




OPEN

A multistart tabu search-based method for feature selection in medical applications

Joaquín Pacheco , Olalla Saiz, Silvia Casado & Silvia Ubbillos

In the design of classification models, irrelevant or noisy features are often generated. In some cases, there may even be negative interactions among features. These weaknesses can degrade the performance of the models. Feature selection is a task that searches for a small subset of relevant features from the original set that generate the most efficient models possible. In addition to improving the efficiency of the models, feature selection confers other advantages, such as greater ease in the generation of the necessary data as well as clearer and more interpretable models. In the case of medical applications, feature selection may help to distinguish which characteristics, habits, and factors have the greatest impact on the onset of diseases. However, feature selection is a complex task due to the large number of possible solutions. In the last few years, methods based on different metaheuristic strategies, mainly evolutionary algorithms, have been proposed. The motivation of this work is to develop a method that outperforms previous methods, with the benefits that this implies especially in the medical field. More precisely, the present study proposes a simple method based on tabu search and multistart techniques. The proposed method was analyzed and compared to other methods by testing their performance on several medical databases. Specifically, eight databases belong to the well-known repository of the University of California in Irvine and one of our own design were used. In these computational tests, the proposed method outperformed other recent methods as gauged by various metrics and classifiers. The analyses were accompanied by statistical tests, the results of which showed that the superiority of our method is significant and therefore strengthened these conclusions. In short, the contribution of this work is the development of a method that, on the one hand, is based on different strategies than those used in recent methods, and on the other hand, improves the performance of these methods.

The introduction is organized into 4 sub-sections: “[Motivation](#)”, “[Literature on feature selection methods](#)”, “[Literature on feature selection in medicine](#)”, and “[Contribution](#)”.

Motivation

Diagnosis is a process in which a disease, injury, or other adverse effect is identified based on its signs and symptoms. To aid the establishment of a diagnosis, the clinician may use the patient’s medical history or conduct a physical examination and various tests, such as blood tests, imaging tests, and biopsies. In most pathologies, diagnosis has evolved toward faster and more accurate processes¹.

In medicine, early diagnosis is vital for the effective treatment of some pathologies. For example, early detection improves patient survival in pathologies such as pancreatic cancer². Moreover, early diagnosis is crucial to improve the prognosis of many diseases, e.g., tumors and melanomas³. In other diseases, such as diabetes, early diagnosis has become essential to alleviate the secondary complications of the disease, e.g., arterial lesions, hypercholesterolemia, hypertension, obesity, and myocardial infarction⁴. The diagnosis of Alzheimer’s disease in its earliest stages is very important, as it allows this disease to be differentiated from other neurodegenerative disorders with dementia⁵.

Data mining^{6,7} is a set of tools that have enabled the analysis of large amounts of information, generating patterns and norms that help to understand the behavior of a system. Within this set of tools, those used in prediction—and especially classification—have been adapted to the diagnosis of diseases. Their use in this scope has become an important emerging line of research⁸. With different models, such as those based on logistic regression⁹, discriminant analysis¹⁰, support vector machine¹¹, neural networks¹², classification trees¹³, nearest neighbor search¹⁴, and Bayesian classifiers¹⁵, among other approaches, it is possible to establish an early diagnosis.

University of Burgos, Burgos, Spain.  email: jpacheco@ubu.es

Furthermore, in addition to good diagnostic accuracy, the models must be able to identify the most influential variables and features—that is, the ones that are truly relevant for diagnosis. In other words, it is useful to know which set of features optimize precision and other performance parameters in a diagnostic task. In previous works, the goal has been to determine which set of features, among the original ones, performs the diagnostic task in an optimal way ("optimal set").

In practice, this optimal set of features is generally unknown a priori, and the initial model usually contains irrelevant or redundant features¹⁶. Identifying the most relevant set of features and separating them from the redundant and/or irrelevant features is critical for successful diagnosis¹⁷. Therefore, it is important to incorporate feature selection in model creation to ensure the generation of simpler and more understandable models. Being able to reduce the complexity of a model may help to improve its precision and performance¹⁸ as well as the economy and speed with which data can be acquired and prepared^{19,20}.

Literature on feature selection methods

Feature selection methods can be classified into three types: filter, wrapper, and embedded¹⁹. Filter methods select a certain number of variables based on criteria such as correlation, likelihood, and information gain without the intervention of a classifier^{19,20}. Some examples of filter methods are the following algorithms: correlation feature selection (or the correlation algorithm)⁶, mutual information¹⁶, the ReliefF algorithm⁷, the chi-square algorithm²¹, the Fisher score algorithm²², and the fast correlation-based filter²³. Additionally, Hancer²⁴ proposed a filter method based on differential evolution.

Wrapper methods explore different combinations or subsets of variables to evaluate the usefulness of each of these subsets through the predictive efficacy of a specific classifier. The aim is to find the best of these subsets. Wrapper methods usually generate better results than filter methods, since the latter do not evaluate the performance of the selected variables with a certain classifier¹⁷, although wrapper methods require longer computation times than filter methods. Due to the robustness of metaheuristic techniques in several complex applications, some of these techniques have been used to create wrapper methods for feature selection, such as genetic algorithms^{25–27}, the gray wolf optimizer^{28,29}, the flower pollination algorithm^{30,31}, the bat algorithm^{32,33}, the ant colony optimization algorithm^{34,35}, the whale optimization algorithm^{36,37}, particle swarm optimization^{38,39}, the harmony search algorithm^{40,41}, and the Harris hawk optimization algorithm^{18,42}.

Finally, embedded methods integrate feature selection and classifier learning in a single process. These methods have been used in studies as^{43–46}, and⁴⁷.

Literature on feature selection in medicine

Different approaches have been used to select variables/features and obtain models that establish diagnoses with high precision⁴⁸. For example⁴⁹, extracted different features of magnetic resonance imaging (MRI) data to select the most important characteristics for the diagnosis of brain tumors. Additionally, from MRI data⁵⁰, defined biomarkers by quantifying the precision of different sets of morphological features to diagnose Alzheimer's disease. Liu et al.⁵¹ developed a method to measure the performance of both individual features and different subsets of features to identify lung diseases. Similarly, Chong et al.⁵² created a classifier that enabled the detection of fibrotic interstitial lung disease using 3D texture features of computed tomography images. Shi et al.⁵³ used a method based on feature selection that incorporates the clinician's knowledge to achieve an accurate diagnosis of prostate problems. Additionally, Guinin et al.⁵⁴ developed an automatic diagnostic tool for prostate pathology, and Sahran et al.⁵⁵ proposed a new feature selection method from prostate histopathological images. Furthermore, feature selection studies have been conducted for different types of cancers. Thus, Jain et al.⁵⁶ proposed a model to improve the diagnosis and the identification of cancer types. Wang et al.⁵⁷ used a feature selection strategy for the diagnosis and classification of cancer with different gene expression profiles. Peng et al.⁵⁸ used a forward feature selection algorithm for tumor classification. Finally, Kang et al.⁵⁹ combined feature selection, logistic regression models, and support vector regression for tumor classification. Feature selection algorithms have also been combined with the AdaBoost classifier for the detection of glaucoma⁶⁰.

An interesting and recent line of work in this field is that of methods based on evolutionary strategies inspired by biological metaphors. Thus, in Awadallah et al.^{61,62}, strategies based on Rat Swarm and Horse Herd behavior were used. In order to improve these methods, different operators are analyzed. On the other hand, Braik et al.⁶³ proposed three methods based on the Capuchin Search Algorithm, considering the k-Nearest Neighbor (k-NN) classifier.

One interesting field of application of feature selection in bio-medicine is gene interaction detection, such as single nucleotide polymorphisms (SNPs) epistatic interaction. In many cases, different metaheuristic strategies have been used. Thus, in the works of Tuo et al.^{64,65} this problem was approached from a multi-objective and multi-task optimization point of view. In both studies, methods based on harmony search optimization were proposed. In⁶⁶ a similar approach was taken and a method based on ant colony optimization was proposed. Shang et al.⁶⁷ proposed a simulation method combined with resampling methods.

In this paper, we propose a method based on tabu search and multi-start strategies. This is a stark difference with recent work based on evolutionary methods (such as those mentioned above). Evolutionary methods are, in general, more intuitive and easier to adapt and implement. Methods based on local search, such as the one proposed in this work, require greater adaptation to each specific problem and a greater implementation effort. On the other hand, our experience in different fields^{68–70} demonstrates that these local-search based methods (and specifically those based on tabu search) perform better than evolutionary methods. Indeed, as will be seen below, different computational tests show that our proposed method outperformed other recent methods as gauged by various metrics and classifiers.

Method *MultiStartTabu*

$iter = 0, iterbest = 0, f^{best} = 0$

Repeat

1. $iter = iter + 1$
2. Execute *Constructive*(α, S)
3. Execute *TabuSearch*(*tenure*, *maxiter*, S, S^*)
4. **if** ($f(S^*) > f^{best}$) **then** $S^{best} = S^*, f^{best} = f(S^*)$ and $iterbest = iter$

until ($iter - iterbest \geq maxiterMS$)

Pseudocode 1. *MultiStartTabu* method.

Contribution

The present study proposes a wrapper method of variable/feature selection that combines two metaheuristic strategies of combinatorial optimization: tabu search (TS) and multistart. Computational tests including several different classifiers were carried out with different diagnostic databases. These tests showed that our tool outperformed other known wrapper methods in the recent literature in terms of different metrics and classifiers. Our method obtains more efficient models—i.e. a better balance between the number of variables selected in the model and its precision. The results of various statistical tests support these conclusions.

The rest of the study is organized in the following manner: Section “[Notation and formulation of the research problem](#)” formulates the research problem formally; Section “[Solution method for the proposed problem](#)” describes the proposed method in detail; Section “[Computational tests](#)” describes the computational tests conducted to analyze the performance of our method against the other methods; and finally, Section “[Conclusions](#)” presents the conclusions.

Notation and formulation of the research problem

To formulate the research problem that is addressed in this study (selection of variables for diagnosis), the data used to select the variables and to generate the models were defined as the training set. This dataset is X . The number of individuals/cases of X is n . The number of variables is m . Finally, the set of variables is V , and the j th variable is v_j , that is:

$$V = \{v_1, v_2, v_3, \dots, v_j, \dots, v_m\}$$

For each individual of X , both the value of his or her features and his or her class (the presence or absence of the disease) are known. The problem lies in finding the $S \subset V$ subset that maximizes the objective function $f(S)$. This function is defined in the following manner:

$$f(S) = \beta \cdot Rat(S) + (1 - \beta) \cdot \left(1 - \frac{|S|}{m}\right)$$

where $Rat(S)$ is defined as the rate of cases of X that are well classified by the model obtained with the variables of S and the classifier considered. Parameter $\beta \in [0, 1]$ controls for the rate of correct diagnoses and the size of $|S|$. Therefore, the objective function f balances the classifying capacity with the size of the set obtained. This objective function has been used in similar studies, such as^{26,29,71}, and³⁸. Usually, the β values are equal or approximately equal to 0.99, which is the value we used in the present study.

Solution method for the proposed problem

The solution method is a procedure (*MultiStartTabu*) that combines the multistart and TS strategies. In each iteration, two procedures are performed: a constructive procedure (*Constructive*) that generates an initial solution, and a second procedure that improves the solution generated by the constructive procedure. This improvement procedure (*TabuSearch*) is based on the TS strategy. The method ends when the solution does not improve after a series of iterations. Pseudocode 1 presents the *MultiStartTabu* method.

As shown in the pseudocode, *iter* is an auxiliary variable that indicates the number of the current iteration; *iterbest* indicates where the best solution was found; S^{best} and f^{best} are the best solution and its classifying capacity, respectively; finally, *maxiterMS* is a previously defined parameter indicating the number of iterations that must take place, with no improvement of f^{best} , in order for the method to conclude.

Next, we explain the *Constructive* and *TabuSearch* procedures. The *Constructive* procedure begins with $S = \emptyset$; in each step, it selects an element of $v_{j^*} \in V - S$ among those that would best improve the objective function if added to S , and it is then added to S . The process ends when there are no elements in $V - S$ that improve the objective function f . Pseudocode 2 shows the *Constructive* procedure.

The procedure is relatively simple. Initially, we execute $S = \emptyset$; in the following steps, for the v_j elements that are not in S , the value of the objective function of S is calculated if v_j is added ($g(j) = f(S \cup \{v_j\})$); then, the list L' is created with the indices of the elements that improve the objective function of the current solution ($g(j) \geq f(S)$). Next, from the elements of the list L' , another list (L , a list of candidates) is created from the indices

Procedure *Constructive*(α ; output: S)

1. Execute $S = \emptyset$
 2. $\forall v_j \in V$: calculate $g(j) = f(S \cup \{v_j\})$
 3. Execute $L' = \{j : j = 1..m\}$
- Repeat**
4. Determine: $gmin = \min\{g(j) : j \in L'\}$ and $gmax = \max\{g(j) : j \in L'\}$
 5. Construct $L = \{j : j \in L', g(j) \geq \alpha \cdot gmax + (1 - \alpha) \cdot gmin\}$
 6. Randomly select $j^* \in L$ and execute $S = S \cup \{v_{j^*}\}$
 7. $\forall v_j \in V - S$: calculate $g(j) = f(S \cup \{v_j\})$
 8. Execute $L' = \{j : v_j \in V - S, g(j) \geq f(S)\}$
- Until** $L' = \emptyset$

Pseudocode 2. *Constructive procedure.*

with the highest $g(j)$ values, and one of them is randomly selected (j^*). Finally, the corresponding variable v_{j^*} is added to S . The process ends when $L' = \emptyset$ (there are no elements that improve the current solution). As can be observed, L' is initially composed of all the indices of the elements of V (considering $f(\emptyset) = 0$). Parameter α regulates the size of L . It takes values between 0 and 1; thus, if $\alpha = 0$, then $L = \{j : v_j \in V - S\}$, and the process is completely random; on the other hand, if $\alpha = 1$, then L consists exclusively of index j , which corresponds to $gmax$, and the process is therefore deterministic. It is important to select an adequate value of α that allows different high-quality solutions to be obtained.

TS⁷² is a metaheuristic strategy that, in its basic version, consists of a neighbor search procedure. Each step analyses all the possible movements that can be made from the current solution, and the best movement is selected. Simple movements are used in order to ensure that each movement results in a solution that is relatively similar to the current solution (a nearby or “neighboring” solution). The procedure allows movements to solutions that do not improve the current solution. Moreover, to prevent the algorithm from cycling, some movements are declared “tabu” and are initially not considered.

In our case, we considered three types of movements: a) adding an element $v_j \in V - S$; b) removing an element $v_j \in S$; and c) exchanging an element $v_j \in S$ with another element $v_j \in V - S$. The set of neighbor solutions of S (i.e., those that are reached through these movements) is defined as $N(S)$.

To avoid cycles, the output from S (throughout a series of iterations) of elements that recently entered S is declared “tabu”. Similarly, the input into S of elements that recently left S is also declared “tabu”. To verify the tabu status of the entry or exit of an element $v_j \in V$, we defined:

VectorIn(j): Number of the iteration in which the element v_j entered S .

VectorOut(j): Number of the iteration in which the element v_j left S .

Thus, the entry of an element $v_j \in V - S$ is tabu if

$$iter \leq VectorOut(j) + tenure \quad (1)$$

Additionally, the exit of an element $v_j \in S$ is tabu if

$$iter \leq VectorIn(j) + tenure \quad (2)$$

Finally, the exchange of an element $v_j \in S$ with an element $v_j \in V - S$ is tabu if either of the two aforementioned conditions is confirmed to be present.

The parameter *tenure* indicates the number of iterations in which an output or input is tabu. The auxiliary variable *iter* represents the number of iterations. On the other hand, the tabu status of a movement can be ignored (and thus the movement can be considered) if such movement results in a solution with a greater value of the objective function f than the previous solutions visited (“aspiration criterion”). Pseudocode 3 shows the *TabuSearch* procedure.

As can be observed in Pseudocode 3, each iteration considers all the movements that are not tabu or that meet the aspiration criterion. The best neighbor solution considered is stored in the variable S^b . This change is executed ($f^b = f(S)$ and $S = S^b$), and the values of *VectorIn* and/or *VectorOut* are updated according to the type of movement performed and the elements involved. After each iteration, S^* and f^* , which are the best solution found during the search and its value of the objective function f , respectively, are updated. The procedure ends after a preestablished number of iterations (*maxiterTS*) have taken place with no improvement of f^* . In this procedure, the parameter *tenure* plays an important role: high values result in many movements being declared tabu, thus reducing the flexibility of the process; low values may not prevent cycles. Therefore, adequate selection is critical.

Procedure *TabuSearch* (*tenure*, *maxiterTS*, *S*; output: S^*)

1. Execute: $S^* = S, f^* = f(S), iter = 0, iterbest = 0$
 2. Execute: $VectorIn(j) = -tenure, VectorOut(j) = -tenure, \forall j = 1, \dots, m$
- Repeat**
3. Execute $iter = iter + 1$
 4. Execute $f^b = -\infty$
 - $\forall S' \in N(S)$ Execute:

begin

 5. Determine the tabu status of the corresponding movement
 6. Determine whether the “aspiration criterion” is met, i.e., verify whether $f(S') > f^*$
 7. If the movement is not tabu or meets the aspiration criterion, then if $f(S') > f^b$ execute: $f^b = f(S')$ and $S^b = S'$

end
 8. Execute $S = S^b$
 9. Update *VectorIn* and/or *VectorOut*
 10. If $f(S) > f^*$ then: $S^* = S, f^* = f(S)$ and $iterbest = iter$
- Until** $iter > iterbest + maxiterTS$

Pseudocode 3. *TabuSearch* procedure.

Computational tests

This section describes different computational tests. The first group of tests was performed to adjust the parameters of the proposed model (Section “[Fine-tuning of parameters](#)”). The second group of tests was conducted to compare our method with other popular variable selection methods in the recent literature (Section “[Comparison with other methods](#)”). As can be observed in Section “[Comparison with other methods](#)”, our method, in general, obtained better results according to several metrics. Multiple databases of medical diagnoses were used.

To carry out these tests, nine databases were used: eight databases from the well-known repository of the University of California, Irvine (USA) (<https://archive.ics.uci.edu/>) and another database of Alzheimer’s diagnoses, which is presented for the first time in this study (<https://www.ubu.es/metaheuristicos-grinubumet/ejemplos-y-datos-de-problemas>). Because these databases concern diagnoses, the data were divided into two different types (presence of the disease, or “positive,” and absence of the disease, or “negative”). Table 1 shows descriptions of the different databases and their characteristics.

It should be noted that in the Cervical database, the two variables with the highest number of missing values were initially eliminated, and then the cases with missing values were eliminated.

Fine-tuning of parameters

To adjust the parameters, three of the nine databases were considered: one with few features (*Parkinson*), another with many features (Quality Assessment of Digital Colposcopies, or *QADC*), and another with an intermediate number of features (Wisconsin Breast Cancer – Prognosis, or *WPBC*). Discriminant analysis was used as a classifier due to its rapid calculation capacity.

Four parameters in our *MultiStartTabu* (MST) method were analysed: α , *tenure*, *maxiterTS* and *maxiterMS*. The parameter *maxiterTS* was used as the stopping criterion in the *TabuSearch* procedure, whereas *maxiterMS*

Database name	Cases (pos/neg)	Features
Parkinson	195 (46/146)	22
Quality assessment of digital colposcopies (QADC)	287 (72/215)	67
SPECTF Heart	267 (56/211)	44
Wisconsin breast cancer—diagnosis (WDBC)	569 (212/357)	30
Wisconsin breast cancer—prognosis (WPBC)	198 (151/47)	31
Alzheimer	1548 (630/918)	28
Cervical cancer	668 (605/63)	28
Glioma	838 (486/352)	23
Mesothelioma	324 (228/96)	33

Table 1. Databases used.

was used as the stopping criterion in the general MST procedure. To analyze the parameters α and *tenure*, we set the values $\text{maxiterTS} = 10 \cdot n$ and $\text{maxiterMS} = 20$. For α , we considered the following values: $\alpha = 0, 0.1, 0.5, 0.9, 0.99$, and 1 . For *tenure*, we considered the following values: $\text{tenure} = n/2, n, 2 \cdot n$ and $5 \cdot n$. After analyzing the 20 combinations, we found the best results at $\alpha = 0.99$ and $\text{tenure} = n/2$. Subsequently, the parameters *maxiterTS* and *maxiterMS* were analyzed using these values for α and *tenure*. It was observed that with values higher than $\text{maxiterTS} = 10 \cdot n$ and $\text{maxiterMS} = 10$, there were no significant improvements. Therefore, we selected these values. Figure 1 below shows the evolution of the objective function f over a series of iterations that make up the *MultiStartTabu* method for the WPBC database. The blue line indicates the values obtained in each iteration, and the red line shows the evolution of the best value. It can be seen that with a relatively low number of iterations the best final value is already reached.

Comparison with other methods

Baseline methods, classifiers, metrics, and experimental details

This subsection describes the tests and the corresponding results that were used to compare our method with five recent wrapper methods (baseline methods): *genetic algorithm*, or GA²⁶; *gray wolf optimizer*, or GWO⁷¹; *particle swarm optimization*, or PSO³⁹; *whale optimization algorithm*, or WOA³⁶; and *flower pollination algorithm*, or FPA³¹.

As classifiers, we used discriminant analysis (DA), logistic regression (LR), and a support vector machine (SVM). For the experiments, we used a k-fold cross-validation design ($k = 10$), which allowed us to conduct statistical tests. In addition to the value of the objective function, the following metrics were used:

$$\text{ACC}(\text{Accuracy}) = \frac{TP + TN}{TP + FP + TN + FN}$$

$$\text{AUC}(\text{Area Under the Curve}) = \left(1 + \frac{TP}{TP + FN} - \frac{FP}{TN + FP}\right)/2$$

$$\text{Gmean}(\text{Geometric Mean}) = \sqrt{\frac{TP}{TP + FN} \cdot \frac{TN}{TN + FP}}$$

$$\text{F1}(\text{F1 Score}) = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}$$

where *TP*, *TN*, *FP*, and *FN* represent true positive, true negative, false positive, and false negative results, respectively. As can be observed, the definition of *ACC* is equivalent to that of *Rat*, although it refers to the training set. The three metrics are also frequently employed and very useful when the subsets of each class are unbalanced or poorly balanced.

For each database, the following pre-processing steps were carried out: the data of the features were normalized, and the database was divided into 10 folds, resulting in 10 training set–test set pairs. Subsequently, for each database and classifier, the following process was performed:

- For each training set, our MST algorithm was executed with the previously considered parameters, and the computation time was recorded. Then, each of the other 5 methods was executed, with the computation time of the MST as the stopping criterion.
- The results obtained for each fold of each of the methods were recorded (objective function on the training set, metrics on the validation set, and mean difference tests).
- All algorithms were implemented with the Object-Pascal programming language using the Delphi compiler and the Rad Studio development environment (version 10.3). All tests were performed on a computer with

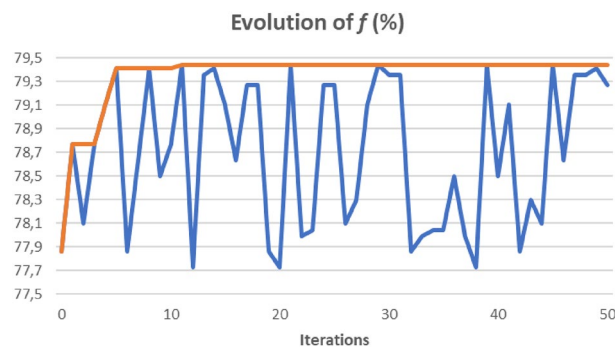


Figure 1. Evolution of the value of f for WPBC.

- an i9-10920X processing unit and 128 GB RAM. The parameters used by the baseline methods were the ones recommended by the corresponding articles, except for the stopping criterion, as was commented on above.
- For the estimation of the logistic regression models, we used the algorithm of Lin et al.⁷³, whereas for the SVM models, we used the method proposed by Hsieh et al.⁷⁴.

Table 2 shows the computation times used by our MST (and thus by the rest of the methods) for each of the databases and classifiers. The mean time and standard deviation on the set of the 10 folds are indicated.

As can be observed, the computation times were longer in the databases with a larger number of variables (QADC) than in those with a smaller number of variables (Parkinson). Moreover, for the calculation of the objective function $f(S)$ and, more specifically, the rate of correct diagnoses $Rat(S)$, it was necessary to generate the model with the corresponding classifier using the variables of S . In the case of DA, the models were obtained easily and immediately, whereas in the case of LR and SVM, the models were obtained through a more complex optimization process, as mentioned above.

Results of DA

Next, we show the results obtained for each classifier. First, the results obtained for DA are presented. Table 3 displays the results obtained for the objective function f , with the mean and standard deviation for each database and method. The method with the best mean is indicated in bold text. From these results, a paired two-tailed t test was carried out with each of the other methods. Significant differences in favor of our method are indicated with “+” after the result of the method that it is compared with; negative differences in favor of the other method are indicated with “-”; and “=” represents the absence of significant differences. The last row (W/T/L) indicates, for each baseline method, the number of “wins” (W) (i.e., the number of significant differences in favor of our method compared to the other method), “ties” (T) and “losses” (L).

From Table 3, it can be observed that, in all databases, our method obtained the best results in terms of the objective function f . Moreover, in all cases, the difference with respect to each of the other methods was significant. Figure 2 shows nine radial plots (one per database) representing the mean results of the different methods.

Similarly, Table 4 and Fig. 3 present the results of the ACC metric on the test set. Table 4 shows the means, standard deviations, and results of t tests (as in Table 3). Figure 3 displays radial plots of the mean results. From Table 4 and Fig. 3, the following can be observed:

- In all databases, our method obtained better mean results than any of the other methods tested. The PSO method obtained the same mean result only in the WPBC database.

	DA	LR	SVM
Parkinson	0.529 ± 0.033	17.761 ± 1.095	10.412 ± 0.690
QADC	71.748 ± 6.488	1906.915 ± 209.946	1513.337 ± 177.693
SPECTF heart	9.091 ± 0.620	642.467 ± 80.625	383.022 ± 47.990
WDBC	1.187 ± 0.020	171.111 ± 13.576	176.336 ± 13.383
WPBC	1.019 ± 0.045	86.006 ± 11.781	37.534 ± 1.407
Alzheimer	1.630 ± 0.089	260.755 ± 37.853	274.872 ± 35.308
Cervical cancer	1.386 ± 0.021	392.119 ± 23.511	227.206 ± 24.201
Glioma	1.463 ± 0.062	162.417 ± 10.001	156.452 ± 11.901
Mesothelioma	3.125 ± 0.099	763.716 ± 71.011	396.045 ± 42.541

Table 2. Computation time in seconds (mean and standard deviation) for each method.

	MST	GA	GWO	PSO	WOA	FPA
Parkinson	83.76 ± 0.36	83.04 ± 0.43 (+)	83.31 ± 0.24 (+)	82.69 ± 0.40 (+)	82.56 ± 0.72 (+)	82.84 ± 0.54 (+)
QADC	83.95 ± 0.51	82.96 ± 0.23 (+)	83.25 ± 0.50 (+)	82.68 ± 0.41 (+)	82.27 ± 0.30 (+)	82.60 ± 0.40 (+)
SPECTF Heart	81.49 ± 0.52	80.53 ± 0.37 (+)	80.94 ± 0.50 (+)	80.15 ± 0.56 (+)	79.63 ± 0.45 (+)	80.01 ± 0.52 (+)
WDBC	95.58 ± 0.16	95.21 ± 0.32 (+)	95.35 ± 0.15 (+)	94.98 ± 0.26 (+)	94.86 ± 0.44 (+)	95.21 ± 0.32 (+)
WPBC	79.44 ± 0.48	78.50 ± 0.23 (+)	78.77 ± 0.42 (+)	78.10 ± 0.47 (+)	77.73 ± 0.35 (+)	77.99 ± 0.40 (+)
Alzheimer	82.38 ± 0.45	81.73 ± 0.46 (+)	81.83 ± 0.32 (+)	81.35 ± 0.33 (+)	81.41 ± 0.58 (+)	81.72 ± 0.47 (+)
Cervical Cancer	91.53 ± 0.19	91.27 ± 0.30 (+)	91.26 ± 0.31 (+)	90.87 ± 0.43 (+)	90.65 ± 0.45 (+)	91.08 ± 0.31 (+)
Glioma	89.56 ± 0.70	88.18 ± 0.51 (+)	88.60 ± 0.67 (+)	87.74 ± 0.54 (+)	87.51 ± 0.77 (+)	87.96 ± 0.60 (+)
Mesothelioma	75.03 ± 0.64	74.16 ± 0.6 (+)	74.27 ± 0.56 (+)	73.72 ± 0.54 (+)	73.66 ± 0.82 (+)	74.00 ± 0.68 (+)
W/T/L		9/0/0	9/0/0	9/0/0	9/0/0	9/0/0

Table 3. Results of f (%) using DA. The best values are in bold.

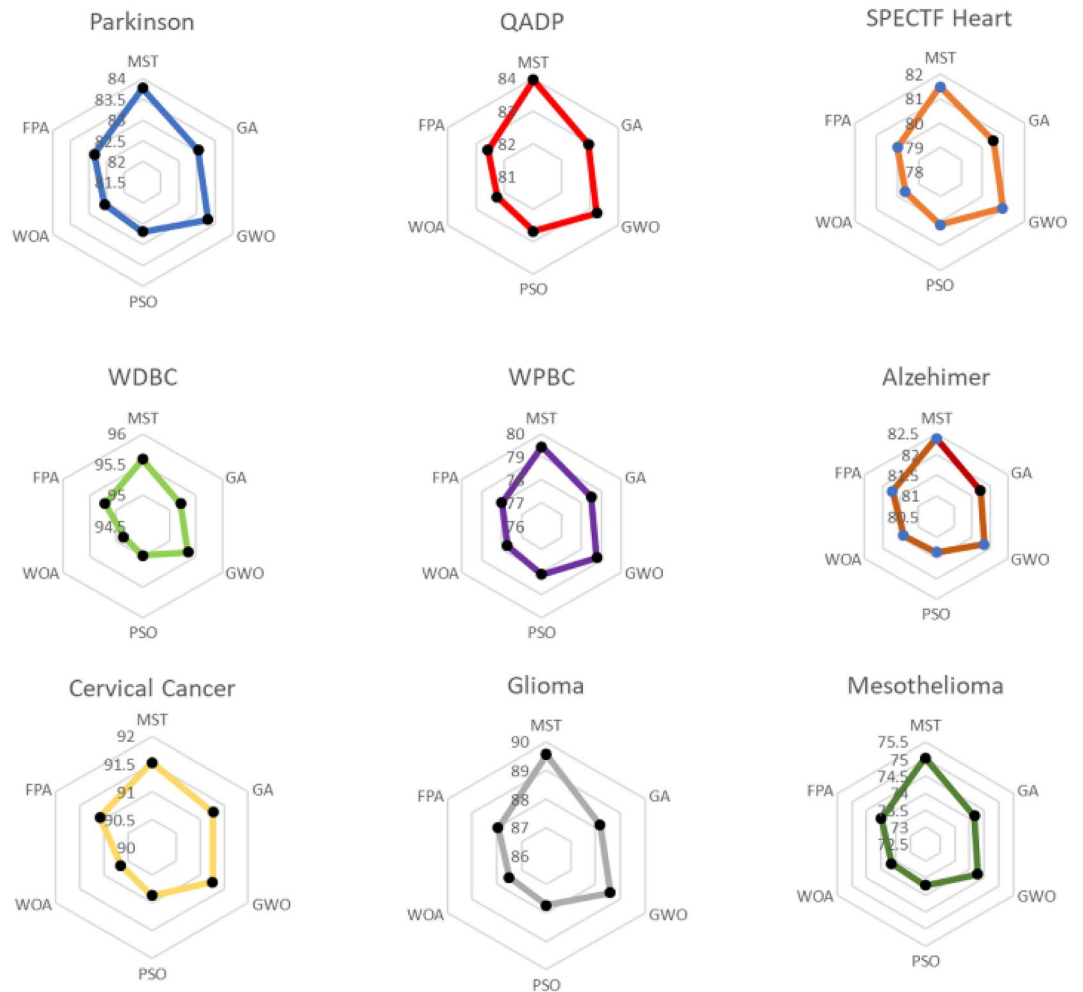


Figure 2. Radial plots of the mean values of the objective function f (%) of the different methods using DA.

	MST	GA	GWO	PSO	WOA	FPA
Parkinson	81.55 ± 2.50	78.45 ± 2.24 (+)	80.50 ± 2.23 (=)	77.47 ± 2.15 (+)	76.89 ± 2.90 (+)	77.95 ± 2.41 (+)
QADC	78.76 ± 2.44	77.34 ± 2.55 (+)	78.07 ± 2.69 (=)	75.96 ± 1.90 (+)	76.64 ± 2.49 (+)	76.65 ± 2.36 (+)
SPECTF Heart	79.02 ± 2.01	77.15 ± 1.15 (+)	78.63 ± 2.00 (=)	77.14 ± 1.47 (+)	76.03 ± 2.61 (+)	76.41 ± 2.41 (+)
WDBC	95.96 ± 0.84	95.26 ± 0.84 (+)	95.43 ± 0.90 (+)	95.08 ± 1.10 (+)	94.56 ± 1.28 (+)	95.26 ± 0.84 (+)
WPBC	77.76 ± 3.59	76.26 ± 3.36 (=)	77.26 ± 4.30 (=)	77.76 ± 2.71 (=)	72.74 ± 4.80 (+)	72.24 ± 4.80 (+)
Alzheimer	79.01 ± 3.40	75.45 ± 2.09 (+)	76.87 ± 3.27 (+)	75.00 ± 2.38 (+)	74.10 ± 0.90 (+)	74.35 ± 0.57 (+)
Cervical Cancer	91.62 ± 0.77	91.17 ± 0.83(+)	91.17 ± 1.09 (+)	89.52 ± 1.00 (+)	89.37 ± 1.01 (+)	90.27 ± 1.26 (+)
Glioma	88.30 ± 0.52	87.35 ± 0.83(+)	87.71 ± 0.79 (+)	86.64 ± 0.74 (+)	86.28 ± 1.00 (+)	86.99 ± 0.89 (+)
Mesothelioma	74.07 ± 1.58	73.47 ± 1.36 (=)	73.47 ± 1.36 (=)	73.47 ± 1.36 (=)	72.85 ± 2.27 (+)	73.47 ± 1.36 (=)
W/T/L		7/2/0	4/5/0	7/2/0	9/0/0	8/1/0

Table 4. Results of ACC (%) using DA. The best values are in bold.

– Moreover, in most cases (combinations of databases/baseline method), these differences were significant. Only in 10 of the 45 cases were the differences not significant. Specifically, in the t tests compared to GWO, our method “tied” in 5 databases and “won” in 4 of them. In general, the results obtained by the GWO method were relatively similar to those obtained by our MST method.

Tables 5, 6, and 7 show the results of the AUC, GMean, and F1 metrics, respectively, as in the previous tables. The conclusions of these results are very similar to those drawn for the ACC metric:

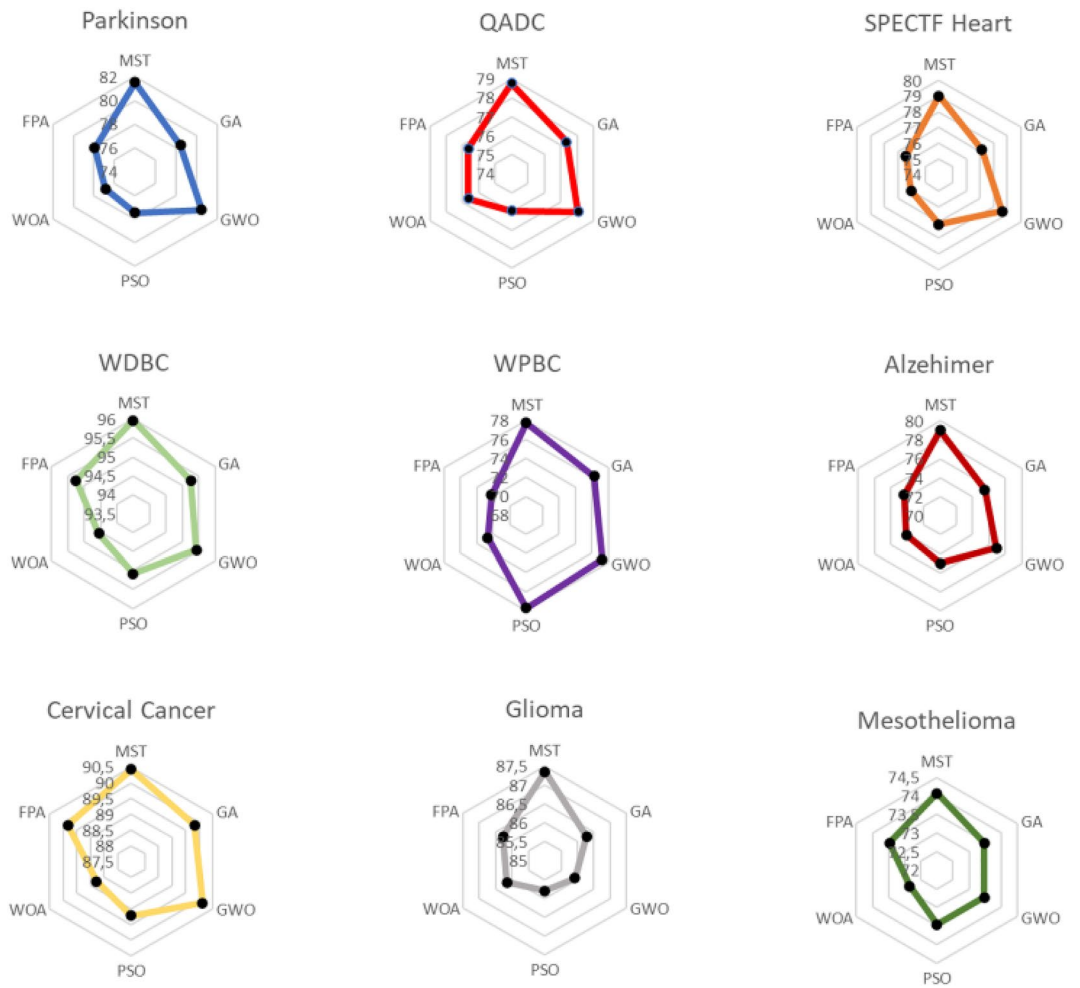


Figure 3. Radial plots of the mean ACC values (%) of the different methods using DA.

	MST	GA	GWO	PSO	WOA	FPA
Parkinson	80.85 ± 1.56	78.77 ± 1.57 (+)	80.15 ± 1.73 (=)	78.13 ± 1.41 (+)	77.73 ± 1.90 (+)	78.44 ± 1.63 (+)
QADC	64.15 ± 3.50	64.10 ± 4.43 (=)	63.71 ± 3.94 (=)	61.33 ± 3.26 (+)	62.67 ± 4.49 (=)	62.76 ± 3.83 (=)
SPECTF Heart	86.72 ± 1.31	85.55 ± 0.51 (+)	86.48 ± 1.25 (=)	85.54 ± 0.80 (+)	84.85 ± 1.45 (+)	85.09 ± 1.37 (+)
WDBC	94.86 ± 0.66	94.22 ± 0.74 (+)	94.44 ± 0.73 (+)	94.07 ± 1.04 (+)	93.17 ± 1.61 (+)	94.22 ± 0.74 (+)
WPBC	77.31 ± 3.59	75.65 ± 4.29 (=)	76.31 ± 4.93 (=)	77.31 ± 3.23 (=)	69.75 ± 7.50 (+)	68.50 ± 7.56 (+)
Alzheimer	76.85 ± 3.67	73.08 ± 2.30 (+)	74.60 ± 3.52 (+)	72.65 ± 2.49 (+)	71.66 ± 0.81 (+)	71.90 ± 0.47 (+)
Cervical Cancer	57.07 ± 0.70	55.99 ± 2.83 (+)	56.90 ± 0.72 (=)	55.66 ± 2.68 (+)	53.91 ± 3.59 (+)	55.99 ± 2.83 (+)
Glioma	88.19 ± 1.47	87.16 ± 1.05 (+)	86.77 ± 0.86 (+)	86.62 ± 1.10 (+)	87.05 ± 1.06 (+)	87.16 ± 1.05 (+)
Mesothelioma	58.69 ± 0.88	58.25 ± 0.59 (+)	58.25 ± 0.59 (+)	58.25 ± 0.59 (+)	57.81 ± 1.30 (+)	58.25 ± 0.59 (+)
W/T/L		7/2/0	4/5/0	8/1/0	8/1/0	8/1/0

Table 5. Results of AUC (%) using DA. The best values are in bold.

- Our MST model obtained a better mean result than the other methods in almost all cases. There was only one case in which the mean value of a baseline method was the same (the AUC metric, WPBC database, and PSO method), and there were another two cases in which the result of a baseline method was slightly better (the Gmean metric, WPBC database, and PSO method; the Gmean metric, QADC database and GA method), although the differences were not significant.
- Moreover, in most cases, the differences in favor of our method were significant. Only the results of the GWO method were relatively similar to those of our method in terms of the three metrics, according to the t tests (5 not significant differences and 4 significant differences in favor of our method in AUC and 6 not significant differences and 3 significant differences in favor of our method in Gmean and F1).

	MST	GA	GWO	PSO	WOA	FPA
Parkinson	80.81 ± 1.54	78.75 ± 1.61 (+)	80.14 ± 1.71 (=)	78.10 ± 1.44 (+)	77.67 ± 1.95 (+)	78.41 ± 1.68 (+)
QADC	56.72 ± 5.98	57.94 ± 7.02 (=)	56.47 ± 6.23 (=)	53.60 ± 5.31 (+)	55.65 ± 7.08 (=)	55.90 ± 6.17 (=)
SPECTF Heart	85.69 ± 1.52	84.32 ± 0.61 (+)	85.41 ± 1.45 (=)	84.31 ± 0.96 (+)	83.47 ± 1.76 (+)	83.75 ± 1.66 (+)
WDBC	94.76 ± 0.63	94.12 ± 0.74 (+)	94.36 ± 0.71 (+)	93.99 ± 1.04 (+)	92.99 ± 1.71 (+)	94.12 ± 0.74 (+)
WPBC	77.20 ± 3.75	75.51 ± 4.46 (=)	76.18 ± 5.09 (=)	77.22 ± 3.41 (=)	69.20 ± 8.16 (+)	67.83 ± 8.23 (+)
Alzheimer	75.94 ± 3.93	71.94 ± 2.50 (+)	73.57 ± 3.77 (+)	71.53 ± 2.63 (+)	70.45 ± 0.75 (+)	70.68 ± 0.39 (+)
Cervical Cancer	39.55 ± 1.44	35.44 ± 12.54 (+)	39.49 ± 1.47 (=)	35.30 ± 12.49 (+)	27.17 ± 18.80 (+)	35.44 ± 12.54 (+)
Glioma	88.03 ± 1.49	86.97 ± 1.09 (+)	86.60 ± 0.89 (+)	86.47 ± 1.13 (+)	86.86 ± 1.09 (+)	86.97 ± 1.09 (+)
Mesothelioma	44.87 ± 1.13	44.68 ± 1.23 (=)	44.68 ± 1.23 (=)	44.68 ± 1.23 (=)	44.47 ± 1.46 (=)	44.68 ± 1.23 (=)
W/T/L		6/3/0	3/6/0	7/2/0	7/2/0	7/2/0

Table 6. Results of *Gmean* (%) using DA. The best values are in bold.

	MST	GA	GWO	PSO	WOA	FPA
Parkinson	86.92 ± 2.09	84.39 ± 2.02 (+)	86.10 ± 1.76 (=)	83.58 ± 1.86 (+)	83.04 ± 2.66 (+)	83.97 ± 2.15 (+)
QADC	86.82 ± 1.62	85.72 ± 1.50 (+)	86.34 ± 1.69 (=)	84.97 ± 1.17 (+)	85.34 ± 1.45 (+)	85.34 ± 1.45 (+)
SPECTF Heart	84.66 ± 1.73	83.11 ± 0.71 (+)	84.35 ± 1.65 (=)	83.09 ± 1.12 (+)	82.11 ± 2.06 (+)	82.44 ± 1.95 (+)
WDBC	96.85 ± 0.69	96.29 ± 0.68 (+)	96.43 ± 0.72 (+)	96.15 ± 0.89 (+)	95.79 ± 0.96 (+)	96.29 ± 0.68 (+)
WPBC	61.98 ± 5.56	59.70 ± 5.32 (=)	60.82 ± 6.74 (=)	61.88 ± 4.79 (=)	52.37 ± 9.32 (+)	50.91 ± 10.05 (+)
Alzheimer	71.62 ± 4.90	66.64 ± 3.10 (+)	68.67 ± 4.68 (+)	66.13 ± 3.27 (+)	64.80 ± 0.93 (+)	65.09 ± 0.47 (+)
Cervical Cancer	23.97 ± 2.13	21.02 ± 7.84 (+)	23.39 ± 1.82 (=)	19.89 ± 7.22 (+)	14.16 ± 10.35 (+)	21.02 ± 7.84 (+)
Glioma	86.14 ± 1.55	85.01 ± 1.14 (+)	84.60 ± 0.91 (+)	84.45 ± 1.18 (+)	84.90 ± 1.12 (+)	85.01 ± 1.14 (+)
Mesothelioma	32.31 ± 1.32	31.79 ± 1.32 (=)	31.79 ± 1.32 (=)	31.79 ± 1.32 (=)	31.36 ± 1.90 (+)	31.79 ± 1.32 (=)
W/T/L		7/2/0	3/6/0	7/2/0	9/0/0	8/1/0

Table 7. Results of *F1* (%) using DA. The best values are in bold.

Finally, Table 8 shows a summary of the test results (“wins”, “ties” and “losses”), in terms of both the objective function f and each of the metrics, for each baseline method. The last row indicates the total sum of “wins”, “ties” and “losses” for each baseline method. From these results, it can be concluded that our method obtained better results than the rest of the methods. With respect to the objective function f , all differences were significant in favor of our method. Regarding the metrics, compared to GA, PSO, WOA, and FPA, significant differences were obtained in most cases (of 36 cases, there were 27 significant differences compared to GA, 29 compared to PSO, 33 compared to WOA and 31 compared to FPA). Compared to GWO alone, fewer significant differences were observed (13 out of 36). There were no significant differences in favor of any of the other methods.

Results of LR

Next, we analyze the results obtained for LR. To avoid an excessive number of tables and figures, we present only Table 9, which shows the results of the accuracy (*ACC*) metric (means, standard deviations, and t tests), and Table 10, which displays a summary of the t tests (“Wins”, “Ties” and “Losses”), both for the objective function f and for each of the metrics, as applied to each baseline method.

	GA	GWO	PSO	WOA	FPA
Objective function f	9/0/0	9/0/0	9/0/0	9/0/0	9/0/0
Metrics on test sets					
ACC	7/2/0	3/6/0	7/2/0	9/0/0	8/1/0
AUC	7/2/0	4/5/0	8/1/0	8/1/0	8/1/0
Gmean	6/3/0	3/6/0	7/2/0	7/2/0	7/2/0
F1	7/2/0	3/6/0	7/2/0	9/0/0	8/1/0
Total (W/T/L)	27/9/0	13/23/0	29/7/0	33/3/0	31/5/0

Table 8. Summary of the t test results for f and the different metrics with DA.

	MST	GA	GWO	PSO	WOA	FPA
Parkinson	87.16 ± 3.66	85.13 ± 1.58 (+)	85.13 ± 1.58 (+)	85.13 ± 1.58 (+)	85.13 ± 1.58 (+)	85.13 ± 1.58 (+)
QADC	84.68 ± 2.33	83.28 ± 2.17 (+)	84.34 ± 2.33 (=)	83.28 ± 2.17 (+)	81.54 ± 1.56 (+)	82.93 ± 2.51 (+)
SPECTF Heart	87.29 ± 3.50	85.80 ± 2.16 (+)	87.29 ± 3.03 (=)	85.80 ± 2.16 (+)	85.03 ± 1.66 (+)	84.64 ± 2.14 (+)
WDDBC	97.71 ± 1.68	96.31 ± 1.55 (+)	96.31 ± 1.55 (+)	96.31 ± 1.55 (+)	95.43 ± 1.23 (+)	95.96 ± 1.45 (+)
WPBC	86.39 ± 2.25	85.39 ± 3.55 (=)	85.89 ± 3.02 (=)	84.89 ± 3.94 (=)	83.39 ± 3.93 (+)	83.39 ± 3.93 (+)
Alzheimer	81.20 ± 2.15	79.46 ± 1.22 (+)	79.97 ± 1.49 (+)	79.13 ± 1.25 (+)	78.29 ± 1.17 (+)	78.94 ± 0.99 (+)
Cervical Cancer	91.17 ± 1.09	90.12 ± 1.04 (+)	90.57 ± 0.70 (+)	90.12 ± 1.24 (+)	89.82 ± 0.93 (+)	90.12 ± 1.02 (+)
Glioma	89.15 ± 0.37	87.84 ± 0.50 (+)	88.56 ± 0.84 (+)	87.49 ± 0.83 (+)	86.65 ± 0.92 (+)	87.37 ± 0.82 (+)
Mesothelioma	76.54 ± 1.55	75.30 ± 2.12 (+)	75.30 ± 2.12 (+)	74.08 ± 2.03 (+)	73.77 ± 1.50 (+)	74.08 ± 2.02 (+)
W/T/L		8/1/0	6/3/0	8/1/0	9/0/0	9/0/0

Table 9. Results of ACC (%) using LR. The best values are in bold.

	GA	GWO	PSO	WOA	FPA
Objective function f	9/0/0	9/0/0	9/0/0	9/0/0	9/0/0
Metrics on test sets					
ACC	8/1/0	6/3/0	8/1/0	9/0/0	9/0/0
AUC	7/2/0	5/4/0	8/1/0	9/0/0	9/0/0
Gmean	6/3/0	5/4/0	6/3/0	7/2/0	7/2/0
F1	7/2/0	5/4/0	8/1/0	7/2/0	7/2/0
Total (W/T/L)	28/8/0	21/15/0	30/6/0	32/4/0	32/4/0

Table 10. Summary of the t test results for f and the different metrics with LR. The best values are in bold.

As can be observed from Table 9, our method obtained the best mean results in all databases. The GWO method obtained the same mean result only on the *SPECTF Heart* database. Moreover, in most cases (40 out of the 45 cases), these differences were significant. Specifically, in the t tests against the GWO method, our method “tied” on 3 databases and “won” on the other 6. As in the case of DA, it seems that the results obtained by GWO were relatively similar to those obtained by our MST.

From Table 10, it can be observed that our method obtained better results than the rest of the methods. Regarding the objective function f , all differences were significant in favor of our method. With respect to the metrics, compared to WOA and FPA, there were 32 significant differences in favor of MST; compared to PSO, there were 30 significant differences in favor of MST; compared to GA, there were 28 significant differences; and compared to GWO, there were 21 significant differences. As in the case of DA, there were no significant differences in favor of any of the other methods.

Results of SVM analysis

Finally, we analyze the results obtained for SVM. As in the case of LR, Table 11 shows the results of the ACC metric, and Table 12 presents the summary of the t tests in terms of both the objective function f and each of the metrics (Table 12).

	MST	GA	GWO	PSO	WOA	FPA
Parkinson	86.13 ± 3.49	85.13 ± 1.58 (=)	85.13 ± 1.58 (=)	85.13 ± 1.58 (=)	84.61 ± 0.42 (+)	85.13 ± 1.58 (=)
QADC	82.23 ± 1.07	79.78 ± 2.84 (+)	81.88 ± 1.39 (=)	79.78 ± 2.84 (+)	78.04 ± 1.74 (+)	80.12 ± 2.50 (+)
SPECTF heart	86.52 ± 1.86	85.40 ± 1.10 (+)	85.78 ± 1.42 (=)	83.90 ± 3.03 (+)	84.63 ± 1.38 (+)	84.27 ± 2.30 (+)
WDDBC	97.53 ± 1.50	96.13 ± 1.12 (+)	96.31 ± 1.55 (+)	96.31 ± 1.55 (+)	95.43 ± 1.23 (+)	95.60 ± 1.25 (+)
WPBC	86.39 ± 3.26	83.37 ± 3.23 (+)	84.37 ± 3.60 (+)	84.37 ± 3.60 (+)	82.34 ± 2.49 (+)	82.87 ± 3.34 (+)
Alzheimer	80.88 ± 1.64	78.36 ± 1.93 (+)	79.65 ± 1.78 (+)	78.10 ± 2.17 (+)	76.42 ± 2.26 (+)	77.20 ± 2.08 (+)
Cervical cancer	90.27 ± 0.76	89.82 ± 1.15 (+)	90.12 ± 1.02 (=)	89.37 ± 1.08 (+)	88.93 ± 1.41 (+)	89.82 ± 1.15 (+)
Glioma	87.37 ± 1.50	86.30 ± 1.13 (+)	86.41 ± 1.60 (+)	85.82 ± 1.17 (+)	86.18 ± 1.12 (+)	86.30 ± 1.13 (+)
Mesothelioma	72.52 ± 3.75	71.29 ± 2.14 (=)	70.98 ± 2.13 (+)	70.68 ± 1.50 (+)	70.37 ± 2.57 (+)	70.68 ± 1.50 (+)
W/T/L		7/2/0	5/4/0	8/1/0	9/0/0	8/1/0

Table 11. Results of ACC (%) using SVM. The best values are in bold.

	GA	GWO	PSO	WOA	FPA
Objective function f	9/0/0	9/0/0	9/0/0	9/0/0	9/0/0
Metrics on test sets					
ACC	7/2/0	5/4/0	8/1/0	9/0/0	8/1/0
AUC	7/2/0	6/3/0	8/1/0	9/0/0	8/1/0
Gmean	6/3/0	6/3/0	7/2/0	8/1/0	7/2/0
F1	6/3/0	5/4/0	7/2/0	8/1/0	8/1/0
Total (W/T/L)	26/10/0	22/14/0	30/6/0	34/2/0	31/5/0

Table 12. Summary of the t test for f and the different metrics with SVM.

Table 11 shows that, in all the databases, our method obtained better mean results than any of the baseline methods. Moreover, in most (37 out of the 45 cases), these differences were significant. Specifically, as in the cases of LR, in the t tests compared to GWO, our method “tied” in 4 databases and “won” in the other 5.

As can be observed in Table 12, our method obtained better results than any of the other methods. Regarding the objective function f , all the differences were significant in favor of our method. With respect to the metrics, compared to the WOA, there were 34 significant differences in favor of MST; compared to FPA, there were 31 significant differences in favor of MST; compared to PSO, there were 30 significant differences; compared to GA, there were 26 significant differences; and compared to GWO, there were 22 significant differences. As in the cases of DA and LR, there were no significant differences in favor of any of the other methods.

Comments on the set of results

Ultimately, the obtained results indicate the following:

- Considering all the databases, our method obtained better results than any of the other classifiers in terms of the objective function f used in the variable selection process. Furthermore, with respect to the other methods, these differences were significant in all cases.
- In terms of all the metrics considered, the models obtained by our method in this process achieved better global results on the test set than the models obtained by any of the other methods. In fact, our method yielded the best mean results in most of the specific cases (that is, considering all the metrics, all the databases and all the classifiers used). Similar mean results were obtained in a very small number of cases, and there was one isolated case of a difference in favor of another method, although that difference was not significant.
- In addition, these differences in the metrics in favor of our method were significant in most of the cases. Thus, of the 108 tests that were conducted with each of the baseline methods (considering 4 metrics, 9 databases, and 3 classifiers), the differences in favor of our method were significant in 81 cases compared to GA, in 56 cases compared to GWO, in 89 cases compared to PSO, in 99 cases compared to WOA, and in 94 cases compared to FPA. In the rest of the cases, there were no significant differences in favor of any method. Therefore, it seems that only the results of the GWO method were similar to those of MST, and only in half of the cases.

Conclusions

The use of classification models is increasing in the field of medicine, since they help to improve the diagnosis of diseases. Advances in these models are necessary and useful to increase the precision and reduce the uncertainty of diagnoses. Currently, there are databases with several features that aid in the creation of models. However, not all features contribute to this task in the same way. Some features may be irrelevant or noisy, and they can deteriorate the performance of the models. There may also be negative interactions among features, which are not always easy to distinguish and can also reduce the performance of the models. Thus, feature selection is critical in the field of medicine: it improves the diagnosis of diseases, identifies which features lead to a better diagnosis, and identifies which risk factors (personal traits, habits, etc.) have the largest impacts on the disease. From a computational perspective, feature selection is a hard problem due to the complexity of the search space. The most commonly used feature selection methods are based on evolutionary strategies, such as GAs, swarm optimization techniques, and bioinspired algorithms. This study proposes a wrapper method that creates solutions and improves them through a tabu search process in a multistart framework. Therefore, the strategy it follows is not among the ones that are most frequently used in this field. The results of different computational tests in medical databases show that our method outperforms other recently developed wrapper methods in the literature. The tests were conducted using different classifiers and metrics, and the results are accompanied by statistical tests that strengthen the conclusions.

In general, the main limitation of our method is that it requires significant adaptation to each specific problem and costs a great deal of effort to implement. Other recent methods, such as evolutionary methods, seem to be more intuitive and easier to implement. On the other hand, our method demonstrates better performance considering different metrics, databases, statistical tests, etc.

In short, the contribution of this work is the development of a method that, on the one hand, is based on different strategies than those used in recent methods, and on the other hand, significantly improves the performance of these methods.

Data availability

The datasets used for the current study are available in the following links/repositories: <https://archive.ics.uci.edu/>, <https://www.ubu.es/metaheuristicos-grinubumet/ejemplos-y-datos-de-problemas>. The code is available on the following GitHub link: <https://github.com/japachecob/MSTabuFeatureSelection>.

Received: 14 March 2023; Accepted: 8 October 2023

Published online: 10 October 2023

References

- Kapetanovic, J. C., Patrício, M. I. & MacLaren, R. E. Progress in the development of novel therapies for choroideremia. *Expert Rev. Ophthalmol.* **14**(6), 277–285 (2019).
- Kaur, S., Baine, M. J., Jain, M., Sasson, A. R. & Batra, S. K. Early diagnosis of pancreatic cancer: Challenges and new developments. *Biomark. Med.* **6**(5), 597–612 (2012).
- Richard, M. A. *et al.* Melanoma and tumor thickness: Challenges of early diagnosis. *Arch. Dermatol.* **135**(3), 269–274 (1999).
- Tschoepe, D. & Roesen, P. Heart disease in diabetes mellitus: A challenge for early diagnosis and intervention. *Exp. Clin. Endocrinol. Diabetes* **106**(01), 16–24 (1998).
- Nordberg, A. Towards early diagnosis in Alzheimer disease. *Nat. Rev. Neurol.* **11**(2), 69–70 (2015).
- Hall, M. A. *Correlation-Based Feature Selection for Machine Learning*. Doctoral dissertation, The University of Waikato (1999).
- Kononenko, I. Estimating attributes: Analysis and extensions of RELIEF. in *European Conference on Machine Learning*, 171–182. (Springer, 1994).
- Alonso, S. G. *et al.* Data mining algorithms and techniques in mental health: A systematic review. *J. Med. Syst.* **42**(9), 1–15 (2018).
- Shipe, M. E., Deppen, S. A., Farjah, F. & Grogan, E. L. Developing prediction models for clinical use using logistic regression: an overview. *J. Thorac. Dis.* **11**(Suppl 4), S574 (2019).
- Adams, K. M. Linear discriminant analysis in clinical neuropsychology research. *J. Clin. Exp. Neuropsychol.* **1**(3), 259–272 (1979).
- Pisner, D. A. & Schnyer, D. M. Support vector machine. in *Machine learning*, 101–121. (Academic Press, 2020).
- Penny, W. & Frost, D. Neural networks in clinical medicine. *Med. Decis. Making* **16**(4), 386–398 (1996).
- Podgorelec, V., Kokol, P., Stiglic, B. & Rozman, I. Decision trees: an overview and their use in medicine. *J. Med. Syst.* **26**(5), 445–463 (2002).
- Khamis, H. S., Cheruiyot, K. W. & Kimani, S. Application of k-nearest neighbour classification in medical data mining. *Int. J. Inf. Commun. Technol. Res.* **4**(4), 1–10 (2014).
- Soni, J., Ansari, U., Sharma, D. & Soni, S. Predictive data mining for medical diagnosis: An overview of heart disease prediction. *Int. J. Comput. Appl.* **17**(8), 43–48 (2011).
- Estévez, P. A., Tesmer, M., Perez, C. A. & Zurada, J. M. Normalized mutual information feature selection. *IEEE Trans. Neural Netw.* **20**(2), 189–201 (2009).
- Xue, B., Zhang, M., Browne, W. N. & Yao, X. A survey on evolutionary computation approaches to feature selection. *IEEE Trans. Evol. Comput.* **20**(4), 606–626 (2015).
- Chantar, H., Thaher, T., Turabieh, H., Mafarja, M. & Sheta, A. BHHO-TVS: A binary harris hawks optimizer with time-varying scheme for solving data classification problems. *Appl. Sci.* **11**(14), 6516 (2021).
- Li, J. *et al.* Feature selection: A data perspective. *ACM Comput. Surv. (CSUR)* **50**(6), 1–45 (2017).
- Liu, M. & Zhang, D. Feature selection with effective distance. *Neurocomputing* **215**, 100–109 (2016).
- Su, C. T. & Hsu, J. H. An extended chi2 algorithm for discretization of real value attributes. *IEEE Trans. Knowl. Data Eng.* **17**(3), 437–441 (2005).
- Gu, Q., Li, Z. & Han, J. Generalized fisher score for feature selection. [arXiv:1202.3725](https://arxiv.org/abs/1202.3725) (2012).
- Yu, L. & Liu, H. Feature selection for high-dimensional data: A fast correlation-based filter solution. in *Proceedings of the 20th international conference on machine learning (ICML-03)*, 856–863 (2003).
- Hancer, E. New filter approaches for feature selection using differential evolution and fuzzy rough set theory. *Neural Comput. Appl.* **32**, 2929–2944 (2020).
- Mehanović, D. *et al.* Feature selection using cloud-based parallel genetic algorithm for intrusion detection data classification. *Neural Comput. Appl.* **33**, 11861–11873 (2021).
- Too, J. & Abdullah, A. R. A new and fast rival genetic algorithm for feature selection. *J. Supercomput.* **77**(3), 2844–2874 (2021).
- Sainin, M. S. & Alfred, R. A genetic based wrapper feature selection approach using nearest neighbour distance matrix. in *2011 3rd Conference on Data Mining and Optimization (DMO)*, 237–242. (IEEE, 2011).
- Emary, E., Zawbaa, H. M. & Hassanien, A. E. Binary grey wolf optimization approaches for feature selection. *Neurocomputing* **172**, 371–381 (2016).
- Too, J. & Abdullah, A. R. Opposition based competitive grey wolf optimizer for EMG feature selection. *Evol. Intell.* **14**(4), 1691–1705 (2021).
- Rodrigues, D., Yang, X. S., Souza, A. N. D. & Papa, J. P. Binary flower pollination algorithm and its application to feature selection. in *Recent Advances in Swarm Intelligence and Evolutionary Computation*, 85–100. (Springer, 2015).
- Sayed, S. A. F., Nabil, E. & Badr, A. A binary clonal flower pollination algorithm for feature selection. *Pattern Recogn. Lett.* **77**, 21–27 (2016).
- Rodrigues, D. *et al.* A wrapper approach for feature selection based on bat algorithm and optimum-path forest. *Expert Syst. Appl.* **41**(5), 2250–2258 (2014).
- Nakamura, R. Y. *et al.* BBA: A binary bat algorithm for feature selection. in *2012 25th SIBGRAPI Conference on Graphics, Patterns and Images*, 291–297. (IEEE, 2012).
- Jameel, S. & Rehman, S. U. An optimal feature selection method using a modified wrapper-based ant colony optimisation. *J. Natl. Sci. Found. Sri Lanka* **46**(2), 143–151 (2018).
- Dwivedi, R., Kumar, R., Jangam, E. & Kumar, V. An ant colony optimization based feature selection for data classification. *Int. J. Recent Technol. Eng.* **7**, 35–40 (2019).
- Too, J., Mafarja, M. & Mirjalili, S. Spatial bound whale optimization algorithm: An efficient high-dimensional feature selection approach. *Neural Comput. Appl.* **33**(23), 16229–16250 (2021).
- Mafarja, M. & Mirjalili, S. Whale optimization approaches for wrapper feature selection. *Appl. Soft Comput.* **62**, 441–453 (2018).
- Too, J., Abdullah, A. R. & Mohd Saad, N. A new co-evolution binary particle swarm optimization with multiple inertia weight strategy for feature selection. *Informatics* **6**(2), 21 (2019).
- Xue, B., Zhang, M. & Browne, W. N. Particle swarm optimisation for feature selection in classification: Novel initialisation and updating mechanisms. *Appl. Soft Comput.* **18**, 261–276 (2014).
- Das, S., Singh, P. K., Bhowmik, S., Sarkar, R. & Nasipuri, M. (2016). A harmony search based wrapper feature selection method for holistic bangla word recognition. *Procedia Computer Science*, **89**, 395–403
- Zheng, L., Diao, R. & Shen, Q. Efficient feature selection using a self-adjusting harmony search algorithm. in *13th UK Workshop on Computational Intelligence (UKCI)*, 167–174. (IEEE, 2013).

42. Abdel-Basset, M., Ding, W. & El-Shahat, D. A hybrid Harris Hawks optimization algorithm with simulated annealing for feature selection. *Artif. Intell. Rev.* **54**(1), 593–637 (2021).
43. Queen, O. & Emrich, S. J. LASSO-based feature selection for improved microbial and microbiome classification. in *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2301–2308. (IEEE, 2021).
44. Liu, H., Zhou, M. & Liu, Q. An embedded feature selection method for imbalanced data classification. *IEEE/CAA J. Autom. Sin.* **6**(3), 703–715 (2019).
45. da Costa, F., Chaves, A., Vellasco, M. M. B. & Tanscheit, R. Fuzzy rules extraction from support vector machines for multi-class classification. *Neural Comput. Appl.* **22**, 1571–1580 (2013).
46. Pacheco, J., Casado, S. & Nunez, L. Use of VNS and TS in classification: Variable selection and determination of the linear discrimination function coefficients. *IMA J. Manag. Math.* **18**(2), 191–206 (2007).
47. Guyon, I., Weston, J., Barnhill, S. & Vapnik, V. Gene selection for cancer classification using support vector machines. *Mach. Learn.* **46**, 389–422 (2002).
48. Remeseiro, B. & Bolon-Canedo, V. A review of feature selection methods in medical applications. *Comput. Biol. Med.* **112**, 103375 (2019).
49. Jothi, G. & Inbarani, H. H. Hybrid tolerance rough set-firefly based supervised feature selection for MRI brain tumor image classification. *Appl. Soft Comput.* **46**, 639–651 (2016).
50. Dimitriadis, S. I., Liparas, D., Tsolaki, M. N. (2018) Random forest feature selection, fusion and ensemble strategy: Combining multiple morphological MRI measures to discriminate among healthy elderly, MCI, cMCI and alzheimer's disease patients: From the alzheimer's disease neuroimaging initiative (ADNI) database. *Journal of neuroscience methods*, **302**, 14–23.
51. Liu, X. *et al.* Recognizing common CT imaging signs of lung diseases through a new feature selection method based on Fisher criterion and genetic optimization. *IEEE J. Biomed. Health Inform.* **19**(2), 635–647 (2015).
52. Chong, D. Y. *et al.* Robustness-driven feature selection in classification of fibrotic interstitial lung disease patterns in computed tomography using 3D texture features. *IEEE Trans. Med. Imaging* **35**(1), 144–157 (2015).
53. Shi, Y. *et al.* A learning-based CT prostate segmentation method via joint transductive feature selection and regression. *Neurocomputing* **173**, 317–331 (2016).
54. Guinin, M., Ruan, S., Dubray, B., Massoptier, L. & Gardin, I. Notice of removal: feature selection and patch-based segmentation in MRI for prostate radiotherapy. in *2016 IEEE International Conference on Image Processing (ICIP)*, 2663–2667. (IEEE, 2016).
55. Sahran, S., Albashish, D., Abdullah, A., Abd Shukor, N. & Pauzi, S. H. M. Absolute cosine-based SVM-RFE feature selection method for prostate histopathological grading. *Artif. Intell. Med.* **87**, 78–90 (2018).
56. Jain, I., Jain, V. K. & Jain, R. Correlation feature selection based improved-binary particle swarm optimization for gene selection and cancer classification. *Appl. Soft Comput.* **62**, 203–215 (2018).
57. Wang, H., Jing, X. & Niu, B. A discrete bacterial algorithm for feature selection in classification of microarray gene expression cancer data. *Knowl. Based Syst.* **126**, 8–19 (2017).
58. Peng, H., Fu, Y., Liu, J., Fang, X. & Jiang, C. Optimal gene subset selection using the modified SFFS algorithm for tumor classification. *Neural Comput. Appl.* **23**, 1531–1538 (2013).
59. Kang, C., Huo, Y., Xin, L., Tian, B. & Yu, B. Feature selection and tumor classification for microarray data using relaxed Lasso and generalized multi-class support vector machine. *J. Theor. Biol.* **463**, 77–91 (2019).
60. Niwas, S. *et al.* Reliable feature selection for automated angle closure glaucoma mechanism detection. *J. Med. Syst.* **39**(3), 1–10 (2015).
61. Awadallah, M. A. *et al.* An enhanced binary Rat Swarm Optimizer based on local-best concepts of PSO and collaborative crossover operators for feature selection. *Comput. Biol. Med.* **147**, 105675 (2022).
62. Awadallah, M. A., Hammouri, A. I., Al-Betar, M. A., Braik, M. S. & Abd Elaziz, M. Binary Horse herd optimization algorithm with crossover operators for feature selection. *Comput. Biol. Med.* **141**, 105152 (2022).
63. Braik, M., Awadallah, M. A., Al-Betar, M. A., Hammouri, A. I. & Alzubi, O. A. Cognitively enhanced versions of capuchin search algorithm for feature selection in medical diagnosis: A COVID-19 case study. *Cogn. Comput.* **1**, 1–38 (2023).
64. Tuo, S., Liu, H. & Chen, H. Multipopulation harmony search algorithm for the detection of high-order SNP interactions. *Bioinformatics* **36**(16), 4389–4398 (2020).
65. Tuo, S. *et al.* MTHSA-DHEI: Multitasking harmony search algorithm for detecting high-order SNP epistatic interactions. *Comput. Intell. Syst.* **9**(1), 637–658 (2023).
66. Tuo, S. *et al.* A novel multitasking ant colony optimization method for detecting multiorder SNP interactions. *Interdiscipl. Sci. Comput. Life Sci.* **14**(4), 814–832 (2022).
67. Shang, J. *et al.* EpiResIM: A resampling method of epistatic model without marginal effects using under-determined system of equations. *Genes* **13**(12), 2286 (2022).
68. Pacheco, J. & Laguna, M. Vehicle routing for the urgent delivery of face shields during the COVID-19 pandemic. *J. Heurist.* **26**(5), 619–635 (2020).
69. Pacheco, J., Cepa, L., Puche, J. & Casado, S. Selection of investment portfolio with social responsibility: A multi-objective model and a tabu search method. *Appl. Intell.* **52**, 15785–15808 (2022).
70. Pacheco, J. & Casado, S. A stepped tabu search method for the clique partitioning problem. *Appl. Intell.* **53**, 16275–16292 (2023).
71. 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).
72. Glover, F. & Laguna, M. *Tabu Search* (Springer, 1997).
73. Lin, C. J., Weng, R. C. & Keerthi, S. S. Trust region newton methods for large-scale logistic regression. in *Proceedings of the 24th International Conference on Machine Learning*, 561–568 (2007).
74. Hsieh, C. J., Chang, K. W., Lin, C. J., Keerthi, S. S. & Sundararajan, S. A dual coordinate descent method for large-scale linear SVM. in *Proceedings of the 25th International Conference on Machine Learning*, 408–415 (2008).

Acknowledgements

The authors are grateful to the following entities and projects: The Spanish Research Agency (Projects PID2019-104263RB-C44 and PDC2021-121021-C22 and PID2022-139543OB-C44). The Regional Government of “Castilla y León” and FEDER funds (Project BU056P20).

Author contributions

J.P. and S.C. implemented the variable selection method proposed in this work as well as all the computational tests. O.S. and S.U. prepared the medical databases used in this work, especially the Alzheimer database. J.P., S.C., O.S. and S.U. wrote the state-of-the-art and revised the final version of the paper.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to J.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023