Palestinian Journal for Open Learning & e-Learning No. (20) – May 2025
https://journals.qou.edu/index.php/jropenres
E-ISSN: 2074-5656 / P-ISSN: 2520-5692

Feature Selection for Serving Medical Datasets Applying Heuristic Algorithms (Scatter Search within Decision Tree Classifier)

Mr. Maher Ibrahim Issa*

lecturer , Computer Information Systems , Al-Quds University, Salfeet, Palestine.

Oricd No: 0009-0003-5742-9772
Email: missa@qou.edu

Received:
6/03/2024

Revised:
6/03/2024

Accepted:
7/04/2024

***Corresponding Author:**
missa@qou.edu

Citation:

Abstract

Objectives: This research presents a feature selection process on different datasets of the medical domain with different aims and sizes using a wrapper approach based on a powerful metaheuristic algorithm which is the Scatter Search Algorithm and J48 decision tree classifier as the selection criteria.

Methods: The paper applied a modified approach of the basic Sequential Scatter Search algorithm called Improved Sequential Scatter Search follows the basic procedures of the original algorithm in addition to an early improvement mechanism choosing decision tree classifier to be the evaluator of the experiments.

Results: The experimental results show competition and superiority in feature selection compared to other metaheuristic algorithms for the same datasets in consideration of number of features selected and accuracy.

Conclusion: This research emphasizes the importance of wrapper approaches using metaheuristic algorithms to select the most dominant attributes in a dataset which is very important in reduction of the cost and complexity of all data analysis areas.

Keywords: Metaheuristic (MH), feature selection (FS), scatter search Algorithm (SSA), decision tree (DT), medical datasets.

اختيار الميزة لخدمة مجموعات البيانات الطبية من خلال تطبيق الخوارزميات الإرشادية (البحث المبعثر ضمن مصنف شجرة القرار)

أ. ماهر ابراهيم عيسى*

[1]محاضر، انظمة معلومات حاسوبية، جامعة القدس المفتوحة، سلفيت، فلسطين.

الملخص

الأهداف: تعرض هذه الورقة عملية اختيار الميزات على مجموعات بيانات مختلفة في المجال الطبي بأهداف وأحجام مختلفة باستخدام نهج مجمّع يعتمد على خوارزمية ارشادية قوية وهي خوارزمية البحث المبعثرة ومصنف شجرة القرار J48 كمعيار للاختيار.

الطرق: طبقت هذه الورقة البحثية طريقة مستحدثة لخوارزمية البحث المبعثر اطلق عليها خوارزمية البحث المبعثر المحسن والذي اتبع خطوات الخوارزمية الرئيسية واضاف الية تحسين بتطبيق مصنف شجرة القرار كاداة تقييم للتجربة.

النتائج: اظهرت الدراسة ان النهج المستخدم اظهر تفوقا احيانا ومنافسة احيانا اخرى مقارنة مع خوارزميات ارشادية اخرى لنفس مجموعة البيانات على اساس معيارين هما: الدقة واختيار الميزة.

في الختام: يؤكد هذا البحث على اهمية النهج المجمع مع الخوارزميات الارشادية في اختيار الميزات السائدة من مجموعة البيانات والذي يعتبر مهما جدا في تقليل التكلفة والتعقيد في جميع مجالات تحليل البيانات.

الكلمات المفتاحية: الميتايورستك (MH)، اختيار الميزات (FS)، خوارزمية البحث المبعثر (SSA)، شجرة القرار (DT)، مجموعة البيانات الطبية.

Introduction

Data mining and knowledge discovery in data (KDD) have been applied successfully in a number of study domains to extract new and useful knowledge from historical data (Ghazal and Hammad, 2022). Knowledge Discovery in Database (KDD) is growing at an unprecedented scale in medicine, industry, government, and civil society. Analysis and distilling knowledge from big data now drive many aspects of our society ،Insights gained from big data have revolutionized how we conduct business, governance, research, design, production, human interactions, and daily life (Shu and Ye, 2023).

High dimensionality seems to be the main challenge for implementation of the data mining techniques in different areas, in addition to noise, outliers, and errors in huge data sets. High-dimensional data presents a number of challenges for pattern identification. Additionally, smaller data sizes typically result in faster model training periods, which in turn speed research (Hancock et al., 2024). One popular technique for reducing data and comprehending feature information is feature selection (García-Pedrajas et al., 2021).

Feature Selection

Feature Selection addresses the issue of high dimensionality. In order to get the best-performing subset of the original characteristics without any changes, it involves choosing the pertinent features and eliminating the redundant, noisy, and irrelevant ones (Bouchlaghem et al., 2022).

Reducing the dimensionality of the data is the initial stage in integrating low-dimensional data into a pattern recognition system. One of the most important tasks in pattern recognition research is creating an accurate system for recognizing patterns in high-dimensional data (Varma et al., 2022). When there are more dimensions than there are observations, this is referred to as high dimensionality. This makes computations extremely challenging. Data from high-dimensional space is reduced to a more manageable low-dimensional space using dimensionality reduction techniques (Probst and Reymond, 2020). Using feature selection, the influence of dimensionality on the dataset is minimized by identifying the subset of features that best captures the data (Abdulrazzaq and Saeed, 2019). It is helpful for identifying a good subset of features that is suited for the given problem since it extracts from the input data the significant and pertinent features for the mining process and eliminates redundant and irrelevant features (Ayesha et al., 2020). Creating a limited subset of features that accurately captures the essential aspects of the entire input data set is the primary goal of feature selection (Velliangiri and Alagumuthukrishnan, 2019). Feature selection reduces the amount of data, lowers the amount of storage required, improves prediction accuracy, prevents over fitting, and shortens the execution and training times for variables that are simple to grasp (Zebari et al., 2020). According to Dash and Liu (1997), in a typical attribute reduction method there are four basic steps (see Figure 1), i.e., (a) a generation procedure to generate the next candidate subset: (b) an evaluation function to evaluate the generated subset; (c) a stopping criterion to decide when to terminate the process; and (d) a validation procedure to check the validity of the subset (Dash and Liu, 1997).

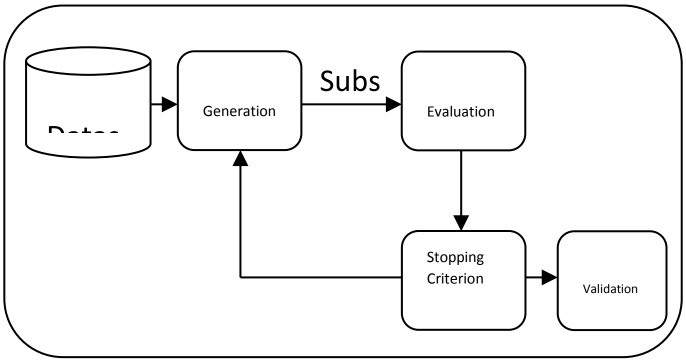


Figure (1): Attribute reduction process with validation

The goal of feature selection is to choose a subset of features based on redundancy and relevance from the original collection of features. Four categories of assessment approaches were first used in feature selection: filter, wrapper, embedding, and hybrid (Abd-Alsabour, 2018). Ensemble feature selection is a new kind of evaluation technique that has been created recently (Zebari et al., 2020). Here, the filter and wrapper models are discussed. Filter methods usually don’t involve induction algorithm and evaluate the goodness of attributes subset cheaply using intrinsic characteristic of the data, while wrapper methods are computationally expensive as they used the induction algorithm to evaluate the attributes subset, but outperform filter methods in terms of predictive accuracy (Lyu et al., 2023).

Medical datasets are frequently utilized in data mining research in area of feature selection. A study by (Chen, C. W, 2020) introduced a combination of different types of feature selection algorithms, then results show that a combination of filter (i.e., principal component analysis) and wrapper (i.e., genetic algorithms) techniques by the union method is a better choice, providing relatively high classification accuracy. Nadimi-Shahraki (2021) used B-MFO in solving the feature selection problem for different medical datasets compared to other comparative algorithms and showed superiority in comparison to BDA, and BSSA algorithms.

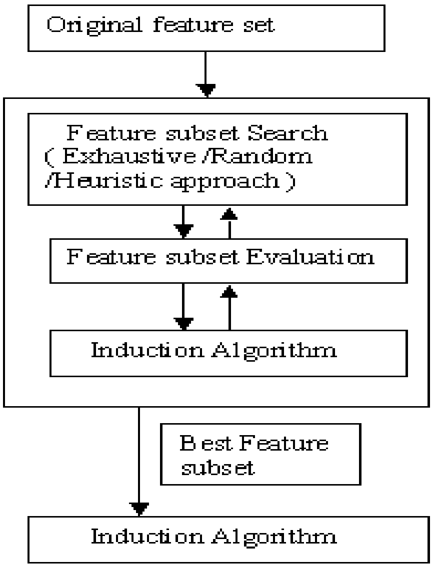
Feature Subset Generation

Feature selection as a search problem starts with feature generation. Feature subsets can be generated in a number of ways (Vandana and Chikkamannur, 2021; Hussain et al., 2021):

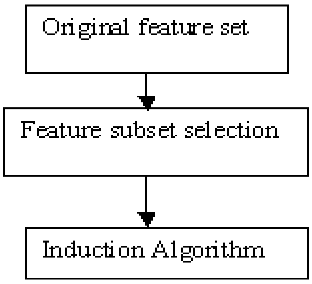
- Sequential Forward Selection (SFS) begins with the selection of the single best feature, which is determined by the objective function. The following stage involves creating a pair of features by combining the best feature and one feature from the remaining initial set. The pair with the best match is chosen. Using one of the remaining features, the next triplet of features is created, and the best triplet is chosen from the best pair that was previously chosen. Until a predetermined number of features are chosen, the process is repeated. When the optimal subset is tiny, it operates at its best. In SFS, a particular feature that might be helpful in the beginning might become unnecessary in later iterations for reasons that are not discoverable.
- Sequential Backward Selection (SBS) begins with an entire feature set as the beginning set. First, this starting collection is used to compute the criteria function. Every iteration involves the deletion of one feature, computation of the criterion feature for every subset, and deletion of the poorest feature. Until a certain number of features remain, this is repeated. It functions best when there is a big optimal subset. In SBS, a feature that is dropped in the first iteration might resurface in later iterations and be more helpful; however, this usefulness cannot be assessed.
- Bidirectional Search (BDS): This method applies SFS and SBS simultaneously to converge to the same answer, bringing a trade-off between the two.
- Random Generation is the fourth approach and performs the search process starting with randomly selected subset, then use sequential or bidirectional strategy.

Selection of some feature during the generation process is taken upon some evaluation measurement, such as classical ones: entropy, Bayesian, and Euclidean distance, or predictive classifier measure i.e. accuracy. The search process is iterative until reaching stopping criteria. The solution can be validated by univariate or multivariate approach. Decision Tree classifier is an example of univariate approaches, while neural network is a multivariate example.

The sequential search techniques in FS suffer from computational complexity since they are exhaustive procedures in their hunt for the best solution. (Elgamal et al., 2020). Current research focuses on meta-heuristic search, which, in contrast to optimum solutions, enables handling large-scale issue instances by providing a near-optimal solution in a reasonable amount of time (Sharma and Kaur, 2021).



a. Wrapper approach



b. Filter approach

FIGURE (2) : WRAPPER VS. FILTER

Meta-heuristic Algorithms

Metaheuristic algorithms have been categorized into two types based on solution search strategies (Sharma and Kaur, 2021): single-solution-based algorithms (S-Metaheuristic) and population-based algorithms (P-Metaheuristic). S-metaheuristics employ local search techniques to enhance a single solution iteratively until a stopping condition is met. Examples of S-metaheuristic algorithms include Simulated Annealing (SA), Hill Climbing (HC), Record-to-Record (RR), and Threshold Accepting. On the other hand, P-metaheuristics, such as Genetic Algorithm (GA) and Scatter Search (SS), utilize evolutionary approaches to improve a population of solutions iteratively. Additionally, algorithms like Ant Colony Optimization, Bee Colony Optimization, and Bat Algorithm are also categorized as P-metaheuristic, drawing inspiration from natural systems (Jaddi and Abdullah, 2020). P-metaheuristics evolve a population of solutions iteratively until a predetermined stopping criterion is satisfied.

Despite the early idea of Scatter Search (SS), the FS problems implementation for this algorithm is still at the beginning. Few studies such as (Wang et al., 2009; Wang et al., 2012; López et al., 2006) handle FS using SSA in different manner. (Wang et al., 2009) Used SSA to enhance the attribute reduction process in rough set theory. The approach of this study applied at different datasets and compared the result to other Computational Intelligence (CI) algorithms tried to address the same problem. The result shows promising and competitive performance in terms of solution quality, and superior performance in terms of computational costs. The other study (Wang et al., 2012) proposed a novel approach based on rough set and SSA, invoking entropy for searching optimal solution. The approach has been applied for two international credit scoring datasets and shows huge saving in computational costs and higher accuracy compared to base classification methods. The previous two studies based on MATLAB coding. López et al. (2006) applies FS using SSA on different 8 datasets with parallel computers for two combination methods, and compared results to basic SSA and Genetic algorithm supported by WEKA tool.

This paper introduces a novel wrapper feature selection methodology that combines the Scatter Search Algorithm (SSA) with the Decision Tree (DT) J48 classifier for application on standard medical datasets. The proposed approach generates an initial set of solutions termed as the RefSet, which undergoes refinement through a local search procedure. Subsequently, these solutions are evaluated using DT to identify the optimal solution. The chosen solution then undergoes further enhancement. This iterative process continues until a predefined stopping criterion is met, typically defined as the best solution in terms of both accuracy and the number of features selected. The approach is systematically applied across various parameters. The primary innovation lies in the inclusion of an early improvement mechanism following the creation of initial solutions, a step not typically addressed in traditional SSA approaches. The improvement strategy is explained later in a separate section.

Decision tree classifier

Decision Tree (DT) is a widespread classification algorithm that partitions the data using information gain until all instances reach uniform class labels (Huan and Hiroshi, 1998). Each time a single feature used as splitting node according to its values, and information gain used to determine the split feature F. DT achieves high accuracy level of prediction despite high computational cost (Wang et al., 2009).

Data in DT approach is divided into two datasets (Tan et al., 2006): Training set (large) and Test set (small) before starting the induction. Basic Induction Algorithm of DT is as follows:

- Initialize by setting variable T to be the training set
- Apply the following steps to T:

1. If all elements in T are of class X, create node X and halt.
2. Otherwise, select a feature F with values v1, v2…, vn, and partition T into T1, T2…, Tn according to their values in F, then create a parent node with F and T1, T2…, Tn as child nodes
3. Apply the procedure recursively to each child node.

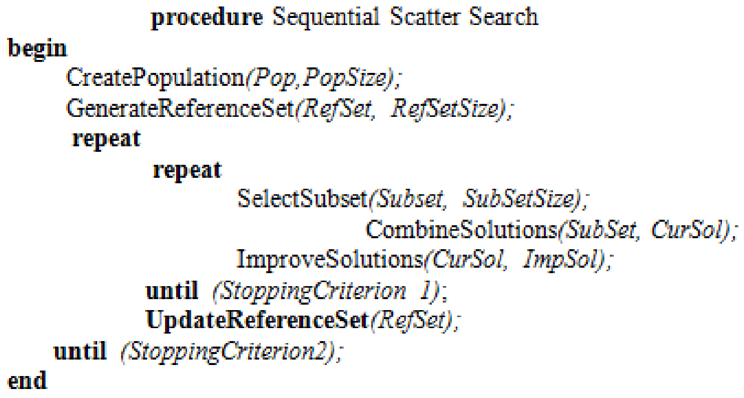
One of the robust algorithms for decision tree (DT) induction is the C4.5 algorithms. It represents an advancement over classical divide-and-conquer approaches (Bahar and Saad, 2024). C4.5 utilizes gain ratio as the basis for splitting the training set and incorporates methods to handle missing values, noisy data, and numeric attributes (Cherfi et al., 2020; Quinlan, 1996; Sugumaran et al., 2007). In our computational analysis, we employed the J48 decision tree, which implements the C4.5 criteria and is provided by the WEKA software suite.

Scatter search algorithm

The foundation of Scatter Search (SS) was introduced by Glover in 1977 as an evolutionary approach for addressing combinatorial and nonlinear optimization problems. SS follows a systematic problem-solving approach outlined in the following steps (Ghamisi and Benediktsson, 2014):

1. Generate starting set of solution vectors by heuristic processes, then choose a subset of the best vectors to be reference solutions.
2. Perform linear combinations of the subsets of the reference solutions and create new solutions.
3. Extract the best solutions in procedure 2, and use it as starting solutions.
4. Repeat steps until specified iteration limit.

It is good to know that heuristic processes in SS are not of uniform design but represent a varied collection of procedures, also SS can start with infeasible solution to reach elite ones. SSA can be reflected to solve FS problems following the same way. Figure (3) shows the pseudo code of the Sequential Scatter Search (SSS) as the improvement mechanism executed sequentially (Garcıa-López et al., 2003).



```
procedure Sequential Scatter Search
begin
    CreatePopulation(Pop,PopSize);
    GenerateReferenceSet(RefSet, RefSetSize);
    repeat
        repeat
            SelectSubset(Subset, SubSetSize);
            CombineSolutions(SubSet, CurSol);
            ImproveSolutions(CurSol, ImpSol);
        until (StoppingCriterion 1);
        UpdateReferenceSet(RefSet);
    until (StoppingCriterion2);
end
```

FIGURE (3): Sequential Scatter Search Algorithm

Solution Representation

Scatter Search Algorithm uses binary representation for solutions. The solution is a ‘0’,’1’ vector (Ghamisi and Benediktsson, 2014), its size equal to the number of conditional attributes in a dataset. If a feature is selected in this subset the corresponding index will have 1, while 0 indicates that the feature is not selected in this solution.

Population Generation

In this step a set of initial solutions vectors are generated. These solutions are divided as diverse and good quality solutions. In our approach we use the vector of weights used by (López et al., 2006) to generate a solution. This strategy considers the vector of weights of features P(X) = P(X1),.,P(Xc) where C is the number of conditional features, and given by P(Xi) = ft({Xi}) where these weights indicate the quality of the features for classifying by itself. Let L be the set of features Xi with the highest weight P(Xi), then select randomly good feature (high P) from L where the presence of this feature improves the set.

Reference Set Generation

Two sets of solutions are added to each other to form the reference set (RefSet). The first one (RefSet1) is good solutions set, while the other (RefSet2) is the set of diverse solutions. In order to achieve diversity, the symmetric difference criteria had been used with the procedures explained as follow: Let S be any solution and C is the set of features belong to any solution in the RefSet, then the diversity of each solution S is given by the following equation:

$\mathrm{Div}(S) = \mathrm{Diff}(S,C) = |(S \cup C)/ (S \cap C)|$

For generating the RefSet, SS usually considers all subsets of two solutions in the current set of solutions to be combined to generate new solutions.

Solutions Combination

This procedure generates new solutions by combining subsets of the original reference solution. In this approach the suggested greedy combination (GC) method by (López et al., 2006) had been used. Let S1 and S2 be solutions in the subset. GC method generates two new solutions S1', and S2'. First, GC starts by adding the common features of S1 and S2 to the new solutions S1' and S2'. Then at each iteration one of the remaining features in S1 or S2 is added to S1' or S2'.

Improvement Method

This method is applied to each solution obtained by the described combination strategy. The purpose of this method is to add some characteristics to the solution that improve it. The solution S accepts to add a feature if this improves its quality. If the solution has not been enhanced, then the output solution is the same as the input solution.

Updating Refrence Set

The reference set is updated after the previous procedures to be formed from two parts, the first one is the RefSet1 obtained by improvement method, where the second one RefSet2 is the produced by the diversity criteria explained earlier.

Medical Datasets Description

The Medical Datasets are widely used in data mining and machine learning researches for serving health care area. In our approach we chose to apply our approach on 7 well-known medical datasets and compared to a research paper that found the best feature selection results for these datasets in last 14 years in terms of accuracy and selected features. Using these medical datasets is justified by their frequent utilization in data mining research, facilitating easy comparison of results with best prior study. Mullins et al. (2006) provided a concise overview of these datasets, which are briefly described below.

1. **Breast-Cancer**: This dataset typically contains features derived from digitized images of breast cancer biopsies. These features can include characteristics like cell nuclei properties and are used to classify whether a tumor is benign or malignant.
2. **Breast-w**: This dataset is also related to breast cancer diagnosis. It may contain different features or be derived from a different source, but its purpose is likely similar: to predict whether a tumor is benign or malignant.
3. **Heart-c:** This dataset is often used for predicting the presence of heart disease. It typically includes various clinical parameters such as blood pressure, cholesterol levels, and electrocardiogram readings, among others.
4. **Heart-stolog:** Another dataset related to heart disease diagnosis. It might contain similar features to the "Heart-c" dataset but could be from a different source or have variations in the features included.
5. **Hepatitis**: This dataset involves data related to the diagnosis and treatment of hepatitis, a liver inflammation caused by viral infection. It may include various clinical and laboratory parameters for patients with hepatitis.
6. **Lung-Cancer:** This dataset pertains to the diagnosis and prognosis of lung cancer. It might include features such as tumor size, histological type, and patient demographics, among others.
7. **Dermatology:** This dataset typically includes various attributes related to dermatological conditions, such as symptoms, patient characteristics, and diagnostic results. It's often used for predicting skin diseases or analyzing patterns in dermatological data.

Table (1): Datasets description

#	Data Set	No. of Features	No. of Objects
1	Breast-cancer	9	286
2	Breast-w	9	699

#	Data Set	No. of Features	No. of Objects
3	Heart-c	13	303
4	Heart-stolog	13	270
5	Hepatitis	19	155
6	Lung-Cancer	56	32
7	Dermatology	34	366

The proposed approach

The Improved Sequential Scatter Search (ISSS) builds upon the foundation of the basic Scatter Search (SSS) method described previously. The key distinction lies is the integration of an early improvement technique following the generation of initial solutions, illustrated in Figure 4. This essentially constitutes a double improvement approach. Consequently, the reference set formed comprises superior solutions, enhancing the potential for generating improved solutions.

```
procedure Improved Sequential Scatter Search
begin
    CreatePopulation(Pop,Popsize);
    ImproveInitialPop;
    GenerateReferenceSet(RefSet, RefSetSize);
    repeat
        repeat
            SelectSubset(Subset, SubSetSize);
            CombineSolutions(Subset, Cursol);
            ImproveSolutions(CurSol, ImpSol);
        Until (StoppingCriterion1);
        UpdateReferenceSet(RefSet);
    Until (StoppingCriterion2);
End
```

FIGURE (4): ISSS pseudo code

In a more expressive manner, the following illustrations Figure 5 and Figure 6 clarify the difference between the basic Sequential Scatter Search and the Improved one which is our approach. The improvement step includes an accuracy evaluation mechanism using Decision Tree. The early improvement step helps to generate an initial strong Reference Set.

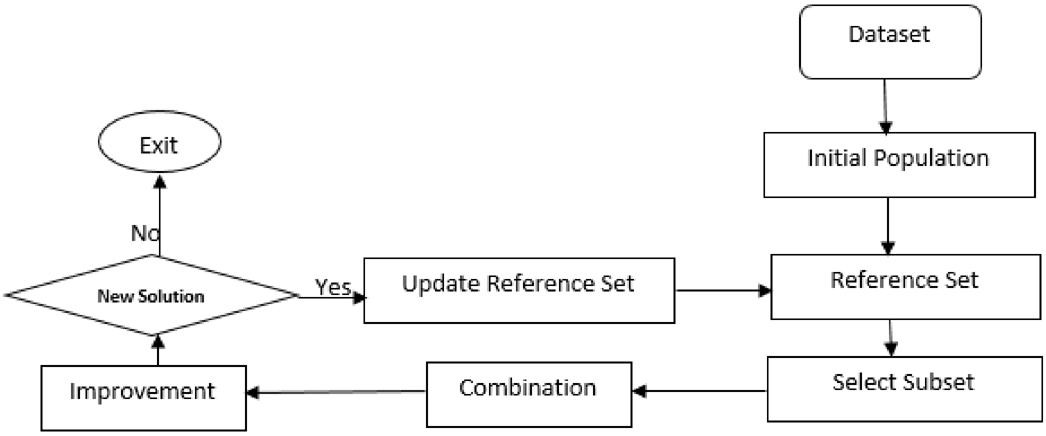


FIGURE (5): Basic SSS

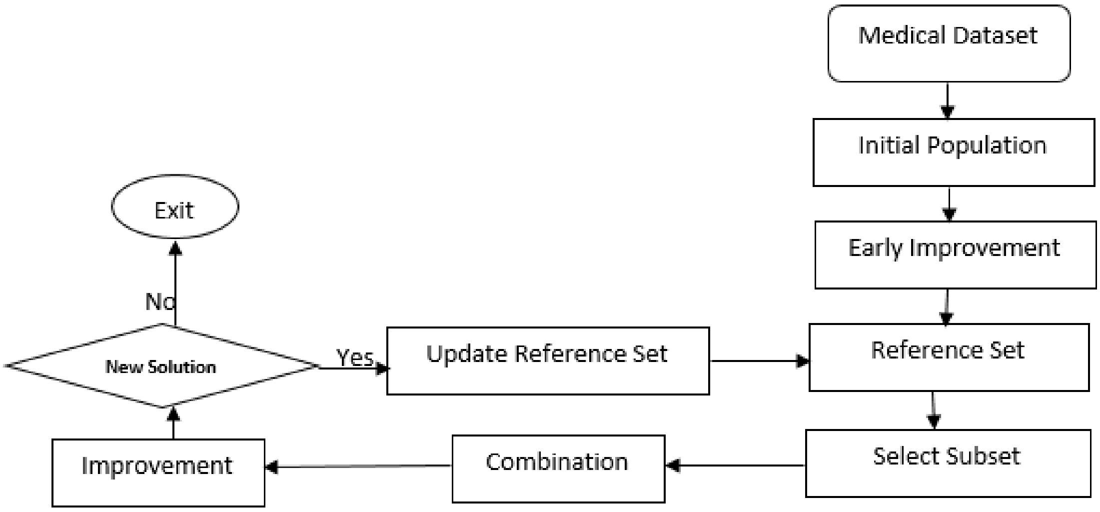


FIGURE (6): Improved ISSS

Experiment and results

The computational experiments performed for this paper aims mainly to show the performance of our wrapper approach (Scatter Search Algorithm and J48 classifier evaluator) in feature selection for serving medical datasets, and the competitively with other MH algorithms. The datasets were taken from UCI repository in (ARFF) format to fit WEKA requirements. The chosen medical datasets are of different branches and different sizes, and most of them already have been used in previous FS studies. Datasets are shown in table 1 above. All experiments had been carried out using a Portable Computer Duo Core CPU with 2.2 GHz speed, and has 4GB memory. We applied the approach on all datasets using fixed and variable parameters.

ISSS Experiments

In this part the size the feature selection process was performed with three different population size which is (|C|/2|, |C|, 2|C|) to see the best result produced according to features and accuracy. The accuracy had been calculated using 10-fold cross validation. Table (2) show the experiment results for 5 or more runs without any change in results.

Table (2): No. of selected features and avg. accuracy for different population size

	PopSize=|C|/2		PopSize=|C|		PopSize=2|C|	
Data Set	Selected Features By ISSS	Avg.Acuracy %	Selected Features By ISSS	Avg.Acuracy %	Selected Features By ISSS	Avg.Acuracy %
Breast-cancer	2	74.8	2	74.8	3	74.5
Breast-w	3	95.4	3	95.6	3	95.6
Heart-c	3	81.5	4	83.5	5	82.5
Heart-stolog	3	78.5	3	85.2	3	85.2
Hepatitis	2	85.1	1	84.5	2	85.1
Lung-Cancer	2	68.8	2	75	No result	No result
Dermatology	8	92.3	8	92.3	No result	No result

From the information given in Table (2), the effect of using different size of population size can be noticed. In average using bigger population size increases the strength of the result according to accuracy and features. But it is obvious that using big population size leads us to unacceptable computational time. So the best results according to our experiment for ISSS approach happens when PopSize=|C|.

In order to recognize the strength of the proposed approach, the results should be compared to other approached. In the study conducted by Kilic, Essiz, and Keles (2023), it was noted that by Palanisamy and Kanmani (2012) achieved the best results by using ABC (Artificial Bee Colony) for feature selection for the same medical datasets used in this paper. However, they tried to improve results using binary ASO (Anarchic Society Optimization) algorithm but their approach has superiority just in Heart-C Dataset as the number of features selected was 5 with the same accuracy.

The following table3 is a comparison between our approach results when PopSize=|C| and the results mentioned by Palanisamy and Kanmani (2012) which is already has been compared to other approaches and shows superiority. It's noteworthy to refer that ABC stands for Artificial Bee Colony Algorithm used for Feature Selection.

Table (3): Results comparison between ISSS and ABC

Dataset	FS by ISSS	accuracy	FS by ABC	accuracy
Heart-C	2	74.8	6-7	86.92
Dermatology	8	92.3	24	98.55
Hepatitis	1	84.5	11	81.26
Lung Cancer	2	75	27	89.25
Breast-w	3	95.6	4	96.99
Heart-stolog	3	85.2	6	84.07

In Table (3), the ISSS approach was implemented with classification accuracy (CA) as the fitness function. The results demonstrate the power and competitiveness of our approach in terms of both the number of selected features and CA. For certain datasets (e.g., heart-c, lung), ISSS excels in choosing fewer features but may yield lower CA. However, for others (e.g., heart-stolog, hepatitis), ISSS demonstrates superiority in both feature selection and accuracy.

Obviously, the wrapper-based approach of ISSS proves its superiority over the filter-based approach in terms of the number of selected attributes. However, despite its efficacy, ISSS does have certain drawbacks. Chief among them is its computational time, particularly when dealing with large datasets. The method demands significant computational resources and can sometimes reach unacceptable levels of time.

It is important to mention that Weka software supports several original algorithms, including Naive Bayes and Support Vector Machine. We applied Naive Bayes to these datasets, but the results were poor compared to our approach and the other studies discussed here.

Conclusion and future work

Scatter Search proves to be a promising approach for feature selection problems. In the medical domain, the ISSS wrapper method effectively conducts feature selection, leveraging its built-in diversification and intensification mechanisms. This method demonstrates clear superiority in both the quantity of selected features and competitiveness in accuracy.

For future studies, we suggest integrating new diversification techniques and improvement methods to enhance the performance and accuracy of solutions, catering to diverse domain datasets. Additionally, exploring different algorithms and approaches for feature selection in medical datasets holds significant potential for advancing research in this area

REFERENCES

- Abd-Alsabour, N. (2018). On the Role of Dimensionality Reduction. J. Comput., 13(5), 571-579.
- Abdulrazzaq, M. B., & Saeed, J. N. (2019, April). A comparison of three classification algorithms for handwritten digit recognition. In 2019 International Conference on Advanced Science and Engineering (ICOASE) (pp. 58-63). IEEE.
- Ayesha, S., Hanif, M. K., & Talib, R. (2020). Overview and comparative study of dimensionality reduction techniques for high dimensional data. Information Fusion, 59, 44-58.
- Bahar, M. H., & Saad, H. (2024). Decision Tree Induction Using Evolutionary Algorithms: A Survey. International Journal of Computing and Digital Systems, 15(1), 99–113. https://doi.org/10.12785/ijcds/150109
- Bouchlaghem, Y., Akhiat, Y., & Amjad, S. (2022). Feature Selection: A Review and Comparative Study. E3S Web of Conferences, 351, 01046. https://doi.org/10.1051/e3sconf/202235101046
- Chen, C. W., Tsai, Y. H., Chang, F. R., & Lin, W. C. (2020). Ensemble feature selection in medical datasets: Combining filter, wrapper, and embedded feature selection results. Expert Systems, 37(5), e12553.

- Chen, J., Yuan, S., Dongdong Lv, & Xiang, Y. (2021). A novel self-learning feature selection approach based on feature attributions. Expert Systems with Applications, 183, 115219–115219. https://doi.org/10.1016/j.eswa.2021.115219
- Cherfi, A., Nouira, K., & Ferchichi, A. (2020). MC4.5 decision tree algorithm: an improved use of continuous attributes. International Journal of Computational Intelligence Studies, 9(1/2), 4. https://doi.org/10.1504/ijcistudies.2020.106485
- Dash, M., & Liu, H. (1997). Feature selection for classification. Intelligent data analysis, 1(1-4), 131-156.
- Elgamal, Z. M., Yasin, N. B. M., Tubishat, M., Alswaitti, M., & Mirjalili, S. (2020). An improved Harris hawks optimization algorithm with simulated annealing for feature selection in the medical field. IEEE access, 8, 186638-186652.
- Erfani, S. M., Rajasegarar, S., Karunasekera, S., & Leckie, C. (2016). High-dimensional and large-scale anomaly detection using a linear one-class.
- Garcıa-López, F., Melián-Batista, B., Moreno-Pérez, J. A., & Moreno-Vega, J. M. (2003). Parallelization of the scatter search for the p-median problem. Parallel computing, 29(5), 575-589.
- García-Pedrajas, N., del Castillo, J. A. R., & Cerruela-García, G. (2021). SI(FS)2: Fast simultaneous instance and feature selection for datasets with many features. Pattern Recognition, 111, 107723. https://doi.org/10.1016/j.patcog.2
- Ghamisi, P., & Benediktsson, J. A. (2014). Feature selection based on hybridization of genetic algorithm and particle swarm optimization. IEEE Geoscience and remote sensing letters, 12(2), 309-313.
- Ghazal, M. M., & Hammad, A. (2022). Application of knowledge discovery in database (KDD) techniques in cost overrun of construction projects. International Journal of Construction Management, 22(9), 1632-1646.
- Glover, F., Laguna, M., & Martí, R. (2003). Scatter search. Advances in evolutionary computing: theory and applications, 519-537.
- Hancock, J. T., Wang, H., Khoshgoftaar, T. M., & Liang, Q. (2024). Data reduction techniques for highly imbalanced medicare Big Data. Journal of Big Data, 11(1), 8.
- Hussain, K., Neggaz, N., Zhu, W., & Houssein, E. H. (2021). An efficient hybrid sine-cosine Harris hawks optimization for low and high-dimensional feature selection. Expert Systems with Applications, 176, 114778.
- Jaddi, N. S., & Abdullah, S. (2020). Global search in single-solution-based metaheuristics. Data Technologies and Applications, 54(3), 275–296. https://doi.org/10.1108/dta-07-2019-0115
- Kaur, N., Singla, J., Mathur, G., Talwani, S., & Malik, N. (2023, November). An Advanced Feature Selection Approach to Improve Intrusion Detection System using Machine Learning. In 2023 7th International Conference on Electronics, Communication and Aerospace Technology (ICECA) (pp. 984-992). IEEE.
- KILIC, U., ESSIZ, E. S., & KELES, M. K. (2023). Binary anarchic society optimization for feature selection. Romanian J. Inf. Sci. Technol, 26, 351-364.
- López, F. G., Torres, M. G., Batista, B. M., Pérez, J. A. M., & Moreno-Vega, J. M. (2006). Solving feature subset selection problem by a parallel scatter search. European Journal of Operational Research, 169(2), 477-489.
- Lyu, Y., Feng, Y., & Sakurai, K. (2023). A survey on feature selection techniques based on filtering methods for cyber-attack detection. Information, 14(3), 191.
- Mullins, I. M., Siadaty, M. S., Lyman, J., Scully, K., Garrett, C. T., Miller, W. G., ... & Knaus, W. A. (2006). Data mining and clinical data repositories: Insights from a 667,000 patient data set. Computers in Biology and Medicine, 36(12), 1351-1377.
- Nadimi-Shahraki, M. H., Banaie-Dezfouli, M., Zamani, H., Taghian, S., & Mirjalili, S. (2021). B-MFO: a binary moth-flame optimization for feature selection from medical datasets. Computers, 10(11), 136.
- Palanisamy, S., & Kanmani, S. (2012). Artificial bee colony approach for optimizing feature selection. International Journal of Computer Science Issues (IJCSI), 9(3), 432.
- Probst, D., & Reymond, J. L. (2020). Visualization of very large high-dimensional data sets as minimum spanning trees. Journal of Cheminformatics, 12(1), 12.
- Quinlan, J. R. (1996). Improved use of continuous attributes in C4. 5. Journal of artificial intelligence research, 4, 77-90.
- Sharma, M., & Kaur, P. (2021). A comprehensive analysis of nature-inspired meta-heuristic techniques for feature selection problem. Archives of Computational Methods in Engineering, 28, 1103-1127.
- Shu, X., & Ye, Y. (2023). Knowledge Discovery: Methods from data mining and machine learning. Social Science Research, 110, 102817.
- Sugumaran, V., Muralidharan, V., & Ramachandran, K. I. (2007). Feature selection using decision tree and classification through proximal support vector machine for fault diagnostics of roller bearing. Mechanical systems and signal processing, 21(2), 930-942.
- Tan, P.-N., Steinbach, M., & Pearson, V. (2006). Introduction to Data Mining.WP CO.

- Vandana, C. P., & Chikkamannur, A. A. (2021). Feature selection: An empirical study. International Journal of Engineering Trends and Technology, 69(2), 165-170.
- Varma, K., Ajmire, P. E., & Rehapande, A. B. (2022). A REVIEW OF DIMENSIONALITY REDUCTION TECHNIQUES FOR HIGH DIMENSIONAL DATA. Journal of the Oriental Institute, 71(4).
- Velliangiri, S., & Alagumuthukrishnan, S. J. P. C. S. (2019). A review of dimensionality reduction techniques for efficient computation. Procedia Computer Science, 165, 104-111.
- Wang, J., Hedar, A. R., Wang, S., & Ma, J. (2012). Rough set and scatter search metaheuristic based feature selection for credit scoring. Expert Systems with Applications, 39(6), 6123-6128.
- Wang, J., Hedar, A. R., Zheng, G., & Wang, S. (2009, April). Scatter search for rough set attribute reduction. In 2009 International Joint Conference on Computational Sciences and Optimization (Vol. 1, pp. 531-535). IEEE.
- WEKA: A Java Machine Learning Package, https://ml.cms.waikato.ac.nz/weka/
- Zebari, R., Abdulazeez, A., Zeebaree, D., Zebari, D., & Saeed, J. (2020). A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction. Journal of Applied Science and Technology Trends, 1(1), 56-70.