NAKHODA

Journal of Algorithms and Computation

journal homepage: http://jac.ut.ac.ir

Intelligent application for Heart disease detection using Hybrid Optimization algorithm

Marzieh Eskandari*[1] and Zeinab Hassani†[2]

[1]Department of computer science, Alzahra University, Tehran, Iran.
[2]Department of computer science, Kosar University of Bojnord, Iran.

ABSTRACT

Prediction of heart disease is very important because it is one of the causes of death around the world. Moreover, heart disease prediction in the early stage plays a main role in the treatment and recovery disease and reduces costs of diagnosis disease and side effects it. Machine learning algorithms are able to identify an effective pattern for diagnosis and treatment of the disease and identify effective factors in the disease. this paper is investigated a new hybrid algorithm of Whale Optimization and Dragonfly algorithm using a machine learning algorithm. the hybrid algorithm employs a Support Vector Machine algorithm for effective Prediction of heart disease. Proposed method is evaluated by Cleveland standard heart disease dataset. The experimental result indicates that the SVM accuracy of 88.89 % and nine features are selected in this respect.

ARTICLE INFO

Article history:
Received 12, March 2018
Received in revised form 8, April 2019
Accepted 19 May 2019
Available online 01, June 2019

Keyword: Hybrid Optimization Algorithm, Support Vector Machine, Whale Optimization Algorithm, Dragonfly Algorithm, Feature Selection.

AMS subject Classification: 05C79.

*Corresponding author: M. Eskandari Email: eskandari@alzahra.ac.ir
†Hassani@kub.ac.ir

1 Introduction

Recent researches show that machine learning and data mining are employed for early Diagnosis of disease. Heart disease is the most common cause of death in the world. Besides, coronary artery disease (CAD) is the most common cardiovascular condition [20]. There are several methods to diagnose CAD, including stress test, Electrocardiography (ECG), echocardiography, coronary angiography or cardiac catheterization. All of the diagnostic methods have demerits such as feeling pain and apart from that it necessitates special processes which involve a lot of time, effort, cost, and should be done by trained people [6]. A computer-aided system can extract and detect effective attributes which in turn utilized in classifier to diagnose the disease. Algorithms of machine learning are employed to diagnose heart diseases that include decision tree [15], SVM [6], and neural network [5]. [6] introduced a methodology for the automatic diagnosis of normal and Coronary Artery Disease This method utilize Heart Rate Variability (HRV) signal extracted from electrocardiogram (ECG). They applied Principal Component Analysis (PCA) for reducing the dimension of the features and Support Vector Machine (SVM) for the classifier. Ritika Chadha et al. studied Artificial Neural Networks (ANN), Naive Bayes and Decision Tree on Cleveland heart disease dataset. Their observations showed that Artificial Neural Networks outperformed Naive Bayes and Decision Tree [4]. El-Bialy et al. considered different decision trees such as: C4.5 and Fast Decision Tree and evaluated their method by datasets of heart disease. The accuracy obtained from Statlog dataset was 78.06% [7]. [11] Presented Support Vector Machines (SVM), Multi-Layer Perceptrons Ensembled (MLPE) and generalized additive model (GAM) that investigated diagnosis and identification important features of heart disease. SVM and MLPE were 84.77% and accuracy on Statlog dataset was 81.82%. The rest of the paper is organized as follows: section 2 describes the concepts of Support Vector Machine, Whale Optimization Algorithm, Dragonfly algorithm, and feature selection. Section 3 explains Hybrid Whale Optimization Algorithm and Dragonfly algorithm. Experiment Setup and Result Analysis are discussed in section 4. Finally, in Section 5, conclusions are given.

2 Methods

2.1 Support Vector Machine (SVM)

Support Vector Machine (SVM) is popular supervised machine learning algorithms [8]. Boser, Guyon, and Vapnik proposed SVM in 1992 that is a state-of-the-art classification method [19]. The idea of structural risk minimization is the cornerstone of the theory of SVM [1]. The Support Vector Machine (SVM) is utilized to make a separating hyper plane as the decision surface which divides two class instances with a maximal margin [14]. The training data set is m objects which each observation xi has n features and a corresponding label , the objective of the SVM problem is to identify a hyper plane that separates the two classes of points with a maximal separation margin as measured by the l2–norm that classifies the objects xi correctly. The optimal separating hyper plane is computed as a decision surface that it can be modeled as follows:

$$z(x) = sgn(\sum_{i=1}^{N} a_i y_i(x_i, x) + b) \tag{1}$$

Where x_i are support vectors, N is the number of support vectors, b is bias, and $y_i(x_i, x)$ is the kernel function which used to map the data from the original dimension to higher dimensions in case the data is linearly separable in the mapped dimension. In this study, a Radial Basis Function (RBF) is selected as the kernel function. RBF can be expressed as:

$$z(x) = exp(-||x - y||^2/2\sigma^2) \tag{2}$$

Where $\sigma > 0$ is a user specified constant which specifies the kernel width. The number and value support vectors in RBF kernel SVM determine the number of kernels and their centers [14]. SVMs belong to the general category of kernel methods. A kernel method is an algorithm that depends on the data only through dot-products. When this is the case, the dot product can be replaced by a kernel function which computes a dot product in some possibly high dimensional feature space. The four basic kernels are linear kernel, polynomial kernel, RBF kernel and sigmoid kernel. The kernel function determines the feature space in which the training set instances will be classified. Therefore the selection of an appropriate kernel function is important [1].

2.2 Whale Optimization Algorithm

Whale Optimization Algorithm (WOA) is introduced by Mirjalili and Lewis, in 2016 [12]. WOA is nature-inspired meta-heuristic optimization algorithm which mimic the social behavior of humpback whales. The algorithm is inspired by the bubble-net hunting strategy. The humpback whales usually prefer to hunt krills or small fishes which are close to the surface. Moreover, the most interesting thing about the humpback whales is their special hunting manner that is called bubble net feeding method. In this method, they swim around the prey and construct a distinctive bubbles along a circle or 9-shaped path. The mathematical model of WOA is defined in Encircling prey, Bubble net hunting method and Search the prey. From theoretical point of view, WOA can be considered a global optimizer because it includes exploration/exploitation ability.

2.2.1 Encircling prey

Humpback whales can identify the location of prey and encircle them. After the best search agent is determined, the other search agents will hence try to update their positions towards the best search agent. Furthermore, X* should be updated in each iteration if there is a better solution.

$$\vec{X}(t+1) = \vec{X}^*(t) - \vec{A}.\vec{D} \tag{3}$$

$$\vec{D} = |\vec{C}.\vec{X}^*(t) - \vec{X}(t)| \tag{4}$$

$$\vec{A} = 2.\vec{a}.\vec{r} - \vec{a} \tag{5}$$

$$\vec{C} = 2.\vec{r} \tag{6}$$

Where $\vec{X}^*(t)$, $\vec{X}(t)$ show the position of best solution and position vector respectively. t is Current iteration.$\vec{A}$,$\vec{C}$ denote coefficient vectors. $\vec{a}$ shows directly diminished from 2 to 0. $\vec{r}$ Refers to a random vector [0, 1].

2.2.2 Bubble-net attacking method

Humpback whales swim around the prey within a shrinking circle and along a spiral-shaped path simultaneously. This manner includes shrinking encircling and spiral updating position approach. Shrinking encircling prey is defined by Eq.9. Spiral updating position first evaluates the distance between the position of the whale and the prey. A spiral equation is then construct between the position of whale and prey to simulate the helix-shaped movement of humpback whales as follows:

$$X(t+1) = D'.ebl.cos(2.\pi.l) + X^*(t) \tag{7}$$

To model this simultaneous behavior, to update the position of whales during optimization that there is a probability of 50% to choose between either the shrinking encircling mechanism or the spiral model. Where p is a random number in [0, 1].

$$X(t+1) = \begin{cases} \vec{X}(t+1) = \vec{X}^*(t) - \vec{A}.\vec{D} & \text{if } p < 0.5 \\ X(t+1) = D'.ebl.cos(2.\pi.l) + X^*(t) & \text{if } p >= 0.5 \end{cases} \tag{8}$$

2.2.3 Search for prey

In this pproach according the variation of the vector is employed to search for prey, this is called exploration phase. random search agent is chosen when $|A| > 1$, while the best solution is chosen when $|A| < 1$ for updating the position of the search agents. The mathematical model is as follows:

$$\vec{X}(t+1) = \vec{X}_{rand}(t) - \vec{A}.\vec{D} \tag{9}$$

$$\vec{D} = |\vec{C}.\vec{X}_{rand}(t) - \vec{X}(t)| \tag{10}$$

Where $\vec{X}_{rand}$ shows a random position vector (a random whale) chosen from the current population. WOA is shown Algorithm 1 [12], [10].

Algorithm 1 Whale Optimization Algorithm

Input: Initialize the whales population Xi (i = 1, 2, ..., n), Calculate the fitness of each search agent, X*=the best search agent

Output: The best solution and value of objective function of best solution.

```
while (t < maximumnumberofiterations) do
    for each search agent do
        Update a, A, C, l, and p
        if p < 0.5 then
            if |A| < 1 then
                Update the position of the current search agent by the Eq.4
            elseIf |A| >= 1
                Select a random search agent ( )
                Update the position of the current search agent by the Eq.9
            end if
        elseIf p >= 0.5
            Update the position of the current search by the Eq.7
        end if
    end for
    Check if any search agent goes beyond the search space and amend it
    Calculate the fitness of each search agent
    Update X* if there is a better solution
    t = t + 1;
end while
return X*
```

2.3 Dragonfly algorithm

Dragonfly algorithm (DA) is introduced by Mirjalili in 2015 that is inspired by the unique and superior swarming behaviour of dragonflies [13]. Two swarming behaviours of the DA, static and dynamic swarming behaviours, are very similar to the two main phases of optimization that are exploration and exploitation. In the exploration phase, Dragonflies construct sub-swarms and move over different areas in a static swarm. In the static swarm, however, dragonflies move in bigger swarms and along one direction which is desirable in the exploitation phase[13] the behaviour of swarms follows five primitive principles of separation, alignment, cohesion, distraction from the enemies and attraction towards the food. Each dragonfly in the swarm is considered to be a solution in the search space. Therefore, DA is defined by five different operators such as Separation, Alignment, Cohesion, and Attraction towards food sources and distraction towards enemy sources [13, 16]. Separation (Si) that refers to the static collision avoidance of individuals from the other adjacent individuals. This behavior is modeled as follows:

$$S_i = \sum_{i=1}^{N} X - X_j \tag{11}$$

Where X denotes the position of the current individual, X_j shows is the position jth neighbouring individual, and N is the number of neighbouring individuals. Alignment (A_i) refers to the velocity matching of individuals to other individuals in neighbourhood. Alignment is calculated as follows:

$$A_i = \sum_{i=1}^{N} v_j/N \tag{12}$$

Where v_j corresponds to the velocity of jth neighbouring individual. Cohesion (C_i) refers to the tendency of individuals towards the center of the mass of the neighbourhood. Cohesion is calculated as follows:

$$C_i = \sum_{i=1}^{N} x_j/N - X \tag{13}$$

Attraction towards food source, and distraction from enemies is calculated using Eqs. (14) and (15)

$$F_i = X^+ - X \tag{14}$$

$$E_i = X^- - X \tag{15}$$

Where X shows the position of the current individual, X^+ is the position of the food source, and X^- is the position of the enemy. DA uses two vectors step $(\Delta X\ )$ and position (X) which updates the position of artificial dragonflies in a search space and simulates their movements. The step vector is comparable to the velocity vector in PSO.

The step vector shows the direction of the movement of the dragonflies and can be defined based on five principles as follows:

$$\Delta X_{t+1} = (s.S_i + a.A_i + c.C_i + f.F_i + +e.E_i) + w.\Delta X_t \tag{16}$$

After calculating the step vector, the position vectors are updated as follows:

$$X_{t+1} = X_t + \Delta X_{t+1} \tag{17}$$

Where t is the current iteration, s, a, c,f, and e correspond separation, alignment, cohesion, food, and enemy factors respectively. If there is no dragonfly in the neighbourhood radius the position of the dragonfly will be updated using Levy Flight equation, this improves the randomness, chaotic behaviour and global search capability of dragonflies. Levy Flight as follows:

$$X_{t+1} = X_t + Levy(d).X_{t+1} \tag{18}$$

Where t shows the current iteration, and d denotes the dimension of the position vectors. The fitness function is then calculated based on the updated position and velocitie that is continued till the stop condition is met. DA is shown Algorithm 2 [13, 16].

Algorithm 2 Dragonfly algorithm

Input: Initialize the dragonflies population X_i $(i \leq i \leq n)$, Initialize step vectors ΔX_i $(1 \leq i \leq n)$

Output: The best solution and value of objective function of best solution.

while $(t < maximumnumberofiterations)$ **do**
    Calculate the objective values of all dragonflies
    Update the food source and enemy
    Update w, s, a, c, f, and e
    Calculate S, A, C, F, and E using Eq.11 to Eq.15
    Update neighbouring radius
    if $adragonflyhasatleastoneneighbouringdragonfly$ **then**
        Update velocity vector using Eq.16
        Update position vector using Eq.17
    else
        Update position vector using Eq.18
    end if
    Check and correct the new positions based on the
    boundaries of variables
    $t = t + 1$;
end while
return X*

2.4 Feature Selection

Feature selection evidently select out of the dataset in a way that it describes the dataset efficiently. In this regard, total information content have hardly unique features and contain maximum information about the classes [3]. Feature selection approaches classified into three methods: filter methods, wrapper methods, and embedded methods. Filter is one of the oldest methods of feature selection which deploys filter approach. This method use variable ranking techniques and is done before the execution of any classification algorithm. Wrapper methods perform features selection according to learning algorithm Thus it seems the most "useful" and optimal features in comparison with filter method. The Wrapper method is classified into Sequential Selection Algorithms and Heuristic Search Algorithms. Embedded methods consist of variable selection as part of the training process that measures the "usefulness" of feature subsets [3, 9]. To select of feature, most researchers use Heuristic Search Algorithms of wrapper methods. This study employs hybrid WOA and DA algorithms for feature selection that has described in the next section.

3 Hybrid WOA and DA

Binary optimization algorithm can apply for feature selection where solutions are limited to the binary values. A solution is displayed in one dimensional of binary vector, where the length of it is according to number of features of the original dataset. In the vector, each value is displayed by 1 and 0 while the selected feature value is set to 1 otherwise is set to 0. The convergence ability to the most excellent solution of the problem near optimal solution is recognized as exploitation of algorithm and capability of an algorithm to locate whole parts of a problem search space is known as exploration. Finally, exploitation and exploration can be balanced by metaheuristics in the search space. Several hybrid nature-inspired methods are developed for improving the exploration and exploitation performance of existing algorithms. Hybrid optimization can represent a more robust behavior and display greater flexibility against complex and difficult problems [2]. Two models can be hybridized in low-level and high-level. In low-level a given function of a heuristic algorithm is exchanged by another heuristic algorithm and high-level model employs the second algorithm after applying the first algorithm and finding the best solution [17]. This method introduces a hybridization between global search (WOA) and local search algorithm (DA) that improves the exploitation ability of WOA algorithm. DA algorithm in this approach acts as an operator in WOA algorithm because the exploration in the WOA algorithm depends on changing the position of each search agent based on a randomly selected solution and the proposed method improves the exploration ability of WOA algorithm. In the following, Pseudo-code of hybrid WOA and DA algorithm is described.

Algorithm 3 Hybrid WOA and DA

Input: Initialize the whales population X_i $(i \leq i \leq n)$, Calculate the fitness of each search agent, X*=the best search agent

Output: The best solution and value of objective function of best solution.

while $(t < maximumnumberofiterations)$ **do**
    for each search agent **do**
        Update a, A, C, l, and p
        if $p < 0.5$ **then**
            if $|A| < 1$ **then**
                Update the position of the current search agent by the Eq.4
            elseIf $|A| >= 1$
                Select a random search agent ( )
                Update the position of the current search agent by **DA algorithm**
            end if
        elseIf $p >= 0.5$
            Update the position of the current search by the Eq.7
        end if
    end for
    Check if any search agent goes beyond the search space and amend it
    Calculate the fitness of each search agent
    Update X* if there is a better solution
    $t = t + 1$;
end while
return X^* and accuracy

Table 1: Attributes of Statlog heart disease dataset

NO	Attribute	Explanation
1	Age	age in years
2	Sex	Sex
3	cp	chest pain type
4	trestbps	resting blood pressure
5	chol	serum cholestoral
6	fbs	fasting blood sugar
7	restecg	resting electrocardiographic
8	thalach	maximum heart rate achieved
9	exang	exercise induced angina
10	oldpeak	ST depression induced by exercise relative to rest
11	slope	the slope of the peak exercise ST segment
12	ca	number of major vessels colored by flourosopy
13	thal	Normal, fixed defect, reversable defectl
14	class	Diagnosis classes

4 Experiment Setup and Result Analysis

In this study, the Statlog standard heart disease dataset is collected from the UCI machine learning repository [18]. The dataset includes 270 records of 14 attributes. 14 attributes of the dataset is shown in Table 1.
Several measuring tools are used to evaluate the performance of the proposed models. Among these tools accuracy, precision and recall .In the flowing, we define evaluation metrics:

$$Accuarcy = \frac{TP + TN}{TP + TN + FP + FN} \tag{19}$$

$$Percision = \frac{TP}{TP + FP} \tag{20}$$

$$Recall = \frac{TP}{TP + FN} \tag{21}$$

Where TP, FP, TN, and FN indicate Number of true-positive, false-positive, true-negative, and false-negative. Experiments were done on a system which has Intel Core i7-7700HQ @ 2.80GHz processor, 16GB memory, Windows10, and experiment software, proposed method was coded by Matlab R2017b software. The results were analyzed using 10-fold cross validation to test the accuracy, precision and recall. Two algorithms HWOA-DA and WOA are evaluated on data set of heart disease for diagnosis. In all experiments, the maximum number of iterations sets 20 and 30, and the population size is 10 and 15. The WOA is compared whit assess the effect of hybridizing WOA and DA (HWOA-DA). The accuracy of SVM classifier is the value of fitness function of HWOA-DA and WOA.

Table 2: Result of WOA on Statlog heart disease dataset

Iteration	**Agent**	Accuracy	Precision	**Recall**
20	10	85.56	88.67	86.45
	15	84.44	87.33	85.46
30	10	84.81	89.33	84.57
	15	85.19	88.67	86.42

Table 3: Result of HWOA-DA on Statlog heart disease dataset

Iteration	Agent	**Accuracy**	**Precision**	Recall
20	10	88.15	89.33	89.98
	15	88.52	88.67	91.02
30	10	88.89	90	90.35
	15	88.15	91.33	88.28

Table 2 and table 3 present the results of hybrid algorithm HWOA-DA and WOA for heart disease diagnosis. Best Accuracy obtained of HWOA-DA and WOA are 88.89 and 85.56 respectively and number features selection of HWOA-DA and WOA are 9 features. Results show that the hybrid algorithms are much better than the native one. Table 4 presents the comparison of the proposed algorithm with other previous methods on Statlog heart disease dataset. In the study, the HWOA-DA algorithm is a higher accuracy than other methods. Compared to previous studies, the proposed algorithm is superior.

5 Conclusion

Heart disease is one of the main causes of death, it should be correctly diagnosed at the early stage to get treatment. Algorithms of machine learning and metaheuristic are one of the tools that use for diagnosis of disease. In this paper, the proposed methods integrate DA algorithm with the global search of WOA. The proposed approach is compared with

Table 4: The comparison between proposed algorithm and other methods on Statlog heart disease dataset

Reference	**Method**	**Accuracy**
[18] 2015	C4.5	77.5
	Fast Decision Tree	78.06
[19] 2017	SVM	81.82
	MLPE	84.77
Proposed algorithm	WOA	85.56
	HWOA-DA	88.89

WOA algorithm. The results show that the hybridizing WOA and DA (HWOA-DA) of the collected dataset is 88.89 % higher than native one which is 85.56 %. The result obtained help to the effective performance of the medical diagnosis and analysis.

References

[1] Arockia Vanitha C, D., Devaraj D, Venkatesulu, M. , Gene Expression Data Classification using Support Vector Machine and Mutual Information-based Gene Selection, Procedia Computer Science. 47(2015) 13-21.

[2] Blum, C. , Roli, A. M., Sampels, Hybrid Metaheuristics - An Emerging Approach to Optimization, Springer, 2008. Doi: 10.1007/978-3-540-78295-7.

[3] Chandrashekar, G. , Sahin, F. , A survey on feature selection methods, Computers and Electrical Engineering, 40(2014) 16-28.

[4] Chadha, R. , Mayank, S., Prediction of heart disease using data mining techniques, Springer, December 2016.

[5] Das, R., Turkoglu, I. and Sengur, A., Effective diagnosis of heart disease through neural networks ensembles, Expert Systems with Applications, 36(2009) 7675-7680.

[6] Davari, A. , Esmael Zadeh Khadem, S., Automated Diagnosis of Coronary Artery Disease (CAD) Patients Using Optimized SVM, Computer Methods and Programs in Biomedicine 2016, http://dx.doi.org/doi: 10.1016/j.cmpb.2016.10.011

[7] El-Bialy, R. , Salamay, M.A. , Omar, H. , Karam3 and M.Essam Khalifa, Feature Analysis of Coronary Artery Heart Disease Data Sets, Procedia Computer Science. 65(2015) 459 - 468.

[8] Ghaddar, B. , Naoum-Sawaya, J., High Dimensional Data Classification and Feature Selection using Support Vector Machines, European Journal of Operational Research (2017), doi: 10.1016/j.ejor.2017.08.040

[9] Jain, D., Singh, V. , Feature selection and classification systems for chronic disease prediction: A review. Egyptian Informatics J (2018), https://doi.org/10.1016/j.eij.2018.03.002

[10] Kaur, G., Arora, S. , Chaotic whale optimization algorithm, Journal of Computational Design and Engineering. 5(3), (2018) 275-284.

[11] Mukherjee, S. , Kapoor S. , Banerjee, P., Diagnosis and Identifiation of Risk Factors for Heart Disease Patients Using Generalized Additive Model and Data Mining Techniques, J Cardiovasc Disease Res. 8(4) (2017) 137-144.

[12] Mirjalili, S., Lewis, A. , the Whale Optimization Algorithm, Advances in Engineering Software 95(2016) 51-67.

[13] Mirjalili, S., Dragonfly algorithm: a new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems, The Natural Computing Applications Forum 2015, DOI 10.1007/s00521-015-1920-1

[14] Sy Ahmad Ubaidillah, S.H., Sallehuddin,R., Mustaffa, N.H, Classification of Liver Cancer using Artificial Neural Network and Support Vector Machine, Proc. of Int. Conf. on Advances in Communication, Network, and Computing, CNC, 2014

[15] Shouman, M., Turner, T., Stocker, R., "Using Decision Tree for Diagnising Heart Disease Patients," AusDM '11 Proceedings of the Ninth Australasian Data Mining Conference, 121(2011) 23-30.

[16] Sree Ranjini, K.S., Murugan, S., Memory based Hybrid Dragonfly Algorithm for numerical optimization problems, Expert Systems With Applications 83 (2017) 63-78

[17] Talbi, E.G., A Taxonomy of Hybrid Metaheuristic. J. Heuristics, 8(2002) 541-546.

[18] UCI Machine Learning Repository. [Online]. Available http://archive.ics.uci.edu/ml/datasets/statlog+(heart)

[19] Vapnik, VN., The Nature of Statistical Learning Theory. New York: Springer-Verlag 1995

[20] Wong, N.D., Epidemiological studies of CHD and the evolution of preventive cardiology, Nat. Rev. Cardiol, 11(2014) 276-289