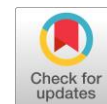


Cuckoo inspired algorithms for feature selection in heart disease prediction



Ali Muhammad Usman ^{a,b,1,*}, Umi Kalsom Yusof ^{a,2}, Syibrah Naim ^{a,3}

^a School of Computer Sciences, Universiti Sains Malaysia, Pulau Penang, Malaysia

^b Department of Computer Sciences, Federal College of Education (Technical) Gombe, Nigeria

¹ muhammadaliusman@student.usm.my; ² umiyusof@usm.my; ³ syibrah@usm.my

* corresponding author

ARTICLE INFO

Article history

Received June 19, 2018

Revised July 16, 2018

Accepted July 17, 2018

Keywords

Heart disease

Cuckoo search

Feature selection

Cuckoo optimization algorithm

Meta-heuristic algorithms

ABSTRACT

Heart disease is a predominant killer disease in various nations around the globe. However, this is because the default medical diagnostic techniques are not affordable by common people. This inspires many researchers to rescue the situation by using soft computing and machine learning approaches to bring a halt to the situation. These approaches use the medical data of the patients to predict the presence of the disease or not. Although, most of these data contains some redundant and irrelevant features that need to be discarded to enhance the prediction accuracy. As such, feature selection has become necessary to enhance prediction accuracy and reduce the number of features. In this study, two different but related cuckoo inspired algorithms, cuckoo search algorithm (CSA) and cuckoo optimization algorithm (COA), are proposed for feature selection on some heart disease datasets. Both the algorithms used the general filter method during subset generation. The obtained results showed that CSA performed better than COA both concerning fewer number of features as well as prediction accuracy on all the datasets. Finally, comparison with the state of the art approaches revealed that CSA also performed better on all the datasets.

This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



1. Introduction

Heart disease is the highest killer disease in many countries in the world including UK, USA, Canada, and Wales. Nearly, 370,000 deaths are recorded in the US each year due to heart disease while, in India, nearly two million peoples are being attacked with heart failure every year, out of which most them are youths. In addition, research showed that in India an average of 1 death in every 33 seconds while in US heart disease attack someone in every 42 seconds [1]. Its projected that near 31% of the global death is due to heart disease and the figure is predicted to rise over 130 million by the year 2035 [1]. There are many features to be considered when doctors diagnose heart diseases, which may be very difficult for the doctors to recognize them and diagnose quickly and accurately. These lead to the reason why soft computing approaches are being used to assist the doctors and rescue the prevailing situation.

Presently, there is some heart disease prediction system (HDPS) that are based on soft computing paradigms. Most of these HDPS models comprises two portions, feature selection (FS) and the classification. In FS, the most relevant features of the heart disease are selected. Whereas, the selected subset features are used as an input in the classification part [2], [3]. Heart disease datasets contain both irrelevant and inessential features that do not contribute at all but rather make noise towards an explanation of the goal class [4]. As such removing these redundant and irrelevant features is imperative

in as much as the classification accuracy must be improved [5]. This will decrease the risk of overfitting data, affords better prediction, and reduced amount of computation time with fewer features [4], [6].

Several researchers have proposed the use of FS in heart disease. Jabbar et al. [7] used Correlation-based Feature Selection (CFS) and Random Forest Classifier (RFC) for heart disease diagnosis and achieved a better result compare to the previous one mentioned in the literature. Verma et al. [2] proposed a non-invasive HDPS using Cleveland data from UCI machine learning and Indira Gandhi college data. The system comprises of dual stages. In the first stage, Particle Swarm Optimization (PSO) algorithm was applied as a search method with CFS plus K-Means Clustering (KMC) for feature selection and extraction. The results obtained in the initial stage was used as an input in the next stage whereby, four different classifiers namely Multi-layer perceptron, (MLP), C4.5, Multinomial Logistic Regression (MLR) as well as the Fuzzy Unordered Rule Induction Algorithm (FURIA) are used to train the model. Experimental outcomes disclosed that MLR has the maximum prediction accuracy of 88.4%. Recently, Shah et al. [8] used Probabilistic Principal Component Analysis (PPCA) to handle omitted values as well as feature extraction with the help of parallel analysis. The feature vectors with reduce dimension are submitted to Radial Basis Function (RBF) based Support Vector Machine (SVM) for classification. Accurateness of 82.18%, 85.82% and 91.30% were achieved on Cleveland, Hungarian, and Switzerland data respectively. Similarly, an accuracy of 83% was obtained on arrhythmia data from UCI machine learning in the work of Vivekanandan and Iyengar [3]. In the paper, modified differential evolution was used as FS and fuzzy feed forward neural network for prediction. Finally, Jabbar [9] applied PSO for FS on heart disease dataset and a reliable correctness is obtained with the support of K-nearest neighbor (KNN) as a classifier.

Cuckoo inspired metaheuristic algorithms are of two types: the cuckoo search algorithm (CSA) by means of levy flight developed by Yang and Deb [10] in 2009 and cuckoo optimization algorithm (COA) established by Rajabioun [11] in 2011. Although, they are not the same but share some common characteristics. In this research, the two algorithms are used as filter-based FS in heart disease prediction. Although, Gadekallu and Khare [12] combined CSA together with Rough Set Theory (RST) for FS on some heart disease datasets. So far none of the research used COA as FS. However, in this research both CSA and COA will be used and compared for feature subset selection.

The key goal of this study is to offer an effective and efficient HDPS that will predict heart disease with fewness features and improve accuracy. In the proposed model, both CSA and COA will be implemented for FS. A comparison of the two cuckoo inspired algorithms will be investigated for FS. Finally, the reduced features will be trained with four well-known classification algorithms namely, naive Bayes (NB), RFC, MLP, and SVM.

2. Method

2.1. Feature Selection

Feature selection (FS) refers to the process of selecting a subset from the actual set of features or attributes from a given data set while ignoring the redundant or irrelevant features [3], [13]. The best feature subset (called the optimal) is measured based on an evaluation condition. However, discovering the optimal feature is generally intractable this is due to the fact that the increase in dimensionality increases the number of features as well [3], [5], [13]. Numerous problems connected to FS are proved to be NP-hard.

FS can be a filter or a wrapper. The filter feature selection uses statistical characteristics to rank features. The highest ranked features are considered for inclusion while the fewer rank features are ignored [14]. They can scale up to a large amount of data, computationally fast, and do not depend on any mining algorithm [5]. The wrapper, on the other hand, used a mining algorithm to determine the goodness of selected features, the subset that provides higher performance are selected [14]. The major drawback of the wrapper model is classifier dependency, computationally expensive and is not suitable for large datasets [3], [13].

In this study, filter-based FS is employed to choose the most relevant attributes or features from the heart disease datasets. The general filter algorithm by Liu and Yu [13] depicted in Fig. 1 is adapted and enhanced with the cuckoo inspired algorithms.

Algorithm 1 General Filter-based Algorithm

Insert: $D (F_0; F_1; \dots; F_{n-1})$ // Dataset with N features for training
 S_0 // Begin search subset
 δ // Terminal condition
 S_{best} // Best and finest subset

Result:

- 1: **Begin**
- 2: Set: $S_{best} = S_0$;
- 3: $\gamma_{best} = \text{eval}(S_0; D; M)$;
- 4: **Do Begin**
- 5: $S = \text{generate}(D)$;
- 6: $\gamma = \text{eval}(S; D; M)$;
- 7: If (γ greater than best γ_{best})
- 8: $\gamma_{best} = \gamma$;
- 9: $S_{best} = S$;
- 10: **End Do** (δ is gotten);
- 11: **Return** S_{best} ;
- 12: **End.**

Fig. 1. The general filter algorithm

2.2. Cuckoo Inspired Algorithms

The two cuckoo inspired algorithms used in this study are briefly describe in the following subsections.

2.2.1. Cuckoo Search Algorithm

The cuckoo search algorithm (CSA) is a swarm-based meta-heuristic optimization algorithm developed by Yang and Deb [10]. The algorithm was inspired based on the lifespan of some birds' chiefly cuckoos. To simplify the description of the cuckoo search algorithm CSA, Yang and Deb [10] listed the following ideal rules:

- 1) Cuckoo places single egg at a while independently, and dump it egg in an arbitrarily preferred nest;
- 2) The finest nest with the highest value of eggs will move to the subsequent generation;
- 3) The quantity of obtainable host nests is static, and the egg placed by a cuckoo and revealed by a probability $P_a[0,1]$. Therefore, the owner either push the cuckoo's egg or build a different nest entirely.

The last rule is approximated by dumping a portion P_a of the eggs and swapping to each generation of n nest. Fundamentally, these three rules offer a variety process for the optimization algorithm. Guaranteeing that the finest eggs endure from generation to generation. In a problem of maximization, the quality or fitness of a solution is relational to the objective function. The CSA can be recapped as shown in Fig. 2.

When developing brand-new solutions x_i^{t+1} for cuckoo i , a levy flight is observed as shown in (1).

$$x_i^{t+1} = x_i^t + \alpha \oplus \text{levy}(\lambda) \quad (1)$$

Where $\alpha > 0$, equal equal to the step or stride size that must be analogous to the scale of the problems of concern. Commonly used as $\alpha = \mathcal{O}(1)$. The symbol product \oplus symbolize entry wise multiplications. Levy flight basically provides a random walk whereas their random steps are obtained after levy distribution of large steps as shown in (2).

$$\text{levy} \sim \mathcal{U} = t^{-\lambda}, (1 < \lambda \leq 3) \quad (2)$$

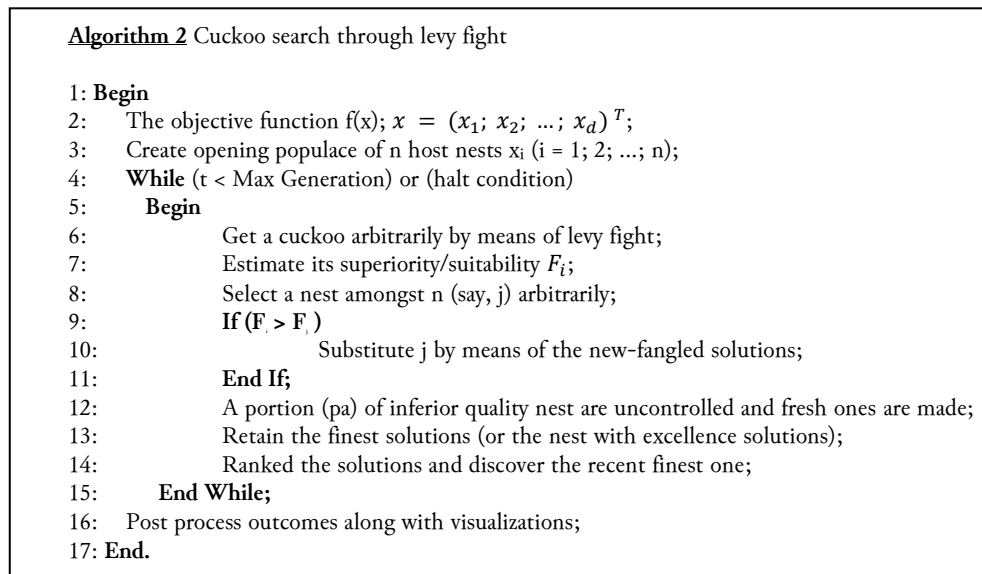


Fig. 2. The cuckoo search algorithm

This has both boundless variance and boundless means. The repeated jumps or steps of a cuckoo principally formulate a random walk progression which follows a power law footstep length movement through the heavy tail. In summary, three parameters are used by the initial CSA, 1) the population size n , 2) the parameter α termed as the stride size scaling point, and 3) the switching or swapping parameter, which is the fraction of the eggs discarded. The bigger the P_a value the more important exploration process is and the less likely chances of being getting trapped top local optima and vice versa. Marichelvam, et al. [15] believes parameters such as P_a and α are the serious values for locating best solution. Similarly, Yang and Deb [10] convergence of the algorithm to the best solution is largely independent of the value of P_a , however, $P_a = 0.25$ returned finest results.

Since the algorithm, Mantegna is widely used to obtain random numbers in levy flights [10], the study applied it to compute the step length S as indicated by (3).

$$S = \frac{\mu}{|v|^{\frac{1}{\beta}}} \quad (3)$$

where, β is the parameter amidst [1,2]. Similarly, μ and v are define are define inform of the standard distribution in (4) through (5) as shown:

$$\mu \sim N(0, \sigma v^2) \text{ and } v \sim N(0, \sigma v^2) \quad (4)$$

$$\left(\frac{\Gamma(1+\beta) \sin(\pi/2)}{\Gamma[(1+\beta)/2] \beta 2^{(\beta-1)/2}} \right)^{\frac{1}{\beta}}, \sigma v = 1 \quad (5)$$

2.2.2. Cuckoo Optimization Algorithm

Rajabioun [11] presented another new evolutionary optimization algorithm named cuckoo optimization algorithm (COA). The foundation behind this innovative optimization algorithm is how cuckoo place egg and their upbringing behavior.

According to the author, the cuckoos can occur in two ways namely, matured or aged cuckoos and eggs. The aged cuckoos placed their eggs in some other birds' nest. If the laid eggs are not recognized

and at the same time are not murder by the host bird. Then, they grow up, turn into matured cuckoos and start making societies. The idea behind this algorithm is that every society has its environment or habitat region where to live. As such matured cuckoos are looking for a better environmental habitat that has higher chances of eggs survival as their terminus for laying eggs. Then, each of the cuckoos starts laying their eggs randomly inside the egg laying radius (ELR) of the nests. The place in which more eggs lived implies that COA is optimized. The algorithm terminates if all the lived matured cuckoos unite to the best environmental position for their upbringing and replica. Now, the better location is the global extreme of the objective functions. A typical cuckoo optimization algorithm is shown in Fig. 3.

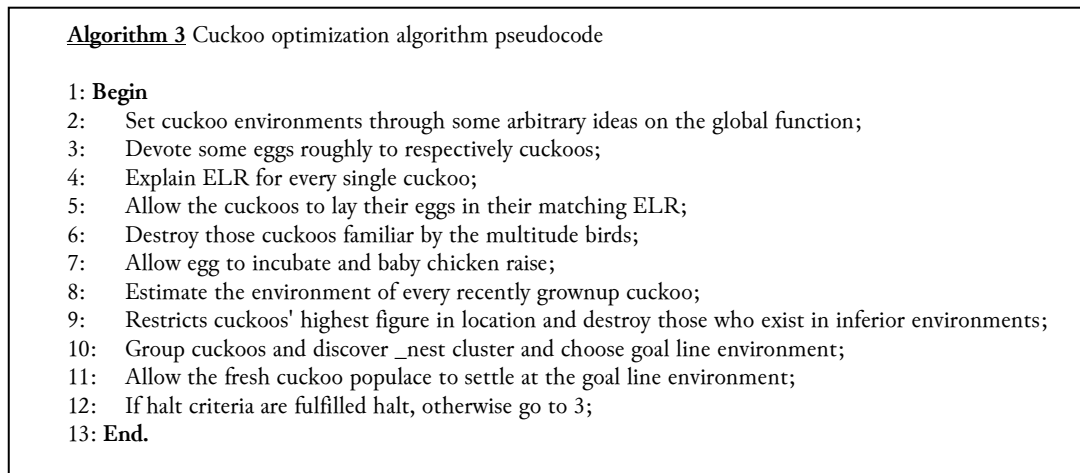


Fig. 3. The cuckoo optimization algorithm

In addition, its basic ideal rules to apply are as follows: The optimization problems variables should be in a form of an array called "habitat". This means that in N_{var} dimensional optimization problem, a territory range of $1 \times N_{var}$, that signify the present active location of the cuckoo is defined as shown in (6).

$$habitat = [x_1, x_2, x_3 \dots, x_{N_{var}}] \quad (6)$$

The profit of the environmental habitation is gotten by the assessment of "profits function" fp in (7) at habitation of $x_1, x_2, x_3 \dots, x_{N_{var}}$ such that:

$$profit = fp(habitation) = fp(x_1, x_2, x_3 \dots, x_{N_{var}}) \quad (7)$$

To apply COA in minimizations problems, simply maximize the fp as shown in (8).

$$profits = -Cost(habitation) = fp(x_1, x_2, x_3 \dots, x_{N_{var}}) \quad (8)$$

To begin the optimization problem, normally a habitat matrix of size $N_{pop} \times N_{var}$ is generated. Since naturally, each cuckoo produced five to twenty eggs. Then, these values will be used for upper and lower limits for each iteration. The eggs are laid inside a range of space from their environment in the ELR. However, the ELR depends on three parameters namely, variable limit, the sum of existing cuckoo's eggs and the entire number of eggs

$$ELR = \alpha x \frac{\text{number of current cuckoos}}{\text{total number of eggs}} x e_{new} \quad (9)$$

In the equation (9), α refers to a whole number that will handle the highest value of ELR and var_{hi} , var_{low} stand for the limits for upper and lower variables respectively. Afterward the cuckoo laid the eggs, usually, $p\%$ of the entire eggs i.e. 10% within fewer profit value and additional cost shall be

eliminated. The categorization of cuckoo's societies is finalized using the KMC technique. Basically, a value of $k = 3$ to 5 proved to be sufficient in the simulation.

Every cuckoo flies a $\lambda\%$ only into the habitat. In addition, it also has an eccentricity of radians. These parameters aid the cuckoo to hunt for more and better strategic spots in the location. They are defined as follows:

$$\lambda \sim U(0, 1), \varphi \sim U(-\omega, \omega) \quad (10)$$

where, $\lambda \sim U(0, 1)$, means that λ is a regularly distributed random number between the array of zero and one. ω stands for the parameter that restricts an alteration on or after the goal line region. An ω of $\pi/6$ rad looks to be sufficient for upright meeting or convergence of the cuckoo populace to universal extreme profit. It must contain all the information about the experimental procedure and materials used to carry out experiments.

2.3. The Proposed Method

2.3.1. CSA Filter-Based Feature Selection

The proposed filter-based FS embed both the general filter-based FS (Algorithm 1, Fig. 1) and cuckoo search algorithm (Algorithm 2, Fig. 2). The detail of the pseudocode is shown in Fig. 4.

Algorithm 4 Filter-based CSA

```

1: Begin
2:   The objective function  $f(x)$ ;  $x = (x_1; x_2; \dots; x_d)^T$ ;
3:   Create opening populace of n host nests  $x_i$  ( $i = 1; 2; \dots; n$ );
4:   Set:  $S_{best} = S_0$ ;
5:    $\gamma_{best} = \text{eval}(S_0; D; M)$ ;
6:   While ( $t < \text{Max Generation}$ ) or (Halt condition)
7:     Begin
8:       Get a cuckoo arbitrarily by means of levy flight;
9:        $S = \text{generate}(D)$ ;
10:       $\gamma_{best} = \text{eval}(S_0; D; M)$ ;
11:      if ( $\gamma$  greater than best)
12:         $\gamma_{best} = \gamma$ ;
13:         $S_{best} = S$ ;
14:      Estimate its superiority/suitability  $F_i$ ;
15:      Select a nest amongst n (say, j) arbitrarily;
16:      If ( $F_i > F_j$ )
17:        Substitute j by means of the new-fangled solutions;
18:      End If
19:      A portion (pa) of inferior quality nest are uncontrolled and fresh ones are made;
20:      Retain the finest solutions (or the nest with excellence solutions);
21:      Ranked the solutions and discover the recent finest one;
22:    End While;
23:    Return  $S_{best}$ ;
24:    Post process outcomes along with visualizations;
25: End.

```

Fig. 4. CSA filter-based feature selection algorithm

2.3.2. COA Filter-Based Feature Selection

The proposed filter-based FS embed both the cuckoo optimization algorithm (Algorithm 3, Fig. 3) and general filter-based FS (Algorithm 1, Fig. 1). The detail of the pseudocode is shown in Fig. 5.

Algorithm 5 Filter-based COA

```

1: Begin
2:   Set cuckoo environments through some arbitrary ideas on the global function;
3:   Set:  $S_{best} = S_0$ ;
4:    $\gamma_{best} = \text{eval}(S_0; D; M)$ ;
5:   Devote some eggs roughly to respectively cuckoos;
6:   Explain ELR for every single cuckoo;
7:   Allow the cuckoos to lay their eggs in their matching ELR;
8:    $S = \text{generate}(D)$ ;
9:    $\gamma = \text{eval}(S_0; D; M)$ ;
10:  if ( $\gamma$  greater than best)
11:    Destroy those cuckoos familiar by the multitude birds;
12:     $\gamma_{best} = \gamma$ ;
13:     $S_{best} = S$ ;
14:  Allow egg to incubate and baby chicken raise;
15:  Estimate the environment of every recently grownup cuckoo;
16:  Restricts cuckoos' highest figure in location and destroy those who exist in inferior environments;
17:  Group cuckoos and discover finest cluster and choose goal line environment;
18:  Allow the fresh cuckoo populace to settle at the goal line environment;
19:  If halt criteria are fulfilled halt, otherwise go to 3;
20:  Return  $S_{best}$ ;
21: End.

```

Fig. 5. CSA filter-based feature selection algorithm**2.3.3. Dataset Description**

The heart disease datasets are inputted. Five heart disease datasets are used in this study. The details of the datasets such as the number of features and instances are shown in Table 1. The data is obtained at the University of California Irvine, UCI machine learning at <https://archive.ics.uci.edu/ml>. Except for Eric dataset which can be obtained at http://eric.univ-lyon2.fr/~ricco/tanagra/fichiers/heart_disease_male.xls. The datasets are sanitized all incomplete instances are deleted. Lastly, few instances that do not contribute significantly are imputed using both the backward and forward fill of the Python Anaconda Navigator.

Table 1. Heart disease datasets description

S/N	Datasets	Instances	Features	Classes
1	Eric	209	7	2
2	Echocardiogram	131	12	2
3	Hungarian	294	13	2
4	Stat log	270	13	2
5	Z-Alizadeh Sani	303	55	2

3. Results and Discussion**3.1. Performance Measure**

The performance of the proposed filter-based algorithms is measured based on the number of features along with the classification accuracy. The detail of them are explained in the following sections:

1) Number of features

The essence of the two proposed filter-based algorithms is to produce fewer features that contribute significantly to heart disease and yet improve prediction accuracy. In this case, each of the algorithms selects the most informative features from each dataset. In addition, the selected features are compared to the ones found in the literature.

2) Classification Accuracy

Accuracy has become very important in classifiers for medical data predictions [5]. As such, in this work, the classification accuracy was employed to find the accuracy, i.e., the number of correct predictions from all predictions made. Four different classifiers such as SVM, MLP, NB and, RFC are employing to measure the capability of the proposed feature selection algorithms. The results obtained are compared among the different classifiers for both CSA and COA. Finally, the results found are also compared to other approaches reported in the literature.

3.2. Analysis and Presentation of Obtained Results

The outcomes gained by the proposed filter-based cuckoo inspired algorithms are evaluated using the least number of features picked by each of the algorithms. Table 2 clearly displays the number of features or attributes chosen by both cuckoo algorithms. CSA filter-based outperform the COA filter-based on the Z-Alizadeh Sani, Stat log and Hungarian data sets. Conversely, the number of features is same in Eric and Echocardiogram datasets.

Table 2. Number of features selected by the proposed filter-based feature selection algorithms

S/N	Datasets	Total Features	COA	CSA
1	Eric	7	4	4
2	Echocardiogram	12	5	4
3	Hungarian	13	6	4
4	Stat log	13	6	4
5	Z-Alizadeh Sani	55	14	7

The four features selected by both CSA and COA for Eric datasets are: (1. chest pain, (2. resting electrocardiographic results, (3. maximum heart rate achieved and (4. exercise angina. While for Echocardiogram data the five selected features are: (1. age at heart attack, (2. fractional -shortening, (3. left ventricular end-diastolic dimension, (4. e point sepal separation EPSS and (5. wall motion score. Similarly, for Hungarian and stat log datasets the features are same with Eric with the addition of fasting blood sugar for CSA. While old peak plus the selected features by CSA represent the six features selected by the COA for both Hungarian and Stat log datasets respectively. On Z-Alizadeh Sani dataset, current smoker feature is added that makes it a total of seven features selected by CSA. Whereas BMI (body mass index), edema, weak peripheral pulse, lung rales, lymph, and VHD makes the fourteen features selected by the COA on the same dataset.

1) Eric

The accuracy attained by the Eric dataset is displayed in Fig. 6. From the figure, one can see clearly that the accuracy obtained by the proposed CSA filter-based FS is higher than the COA filter-based FS both before and after FS, despite they have the same number features chosen by the duo algorithms. On the other hand, SVM has highest classification accuracy of 89.90% after FS.

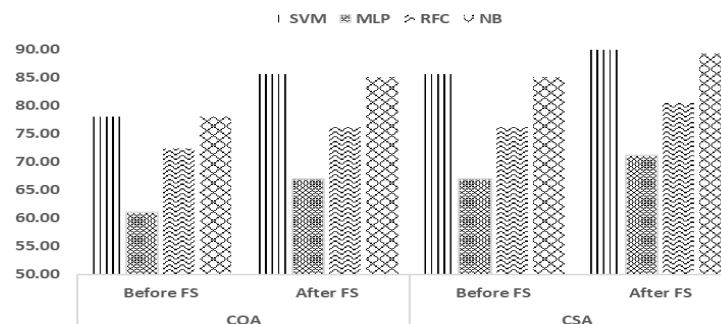


Fig. 6. Accuracy of Eric dataset using different classifiers

2) Echocardiogram

Most of the classifiers attained a better accuracy in favor of CSA both before and after FS on the dataset. Except, in the case of NB that recorded 100% accuracy after FS for both COA and CSA as displayed in Fig. 7. This clearly showed that NB is a good choice for the echocardiogram dataset, especially after FS is conducted.

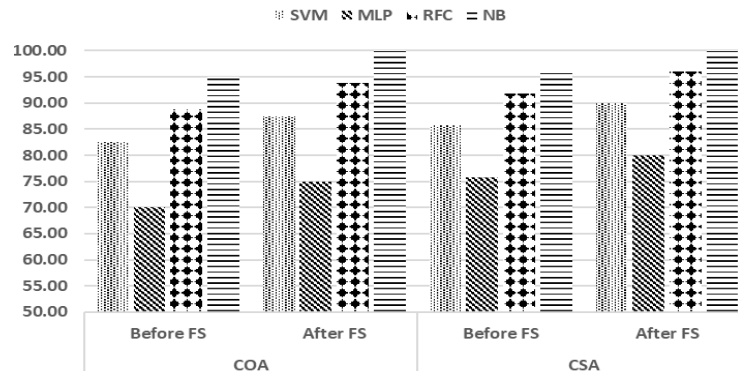


Fig. 7. Accuracy of Echocardiogram dataset using different classifiers

3) Hungarian

Fig. 8 shows clearly that SVM outperformed the rest of the classifiers in terms of accuracy both before and after FS for the respectively proposed algorithms. Therefore, SVM is a good choice for the Hungarian data. Similarly, the performance of the CSA is better than COA both before and after FS.

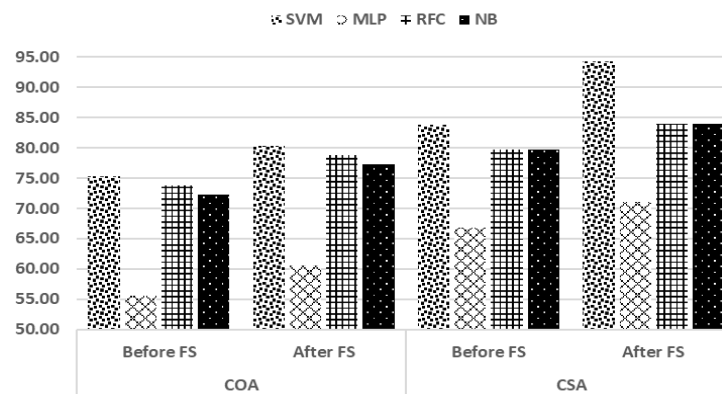


Fig. 8. Accuracy of Hungarian dataset using different classifiers

The results obtained in this study is also compared to the work of Gadekallu and Khare [12] where CSA was combined with RST for FS. The comparison showed that the proposed CSA based FS surpass that one both in terms of fewer number of features and prediction accuracy as depicted in Table 3.

Table 3. Comparison analysis of Hungarian dataset

Reference	Approach	Selected Features	Accuracy
Gadekallu and Khare [12]	CSA + RST	6	91.50%
This Study	Filter-based CSA	12	94.22%

4) Stat log

Also, like the previous datasets, the stat log data has higher accuracy after FS compared to before. In addition, the CSA proved to be more effective with higher accuracy values compared to the COA. Similarly, the SVM recorded high level of accuracy in all the options for both COA and CSA respectively. The detailed diagram of the result showing the accuracy of each classifier can be found in Fig. 9.

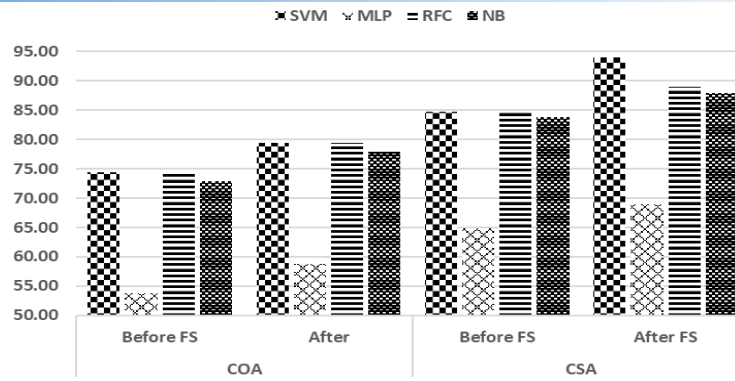


Fig. 9. Accuracy of stat log dataset using different classifiers

Moreover, the results obtained are also compared to Liu et al. [16], where relief f and rough set (RFRS) are used for FS. Similarly, a Bounded Sum of Weighted Fuzzy Membership functions (BSWFM) together with Euclidean distance (ED) was used as FS on the Stat log dataset in the work of Lee [17]. In Tomar and Agarwal [18] least square twin SVM (LSTSVM) based FS was proposed on the same dataset. Buscema et al. [19] used training with input selection and testing (TWIST) algorithm for FS. Finally, Subbulakshmi et al. [20] used extreme learning machine (ELM) to select the most relevant fewer features from the Stat log dataset. But the choice of classifier affects the performance of the prediction accuracy. The comparison is a clear testimony that the proposed CSA based FS performed better in terms of selected features along with classification accuracy. Table 4 summarizes the detailed comparison.

Table 4. Comparison analysis of Hungarian dataset

Reference	Approach	Selected Features	Accuracy
Liu et al. [16]	RFRS	7	92.59%
Lee [17]	BSWFM + ED	10	87.40%
Tomar and Agarwal [18]	LSTSVM	11	85.90%
Buscema et al. [19]	TWIST Algorithm	5	84.14%
Subbulakshmi et al. [20]	ELM	-	87.50%
This Study	Filter-based CSA	4	94.00%

5) Z-Alizadeh Sani

Finally, this dataset also showed that SVM still outperform the rest of the classifiers both before and after FS on the proposed algorithms. Fig. 10 clearly summarized the accuracy attained by each classifier for each of the proposed algorithms.

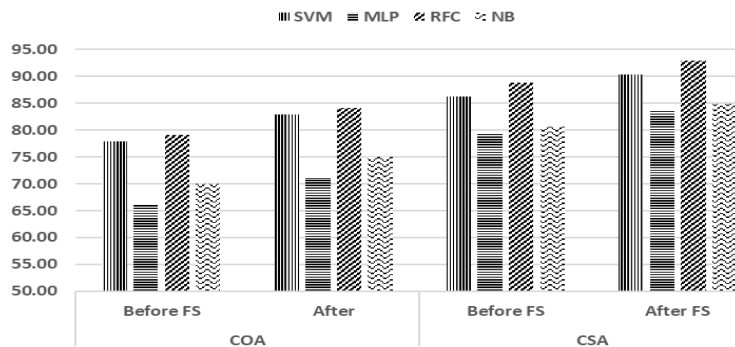


Fig. 10. Accuracy of z-Alizadeh Sani dataset using different classifiers

6) Accuracy of the complete datasets:

The level of accuracy attained by each of the datasets is shown in Fig. 11, shows categorically that CSA after FS performed better than COA on almost all the datasets. In addition, it is observed that 100% accuracy was recorded only once on Echocardiogram dataset using NB classifier. Since then SVM dominated all the remaining classifiers both before and after FS.

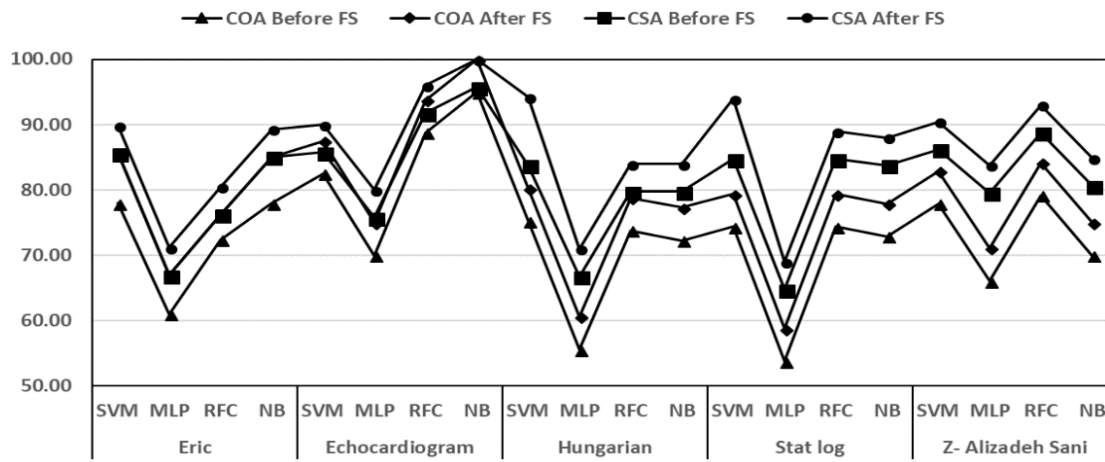


Fig. 11. Accuracy of each dataset using different classifiers

Therefore, SVM can be considered as the best classifier so far in this study. This is due to the majority of accurateness possesses while using it compared to the rest.

4. Conclusion

This paper addresses the inherent problem of feature selection in heart disease prediction. Two cuckoos-inspired algorithms are compared for filter-based FS. The simulation results indicate clearly that CSA filter-based FS outperformed the COA filter-based FS both in terms of reduced features and classification accuracy in most of the datasets. In addition, CSA filter-based feature selection outperform several other ones reported in the literature. However, there is no dependency among the features or attributes selected. As such, future research should consider a cuckoo wrapper-based feature selection as well as hybrid approaches to improve the classification accuracy and feature dependencies. Finally, other metaheuristic algorithms can be used as filter-based FS to compare with the proposed approaches in this study.

Acknowledgment

The authors wish to express gratitude to the Universiti Sains Malaysia (USM) for its support and funding the research through the Fundamental Research Grant Scheme (203/PKOMP/6711463).

References

- [1] E. J. Benjamin *et al.*, "Heart Disease and Stroke Statistics—2018 Update: A Report From the American Heart Association," *Circulation*, vol. 137, no. 12, pp. e67–e492, Mar. 2018, doi: <https://doi.org/10.1161/CIR.0000000000000558>.
- [2] L. Verma, S. Srivastava, and P. C. Negi, "A Hybrid Data Mining Model to Predict Coronary Artery Disease Cases Using Non-Invasive Clinical Data," *J. Med. Syst.*, vol. 40, no. 7, p. 178, Jul. 2016, doi: <https://doi.org/10.1007/s10916-016-0536-z>.
- [3] T. Vivekanandan and N. C. S. N. Iyengar, "Optimal feature selection using a modified differential evolution algorithm and its effectiveness for prediction of heart disease," *Comput. Biol. Med.*, vol. 90, pp. 125–136, Nov. 2017, doi: <https://doi.org/10.1016/j.combiomed.2017.09.011>.
- [4] S. Shilaskar and A. Ghatol, "Feature selection for medical diagnosis : Evaluation for cardiovascular diseases," *Expert Syst. Appl.*, vol. 40, no. 10, pp. 4146–4153, Aug. 2013, doi: <https://doi.org/10.1016/j.eswa.2013.01.032>.

- [5] H. H. Inbarani, M. Bagyamathi, and A. T. Azar, "A novel hybrid feature selection method based on rough set and improved harmony search," *Neural Comput. Appl.*, vol. 26, no. 8, pp. 1859–1880, Nov. 2015, doi: <https://doi.org/10.1007/s00521-015-1840-0>.
- [6] N. C. Long, P. Meesad, and H. Unger, "Attribute Reduction Based on Rough Sets and the Discrete Firefly Algorithm," 2014, pp. 13–22, doi: https://doi.org/10.1007/978-3-319-06538-0_2.
- [7] M. A. Jabbar, B. L. Deekshatulu, and P. Chandra, "Prediction of Heart Disease Using Random Forest and Feature Subset Selection," 2016, pp. 187–196, doi: https://doi.org/10.1007/978-3-319-28031-8_16.
- [8] S. M. S. Shah, S. Batool, I. Khan, M. U. Ashraf, S. H. Abbas, and S. A. Hussain, "Feature extraction through parallel Probabilistic Principal Component Analysis for heart disease diagnosis," *Phys. A Stat. Mech. its Appl.*, vol. 482, pp. 796–807, Sep. 2017, doi: <https://doi.org/10.1016/j.physa.2017.04.113>.
- [9] M. A. Jabbar, "Prediction of heart disease using k-nearest neighbor and particle swarm optimization," *Biomed. Res.*, vol. 28, no. 9, pp. 4154–4158, 2017, available at :<http://www.biomedres.info/biomedical-research/prediction-of-heart-disease-using-knearest-neighbor-and-particle-swarm-optimization.pdf>.
- [10] X.-S. Yang and S. Deb, "Cuckoo Search via Levy flights," in *2009 World Congress on Nature & Biologically Inspired Computing (NaBIC)*, 2009, pp. 210–214, doi: <https://doi.org/10.1109/NABIC.2009.5393690>.
- [11] R. Rajabioun, "Cuckoo Optimization Algorithm," *Appl. Soft Comput.*, vol. 11, no. 8, pp. 5508–5518, Dec. 2011, doi: <https://doi.org/10.1016/j.asoc.2011.05.008>.
- [12] T. R. Gadekallu and N. Khare, "Cuckoo Search Optimized Reduction and Fuzzy Logic Classifier for Heart Disease and Diabetes Prediction," *Int. J. Fuzzy Syst. Appl.*, vol. 6, no. 2, pp. 25–42, Apr. 2017, doi: <https://doi.org/10.4018/IJFSA.2017040102>.
- [13] H. Liu and L. Yu, "Toward integrating feature selection algorithms for classification and clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 4, pp. 491–502, Apr. 2005, doi: <https://doi.org/10.1109/TKDE.2005.66>.
- [14] B. Xue, M. Zhang, W. N. Browne, and X. Yao, "A Survey on Evolutionary Computation Approaches to Feature Selection," *IEEE Trans. Evol. Comput.*, vol. 20, no. 4, pp. 606–626, Aug. 2016, doi: <https://doi.org/10.1109/TEVC.2015.2504420>.
- [15] M. K. Marichelvam, T. Prabakaran, and X. S. Yang, "Improved cuckoo search algorithm for hybrid flow shop scheduling problems to minimize makespan," *Appl. Soft Comput.*, vol. 19, pp. 93–101, Jun. 2014, doi: <https://doi.org/10.1016/j.asoc.2014.02.005>.
- [16] X. Liu *et al.*, "A Hybrid Classification System for Heart Disease Diagnosis Based on the RFRS Method," *Comput. Math. Methods Med.*, vol. 2017, pp. 1–11, 2017, doi: <https://doi.org/10.1155/2017/8272091>.
- [17] S.-H. Lee, "Feature selection based on the center of gravity of BSWFMs using NEWFM," *Eng. Appl. Artif. Intell.*, vol. 45, pp. 482–487, Oct. 2015, doi: <https://doi.org/10.1016/j.engappai.2015.08.003>.
- [18] D. Tomar and S. Agarwal, "Feature Selection based Least Square Twin Support Vector Machine for Diagnosis of Heart Disease," *Int. J. Bio-Science Bio-Technology*, vol. 6, no. 2, pp. 69–82, Apr. 2014, doi: <https://doi.org/10.14257/ijbsbt.2014.6.2.07>.
- [19] M. Buscema, M. Breda, and W. Lodwick, "Training with Input Selection and Testing (TWIST) Algorithm: A Significant Advance in Pattern Recognition Performance of Machine Learning," *J. Intell. Learn. Syst. Appl.*, vol. 05, no. 01, pp. 29–38, 2013, doi: <https://doi.org/10.4236/jilsa.2013.51004>.
- [20] C. V. Subbulakshmi, S. N. Deepa, and N. Malathi, "Extreme Learning Machine for two category data classification," in *2012 IEEE International Conference on Advanced Communication Control and Computing Technologies (ICACCCT)*, 2012, pp. 458–461, doi: <https://doi.org/10.1109/ICACCCT.2012.6320822>.