

# Improvement of BNSGAI to Increase the Accuracy and Reduction of Feature Selection Subsets

**Hojjat Farrahi Farimani**

Azad University: Islamic Azad University

**Davoud Bahrepour**

Azad University: Islamic Azad University

**Seyed Reza Kamel** (✉ [rezakamel@computer.org](mailto:rezakamel@computer.org))

Islamic Azad University Mashhad Branch <https://orcid.org/0000-0003-3043-0207>

**Reza Ghaemi**

Azad University: Islamic Azad University

---

## Research

**Keywords:** Evolutionary Algorithms, Feature Selection, Feature Dimension Reduction, Genetic Algorithm, NSGAI Genetic Algorithm

**Posted Date:** March 17th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-309963/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

## **Improvement of BNSGAI to Increase the Accuracy and Reduction of Feature Selection Subsets**

**Hojjat Farrahi Farimani<sup>1</sup>, Davoud Bahrepour<sup>2</sup>, Seyed Reza Kamel<sup>2</sup>, Reza Ghaemi<sup>3</sup>**

<sup>1</sup> Department of Computer, Neyshabur Branch, Islamic Azad University, Neyshabur, Iran

<sup>2</sup> Department of Computer Engineering, Mashhad Branch, Islamic Azad University, Mashhad, Iran.

<sup>3</sup> Department of Computer Engineering, Quchan Branch, Islamic Azad University, Quchan, Iran.

### **Abstract**

Feature selection is a process for the elimination of irrelevant and redundant features from a dataset in order to improve learning performance in terms of accuracy and time to build a model from the subsets. The conventional techniques in this regard have limitations such as the high computational overhead for training, even in moderate datasets. Although attention has been paid to the development of rapid and accurate detection techniques, finding a dataset of features that could increase detection accuracy is paramount. The issue with feature selection is the NP-hard problem; therefore, an optimal solution cannot be guaranteed. The present study aimed to propose a new solution for the non-dominated sorting genetic algorithm (NSGA II) by making it binary through the Sigmoid transfer function and a thresholding device for binary feature selection in order to improve the performance in feature selection problems in terms of the accuracy and reduction of the subset dimensions. In addition, the efficiency of the proposed algorithm in reducing the mentioned parameters was measured through comparison with other methods in the four datasets of breast cancer, hepatitis, heart, and diabetes.

**Keywords:** Evolutionary Algorithms, Feature Selection, Feature Dimension Reduction, Genetic Algorithm, NSGAI Genetic Algorithm

## 1. Introduction

The rising trend in the volume of data highlights the importance of the multiple applications of feature selection and reduction techniques, including data mining, multimedia information, processing and analysis of medical data, and automatic adjustment of parameters. Given the large volume of medical data and specific features of each database, it could be concluded that decreasing the dimensions of these features and achieving the optimal minimum mode are among the most important challenges in this regard. This is mainly due to the occasional lack of access to sufficient memory and reduction of processing resources. The main goal of feature selection is to reduce the number of features, so that the predictive accuracy would not decrease. As a result, improper features with lower predictive ability will be eliminated (2, 5), and a number of features that contain proper information and have maximum impact and involvement in all the datasets will remain (3). Various methods have been proposed for the

elimination of irrelevant and duplicate features (4, 6, 7, 28).

Feature selection problem-solving is ideally defined as finding a subset with the lowest number of features, which contains the necessary and sufficient information for the predetermined goals. Considering the non-deterministic polynomial-time (NP) nature of the feature selection problem, achieving a complete and accurate response for feature selection within a short period is not expected (1), which is evident in all the existing feature selection methods. In general, feature selection methods could be classified into three categories of filter methods, wrapper methods, and embedded methods (3, 11, 12). Notably, the use of heuristic and meta-heuristic algorithms in large datasets has proven far more efficient than other methods (2, 5).

Meta-heuristic algorithms are the algorithms that are inspired by natural biological and use the fitness function to guide the search processes. In addition, these methods are part a population as a response during the search process and prevent falling into the local optimal trap. Some of the

most frequently used meta-heuristic algorithms in feature selection problem-solving include the genetic algorithm, firefly algorithm, binary ant colony evolutionary algorithm, and bee colony algorithm, which will be covered in Section 2.

The present study aimed to propose a solution for more efficient feature selection by presenting a model of binary NSGA II (26, 30). The genetic NSGAI algorithm is frequently used owing to its high efficiency in feature selection and dataset reduction (15). Given the nature of feature selection issues and the use of meta-heuristic algorithms as a solution, the selected and eliminated features should be determined, which is only possible in discrete space (17). The subsequent challenge is the development of the mentioned algorithm in the binary format in order to achieve accuracy and reduce the subsets in feature selection. The genetic algorithm uses a threshold transfer function (randomizer function) to obtain an output of the function and make the algorithm binary (discrete) (26, 28).

Another objective of the current research was to propose a practical

solution for the feature selection problem through the improvement and discretization of the NSGA II algorithm using four sigmoid transfer functions, while also introducing a threshold transfer function for the selection or lack of selection of features. The second section of the article contains a review of the literature and previous studies, and the third section introduces the genetic NSGA II algorithm, as well as a proposed method for making the algorithm binary. Evaluations have been discussed and compared in the fourth section.

## 2. Literature Review

Feature selection is a process through which the performance of the data mining process is improved by selecting the effective features in large datasets (3). In general, the features are classified as irrelevant, redundant, and related. In the present study, the main objective was to use related features as far as possible (2, 10). Some features of the data were also considered to evaluate the selected features. Accordingly, feature selection methods were categorized as filter methods, wrapper methods,

and embedded (hybrid) methods (2, 4).

Filter methods are most commonly used for feature selection (13) and involve the feature ranking method, along with a criterion for a set of features (3, 4), such as feature ranking and assembly search algorithms. Wrapper methods deploy duplicate techniques (feature subsets) (14) and a classification function to evaluate the fitness of the feature subsets; in addition, the subsets are discovered by a search algorithm. The method has computational features (13, 14), so that the feature selected for each subset would be a cost vector. In addition, the complete search of all the modes is impossible, and solutions could be found only through simple greedy search; however, the method is not optimal. In most cases, the estimation of fitness is not difficult in these methods. Embedded techniques involve a feature selection process, which is carried out in the inductive algorithm. They often optimize a target function, which jointly scores the accuracy of subset formation and finds the use of more features (15).

Nature-inspired meta-heuristic algorithms are used to solve the problems of feature selection considering the NP-hard nature of these issues, yielding more efficient results. Some of these algorithms include the genetic algorithm, cuckoo algorithm, binary ant colony algorithm, bird flocking algorithm, firefly algorithm, and bat algorithm. Meta-heuristic algorithms play a key role in the feature selection problem-solving process considering the discreteness and complexity of the problem (NP-hard). Overall, discrete meta-heuristic algorithms have great potential for solving this type of problem (2, 16). Some of these algorithms have been explained in the following section.

In (17), the cuckoo optimization algorithm was presented, which is inspired by the lifestyle of a bird family known as cuckoo and its characteristics in egg laying. The common cuckoo does not make a nest of its own, but rather destroys the eggs of other birds (hosts) and lays eggs in their nests, and the process is repeated. The binary version of the algorithm is known as

binary cuckoo search (BCS), which is used for solving discrete issues. Since the BCS algorithm is considered appropriated for solving the problems of application optimization, it is also proper for solving feature selection problems. Furthermore, the BCS algorithm is remarkably efficient in industrial datasets.

In (18, 19), the bee colony algorithm was introduced for feature selection. The algorithm was first proposed to optimize math functions. Each response (a place in the search space) shows a potential food region, and the quality of the response equates to the quality of the food source. In this algorithm, the agents (artificial bees) search and exploit the food sources in the search space. Regarding feature selection,  $d$  steps are taken initially, during which it is decided whether a feature should be selected. At the end of the  $d$  trajectory, the bees return to the hive and assess the results. According to the results of the mentioned study, the algorithm had a proper application in feature selection problem-solving. Given the very low improvement in the results, a combination of the bee colony

algorithm and other algorithms (e.g., genetic and neural network algorithms) was recommended.

In (20, 21), the ant colony algorithm was applied and made binary to solve the feature selection problem. Ants communicate based on a substance known as pheromone that they leave behind. After finding food, ants leave pheromones, so that the possibility of finding the optimal path by other ants would be higher by smelling the pheromone. The method of the algorithm is encoding the strain obtained from the trail of each ant, where zero is considered to be the lack of feature selection, and one is defined as selecting the feature. In order to solve the feature selection problem using the binary ant colony algorithm, a strain of zeros and a strain of ones are considered for the total number of the variables. Afterwards, the binary ant colony algorithm is repeated for the strains. In general, data are reduced using the information efficiency and fast correlation filter methods. Following that, the dimensions of the data decreases, and feature selection occurs with the reuse of the binary ant colony algorithm.

The algorithm has been used for solving feature selection problems in image processing (classification) and resetting Persian handwritten figures, yielding extremely beneficial results.

In (2, 22, 24), the firefly algorithm was introduced to accelerate the pace of the feature selection process. The algorithm has been designed by modeling the characteristics of fireflies. Fireflies show their charm by radiating light and attracting other fireflies based on their attractiveness and distance. The parameters of light intensity and attractiveness determine the relationship between the worms. The algorithm is used to solve various optimization problems, such as feature selection. In fact, the algorithm communicates with  $n$  fireflies, each of which has specific features in the state space. Each firefly moves toward a closer worm, thereby increasing the intensity of light, and a worm is eliminated if it is not attracted by any firefly. The firefly algorithm operates faster compared to other methods in terms of reducing the size of features.

The studies in (24, 25) played a key role in feature selection by introducing the flocking algorithm and

making it binary. The algorithm acts based on the random production of an initial population, and each member of the algorithm within the population is defined by the speed vector and position vector in the search space. In the iteration, the new position of the particles is separately updated based on the current speed vector and the optimal position found by the particle in the group. While the nature of the algorithm is continuous, it has been referred to as the binary particle swarm optimization in the discrete state considering the specific uses. In this algorithm, the position of each particle is shown with a binary value of zero or one. In the binary algorithm, the value of each particle may vary from zero to one or from one to zero. In addition, the speed of each particle is defined as the possibility of changing status from one to zero and vice versa. The algorithm is extensively used in optimization functions, artificial intelligence training networks, and fuzzy control systems.

In (26-28, 30), the binary genetic algorithm was introduced, and its use in feature selection problem-solving was explained. The algorithm uses a

set of solutions or chromosomes in a population, producing a better next generation based on the fitness function and dimensional population production using mutation and mating. The application of a discrete genetic algorithm and its combination with other filter and fitness classification methods could result in a better response for the reduction of selected subsets. However, the problem in this regard is that each method alone cannot yield a proper response. Therefore, a combination of various techniques should be used, which requires more memory and processing time. The NSGA II algorithm has temporal complexity for problem-solving, and there may not be a precise answer to the problem at a given time, and only a close approximation to the optimal response could be achieved. The use of these algorithms and their combination with other techniques results in proper responses at better time intervals for genetic algorithms. The steps of the algorithms and the order of their execution process are also extremely important considering the creation of an initial population, selection of parents and carrying out

intersections and establishing a population of children, selection of parents and carrying out mutation and creating a population of mutants, selection of the main new population members among the main population of children and mutants, and continuing the cycle from the second stage if required. The discrete genetic algorithm generally operates efficiently in solving optimization problems and has proper accuracy in addition to the relatively favorable speed. Nevertheless, the findings in this regard have indicated that the combined use of this algorithm with other methods may increase accuracy and speed.

Solving feature selection problems in terms of accuracy indexes and reducing the subset dimensions are paramount. A solution in this regard involves the use of meta-heuristic algorithms. Therefore, binary meta-heuristic algorithms are considered to be proper solutions for feature selection. In addition, the use of each algorithm in various situations yields varied results in the assessment of accuracy and reduction of dimensions depending on their parameters. The following section of the article

discusses the optimization of the accuracy and reduction of the subset dimensions in feature selection by presenting a binary NSGA II algorithm.

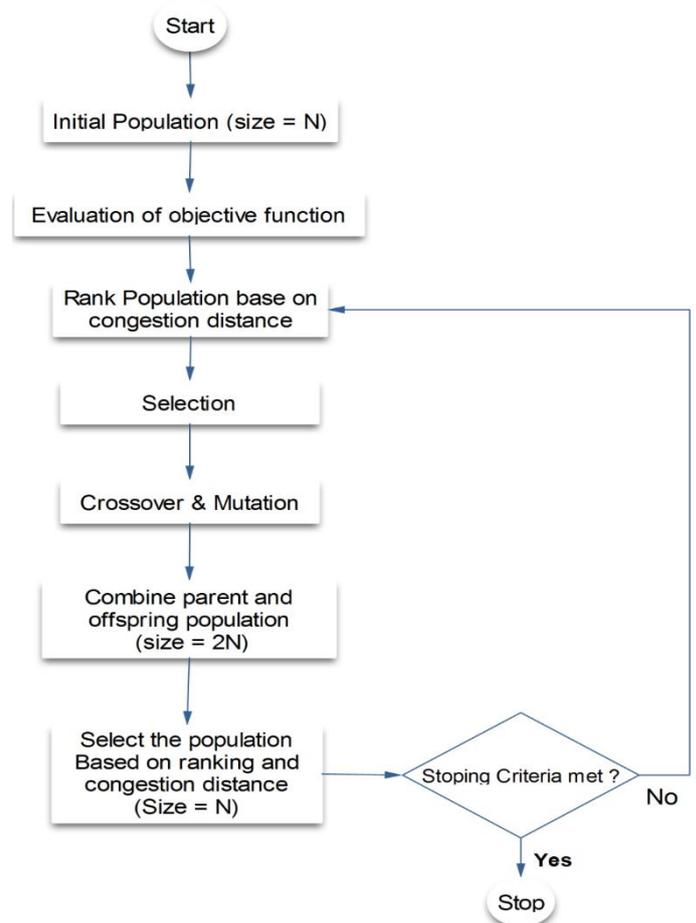
### 3. Proposed Method

In the previous section, the methods and use of meta-heuristic algorithms in solving feature selection problems were explained. In this section, we initially introduce the basic NSGA II algorithm, followed by introducing the proposed algorithm.

#### 3.1. NSGA II Algorithm

The genetic algorithm is an exploratory problem-solving algorithm, which has emerged from the biological modeling of living organisms. In this algorithm, the characteristics of living organisms are created based on the number of the target functions and improved characteristics of each generation over time and through adaptation, and

the next new generation from the previous generation is obtained by mixing and improved by the target function (30)



**Figure 1.** Flowchart of Performance of NSGA II Algorithm

Figure 1 shows the Flowchart of the NSGA II multipurpose genetic

algorithm. The first stage involves the production of the initial population by

generating the  $N$  (random) population. At the second stage, the fitness of each member of the population is estimated using the fitness function. Afterwards, the crowding distance is estimated in the initial population after applying the rank and the crowding distance in the parent selection method when the initial population is arranged based on conditions to overcome, followed by a selection from the population. In addition, two important parameters are involved in feature selection, including the population rank and estimation of distance (26, 27). Stages 1-6 are depicted in Figure 2.

The number of Pareto Fronts was classified based on the ranking method, and the crowding distance method was applied to distinguish the cases with equal ranks in order to determine those that should be selected. If the evaluation was still possible, we entered the seventh stage (i.e., selection), where the algorithm used the roulette wheel method for the selection. The mutation process was carried out in the following stages to produce the population. In the 10<sup>th</sup> and 11<sup>th</sup> stages, the fitness function occurred for the new generation after

adding the generated population to the initial population. In the 12<sup>th</sup> and 13<sup>th</sup> stages, a new population was selected again based on the rank and crowding distance with the  $N$  size (initial population), and the other populations were eliminated. The process continued until the selection and classification of all the Pareto Fronts. The algorithm has improved the problems of multi-objective evolutionary algorithms (e.g., speed). Overall, the method has a higher speed compared to other methods due to the elimination of computational complexities (29, 30). Another advantage is the use of the crowding distance between the similar responses in the ranking method. Notably, the crowding distance is a factor for the better selection of responses in terms of dispersion, which was applied on a number of the responses, and the stages of selection, intersection, and cutting were carried out based on the NSGA II algorithm. Figure 2 depicts the behavior and implementation of the algorithm, along with the name of the applied function (30).

### **3.2. Binary NSGA II Algorithm**

The present study aimed to propose a solution based on the improvement of the accuracy criteria and reduction of the feature dimensions in the filter method by introducing an NSGA II genetic algorithm discretization method. Given the discrete nature of heuristic and meta-heuristic algorithms, some cases required a change from the continuous nature to the discrete nature similar to the feature selection problem in order to be used for solving the issues associated with this nature (26). Since the NSGA II algorithm operates properly in solving discrete problems and has less time order due to eliminating computational complexities (26, 29, 31), it could be more efficient in solving feature selection problems compared to other meta-heuristic algorithms. In the current research, we attempted to resolve the feature selection problem through the discretization of the algorithm. Contrary to continuous problems, discrete problems are mostly in the space between 0-1,

where there is the need for a true mapping from the continuous to the discrete nature (between 0-1) or the selection/lack of selection. Transfer functions were applied for this purpose in the present study.

Due to the discrete nature of feature selection problems, the selected and unselected features must be clearly expressed in the feature selection problem. Meanwhile, this is not possible in the continuous process, and transfer functions should be used in order to discretize the proposed algorithm. There are various transfer functions, such as perceptron and sigmoid, each of which is of various types. Both functions are used in neural networks and discretization using transfer functions. However, the findings in this regard have been indicative of the better output of sigmoid functions for more complicated problems compared to perceptron (7). Furthermore, sigmoid functions have features such as monotony, first-order derivation, and convex output for the values less than zero and concave output for the values greater than zero. With regard to this performance, it could be stated that

the sigmoid function and its other forms have several optimal values

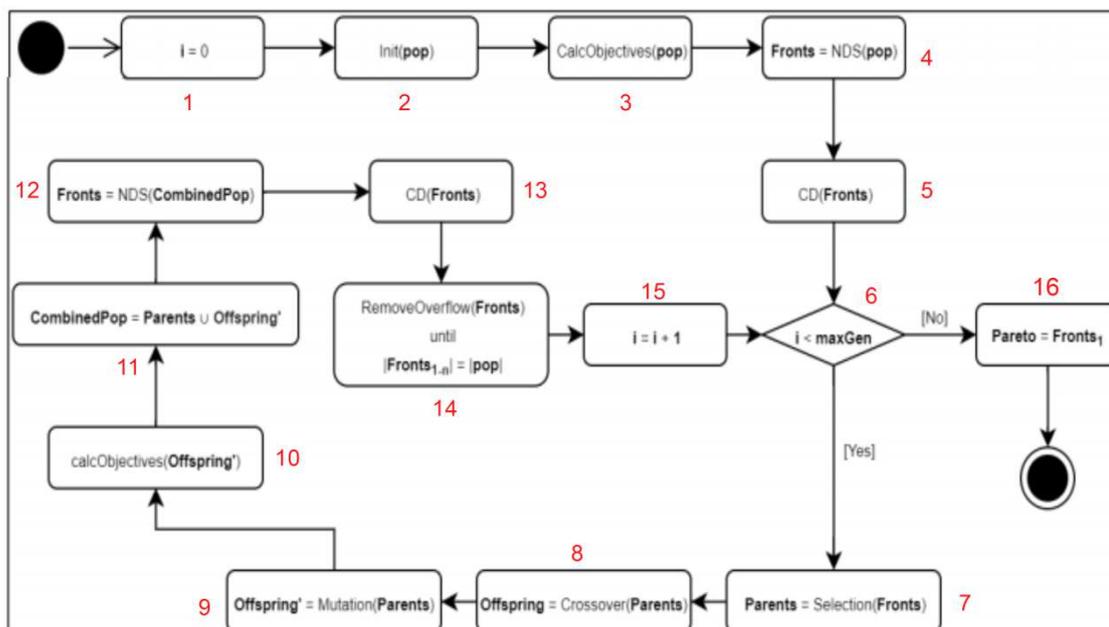


Figure 2. NSGAI I Algorithm activity diagram

Given the nonlinearity of sigmoid function characteristics, the S-shaped function (7) shown in Equations 1-4

was used for transfer in the current research

$$T(x) = \frac{1}{1+e^{-2x}} \quad (1)$$

$$T(x) = \frac{1}{1+e^{-x}} \quad (2)$$

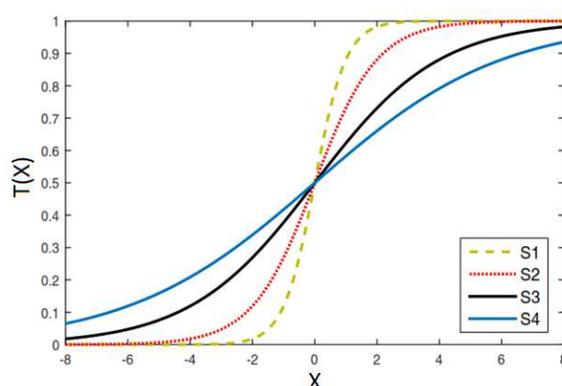
$$T(x) = \frac{1}{1+e^{\frac{x}{2}}} \quad (3)$$

$$T(x) = \frac{1}{1+e^{\frac{x}{3}}} \quad (4)$$

In addition to higher coverage, using the four different forms of sigmoid function, (Equations 1-4) could increase the accuracy of feature selection (=1) rather than the In the present study, the sigmoid transfer function was used to make

selection (n=0). Figure 2 shows an overview of the four sigmoid transfer functions (X). As is observed, the output of the four sigmoid functions is continuous and between 0-1.

continuous solutions, a thresholding device was used for the final



**Figure 3.** Overview of the four transfer functions

the mentioned algorithm binary. As mentioned earlier, thresholding occurred in discretization following the sigmoid function output. The optimal mode of meta-heuristic algorithms involves the use of a randomization function to determine the threshold (23, 25). After applying one of the four sigmoid functions on

transformation of the algorithm to the discrete (binary) form (Equation 5). Since the decision variable in the binary NSGA II algorithm was only within the range of 0-1, a nonlinear function (V) was used to map the current status into another function or space (29).

$$F_s(t+1) = \begin{cases} 1 & \text{if rand}(0,1) < \text{Sig}(T(x)) \\ 0 & \text{if rand}(0,1) \geq \text{Sig}(T(x)) \end{cases} \text{ In}$$

the equation above,  $F_s$  is the set of the features and introducer of the selected (=1) and non-selected features (=0), showing the direction of the speed and transfer vectors in the discrete mode of the NSGA II algorithm.

#### 4. Evaluation and Comparison

In this section, we have analyzed the evaluated data and results. In the present study, the UCI standard dataset was applied (32), and MATLAB software was also exploited to implement the proposed method. Table 1 shows the four evaluated datasets (32). With regard to the number of the datasets, records, and a number of the features (samples), four datasets with a diverse number of records were used to evaluate the proposed method (Table 1). Moreover, we examined the two criteria of accuracy and feature reduction rate in the dataset and compared the evaluation results with other methods.

**Table 1.** Evaluated data sets

<i>Examples</i>	<i>Records</i>	<i>Data Sets</i>	<i>Row</i>
9	788	diabet	1
9	699	brastcancer	2
13	270	heart	3
19	155	hepatitis	4

#### *Accuracy Criterion*

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

Accuracy is the most important criterion for determining the performance of a classification algorithm, which indicates the extent to which a total set of experimental records is categorized correctly. In the current research, the accuracy criterion was determined using Equation 6 (4-6), where  $TP$  is the true positive,  $TN$  is the true negative,  $FP$  shows the false positive, and  $FN$  shows the false positive.

#### *Feature Selection Size Criterion*

$$\text{Size} = \frac{1}{M} \sum_{i=1}^M \frac{d_i^*}{D} \quad (7)$$

This criterion indicated the number of the feature selections (subsets) that were calculated using Equation 7, where  $M$  is the number of the implementation times,  $d_i^*$  shows the number of the selected features in the  $i$ th implementation stage, and  $D$  is the

total number of the features in the tested dataset (4-6).

The accuracy criterion also shows the ability of the algorithm to detect patient categories (4, 5).

$$Accuracy = \frac{TP}{TP+FN} \quad (8)$$

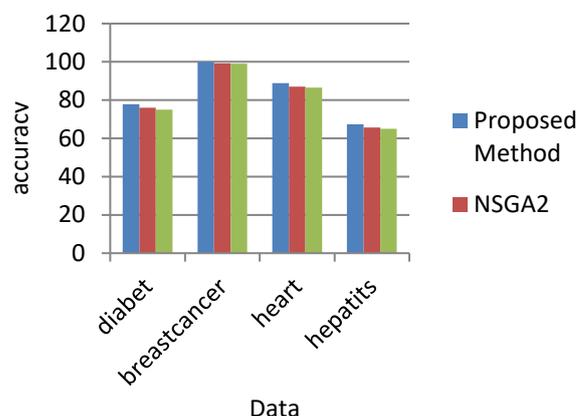
Table 2 shows the evaluation of the four datasets and their outputs, as well. Accordingly, feature selection was estimated with 100% accuracy in the proposed method compared to other techniques in the breast cancer datasets.

**Table 2.** Comparison of accuracy criteria in the studied data sets

Data Setese				Feature Selection Algorithms	Row
Accuracy					
hepatitis	heart	Brast cancer	diabet		
65.01	86.52	99.00	74.02	GA	1
65.79	87.03	99.28	75.97	NSGAI	2
67.40	88.88	100	77.92	Proposed Method (BNSGAI)	3

Figure 4 depicts the highest accuracy percentage of the proposed method (i.e., BNSGA II algorithm) compared to the other two genetic algorithms in the four datasets. Table 3 contains the data on the comparison of the

as the comparison of the proposed method with other algorithms based on the introduction of assessable parameters.



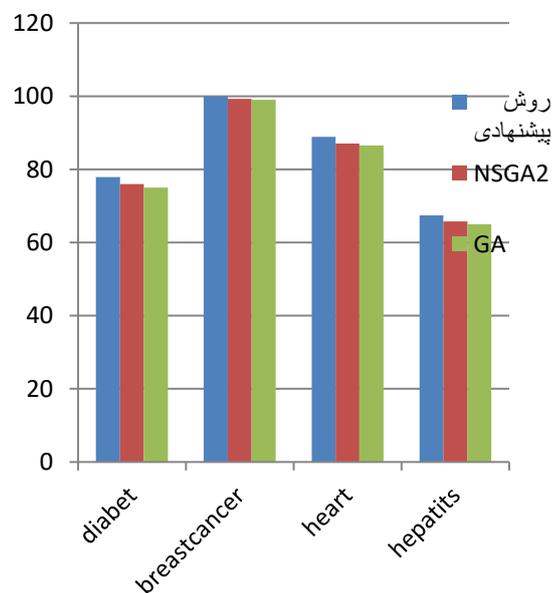
**Figure 4.** Compare accuracy percentage in proposed method and other algorithms with the four datasets in the accuracy criterion (P) and reduction criterion (R). Accordingly, the feature selection reduction and accuracy criterion were used together and presented by Equations 7-8, which encompass the same feature selection concept. Feature selection involves the elimination of unnecessary data and identification of necessary data. Therefore, the accuracy of a model depends on the reduction rate by feature selection. Accuracy is a criterion that requires calibration, which occurs by the reduction rate, the output of which is shown in Table 3. Figure 5 depicts the

reduction of the selected features in each dataset, and Table 4 shows the number of the selected features between the NSGA II algorithm and the presented binary algorithm to the total number of the features in each dataset. The datasets on diabetes had nine features, five of which were

selected by the NSGA II algorithm. However, the proposed method selected three features, and the reduction further involved the selection of the subsets.

**Table 3.** Comparison of Precision and reduction in the studied data sets

Data Setes				Feature Selection Algorithms	Row
Precision/Reduction					
hepatitis	heart	Breast cancer	diabetes		
$P=84.12$ $R=80\%$	$P=94.8$ $R=70\%$	$P=92.56$ $R=58\%$	$P=73.70$ $R=50\%$	GA	1
$P=85.81$ $R=83\%$	$P=95.23$ $R=71\%$	$P=97.96$ $R=63\%$	$P=67.85$ $R=58\%$	NSGAI	2
$P=90.23$ $R=85\%$	$P=95.69$ $R=72\%$	$P=99.8$ $R=64\%$	$P=74.22$ $R=60\%$	Proposed Method (BNSGAI)	3



**Figure 5.** Compare Reduction percentage in feature selection

Table 4. Comparison of the number of selected features in NSGA2 base and BNSGA2 algorithms

<i>Number of selected attributes in each dataset</i>				<i>Methods</i>	رد یف
hepatitis <i>max : 19</i>	heart <i>max : 13</i>	<i>brastcancer</i> <i>max : 9</i>	<i>diabet</i> <i>max : 9</i>		
9	7	5	5	<i>NSGAI</i>	1
8	4	2	3	<b><i>Proposed Metod (BNSGAI)</i></b>	2

The results of the present study confirmed the more efficient performance of the BNSGA II algorithm, and it was concluded that the BNSGA II yielded better results regarding the selection accuracy and reduction of the selected features compared to the basic genetic algorithm and NSGA II algorithm. However, the speed of selecting the features slightly decreased due to the increased accuracy and decreased subset selection. The results of the mentioned algorithms and the presented model by MATLAB software were implemented. For each algorithm and dataset used for the test, the program was executed five times, and the average results were

considered as the output. Therefore, the proposed BNSGA II algorithm could be used for solving scheduling problems, feature selection, automatic adjustment of discrete parameters, clustering, and problems associated with the discrete nature. By improving its performance, time reduction could also be considered along with accuracy and the reduction of subset dimensions.

## 5. Conclusion

This study aimed to propose an effective solution for the feature selection problem in terms of increasing the accuracy criterion and decreasing feature selection and the selected subsets by making the NSGA II algorithm binary. Using four

sigmoid functions and a randomizer function for the selection/lack of selection of the features in the continuous solutions of the filter method and a thresholding device, we attempted to make the algorithm discrete. As a result, four UCI datasets (32) were evaluated in terms of the accuracy criterion and feature selection reduction and compared with other algorithms. According to the results, accuracy in feature selection and a further reduction in the subset selection could be achieved although at a slower rate in significantly larger datasets as the computational volume was slightly larger than the NSGA II algorithm. In addition, the parameters could be fine-tuned to categorize the features and reduce their dimensions to increase the speed of classification in addition to its accuracy.

## Appendix

Feature selection: Feature selection problem-solving is ideally defined as finding a subset with the lowest number of features, which contains the necessary and sufficient information for the predetermined goals.

Ethics approval and consent to participate  
Not applicable.

Consent for publication  
Not applicable.

Competing interests

The authors have no conflict of interest.

Funding

Not applicable

Authors' contributions

Not applicable

## Acknowledgements

Acknowledgements The authors are thankful to anonymous reviewers for their valuable comments and suggestions that helped improving the quality of the paper.

Authors' information

Not applicable

## 6. Resources

- 1- Rashedi, Esmat, Hossein Nezamabadi-Pour, and Saeid Saryazdi. "GSA: a gravitational search algorithm." *Information sciences* 179.13 (2009): 2232-2248.
- 2- Mousavirad, S. J., & Ebrahimpour-Komleh, H. (2013, October). Feature selection using modified imperialist competitive algorithm. In *ICCKE 2013* (pp. 400-405). IEEE.

- 3- Wang, Ling, et al. "Feature selection based on meta-heuristics for biomedicine." *Optimization Methods and Software* 29.4 (2014): 703-719.
- 4- Vieira, Susana M., João MC Sousa, and Uzay Kaymak. "Fuzzy criteria for feature selection." *Fuzzy Sets and Systems* 189.1 (2012): 1-18.
- 5- Alelyani, Salem, J.Tang, and H. Liu. "Feature selection for clustering: A review." In *Data Clustering*, pp. 29-60. Chapman and Hall/CRC, 2018.
- 6- Urbanowicz, Ryan J., M.Meeker, W. La Cava, R.S. Olson, and J. H. Moore. "Relief-based feature selection: Introduction and review." *Journal of biomedical informatics* 85 (2018): 189-203.
- 7- S.A.Mirjalili, and A. Lewis. "S-shaped versus V-shaped transfer functions for binary particle swarm optimization." *Swarm and Evolutionary Computation* 9 (2013): 1-14.
- 8- S.A.Mirjalili, and Siti Zaiton Mohd Hashim. "BMOA: binary magnetic optimization algorithm." *International Journal of Machine Learning and Computing* 2.3 (2012): 204.
- 9- Arora, Sankala, and P. Anand. "Binary butterfly optimization approaches for feature selection." *Expert Systems with Applications* 116 (2019): 147-160.
- 10- Guyon, I.Elisseeff, A. "An introduction to variable and feature selection. *Journal of Machine Learning Research*", Vol. 3, pp.1157–1182, 2003.
- 11- M. Dash and H. Liu, "Feature selection for classification," *Intelligent data analysis*, vol. 1, pp. 131-156, 1997.
- 12- H.Liu and L.Yu, "Toward integrating feature selection algorithms for classification and clustering," *IEEE Transactions on knowledge and data engineering*, vol. 17, pp. 491-502, 2005.
- 13- R.Kohavi and G.H.John, "Wrappers for feature subset selection," *Artificial intelligence*, vol. 97, pp. 273-324, 1997.
- 14- H.Zheng, Y.Zhang, "Feature" selection for high-dimensional data in astronomy", *Advances in Space Research*, Vol. 41, No. 2, pp. 1960–1964, 2008.
- 15- I.S.Oh, J.S. Lee, and B.R. Moon, "Hybrid genetic algorithms for feature selection," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 26, pp. 1424-1437, 2004.
- 16- Baghel, Malti, S. Agrawal, and S. Silakari. "Survey of metaheuristic algorithms for combinatorial optimization." *International Journal of Computer Applications* 58, no. 19 (2012).
- 17- Rodrigues, Douglas, L.A.M Pereira, T. N. S. Almeida, J. Paulo Papa, A. N. Souza, C. C. Ramos, and X.S Yang. "BCS: A binary cuckoo search algorithm for feature selection." In *2013 IEEE international symposium on circuits and systems (ISCAS2013)*, pp. 465-468. IEEE, 2013.
- 18- Schiezero, Mauricio, and H. Pedrini. "Data feature selection based on Artificial Bee Colony algorithm." *EURASIP Journal on Image and Video processing* 2013, no. 1 (2013): 47.
- 19- Ozturk, Celal, E. Hancer, and D. Karaboga. "A novel binary artificial bee colony algorithm based on genetic operators." *Information Sciences* 297 (2015): 154-170.
- 20- Wan, Youchuan, M. Wang, Z. Ye, and X. Lai. "A feature selection method based on modified binary coded ant colony optimization algorithm." *Applied Soft Computing* 49 (2016): 248-258.
- 21- Kashef, Shima, and H. Nezamabadi-pour. "A new feature selection algorithm based on binary ant colony optimization." In *The 5th Conference on Information and Knowledge Technology*, pp. 50-54. IEEE, 2013.
- 22- Sawhney, Ramit, P. Mathur, and R. Shankar. "A firefly algorithm based wrapper-penalty feature selection

- method for cancer diagnosis." In International Conference on Computational Science and Its Applications, pp. 438-449. Springer, Cham, 2018.
- 23- Marie-Sainte, S. Larabi, and N. Alalyani. "Firefly algorithm based feature selection for arabic text classification." *Journal of King Saud University-Computer and Information Sciences* (2018).
- 24- Eberhart, Russell, and J. Kennedy. "A new optimizer using particle swarm theory." In *MHS'95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science*, pp. 39-43. Ieee, 1995.
- 25- Chuang, Li-Yeh, H.W Chang, C.J Tu, and C.H Yang. "Improved binary PSO for feature selection using gene expression data." *Computational Biology and Chemistry* 32, no. 1 (2008): 29-38.
- 26- Deb, Kalyanmoy, A. Pratap, S. Agarwal, and T. A. M. T. Meyarivan. "A fast and elitist multiobjective genetic algorithm: NSGA-II." *IEEE transactions on evolutionary computation* 6, no. 2 (2002): 182-197.
- 27- Deb, Kalyanmoy. "An introduction to genetic algorithms." *Sadhana* 24, no. 4-5 (1999): 293-315.
- 28- Hammami, Marwa, S. Bechikh, C.Cheng Hung, and L. B Said. "A Multi-objective hybrid filter-wrapper evolutionary approach for feature selection." *Memetic Computing* 11, no. 2 (2019): 193-208.
- 29- Senthilkumar, Chinnamuthu, Gowrishankar Ganesan, and R. Karthikeyan. "Optimization of ECM process parameters using NSGA-II." *Journal of Minerals and Materials Characterization and Engineering* 11, no. 10 (2012): 931.
- 30- Johansson, Robin. "NSGA-II DESIGN FOR FEATURE SELECTION IN EEG CLASSIFICATION RELATED TO MOTOR IMAGERY." (2020).
- 31- Binder, Martin, J.Moosbauer, J.Thomas, and B. Bischl. "Multi-Objective Hyperparameter Tuning and Feature Selection using Filter Ensembles." *stat* 1050 (2020): 13.
- 32- M. Lichman. 2013. UCI Machine Learning Repository. (2013). <http://archive.ics.uci.edu/ml>

# Figures

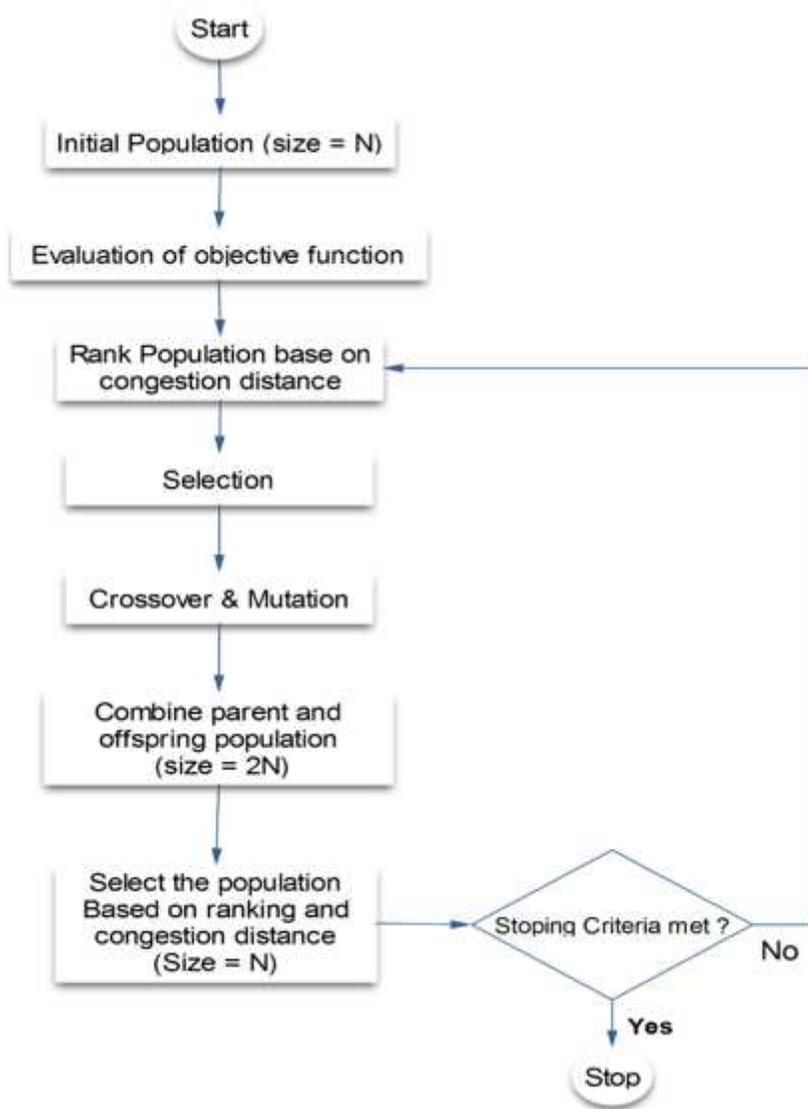


Figure 1

Flowchart of Performance of NSGA II Algorithm

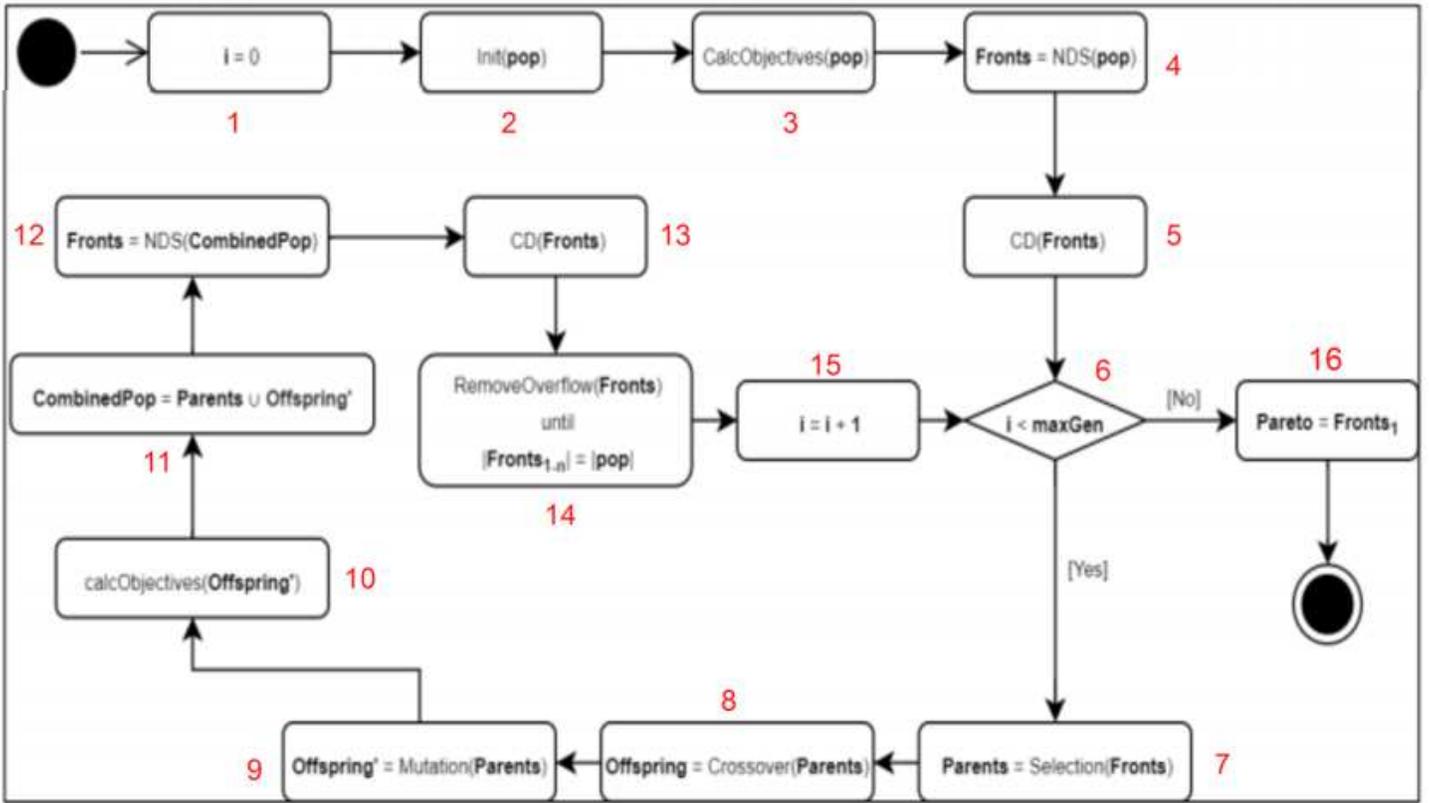


Figure 2

NSGAII Algorithm activity diagram

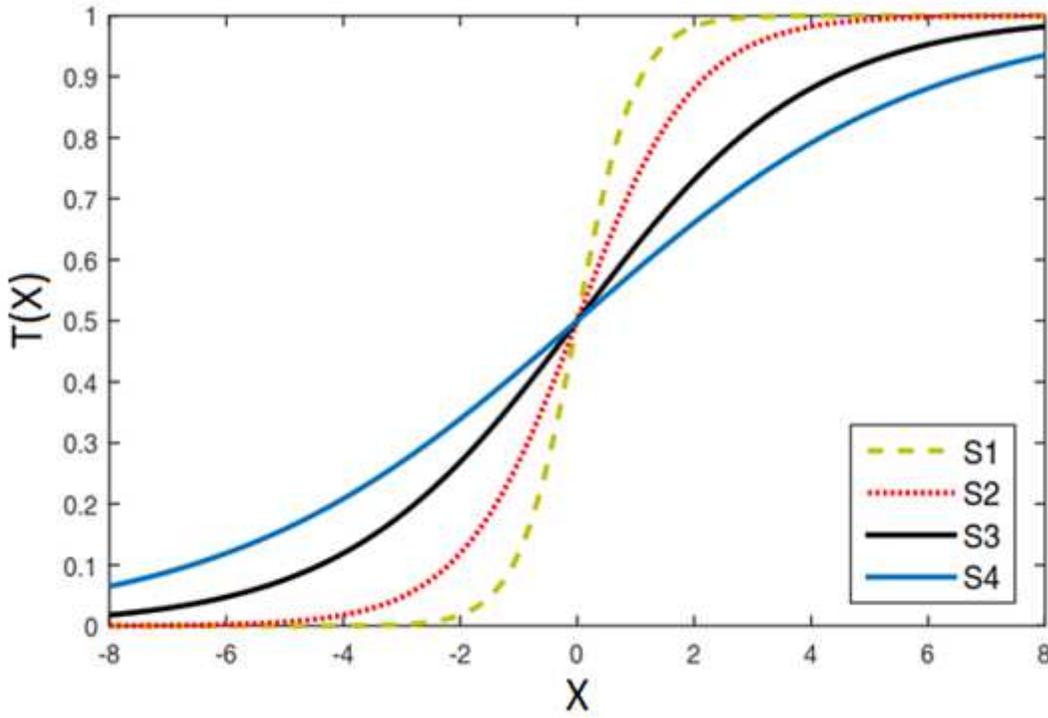


Figure 3

Overview of the four transfer functions

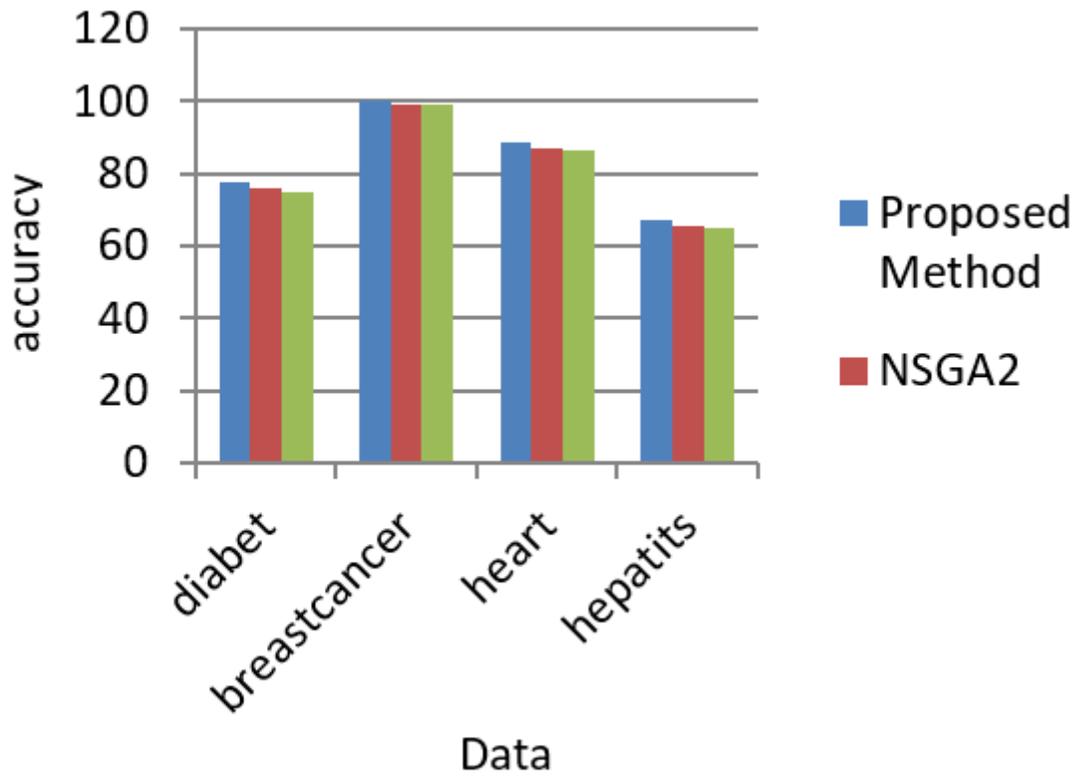


Figure 4

Compare accuracy percentage in feature selection

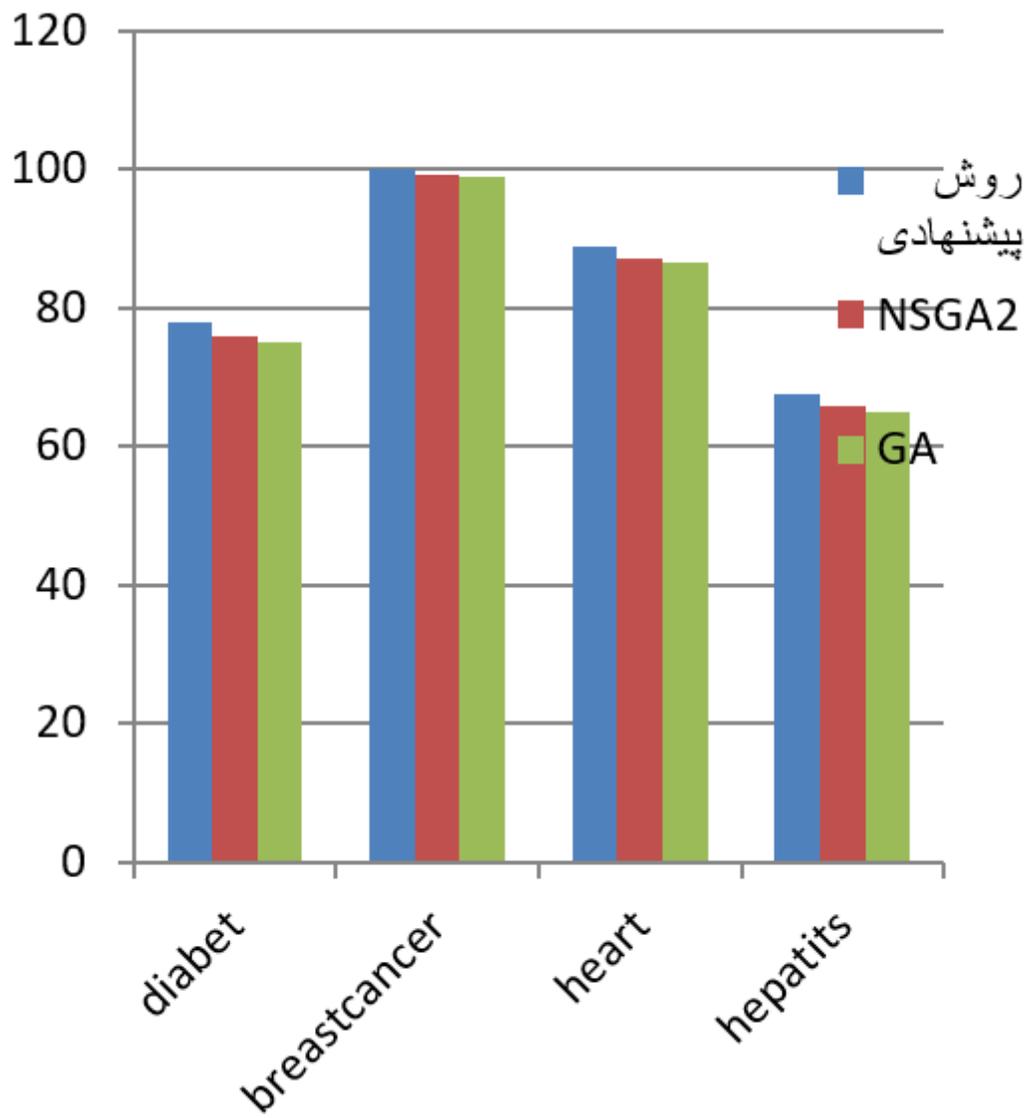


Figure 5

Compare Reduction percentage in feature selection