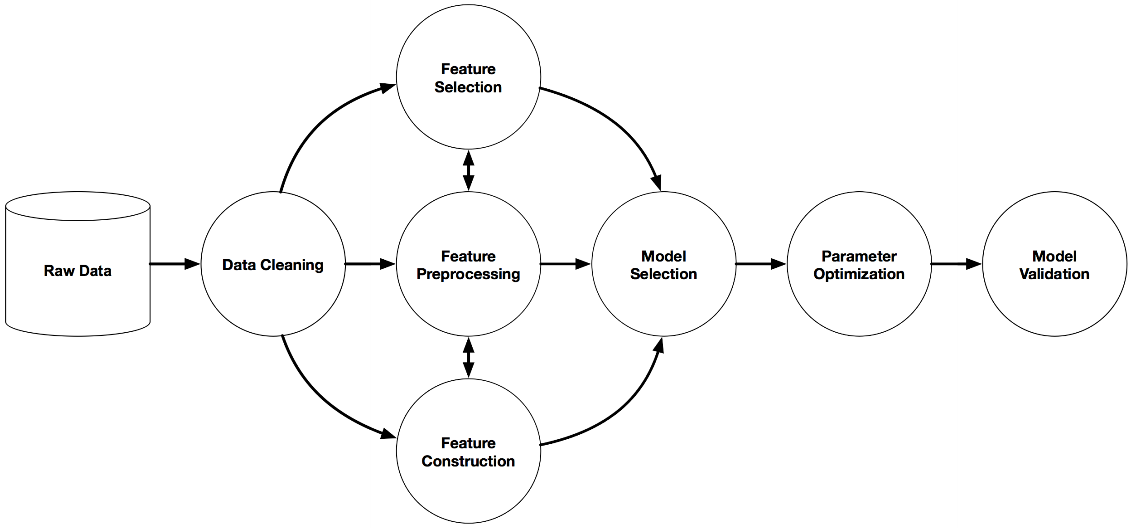


**Figure. 1**: Example of a Regular Machine-learning Pipeline (adapted from [44, 43])

When the idea of feature-selection first came about it was mostly defined as a single objective optimization problem. Since single metrics like classification accuracy are the only objective to be optimized. Nowadays, a more multi-objective oriented approach has been taken regarding this theme. In addition to simple metrics such as accuracy, this class of problems includes multiple objectives. Such as generalization capability for supervised classifiers and bias counterbalancing, with either a greater or smaller amount of features, for unsupervised approaches [48].

One type of solution that as been shown to be successfully used to solve this type of multi-objective problem has been meta-heuristics. Both more classical approaches as well as more recent ones such as nature inspired methods have shown good results in different scenarios [14, 16].

This paper extends the work on [12] and in the following sections the various notions needed to understand the problem in question will be presented and then some of the ways that have been proposed in order to solve it will be pt forward, followed by a brief discussion.

## II. Feature Selection in Data Science

Machine learning and data engineering pipelines regularly include a feature selection step. But how can we define it? Feature selection is the denomination given to the process through which a subset from an original feature set according to selection criterion defined bu the user.

By creating a smaller/more concise subset of relevant feature through an adequate selection criterion it is possible to improve learning accuracy and simplify obtained results while simultaneously reducing the scale of the problem (the amount of data involved), which means better results may be obtained in a shorter learning time [10].

Such effects are desirable since due to the great increased in both number of samples and dimensionality in most ML use cases the volume of high dimensional data has created big problems regarding it processing by existing machine learning methods. Needless is to say that a greater amount of data leads to greater computing times and more complex models. However, much of this dimensional problem relates to the presence of noisy, redundant and irrelevant dimensions. By, in essence, removing them from the problem in question one may expect the aforementioned problems to be solved or, at least, attenuated [39].

Feature selection methods can be classified taking into account different dimensions (Figure 2) [39]:

*1) Label Information*

- Supervised - These methods select and relevant features in order to distinguish samples from different classes.
- Semi-supervised - Similarly to the last case these try to distinguish between different classes. However, since there is little data targeting those labels they take advantage of both labeled data and unlabeled data.
- Unsupervised - Generally, feature selection for unsupervised problems is taken as a more difficult problem than the last two cases. The goal of feature selection for unsupervised learning is to find the feature subsets that reveal natural clusters in the processed data according to the chosen criterion.

*2) Search Strategy*

- Wrapper - These methods use the underlying ML algorithm itself to evaluate the features. That is to say, the selection criterion may be a performance metric used to evaluate the resulting model itself.
- Filter - These methods select the relevant features through specificities of data. Normally, filter methods perform feature selection before the ML algorithm (be it classification or clustering based), and usually fall into a two-step strategy.
- Embedded - In these the process of feature selection is something that is inherent to the process of constructing a model through the ML Algorithm being used. As such, various strategies may fall under this classification.

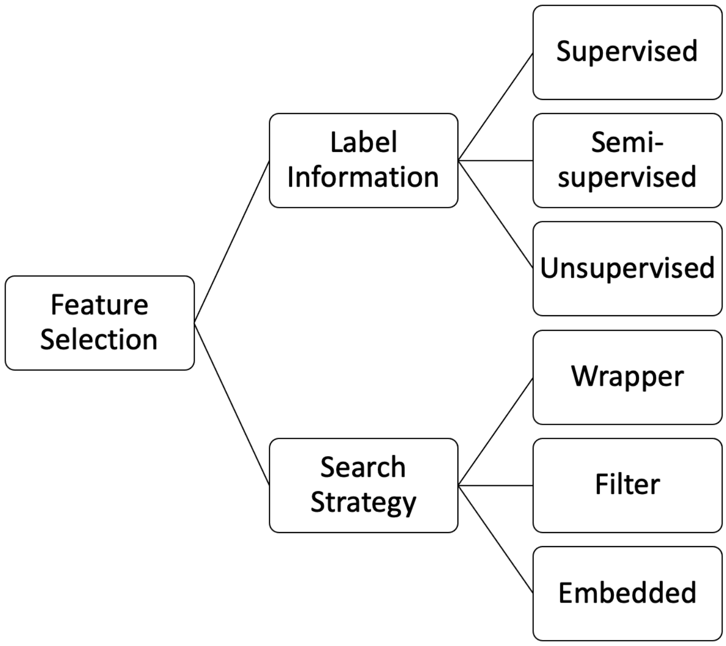


**Figure. 2**: Feature selection classification (adapted from [39])

## III. Multi-Objective Evolutionary Algorithms

Let us understand how Multi-Objective Evolutionary Algorithms (MOEAs) may be defined so then we can build upon that knowledge.

### *A. Definition*

In the context of multi-objective optimization, the principle of finding the most optimal solution cannot be applied to one objective alone, since the various other objectives are equally important. Meaning multi-objective optimization can be expressed by two goals[13]:

1. Convergence: find a (finite) set of solutions which lies on the Pareto-optimal[1] front.
2. Diversity: find a set of solutions which is diverse enough to represent the entire range of the Pareto optimal front.

MOEA try to establish their identity by following both the principles stated above, similar to a posteriori Multiple Criteria Decision-making Method (MCDM) [13]. Figure 49.2 schematically shows the principles followed in an Multi-Objective Evolutionary Procedures (MOEPs). Since MOEPs are heuristic based, they may not guarantee finding exact Pareto-optimal points, as a theoretically provable optimization method would do for tractable problems. However, MOEPs have essential operators to constantly improve the evolving non-dominated points (from the point of view of convergence and diversity mentioned above) similar to how most natural and artificial evolving systems continuously improve their solutions.

[1] Pareto optimality may be roughly defined as a state at which resources in a given system are optimized in a way that one dimension cannot improve without a second worsening. "The main idea of this concept is that a society is enjoying maximum ophelimity when no one can be made better off without making someone else worse off" [36].

The main difference and advantage of using a MOEA compared to a posteriori MCDM is that multiple trade-off solutions can be found in a single run of an MOEA, whereas most a posteriori MCDM methodologies would require multiple independent runs [13]. In Step 1 of the EMO-based multi-objective optimization and decision-making procedure, multiple trade-off, non-dominated points are found. Thereafter, in Step 2, higher-level information is used to choose one of the trade-off points obtained. This process can be seen in Figure 3.

### *B. Types of Algorithms*

Now that a definition for MOEAs has been presented, it is possible to introduce some of the broader notions used in multiple objective evolutionary approaches. In order to present this broader scope the following section serves as a short review of various frameworks used to develop MOEAs. Figure 4 summarizes the different types of MOEAs algorithms.

#### *1) Domination Based Algorithms*

This group of algorithms uses the dominance relation in the fitness assignment process, thus following suggestions presented in [23]. As the dominance relation itself does not preserve the diversity in the population, another techniques, such as niching, are needed to obtain a good spread of solutions. These algorithms have been very popular since mid 1990s [23]. An especially relevant algorithm in this category is refered as NSGA-II. Although NSGA-II is a relatively old algorithm it is still used today in various applications. Its main advantages are its relative effectiveness compared to some newer hypervolume based algorithms. Moreover, it is competitive with the modern algorithms, when optimizing for a low number of objectives [29]. However, the dominance relation is practically useless, if the number of objective functions increases. When the number of objectives

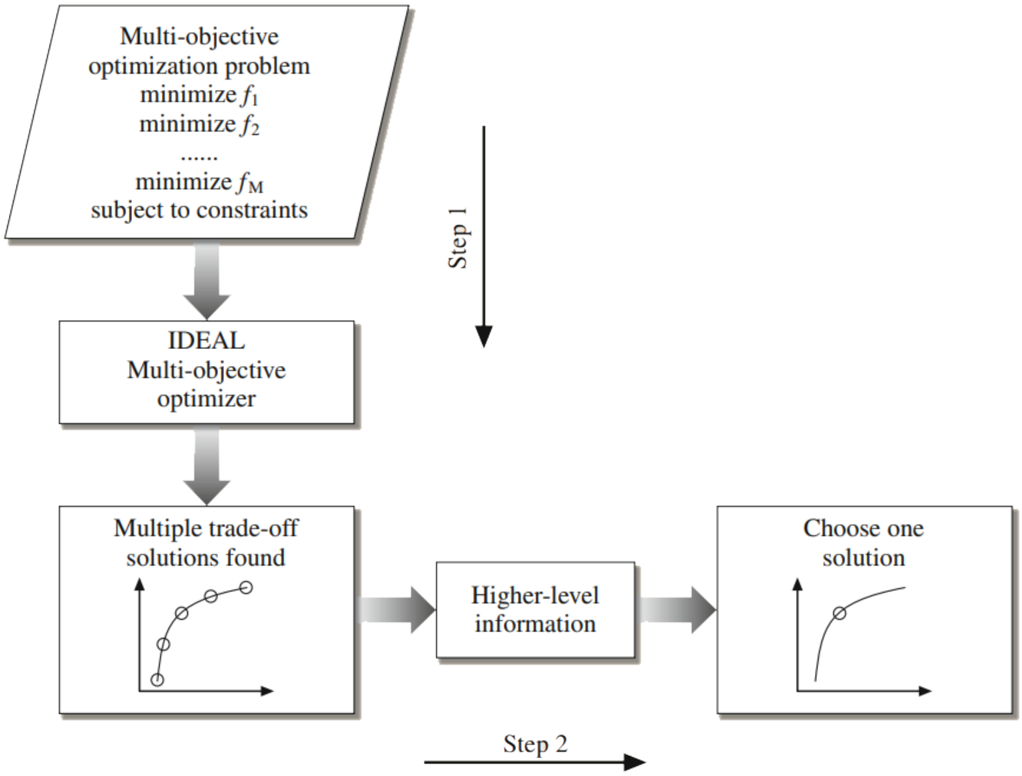


**Figure. 3**: Schematic of a Two-step Multi-criteria Optimization and Decision-making Procedure (adapted from [13])

is about ten most of randomly generated vectors cannot be compared using this relation. In that case, the selection pressure in NSGA-II (and other similar algorithms) is provided only by the niching procedure and does not guide the search towards the Pareto optimal set [29].

### *2) Decomposition Based Algorithms*

A recent framework is that of Multi-Objective Evolutionary Algorithms based on Decomposition (MOEA/D) [60]. It is based on conventional aggregation approaches in which an MOP is decomposed into a number of Scalar Objective optimization Problems (SOPs). The objective of each SOP, also called a sub-problem, is a (linearly or nonlinearly) weighted aggregation of the individual objectives. Neighborhood relations among these sub-problems are defined based on the distances between their aggregation weight vectors. Subproblem $i$ is a neighbor of sub-problem $j$ if the weight vector of sub-problem $i$ is close to that of sub-problem $j$. Each subproblem is optimized in the MOEA/D by using information mainly from its neighboring sub-problems.

In a simple version of the MOEA/D, each individual subproblem keeps one solution in its memory, which could be the best solution found so far for the sub-problem. For each sub-problem, the algorithm generates a new solution by performing genetic operators on several solutions from its neighboring sub-problems, and updates its memory if the new solution is better than old one for the sub-problem. A sub-problem also passes its newly generated solution on to some (or all) of its neighboring sub-problems, which will update their current solutions if the received solution is better. A major advantage of MOEA/Ds is that a scalar objective local search can be used in each sub-problem in a natural way since its task is to optimize a scalar objective [60].

Several improvements on MOEA/Ds have been made recently. Li and Zhang [35] suggested using two different neighborhood structures for balancing exploitation and exploration. Zhang et al. [61] proposed a scheme for dynamically allocating computational efforts to different subproblems in an MOEA/D in order to reduce the overall cost and improve the algorithm performance.

### *3) Preference Based Algorithms*

Due to the conflicts among the objectives in MOPs, the total number of Pareto optimal solutions might be very large or even infinite. However, the decision maker (DM) may be only interested in preferred solutions instead of all Pareto optimal solutions. To find the preferred solutions, the preference information is needed to guide the search towards the region of interest to the DM. Based on the role of the DM in the solution process, multi-objective optimization methods can be classified into priori methods, posteriori methods, and interactive methods [40].

In a priori method, preference information is given by the DM before the solution process. An MOP can be converted into an SOP. Then, a scalar objective solver is applied to find the desired Pareto optimal solution. A posteriori method uses the DM's preference information after the search process. A well distributed approximation is first obtained. Then, the DM selects the most preferred solutions based on its preferences. In an interactive method, the intermediate search results are presented to the DM to investigate; then the DM can understand the problem better and provide more preference

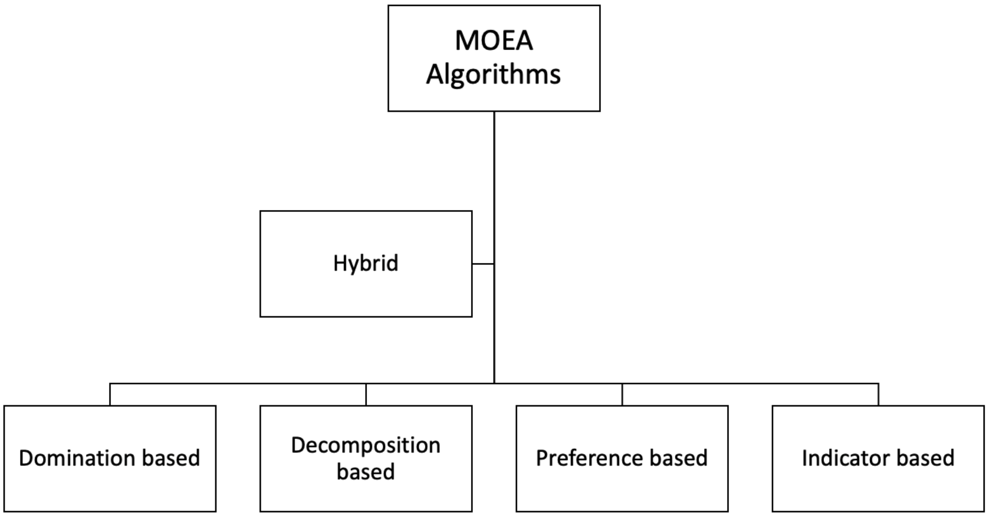


**Figure. 4**: MOEAs algorithms

information for guiding the search [40].
The earliest attempts on MOEAs based on the DM's preference were made by Fonseca and Fleming [18] and Tanino et al. [53] in 1993. In these algorithms, the rank of the members of a population is determined by both the Pareto dominance and the preference information from the DM. In [24], Greenwood et al. used value functions to rank the population, and preference information was also used in the survival criteria.

*4) Indicator Based Algorithms*

Indicator-based MOEAs use an indicator to guide the search, particularly to perform solution selection. Zitzler and Künzli [63] first suggested a general Indicator-Based Evolutionary Algorithm (IBEA). This approach uses an arbitrary indicator to compare a pair of candidate solutions. In comparison to other MOEAs, the IBEA only compares pairs of individuals instead of entire approximation sets.
In [5], Basseur and Zitzler proposed an indicator-based model for handling uncertainty, in which each solution is assigned a probability in the objective space. In an uncertain environment, some methods for computing expected indicator values are discussed, and several variants of their $\epsilon$-indicator-based model are suggested and empirically investigated.
Brockhoff and Zitzler [9] proposed a general approach to incorporate objective reduction techniques into hypervolume-based algorithms. Different objective reduction strategies are studied for improving the performance of hypervolume-based MOEAs.
In [4], Bader and Zitzler suggested a fast hypervolume-based MOEA for many-objective optimization. To reduce the computational overhead in hypervolume computation, a fast method based on Monte Carlo simulations is proposed to estimate the hypervolume value of an approximation set. Therefore, the proposed hypervolume-based MOEA may be applied to problems with many objectives.
Very recently, Bader and Zitzler [3] further investigated the robustness of hypervolume-based multi-objective search methods. Three existing approaches for handling robustness in the area of evolutionary computing, modifying the objective functions, additional objectives, and additional robustness constraints, are integrated into a multi-objective hypervolume-based search. An extension of the hypervolume indicator is also proposed for robust multi-objective optimization.

*5) Hybrid Algorithms*

In MOEAs, there are many techniques which have different characteristics and advantages. Hybridizing these techniques is thus a natural choice to utilize their advantages for dealing with complicated MOPs. What techniques to use and how to hybridize them are two major problems to solve when designing a hybrid MOEA. Some recent work could thus be categorized as follows.
Hybridizing different search methods: A general idea is to combine global search and local search methods, known as the memetic approach [33]. Another widely used idea is to combine the search operators of different algorithms. Particle Swarm Optimization (PSO) and Evolutionary Algorithms (EA) are hybridized in [17]. In each generation, the solutions generated by a PSO (EA) operator are then improved by an EA (PSO) operator. In [34], quantum operators are applied to solutions in binary representation and a genetic operator is then applied to the good solutions in permutation representation.
Hybridizing search and updating methods: This strategy hybridizes different components from different algorithms. For example, in [17], the PSO's operator is inserted into an EA's main loop.
Hybridizing different methods in different search phases: In the above two strategies, the hybrid methods are used in each generation. It is also natural to partition a search process into different phases and to use different search strategies in these phases. For example, in [58], the search is partitioned into three phases to emphasize dominated solutions, to balance dominated and non-dominated solutions, and to focus on non-dominated solutions, respectively. NSGA-II and a local incremental search algorithm are used to achieve the goals.

# IV. MOEA Based Feature Selection

Now that a generalized view of the different types of MOEAs have been presented, and the concept of feature selection has been explained. In the following section, some different approaches used to deal with feature selection will be presented so as to transmit an overview of the state of the art.

### A. *Multi-objective feature selection by using NSGA-II for customer churn prediction in telecommunications*

In this work [28], the authors propose a multi-objective feature selection approach applied to churn prediction in the telecommunication service field, based on the optimization approach NSGA-II. The main idea of this approach is to modify the approach NSGA-II to select local feature subsets of various sizes, and then to use the method of searching non-dominated solutions to select the global non-dominated feature subsets.

### B. *Feature Selection Using Multi-Objective Evolutionary Algorithms: Application to Cardiac SPECT Diagnosis*

In this article [19] the authors introduce an optimization methodology based on the use of MOEAs in order to deal with problems of feature selection in the context of cardiac diagnosis. For that purpose a Support Vector Machines (SVM) classifier was adopted. The aim being to select the best features and optimize the classifier parameters simultaneously while minimizing the number of features necessary and maximize the accuracy of the classifier and/or minimize the errors obtained. That is a A reduced Pareto set genetic algorithm (elitist) (RPSGAe) was adopted in by using an SVM to reduce the size of the Pareto optimal set. The obtained results were favorable to the approach.

### C. *Multi-objective evolutionary algorithms for filter based feature selection in classification*

In their work [57] the authors propose the adaption of NSGAII and SPEA2, in order to create two different filter based feature selection frameworks. Four multi-objective feature selection methods were then developed by applying mutual information and entropy as two different filter evaluation criteria in each of the two proposed frameworks. The results reached by the authors show that the proposed multi-objective algorithms can automatically evolve a set of non-dominated solutions that include a smaller number of features and achieve better classification performance than using all features. Additionally, NSGAII seems to achieves similar performance to SPEA2 for the datasets that consist of a small number of features and slightly better performance when the number of features is larger.

### D. *Feature selection of unreliable data using an improved multi-objective PSO algorithm*

In this work [59] the author proposes the use of an adapted multi-objective feature selection algorithm in order to deal with unreliable data. It accomplishes this by taking an effective multi-objective feature selection algorithm based on bare-bones particle swarm optimization and incorporating two new operators. One is a reinforced memory strategy, which is designed to overcome the degradation phenomenon of particles. Another is a hybrid mutation, which is designed to improve the search ability of the proposed algorithm. Comparison results suggest that the proposed algorithm is highly competitive for the proposed context.

# V. Metaheuristics

Since now a general picture on the topic of feature-selection as been painted, knowledge about metaheuristics: what they are, how they are classified and how they have been used in the past; will be presented, as to obtain a succinct state of the art of how this group of algorithms may be used to solve the same type of multi-objective problems referenced beforehand.

### A. *Definition*

Metaheuristics were first brought forward in order to define heuristic[2] problems that can be applied to a large set of different problems (mainly optimization). What this means is that, generally speaking, a metaheuristic can be taken as a generic algorithm framework which may be applied to various optimization problems with a relative low amount of effort [21, 7].
In recent years experts on the field have tried to improve the capabilities of various metaheuristics by combining them with other concepts/techniques of different fields such as operations research and/or artificial intelligence. These combinations are in large based on the exchange of information between optimization methods ran either sequentially or in parallel to each other. Such synergies of metaheuristics with foreign concepts are generally referred to as hybrid metaheuristics [8].

### B. *Categories*

As previously stated a vast amount of metaheuristics exist. Nevertheless, when discussing the topic, two major categories can be pointed out, namely: metaheuristics based on local search (or single solution based) and population-based metaheuristics [8]. Figure 5 presents some examples of Metaheuristics.
Metaheuristics based on local search have been a target of improvement study for years. The base problem with using a local search algorithm is that the starting point for the search heavily influences the local solution (minimum) that is found. This is a problem since the local optimum for one local search operator is usually not an optimum for another local search operator and this implies that the found local minimums may always be far off from the global minimum [49]. Such metaheuristics include algorithms such as tabu search [20] and simulated annealing [54].
Population based metaheuristics deal with each algorithm iteration with a set of solutions rather than with a single solution. For each iteration of the metaheuristic a new set of solutions is produced from the previous one based on operators. These operators define the rules by which solutions remain the same, are combined or even change [8]. This class

[2] Involving or serving as an aid to learning, discovery, or problem-solving by experimental and especially trial-and-error methods (according to the Merriam-Webster Dictionary)

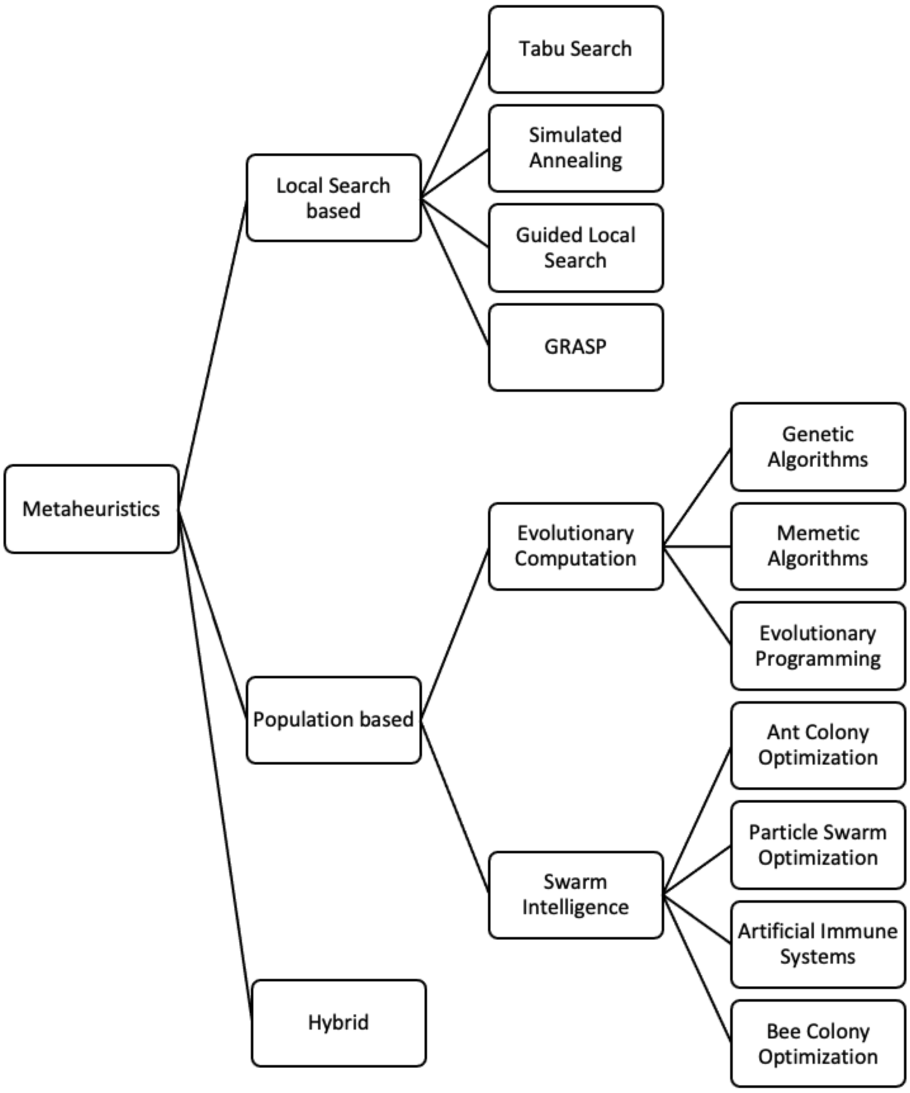


**Figure. 5**: Metaheuristics approaches

of metaheuristics include genetic algorithms [27], ant colony optimization [15] and particle swarm optimization [32].
When it comes to hybrid metaheuristics divisions are not as clear as in regular metaheuristics. Still a division in two categories can be established. These are hybrid metaheuristics combining parts of other metaheuristics and hybrid metaheuristics combining parts of other techniques [8].

### C. Metaheuristics Based Feature Selection

Both metaheuristics and hybrid metaheuristics have been described, and the concept of feature selection has been explained.
In the following section, some different approaches used to deal with feature selection will be described so as to transmit an overview of the state of the art [42, 37, 38, 1]. Take into account that most recent advances in the area rely on the hybridization of metaheuristics, meaning most of the presented works will reflect hybrid metaheuristics.

#### 1) *Feature selection using tabu search with learning memory: learning Tabu Search*

L. Mousin et al. [42] present an approach to feature selection based on a local-search metaheuristic. The authors consider the Feature Selection problem for classification as a combinatorial optimization one. They re-implement a tabu search algorithm firstly developed to solve a railway network problem, and then propose a learning mechanism in order to increase its performance. This learning mechanism works as a map that records the estimation of quality of each combination of features, which are computed from the quality of solutions where those combinations appear. This accelerates future iterations and is related to the pheromones concept of ant colony optimization. Afterwards, various experiments are performed in order to measure its efficiency. According to a data-mining perspective the authors solution ends up being better performing then the base algorithm, which may be explained by the small number of features selected by the proposed algorithm [42].

*2) **Hybrid whale optimization algorithm with simulated annealing for feature selection***

M. Mafarja et al. [37] presents another approach based on a different local-search metaheuristic. Their proposed approach combined a simulated annealing (SA) algorithm with the global search capabilities of a whale optimization algorithm (WO). Two different hybrid models were created, in one SA was used as a local search operator around the selected search agents in WO. By contrast, in the second one, SA was used to search the neighborhood of the best found solution after each iteration of WO. Both approaches performance was measured and compared. Two criteria were reported to evaluate each approach: classification accuracy, average selection size. It was found that the second approach, which used SA to intensify the neighboring region of the best solution found in each iteration of WO and tournament selection to select the search agents, showed the best performance among all proposed models [37].

*3) **Hybrid binary ant lion optimizer with rough set and approximate entropy reducts for feature selection***

M.Mafarja et al. [38] present two implementations of variants of an hybrid ant lion optimizer (ALO) for feature selection plus two different hill-climbing algorithms. One of these hill-climbing algorithms was quick reduct. Quick reduct is a set-based filter method for feature selection that simulates the forward generation method where the algorithm starts from an empty set and only features that improve a fitness value are added. The other hill-climbing method is an algorithm for reduction of knowledge with computing core (CEBARKCC), it works by finding the core features and adding them to the feature subset [38]. Both implementations were tested over various datasets and the approach combining ALO and quick reduct showed best results in terms of accuracy while the one combining ALO and CEBARKCC performed better regarding minimal reducts. Additionally, the authors also claim that both approaches also performed better than other hybridized ALO methods in most case studies [38].

*4) **Binary Optimization Using Hybrid Grey Wolf Optimization for Feature Selection***

Qasem et al. [1] propose a binary version of an hybrid metaheuristic based on grey wolf optimization (GWO) and particle swarm optimization (PSO). The authors argue that this is necessary since feature optimization is inherently a binary problem. They proceed to evaluate the proposed approach. In order to find the best solutions, the wrapper-based method K-nearest neighbors classifier with Euclidean separation metric is used. A set of evaluation measures over eighteen datasets were used to assess the proposed method. The results show that the proposed binary hybrid approach significantly outperformed the binary GWO, the binary PSO, the binary genetic algorithm, and the whale optimization algorithm with simulated annealing when using various performance measures including accuracy while selecting the best optimal features. Additionally, it presented better computational times [1].

*5) Other Techniques*

Other recent works, describing techniques for each metaheuristics based approach, relating to the problem of feature selection can be found through the following systematization:

*6) Filter Based:*

*"Differential evolution for filter feature selection based on information theory and feature ranking"*: new filter criterion inspired by concepts of mutual information such as ReliefF and Fisher Score [25].
*"Relevance–redundancy feature selection based on ant colony optimization"*: unsupervised and multivariate filter-based feature selection methods are proposed by analyzing the relevance and redundancy of features [51].
*"Weighted bee colony algorithm for discrete optimization problems with application to feature selection"*: improving the exploitation power of bee colony optimization (BCO), via allowing the bees to search in the solution space deliberately [41].
*"Optimizing Cuckoo Feature Selection Algorithm with the New Initialization Strategy and Fitness Function"*: new feature selection algorithm FS_CSO, which adopts the chaotic properties of the Chebyshev as a new initialization strategy to get the better original populations [56].

*7) Wrapper Based:*

*"A novel wrapper feature selection algorithm based on iterated greedy metaheuristic for sentiment classification"*: new wrapper feature selection algorithm based on an Iterated Greedy (IG) metaheuristic [22].
*"Bare bones particle swarm optimization with adaptive chaotic jump for feature selection"*: new Bare bones particle swarm optimization (BBPSO) inspired approach to feature selection [46].
*"A feature selection method based on modified binary coded ant colony optimization algorithm"*: new feature selection method based on a modified binary coded ant colony optimization algorithm [55].
*"Pareto front feature selection based on artificial bee colony optimization"*: new feature selection approach on a multi-objective artificial bee colony algorithm integrated with non-dominated sorting procedure and genetic operators [26].

*8) Hybrid Approaches:*

*"An Efficient hybrid filter-wrapper metaheuristic-based gene selection method for high dimensional datasets"*: an hybrid method based on the Incremental Wrapper Subset Selection with replacement (IWSSr) method and the Shuffled Frog Leaping Algorithm (SFLA) [45].
*"Modified binary cuckoo search for feature selection: a hybrid filter-wrapper approach"*: a feature selection method based on hybridization of mutual information feature selection (MIFS) filter and modified binary cuckoo search (MBCS) [30].
*"A new hybrid algorithm based on Grey wolf optimization and crow search algorithm for unconstrained function optimization and feature selection"*: an hybrid which combines

the strengths of both Grey wolf optimizer (GWO) and crow search algorithm (CSA) is proposed [2].

## VI. Discussion

Nowadays, most work being produced in the meta-heuristics research area is concerned with hybrid approaches. The greatest problem with these, is not one of performance or limitations but one of comprehension and lack of information.

The statement made by Christian Blum [6] in 2010 remain mostly true: *"the process of designing and implementing hybrid metaheuristics is rather complicated and involves knowledge about a broad spectrum of algorithmic techniques, programming and data structures, as well as algorithm engineering and statistics"*. This was true in the past and continues to be true nowadays. This is mostly due to the great amount of metaheuristics that exist that could contribute with as a positive feature for a given hybrid approach. Not to mention it is even possible to make combinations of hybrid approaches. In a way this problem to is one inherently related to combinatorial optimization. It could be said that when in the realm of hybrid approaches the solutions obtained start to rely more on the general planning and structure of the framework used then on the base algorithms being used.

This fact, however, should not dissuade researchers from continuing working on the area, as with such a great amount of metaheuristics there is always some extra component that might show positive contributions if only applied to an hybrid under the right conditions.

Similarly, the research and application in evolutionary multi-objective optimization over the in recent years has resulted in a number of efficient algorithms. MOEAs are now regularly applied to different problems in most areas of science, engineering, and commerce using in many cases metaheuristic based approaches. One area that currently seems especially enticing for researchers are collaborative EMO-MCDM (Evolutionary Multi-objective Optimization - Multiple-Criteria Decision-Making) algorithms for achieving a complete multi-objective optimization task to find a set of trade-off solutions and finally arriving at a single preferred solution. Another direction taken by researchers is to address guaranteed convergence and diversity of EMO algorithms through hybridizing them with mathematical and numerical optimization techniques similar to the current trend with metaheuristics [13].

When it comes to the MOEAs currently known as the bleeding-edge, these are the ones based on hybridization approaches, such as the ones used in [17, 34]. Even then more regular approaches such as NSGA and its variants, or algorithms as common as PSO can be found being applied to the problematic of feature-selection as seen in [31, 47] and [62], respectively.

As is possible to see the problematic of feature-selection continues to be recognized as a challenging multi-objective problem to solve. However, the creation of new metaheuristics as well as merging or inter-operation of existing ones, through hybridization, continues to produce better and more accurate solutions to face it.

## VII. Conclusion

Feature selection is an essential part of any machine-learning pipeline. This stage is important since the number of features of a model as well as their quality have been proven to affect model performance [50].

Feature selection has been one of the multi-objective problems long tackled by metaheuristics, since it is essentially a combinatorial optimization problem. Nowadays, most work in meta-heuristics is produced via hybridization of metaheuristics, between themselves, and other various techniques. That is to say, it refers creation of new and more effective approaches by combining older metaheuristics. This being the case it is natural that various hybrid algorithms have tried to tackle the long existing problem that is feature selection.

As we have seen throughout this work, results presented by solutions to this problem have been very promising and in the future even better performances seem all but assured.

## Acknowledgments

This article is a result of the project “Criação de um Núcleo de I&D para a geração de novo conhecimento nas áreas de Inteligência Artificial, Machine Learning, Intelligent Marketing e One-2-One Marketing”, supported by Operational Programme for Competitiveness and Internationalization (COMPETE 2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF).

## References

[1] Al-Tashi, Q., Kadir, S.J.A., Rais, H.M., Mirjalili, S., Alhussian, H.: Binary optimization using hybrid grey wolf optimization for feature selection. IEEE Access **7**, 39496–39508 (2019)

[2] Arora, S., Singh, H., Sharma, M., Sharma, S., Anand, P.: A new hybrid algorithm based on grey wolf optimization and crow search algorithm for unconstrained function optimization and feature selection. IEEE Access **7**, 26343–26361 (2019)

[3] Bader, J., Zitzler, E.: Robustness in hypervolume-based multiobjective search. Computer Engineering and Networks Laboratory (TIK), ETH Zurich, TIK Report **317** (2010)

[4] Bader, J., Zitzler, E.: Hype: An algorithm for fast hypervolume-based many-objective optimization. Evolutionary computation **19**(1), 45–76 (2011)

[5] Basseur, M., Zitzler, E.: Handling uncertainty in indicator-based multiobjective optimization. International Journal of Computational Intelligence Research **2**(3), 255–272 (2006)

[6] Blum, C., Puchinger, J., Raidl, G., Roli, A., et al.: A brief survey on hybrid metaheuristics. Proceedings of BIOMA pp. 3–18 (2010)

[7] Blum, C., Roli, A.: Metaheuristics in combinatorial optimization: Overview and conceptual comparison. ACM computing surveys (CSUR) **35**(3), 268–308 (2003)

[8] Blum, C., Roli, A., Sampels, M.: Hybrid metaheuristics: an emerging approach to optimization, vol. 114. Springer (2008)

[9] Brockhoff, D., Zitzler, E.: Improving hypervolume-based multiobjective evolutionary algorithms by using objective reduction methods. In: 2007 IEEE congress on evolutionary computation. pp. 2086–2093. IEEE (2007)

[10] Cai, J., Luo, J., Wang, S., Yang, S.: Feature selection in machine learning: A new perspective. Neurocomputing **300**, 70–79 (2018)

[11] Chotchantarakun, K., Sornil, O.: An adaptive multi-levels sequential feature selection. International Journal of Computer Information Systems and Industrial Management Applications **13**, 010–019 (2021)

[12] Coelho, D., Madureira, A., Pereira, I., Gonçalves, R.: A review on moea and metaheuristics for feature-selection. In: Proceedings of the 12th International Conference on Innovations in Bio-Inspired Computing and Applications (IBICA 2021) (2021)

[13] Deb, K.: Multi-Objective Evolutionary Algorithms, pp. 995–1015. Springer Berlin Heidelberg, Berlin, Heidelberg (2015)

[14] Diao, R., Shen, Q.: Nature inspired feature selection meta-heuristics. Artificial Intelligence Review **44**(3), 311–340 (2015)

[15] Dorigo, M., Birattari, M., Stutzle, T.: Ant colony optimization. IEEE computational intelligence magazine **1**(4), 28–39 (2006)

[16] Eid, H.F., Abraham, A.: Adaptive feature selection and classification using modified whale optimization algorithm. International Journal of Computer Information Systems and Industrial Management Applications **10**, 174–182 (2018)

[17] Elhossini, A., Areibi, S., Dony, R.: Strength pareto particle swarm optimization and hybrid ea-pso for multi-objective optimization. Evolutionary Computation **18**(1), 127–156 (2010)

[18] Fonseca, C.M., Fleming, P.J., et al.: Genetic algorithms for multiobjective optimization: Formulationdiscussion and generalization. In: Icga. vol. 93, pp. 416–423. Citeseer (1993)

[19] Gaspar-Cunha, A.: Feature selection using multi-objective evolutionary algorithms: application to cardiac spect diagnosis. In: Advances in bioinformatics, pp. 85–92. Springer (2010)

[20] Glover, F., Laguna, M.: Tabu search. In: Handbook of combinatorial optimization, pp. 2093–2229. Springer (1998)

[21] Glover, F.W., Kochenberger, G.A.: Handbook of metaheuristics, vol. 57. Springer Science & Business Media (2006)

[22] Gokalp, O., Tasci, E., Ugur, A.: A novel wrapper feature selection algorithm based on iterated greedy metaheuristic for sentiment classification. Expert Systems with Applications p. 113176 (2020)

[23] Goldenberg, D.E.: Genetic algorithms in search, optimization and machine learning (1989)

[24] Greenwood, G.W., Hu, X., D'Ambrosio, J.G.: Fitness functions for multiple objective optimization problems: Combining preferences with pareto rankings. In: Foga. vol. 96, pp. 437–455 (1996)

[25] Hancer, E., Xue, B., Zhang, M.: Differential evolution for filter feature selection based on information theory and feature ranking. Knowledge-Based Systems **140**, 103–119 (2018)

[26] Hancer, E., Xue, B., Zhang, M., Karaboga, D., Akay, B.: Pareto front feature selection based on artificial bee colony optimization. Information Sciences **422**, 462–479 (2018)

[27] Holland, J.H.: Genetic algorithms. Scientific american **267**(1), 66–73 (1992)

[28] Huang, B., Buckley, B., Kechadi, T.M.: Multi-objective feature selection by using nsga-ii for customer churn prediction in telecommunications. Expert Systems with Applications **37**(5), 3638–3646 (2010)

[29] Ishibuchi, H., Tsukamoto, N., Nojima, Y.: Evolutionary many-objective optimization: A short review. In: 2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence). pp. 2419–2426. IEEE (2008)

[30] Jiang, Y., Liu, X., Yan, G., Xiao, J.: Modified binary cuckoo search for feature selection: a hybrid filter-wrapper approach. In: 2017 13th International Conference on Computational Intelligence and Security (CIS). pp. 488–491. IEEE (2017)

[31] Jiménez, F., Jodár, R., Sánchez, G., Martín, M., Sciavicco, G.: Multi-objective evolutionary computation based feature selection applied to behaviour assessment of children. In: Proceedings of the 9th International Conference on Educational Data Mining EDM. pp. 1888–1897 (2016)

[32] Kennedy, J., Eberhart, R.: Particle swarm optimization. In: Proceedings of ICNN'95-International Conference on Neural Networks. vol. 4, pp. 1942–1948. IEEE (1995)

[33] Lara, A., Sanchez, G., Coello, C.A.C., Schutze, O.: Hcs: A new local search strategy for memetic multiobjective evolutionary algorithms. IEEE Transactions on Evolutionary Computation **14**(1), 112–132 (2009)

[34] Li, B.B., Wang, L.: A hybrid quantum-inspired genetic algorithm for multiobjective flow shop scheduling. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) **37**(3), 576–591 (2007)

[35] Li, H., Zhang, Q.: Multiobjective optimization problems with complicated pareto sets, moea/d and nsga-ii. IEEE transactions on evolutionary computation **13**(2), 284–302 (2008)

[36] Luc, D.T.: Pareto Optimality, pp. 481–515. Springer New York, New York, NY (2008)

[37] Mafarja, M.M., Mirjalili, S.: Hybrid whale optimization algorithm with simulated annealing for feature selection. Neurocomputing **260**, 302–312 (2017)

[38] Mafarja, M.M., Mirjalili, S.: Hybrid binary ant lion optimizer with rough set and approximate entropy reducts for feature selection. Soft Computing **23**(15), 6249–6265 (2019)

[39] Miao, J., Niu, L.: A survey on feature selection. Procedia Computer Science **91**, 919–926 (2016)

[40] Miettinen, K.: Nonlinear multiobjective optimization, vol. 12. Springer Science & Business Media (2012)

[41] Moayedikia, A., Jensen, R., Wiil, U.K., Forsati, R.: Weighted bee colony algorithm for discrete optimization problems with application to feature selection. Engineering Applications of Artificial Intelligence **44**, 153–167 (2015)

[42] Mousin, L., Jourdan, L., Marmion, M.E.K., Dhaenens, C.: Feature selection using tabu search with learning memory: learning tabu search. In: International Conference on Learning and Intelligent Optimization. pp. 141–156. Springer (2016)

[43] Olson, R.S., Moore, J.H.: Tpot: A tree-based pipeline optimization tool for automating machine learning. In: Automated Machine Learning, pp. 151–160. Springer (2019)

[44] Olson, R.: Tpot (tree-based pipeline optimization tool) (2017), `https://github.com/EpistasisLab/tpot`

[45] Pirgazi, J., Alimoradi, M., Abharian, T.E., Olyaee, M.H.: An efficient hybrid filter-wrapper metaheuristic-based gene selection method for high dimensional datasets. Scientific Reports **9**(1), 1–15 (2019)

[46] Qiu, C.: Bare bones particle swarm optimization with adaptive chaotic jump for feature selection in classification. International Journal of Computational Intelligence Systems **11**(1), 1–14 (2018)

[47] Sahoo, A., Chandra, S.: Multi-objective grey wolf optimizer for improved cervix lesion classification. Applied Soft Computing **52**, 64–80 (2017)

[48] Sheth, P., Patil, S.: A review on feature selection problem solving using multiobjective evolutionary optimization algorithms. International Journal of Engineering Applied Sciences and Technology **2**(9), 42–54 (2018)

[49] Sorensen, K., Sevaux, M., Glover, F.: A history of metaheuristics. arXiv preprint arXiv:1704.00853 (2017)

[50] Sugumaran, V., Ramachandran, K.: Effect of number of features on classification of roller bearing faults using svm and psvm. Expert Systems with Applications **38**(4), 4088–4096 (2011)

[51] Tabakhi, S., Moradi, P.: Relevance–redundancy feature selection based on ant colony optimization. Pattern recognition **48**(9), 2798–2811 (2015)

[52] Talbi, E.G., et al.: Hybrid metaheuristics, vol. 166. Springer (2013)

[53] Tanino, T., Tanaka, M., Hojo, C.: An interactive multicriteria decision making method by using a genetic algorithm. In: 2nd International Conference on Systems Science and Systens Engineering (1993)

[54] Van Laarhoven, P.J., Aarts, E.H.: Simulated annealing. In: Simulated annealing: Theory and applications, pp. 7–15. Springer (1987)

[55] Wan, Y., Wang, M., Ye, Z., Lai, X.: A feature selection method based on modified binary coded ant colony optimization algorithm. Applied Soft Computing **49**, 248–258 (2016)

[56] Wang, Y., Li, Z., Yu, H., Deng, L.: Optimizing cuckoo feature selection algorithm with the new initialization strategy and fitness function. In: International Conference on Cloud Computing and Security. pp. 719–730. Springer (2018)

[57] Xue, B., Cervante, L., Shang, L., Browne, W.N., Zhang, M.: Multi-objective evolutionary algorithms for filter based feature selection in classification. International Journal on Artificial Intelligence Tools **22**(04), 1350024 (2013)

[58] Yang, D., Jiao, L., Gong, M.: Adaptive multi-objective optimization based on nondominated solutions. Computational Intelligence **25**(2), 84–108 (2009)

[59] Yong, Z., Dun-wei, G., Wan-qiu, Z.: Feature selection of unreliable data using an improved multi-objective pso algorithm. Neurocomputing **171**, 1281–1290 (2016)

[60] Zhang, Q., Li, H.: Moea/d: A multiobjective evolutionary algorithm based on decomposition. IEEE Transactions on evolutionary computation **11**(6), 712–731 (2007)

[61] Zhang, Q., Liu, W., Li, H.: The performance of a new version of moea/d on cec09 unconstrained mop test instances. In: 2009 IEEE congress on evolutionary computation. pp. 203–208. IEEE (2009)

[62] Zhang, Y., Gong, D.w., Cheng, J.: Multi-objective particle swarm optimization approach for cost-based feature selection in classification. IEEE/ACM transactions on computational biology and bioinformatics **14**(1), 64–75 (2015)

[63] Zitzler, E., Künzli, S.: Indicator-based selection in multiobjective search. In: International conference on parallel problem solving from nature. pp. 832–842. Springer (2004)

[64] Zorarpacı, E., Özel, S.A.: A hybrid approach of differential evolution and artificial bee colony for feature selection. Expert Systems with Applications **62**, 91–103 (2016)

## Author Biographies

**Duarte Coelho** was born in Portugal, in 1995. He got his BSc degree in Computer Engineering in 2016 from ISEP, master's degree in Computational Systems, in 2018, from ISEP. He is currently a Data Scientist at E-goi, Matosinhos, Portugal, Researcher on the Interdisciplinary Studies Research Center. His research interests involve artificial intelligence, machine learning, optimization algorithms, intelligent marketing and computational intelligence.

**Ana Madureira** was born in Mozambique, in 1969. She got his BSc degree in Computer Engineering in 1993 from ISEP, master's degree in Electrical and Computers Engineering–Industrial Informatics, in 1996, from FEUP, and the PhD degree in Production and Systems, in 2003, from University of Minho, Portugal. She became IEEE Senior Member in 2010. She was Chair of IEEE Portugal Section (2015-2017), and Chair/Vice-Chair of IEEE-CIS Portuguese chapter. She was Chair of University Department of IEEE R8 Educational Activities Sub-Committee (2017-2018). She is IEEE R8 Secretary (2019-2020). She is External Member Evaluation Committee of the Agency for Assessment and Accreditation of Higher Education - A3ES for the scientific area of Informatics of Polytechnic Higher Education (since 2012). Currently she is Coordinator Professor at the Institute of Engineering–Polytechnic of Porto (ISEP/IPP) and Director of the Interdisciplinary Studies Research Center (ISRC). In the last few years, she was author of more than 100 scientific papers in scientific conference proceedings, journals, and books.

**Ivo Pereira** was born in Porto, Portugal, in 1984. He got his BSc degree in Informatics Engineering by Institute of Engineering-Polytechnic of Porto in 2006, MSc degree in Informatics Engineering by Institute of Engineering-Polytechnic of Porto in 2009 and PhD in Electronics and Computers Engineering by University of Trás-os-Montes and Alto Douro in 2014. He is currently an Assistant Professor at University Fernando Pessoa, Porto, Portugal, Researcher on the Interdisciplinary Studies Research Center and Consultant at E-goi. His research interests involve artificial intelligence, machine learning, optimization algorithms, intelligent marketing and computational intelligence.

**Ramiro Gonçalves** is an Associate Professor with Habilitation at the University of Trás-os-Montes e Alto Douro, in Vila Real, Portugal, Senior Researcher at INESC TEC Associated Laboratory, in Porto, Portugal, and Executive Director of AquaValor - Centro de Valorização e Transferência de Tecnologia da Água (CoLAB), in Chaves, Portugal. Ramiro has over 200 publications (including book chapters, Scientific Citation Index journal articles, as well as publications in refereed conference proceedings). Ramiro has also been linked to industry throughout his career, and more recently as a manager in the public sector.