# Feature Selection Method Based on Grey Wolf Optimization for Coronary Artery Disease Classification

# Feature Selection Method Based on Grey Wolf Optimization for Coronary Artery Disease Classification

Qasem Al-Tashi[1,2](✉), Helmi Rais[1], and Said Jadid[1]

[1] Department of Computer and Information Sciences, Universiti Teknologi PETRONAS, 32610 Bandar Seri Iskandar, Perak, Malaysia
qasemacc22@gmail.com, {qasem_17004490, helmim, saidjadid.a}@utp.edu.my
[2] University of Albydha, Al Bayda, Yemen

**Abstract.** Cardiovascular disease has been declared as one of the deadly illness that affects humans in the Middle and Old ages across the globe. One of the cardiovascular disease known as Coronary artery, has recorded the highest number of motility rates in the recent years. Machine learning tools have been very effective in investigating the causes of such lethal disease which involve analyzing large amount of dataset. Such datasets might contain redundant and irrelevant features which affect the classification accuracy and processing speed. Hence, applying feature selection technique for the elimination of the said redundant and irrelevant features is necessary. In this paper, a novel wrapper feature selection method is proposed to determine the optimal feature subset for diagnosing coronary artery disease. This proposed method consists of two main stages feature selection and classification. In the first stage, Grey Wolf Optimization (GWO) is used to find the best features in the disease identification dataset. In the second stage, the fitness function of GWO is evaluated using Support Vector Machine classifier (SVM). Cleveland Heart disease dataset is used for performance validation of the proposed method. The experimental results showed that, the proposed GWO-SVM outperforms current existing approaches with an achievement of 89.83% in accuracy, 93% in sensitivity and 91% in specificity rates.

**Keywords:** Feature selection · Grey wolf optimization
Support vector machine

## 1 Introduction

In recent years, machine learning and data mining has been the most important techniques in medical diagnosis as well as intelligent decision-making processes [1]. In medical industry, data mining techniques are tremendously used in the various diseases prediction models such as heart disease [2], brain tumor [3], skin cancer [4], breast cancer [5] and etc. Heart disease prediction techniques are more essential among other prediction since heart disease is significantly increased in most countries around the globe [6]. Hence, such techniques are used to assist doctors with faster and accurate

F. Saeed et al. (Eds.): IRICT 2018, AISC 843, pp. 257–266, 2019.
https://doi.org/10.1007/978-3-319-99007-1_25

heart disease prediction. Also, the heart disorders are called as Cardio Vascular Diseases (CVD) that describes the blood vessels and diseases of heart [7]. CVD includes coronary heart disease (CHD) or Coronary Artery Disease (CAD), Rheumatic heart disease, Cerebrovascular disease (stroke), hypertension (high blood pressure) and heart failure. Among these various heart disorders, CAD is the commonly occurring heart disorder because of deposition of fatty and cholesterol within the internal wall of the arteries that blocks the required blood flow to the heart [8].

Generally, Coronary Artery Disease (CAD) prediction technique consists of two stages such as feature selection stage and classification stage. Feature selection stage is used to select the feature subsets and then use them for the classification stage [9, 10]. The input datasets of CAD include relevant, irrelevant and redundant features. However, the redundant features as well as irrelevant features create noise to the target class. In addition, these features not only affects the classification performance but also reduce the system response time [11]. Hence, feature selection technique is necessary for removing those features, which in turn reduces the risk of over fitting, improves accuracy and requires less computation [6].

Currently, various feature selection approaches have been proposed. In [12], a feature selection method proposed based on genetic algorithm (GA) algorithm and Bayes Naïve (BN) to obtian the optimal features in the disease identification. This method contains of two stages first generation of a subset of features using GA and the evaluation of the system using BN in the second stage.

In [13], a feature selection method proposed based on Artificial Bee Colony (ABC) algorithm and Support Vector Machine (SVM) to obtain the optimal features in the disease identification. SVM classification is used to evaluate the fitness of ABC.

In [14], to predict the risk level of heart disease, genetic algorithm based fuzzy decision support system has been proposed. The method consists of pre-processing the dataset, features are obtained based on different approaches, and weighted fuzzy rules are produced based on selected attributes using GA and finally fuzzy decision support system (FDSS) used for heart disease prediction.

In [15], the authors proposed a hybrid method for the diagnosis of coronary artery disease. In this method the initial weights of neural network were identified via genetic algorithm. Then, the neural network was learned using training data.

In [16], an automatic fuzzy diagnostic system based a modified dynamic multi-swarm particle swarm optimization (MDMS-PSO) and on genetic algorithm (GA) was proposed for predicting the risk level of heart disease.

The previous studies all of them use the same disease dataset which is Cleveland dataset and were selected for impartial comparison with our proposed method. Particle Swarm Optimization (PSO), ABC, and GA are popular meta-heuristic optimization techniques have been used to search and find relevant and optimal features with some drawbacks such as trapped in local optima [17]. The Grey Wolf Optimizer (GWO) is a new optimization algorithm developed by [18], which mimics the leadership hierarchy of wolves which are well known for their group hunting, the gray wolf optimization method demonstrates much robustness in comparison with PSO and GA optimizers against initialization [19]. Further, this algorithm has a few parameters only and easy to implement, which makes it superior than earlier ones. In this paper a wrapper based feature selection based on gray wolf optimizer to find optimal feature subset for Cleveland dataset is proposed.

For classification, there are several classification techniques have been used for heart disease prediction such as Naive Bayes, linear regression, neural networks, support vector machine, and Fuzzy classifier [1, 20, 21]. Several researches have reported with highly confirmation that the support vector machines (SVM) have greater accurate diagnosis ability [22]. Therefore, SVM will be used for classification stage.

The rest of the paper is organized as follows: a summary of the grey wolf optimization and its mathematical model are provided in the second section. Third Section illustrates the proposed wrapper feature selection method GWO-SVM. Section Fourth, then discusses the experimental setup used with the dataset. In the fifth section experimental results are compared with those of existing approaches. Conclusions are finally drawn in the last Section.

## 2 Gray Wolf Optimization (GWO)

GWO is inspired by social hierarchy and the hunting approach of grey wolves proposed by [18]. Grey wolves typically prefer to live in a pack of 5–12 individuals and have a strict social hierarchy. As shown in Fig. 1, GWO Consist of four levels as follows:

(1) **Alpha (α)**: male and female are the leaders of a pack of wolves that are responsible for making decisions such as wake-up time, hunting, and sleep place.
(2) **Beta (β):** either male or female wolves, beta probably the best candidate of replacement for alpha. Assisting α in decisions making and suggesting feedbacks are the main roles of β.
(3) **Delta (δ):** The wolves at this level obey α and β wolves and control ω wolves. Delta acts as sentinels, scouts, elders, sentinels, caretakers in the pack, and hunters.
(4) **Omega (ω)**: the wolves at this level are the weakest. Omega (ω) plays a role of scapegoat. Omega (ω) should obey other individuals' orders.

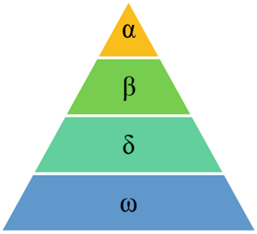

**Fig. 1.** Grey wolves' social hierarchy represented by [17].

## 2.1 Mathematical Model of the GWO

In the GWO algorithm, hunting is guided by α, β, and δ, and ω wolves follow them. Mathematically, the grey wolves' encircling behavior can be denoted as:

$$\vec{\mathrm{X}}(\mathrm{t}+1) = \vec{\mathrm{X}}_p(t) + \vec{\mathrm{A}}\,.\,\vec{\mathrm{D}} \tag{1}$$

Where $\vec{\mathrm{X}}_p$ is the vector of the prey's positions, $\vec{\mathrm{X}}$ is the vector of the grey wolf's positions, $(t)$ is the number of iterations, the coefficient vectors are $\vec{\mathrm{A}}$, $\vec{\mathrm{C}}$. Where $\vec{\mathrm{D}}$ is defined as follows:

$$\vec{\mathrm{D}} = \left|\vec{\mathrm{C}}\,.\,\vec{\mathrm{X}}_p(t) - \bar{\mathrm{X}}(t)\right| \tag{2}$$

The coefficient vectors $\vec{\mathrm{A}}$, $\vec{\mathrm{C}}$ can be calculated as following:

$$\vec{\mathrm{A}} = 2\vec{a}\,.\,\overrightarrow{r_1} - \vec{a} \tag{3}$$

$$\vec{\mathrm{C}} = 2\,.\,\overrightarrow{r_2} \tag{4}$$

Where $\overrightarrow{a}$ are a vector set decrease over iterations linearly from 2 to 0, $\overrightarrow{r_1}$ and $\overrightarrow{r_2}$ are random vectors in [0, 1]. The hunting behaviour of grey wolves to be mathematically simulated, alpha (α) is assumed to be the best candidate for the solution, beta (β), and delta (δ) assumed to have more information about the possible position of the prey. Accordingly, three best solutions obtained so far are saved and forces others i.e., (ω) to update their positions according to the best place in the decision space. Such a hunting behaviour can be represented as the following equations:

$$\overrightarrow{D_\alpha} = \left|\overrightarrow{\mathrm{C}_1}\,.\vec{\mathrm{X}}_\alpha - \vec{\mathrm{X}}\right|, \overrightarrow{D_\beta} = \left|\overrightarrow{\mathrm{C}_2}\,.\vec{\mathrm{X}}_\beta - \vec{\mathrm{X}}\right|, \overrightarrow{D_\delta} = \left|\overrightarrow{\mathrm{C}_3}\,.\vec{\mathrm{X}}_\delta - \vec{\mathrm{X}}\right| \tag{5}$$

$$\overrightarrow{X_1} = \overrightarrow{X_\alpha} - A_1.\left(\overrightarrow{D_\alpha}\right), \overrightarrow{X_2} = \overrightarrow{X_\beta} - A_2.\left(\overrightarrow{D_\beta}\right), \overrightarrow{X_3} = \overrightarrow{X_\delta} - A_3.\left(\overrightarrow{D_\delta}\right) \tag{6}$$

$$\overrightarrow{X}(t+1) = \frac{\overrightarrow{X_1 + X_2 + X_3}}{3}. \tag{7}$$

Finally, the trade-off between exploration and exploitation is controlled by the updating of the $\vec{a}$ parameter. In each iteration $\vec{a}$ parameter is updated linearly to range from 2 to 0 as according to the equation below:

$$\vec{a} = 2 - t\,.\,\frac{2}{\underset{i}{max}\,ter} \tag{8}$$

Where $max_i\,ter$ indicates the total number of iterations allowed for the optimization and $t$ is the number of iteration.

## 3 Proposed Method

This study proposed an effective feature selection method, GWO-SVM, to diagnosis a Coronary artery disease. GWO-SVM is consisting of two main phases. In the first phase, GWO proposed by [17] is used to eliminate the redundant and irrelevant features by searching for the best feature in the Cleveland heart dataset. Firstly, GWO produce the initial positions of population, and then update the current positions of population in the discrete searching space. In the second phase, the highly effective SVM classifier is conducted based on the optimal feature subset gained in the first phase. Figure 3 presents a general structure of the proposed GWO-SVM method.

The GWO is used to effectively search the feature space for best feature. The optimal and best feature is the one with high classification accuracy and less number of selected features. The fitness function is used to maximize the classification accuracy also used in GWO to evaluate the selected features is denoted as in Eq. 9:

$$Fitness = \alpha P + \beta \frac{N - L}{L} \tag{9}$$

Where $P$ is the classification accuracy, $L$ is the length of the selected feature, $N$ is the total number of features in the overall dataset, where $\alpha$ and $\beta$ are the parameters corresponding to the classification accuracy weight and quality of feature selection, $\alpha \in [0, 1]$ and $\beta = 1 - \alpha$. SVM classifier is used for calculating the value (P) in Eq. 9.

The first and primary step for solving a feature selection problem utilizing GWO is to illustrate a feature subset in a solution representation. Figure 2 shows the solution representation. For the proposed feature selection method, we utilized a binary chromosome to illustrate a feature subset. The chromosome's length is represented as $d$, where $d$ indicates the whole numbers of features. The chromosome's position can take either a '1' or a '0' value. If the value of the $i$th bit equals one then the feature is selected; otherwise, this feature is not selected ($i$ = 1, 2, …). The number of bits whose values are one are therefore representing the size of a feature subset.

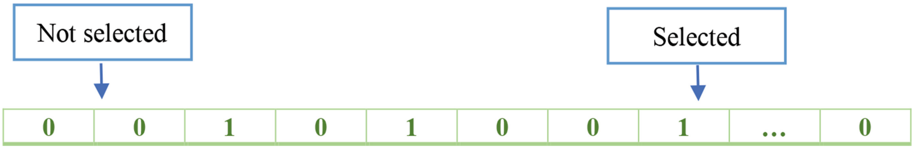

**Fig. 2.** Solution representation of feature selection

## 4 Experimental Setup

The proposed method is implemented using MatLab R2018a running on a machine equipped with Intel core i5 processor and a 4 GB RAM capacity and the operating system is Windows 10 Professional 64 bit.

### 4.1 Dataset Used

Cleveland dataset freely available and can be downloaded from UCI repository [23]. The attributes description of the Cleveland dataset is given in Table 2.

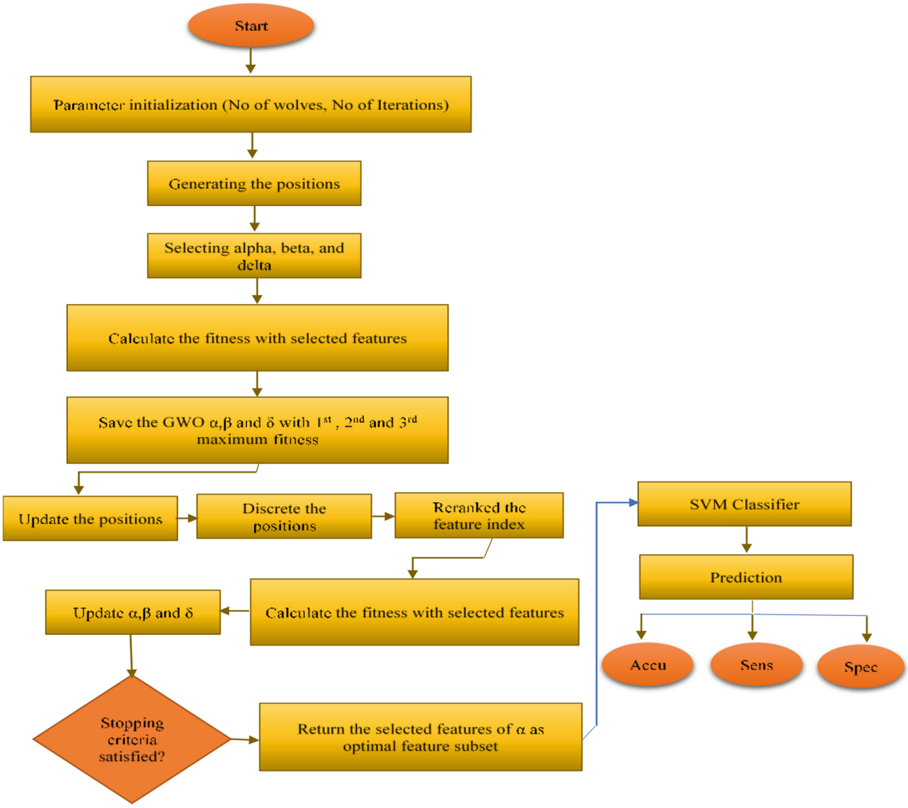
Start
Parameter initialization (No of wolves, No of Iterations)
Generating the positions
Selecting alpha, beta, and delta
Calculate the fitness with selected features
Save the GWO α,β and δ with 1st , 2nd and 3rd maximum fitness
Update the positions
Discrete the positions
Reranked the feature index
Calculate the fitness with selected features
Update α,β and δ
Stopping criteria satisfied?
Return the selected features of α as optimal feature subset
SVM Classifier
Prediction
Accu
Sens
Spec

**Fig. 3.** The proposed feature selection method.

### 4.2 Parameter Control Setting

Table 1 shows the parameter setting used for the proposed method. Where the iterations, wolves, dimension numbers and search domain are identified. α and β parameters for the fitness function are declared.

**Table 1.** Parameter setting for the proposed method

| Parameter | Numbers |
|---|---|
| Iterations no. | 100 |
| Wolves no. | 5 |
| Dimensions no. | 14 |
| Search domain | [0 1] |
| $\alpha$ in fitness function | 0.99 |
| $\beta$ in fitness function | 0.01 |

**Table 2.** Attributes of Cleveland dataset

| No | Attributes | Description |
|---|---|---|
| 1 | Age | Age in year |
| 2 | Sex | 0 for female and 1 for male |
| 3 | Cp | Chest pain type<br>Value 1: typical angina<br>Value 2: atypical angina<br>Value 3: non-anginal pain<br>Value 4: asymptomatic |
| 4 | Trestbps | Resting blood sugar in mm Hg on admission to the hospital |
| 5 | Chol | Serum cholesterol in mg/dl |
| 6 | Fbd | (Fasting blood sugar > 120 mg/dl) (1 = true; 0 = false) |
| 7 | Restecg | Resting ECG result |
| 8 | Thalach | Maximum heart rate achieved |
| 9 | Exang | Exercise induced angina |
| 10 | Oldpeak | ST depression induced by exercise relative to rest |
| 11 | Slope | Slope or peak exercise ST segment |
| 12 | Ca | Number of major vessels colored by fluoroscopy |
| 13 | Thal | Defect type |
| 14 | num | The predicted attribute |

### 4.3 Performance Evaluation

The performance of the proposed method has been evaluated based on sensitivity, specificity, and accuracy tests, which use the true positive (TP), true negative (TN), false negative (FN), and false positive (FP) terms. These measures are calculated as follows:

$$Sensitivity = \frac{TP}{TP + FN} \times 100\%, \tag{10}$$

$$Specificity = \frac{TN}{FP + TN} \times 100\%, \tag{11}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%. \qquad (12)$$

## 5 Result and Discussion

As shown in Table 3, we have conducted two experiments using the same dataset as presented in Sect. 4.1. In the first experiment 10-run time have been conducted and the best result have been chosen, the proposed method GWO-SVM yielded a good result after 10 run time where six features are obtained and achieved accuracy, sensitivity and specificity rates of 87.65%, 88% and 90% respectively. In the second experiment also a 10-run time have been conducted and the best result have been chosen where the number of feature obtained is 7 and the results achieved for accuracy, sensitivity and specificity rates are 89.83%, 93% and 91% respectively. It is found that, Restecg and age features are the major contributing features in cardiovascular disease. Note that, the actual accuracy obtained when all features of Cleveland dataset used as input to SVM is equal to 76.57%.

**Table 3.** Feature obtained by the proposed method (GWO-SVM)

| Experiment | Feature selected | Accuracy | Sensitivity | Specificity |
|---|---|---|---|---|
| First | Fbs, Thalach, Oldpeak, Slope, Chol, ca | 87.65% | 88% | 90% |
| Second | Age, Chol, Fbs, Restecg, Thalach, slope, ca | 89.83% | 93% | 91% |

A comparison of results between our proposed method and state-of-art methods is shown in Table 4. Studies that used the same datasets (Cleveland dataset) were selected for impartial comparison with our proposed method.

**Table 4.** Comparison between GWO-SVM and state of art methods.

| Study | Method | Accuracy | Sensitivity | Specificity |
|---|---|---|---|---|
| [13] | ABC-SVM | 86.76% | N/A | N/A |
| [14] | GA-FBSS | 80% | 84% | 75% |
| [15] | GA-NN | 89.4% | 88% | 91% |
| **Our study** | **GWO-SVM** | **89.83%** | **93%** | **91%** |

Referring to Table 4. we can see that, the accuracy produced by GWO-SVM outperforms ABC-SVM method proposed by [13]. In addition to accuracy, the ABC-SVM method did not employ the use of sensitivity and specificity evaluation. Moreover, in the second comparison, our method also outperforms the GA-FDSS method proposed by [14], in the areas such as accuracy, sensitivity, and specificity. Finally, we compare our proposed method with the hybrid neural network-Genetic algorithm proposed by [15], different heart datasets have been used such as, Z-Alizadeh Sani dataset, Cleveland, Hungarian and Switzerland. However, the comparison is done only on Cleveland dataset based on which our method shows better performance in terms of accuracy, sensitivity, and specificity.

## 6 Conclusion

A wrapper feature selection method based on grey wolf optimization and support vector machine has been proposed in this paper. The proposed method has the capability to select relevant features and eliminate redundant and irrelevant features from Cleveland dataset. A comparison was conducted between the proposed method and three competitive counterparts feature selection methods. Experimented results demonstrated that the proposed method performed greatly in terms of accuracy, sensitivity, and specificity and outperformed the state-of-art methods. On the other hand, different classifiers could be used with GWO to further enhance the results. Moreover, some other datasets can be applied in the future to further investigating the robustness of the proposed method.

## References

1. Shouman, M., Turner, T., Stocker, R.: Using data mining techniques in heart disease diagnosis and treatment. In: Japan-Egypt Conference on Electronics, Communications and Computers, pp. 173–177 (2012)
2. Dereli, T., Seckiner, S.U., Das, G.S., Gokcen, H., Aydin, M.E.: An exploration of the literature on the use of ‘swarm intelligence-based techniques’ for public service problems. Eur. J. Ind. Eng. **3**(4), 379 (2009)
3. Arakeri, M.P., Reddy, G.R.M.: Computer-aided diagnosis system for tissue characterization of brain tumor on magnetic resonance images. Signal Image Video Process. **9**(2), 409–425 (2015)
4. Xie, F., Fan, H., Li, Y., Jiang, Z., Meng, R., Bovik, A.: Melanoma classification on dermoscopy images using a neural network ensemble model. IEEE Trans. Med. Imaging **36** (3), 849–858 (2017)
5. Rasti, R., Teshnehlab, M., Phung, S.L.: Breast cancer diagnosis in DCE-MRI using mixture ensemble of convolutional neural networks. Pattern Recognit. **72**, 381–390 (2017)
6. Long, N.C., Meesad, P., Unger, H.: A highly accurate firefly based algorithm for heart disease prediction. Expert Syst. Appl. **42**(21), 8221–8231 (2015)

7. Krishnaiah, V., Narsimha, G., Chandra, N.S.: Heart disease prediction system using data mining technique by fuzzy K-NN approach. In: Emerging ICT for Bridging the Future-Proceedings of the 49th Annual Convention of the Computer Society of India (CSI), vol. 1, pp. 371–384 (2015)
8. Patidar, S., Pachori, R.B., Acharya, U.R.: Automated diagnosis of coronary artery disease using tunable-Q wavelet transform applied on heart rate signals. Knowl. Based Syst. **82**, 1–10 (2015)
9. Khemphila, A., Boonjing, V.: Heart disease classification using neural network and feature selection. In: 2011 21st International Conference on Systems Engineering (ICSEng), pp. 406–409 (2011)
10. Sanz, J.A., Galar, M., Jurio, A., Brugos, A., Pagola, M., Bustince, H.: Medical diagnosis of cardiovascular diseases using an interval-valued fuzzy rule-based classification system. Appl. Soft Comput. **20**, 103–111 (2014)
11. Shilaskar, S., Ghatol, A.: Feature selection for medical diagnosis: evaluation for cardiovascular diseases. Expert Syst. Appl. **40**(10), 4146–4153 (2013)
12. Mokeddem, S., Atmani, B., Mokaddem, M.: Supervised feature selection for diagnosis of coronary artery disease based on genetic algorithm (2013). arXiv Preprint arXiv:1305.6046
13. Subanya, B., Rajalaxmi, R.R.: Feature selection using artificial bee colony for cardiovascular disease classification. In: 2014 International Conference on Electronics and Communication Systems (ICECS) (2014)
14. Paul, A.K., Shill, P.C., Rabin, M.R.I., Akhand, M.A.H.: Genetic algorithm based fuzzy decision support system for the diagnosis of heart disease. In: 2016 5th International Conference on Informatics, Electronics and Vision (ICIEV), pp. 145–150 (2016)
15. Arabasadi, Z., Alizadehsani, R., Roshanzamir, M., Moosaei, H., Yarifard, A.A.: Computer aided decision making for heart disease detection using hybrid neural network-Genetic algorithm. Comput. Methods Programs Biomed. **141**, 19–26 (2017)
16. Paul, A.K., Shill, P.C., Rabin, M.R.I., Murase, K.: Adaptive weighted fuzzy rule-based system for the risk level assessment of heart disease. Appl. Intell. **48**, 1739–1756 (2017)
17. Xue, B., Zhang, M., Browne, W.N., Yao, X.: A survey on evolutionary computation approaches to feature selection. IEEE Trans. Evol. Comput. **20**(4), 606–626 (2016)
18. Mirjalili, S., et al.: Grey Wolf Optimizer. Adv. Eng. Softw. **69**, 46–61 (2014)
19. Emary, E., Zawbaa, H.M., Grosan, C., Hassenian, A.E.: Feature subset selection approach by gray-wolf optimization. In: Afro-European Conference for Industrial Advancement, pp. 1–13 (2015)
20. Srinivas, K., Rao, G.R., Govardhan, A.: Analysis of coronary heart disease and prediction of heart attack in coal mining regions using data mining techniques. In: 2010 5th International Conference on Computer Science and Education (ICCSE), pp. 1344–1349 (2010)
21. Das, R., Turkoglu, I., Sengur, A.: Effective diagnosis of heart disease through neural networks ensembles. Expert Syst. Appl. **36**(4), 7675–7680 (2009)
22. Akay, M.F.: Support vector machines combined with feature selection for breast cancer diagnosis. Expert Syst. Appl. **36**(2), 3240–3247 (2009)
23. Cleveland dataset. http://archive.ics.uci.edu/ml/machine-learning-databases/heart-disease/processed.cleveland.data. Accessed 27 May 2018