International Journal of Computer Applications (0975 – 8887)
Volume 151 – No.1, October 2016

Swarm Intelligence Approach for Breast Cancer Diagnosis

Hoda Zamani
Faculty of Computer Engineering, Najafabad branch, Islamic Azad University, Najafabad, Iran

Mohammad-Hossein Nadimi-Shahraki*
Faculty of Computer Engineering, Najafabad branch, Islamic Azad University, Najafabad, Iran

ABSTRACT
Since the breast cancer has been become one of the main reasons of death in women especially in the developed countries, there have been done many research for breast cancer diagnosis. Although researchers have recently proposed many methods by using intelligent approaches for diseases diagnosis, a few of them fulfill the need of high accuracy. In this paper, the most popular swarm intelligence algorithms PSO, ICA, FA and IWO are applied to diagnosis the breast cancer. The experimental results show that swarm intelligence approach can be applied for breast cancer diagnosis with high accuracy. Moreover, FA can diagnose the breast cancer more accurate than other swarm intelligence methods compared in this paper.

Keywords
Swarm Intelligence, Diseases diagnosis, Breast cancer

1. INTRODUCTION
Breast cancer is a malignant growth of cancer lumps in breast tissue. In this disease, lumps in the breast tissue are grown abnormally and spread to other tissues and damage them. Nowadays, breast cancer has been become one of the main reasons of death in women. The main cause of developing this disease still remains a mystery, but researchers announced some of the factors that may increase a woman's risk developing breast cancer as follows: aging, family history, previous diagnosis of breast cancer, previous benign breast lump, breast density, exposure to oestrogen, being overweight or obese, being tall, alcohol, radiation, hormone replacement therapy, personal history, breast cancer gene mutation, gynecologic history, breast changes [1]. Generally breast cancer shows no pain and symptoms in the early stage and when the patient is aware of their condition that there is a few chance of survival. Obviously, accurate and timely diagnosis of breast cancer can increase the chances of survival and also reduce the treatment costs. Although evaluation of patient data and decision of experts are the most important factor in diagnosis, many machine learning approaches and pattern recognition are proposed to support experts in decision-making process. Extract informative knowledge of patient data and reduce time and cost of diagnosis are the main purpose of these approaches. In this regard, many methods have been proposed base on swarm intelligence approach.

Swarm intelligence (SI), which is an artificial intelligence discipline, is concerned with the design of intelligent multi-agent systems by taking inspiration from the collective behavior of social insects such as ants, termites, bees and wasps, as well as from other animal societies such as flocks of birds or schools of fish [2]. Swarm intelligence approach refers to exploring the problem space and extracting optimal solutions by the intelligent agents. Recently, SI is used in optimization problems, especially in the medical fields such as diagnosis, predication, treatment and screening. Examples of notable swarm intelligence optimization approaches are applied in medical diagnosis as follows: Particle Swarm Optimization (PSO) [3, 4], Ant Colony Optimization (ACO) [5], Artificial Immune Recognition System (AIRS) [6], Artificial Bee Colony (ABC) [7], Firefly algorithm (FA) [16] and invasive weed optimization (IWO) [18]. All of these approaches aim to extract the optimal solutions, especially in NP-hard problems. They start with initial population of agents, then distributed in the problem space. Each agent depending on its cooperation and competition theory began to explore and exploit. Exploration and exploitation process are done in iteratively finite number of steps.

In this study, the swarm intelligence algorithms PSO, ICA, FA and IWO are applied for breast cancer diagnosis. By using any of these algorithms, the efficient patterns are extracted for diagnosis of breast cancer. Then, the artificial neural networks are applied to classification and evaluation the extracted patterns. The experimental results show that particle swarm optimization (PSO) achieve a high accurate diagnosis for breast cancer comparing with other algorithms.

The rest of this paper is organized as follows. Section 2 is to review the most popular swarm intelligence algorithms applied in this study. Then, in section 3 these approaches are experimentally evaluated for breast cancer diagnosis. Finally, section 4 discusses this study and future works.

2. SWARM INTELLIGENCE ALGORITHMS
In this section, the swarm intelligence algorithms PSO, ICA, FA and IWO are briefly explained in order to apply them for diagnosis of breast cancer. These algorithms are stochastic, nature-inspired and based on population that can be applied for solving the optimization problems. The main strategy for solving this problems is cooperation and competition between agents. They try to explore the problem space and exploit the optimal solutions. In this situation, many concepts are formed and different strategies are proposed to control them such as balance between exploration and exploitation, premature convergence and local minimum.

2.1 Particle Swarm Optimization
Particle swarm optimization (PSO) algorithm was proposed by Kennedy and Eberhart in 1995 [12]. This algorithm is a population-based stochastic optimization technique modelled on the social behaviors observed in animals or insects, e.g., bird flocking, fish schooling and animal herding [2, 13]. In PSO, each particle have five properties which are: $\overrightarrow{X_{\imath d}}(t), J^{\overrightarrow{X_\imath} d(t)}$, $\overrightarrow{V_{\imath d}}(t), \overrightarrow{P_{\imath d}}(t)$ and $J^{\overrightarrow{P_{\imath d}}(t)}$. These parameters in a d-dimensional search space for the i-th particle are defined as follows: $\overrightarrow{X_{\imath d}}(t)$ is a position vector, $J^{\overrightarrow{X_\imath} d(t)}$ is an objective function, $\overrightarrow{V_{\imath d}}(t)$ is a velocity, $\overrightarrow{P_{\imath d}}(t)$ and $J^{\overrightarrow{P_{\imath d}}(t)}$ are the personal and the objective function of the best position

* Corresponding Author

this particle was experienced respectively. At first, for each particle a velocity vector is generated by using a random float number generator that limited with lower and upper bound of d-dimensional search space. Iteratively, the velocity of each particle updated by its personal best position, and the best position found by particle in its neighborhood. Finally, each particle changes its position according to the velocity. The velocity and the next position of a particle are computed by Eq. (1) and (2) where $\boldsymbol{C_1}$ and $\boldsymbol{C_2}$ are personal and global learning coefficient respectively and $\boldsymbol{r_1}$ and $\boldsymbol{r_1}$ are uniformly random number in the range of [0, 1]. In addition, $\boldsymbol{w(t)}$ is inertia weight and $\boldsymbol{P_{gd}(t)}$is the best position achieved by particle in the neighborhood of the i-th particle. Meanwhile, the neighborhood can be defined by different topology such as star, ring and von Neumann.

$$V_{id}(t+1) = w(t) \times V_{id}(t) + C_1 \times r_1 \times (P_{id}(t) - X_{id}(t)) + C_2 \times r_2 \times (P_{gd}(t) - X_{id}(t)) \quad (1)$$

$$X_{id}(t+1) = X_{id}(t) + V_{id}(t+1) \quad (2)$$

2.2 Imperialistic Competition Algorithm

Imperialist competitive algorithm (ICA) is mainly inspired by the imperialist competitive. Imperialism was greatly influenced by an economic theory known as mercantilism which inspired the government to extend their power and the rule beyond its own boundaries [14, 15]. ICA is a population-based algorithm that each individual based on their power is divided into two groups, colonies and imperialists. At first, ICA starts with an initial population called countries. Then, some of the countries with the best fitness value are selected to be the imperialists and others are sets to colonies of these imperialists. These colonies according to their cost are divided between imperialists. Therefore, each empire is comprised of imperialist and their colonies. After that, the imperialistic competition is started until that collapse weak empires and remain powerful empire. Therefore, the main steps of ICA can be summarized as follows [8]:

- Step 1: Generating initial empires.
- Step 2: Move the colonies of an empire toward the imperialist.
- Step 3: Exchanging position of the imperialist and a colony.
- Step 4: Computing the total cost of all empires.
- Step 5: Imperialistic competition and picking the weakest colony from the weakest empire.
- Step 6: Eliminating the powerless empires and the colonies will be divided among other empires.

2.3 Firefly Algorithm

Firefly algorithm (FA) is a kind of meta-heuristic algorithm that developed by Yang in 2008 [16]. FA inspired by fireflies in the nature and can be used for solving a large number of NP-hard problems. All fireflies are unisex therefore they used flashing light to attract partners and it's called attractiveness. The amount of attractiveness or intensity of lights is associated with the objective function. The best individual with high intensity of lights is selected. Then the fireflies are moved towards the more attractive individuals. In summary, FA is controlled by three parameters: the randomization parameter α, the attractiveness β, and the absorption coefficient γ [17]. The next movement of i-th firefly is computed by Eq. (3) where $X_{id}(t)$ and $X_{jd}(t)$ are the current position of i-th and j-th firefly in the d-dimensional search space, ϵ_i is a random number that calculated by Gaussian distribution, r_{ij} is the Euclidean distance between two fireflies and β_0 is the attractiveness at r_0.

$$X_{id}(t+1) = X_{id}(t) + \beta_0 \times e^{-\gamma r_{ij}^2} \times \left(X_{jd}(t) - X_{id}(t) \right) + \alpha \times \epsilon_i \quad (1)$$

2.4 Invasive weed optimization

Invasive weed optimization (IWO) is a stochastic algorithm that inspired from colonizing weeds in the real world. IWO algorithm was proposed by Mehrabian and Lucas in 2006 [18]. This algorithm can be summarized as follows:

- Step 1: Generating a population of initial weeds.
- Step 2: Evaluating the fitness of the whole population members.
- Step 3: Each member of the population is allowed to produce seeds (Reproduction).
- Step 4: The generated seeds are distributed over the search space by normally distributed random numbers with mean equal to zero but varying variance (Spatial distribution).
- Step 5: A competitive mechanism is activated for eliminating undesirable plants with poor fitness and allowing fitter plants to reproduce more sees as expected (Competitive exclusion).
- Step 6: Checking the termination criteria.

3. EXPERIMENTAL EVALUATION

In this section, the performance of swarm intelligence algorithms PSO, ICA, FA and IWO is evaluated. The main focus of this evaluation is to find more accurate diagnosis of breast cancer.

3.1 Wisconsin Breast Cancer Dataset (WBCD)

In the experimental evaluation, all algorithms are run on the original Wisconsin breast cancer dataset downloaded from UCI machine learning repository [19]. This dataset includes 10 features and 699 samples taken from Fine Needle Aspirates (FNA) of human breast tissue. The samples are divided into two classes: Benign and Malignant. About 65.5 Percent of the samples are benign and 34.5 Percent have been diagnosed malignant. The total number of missing values in this dataset are 16. The statistical information of WBCD was analyzed and mean, median and standard deviation (STD) are shown in Table 1.

Table 1. Statistical information of WBCD

Features description	Mean	Median	STD	Min-Max
Clump Thickness	4.44	4	2.82	[1,10]
Uniformity of Cell Size	3.15	1	3.07	[1,10]
Uniformity of Cell	3.21	1	2.99	[1,10]

Shape				
Marginal Adhesion	2.83	1	2.86	[1,10]
Single Epithelial Cell Size	3.23	2	2.22	[1,10]
Bare Nuclei	3.54	1	3.64	[1,10]
Bland Chromatin	3.45	3	2.45	[1,10]
Normal Nucleoli	2.87	1	3.05	[1,10]
Mitoses	1.60	1	1.73	[1,10]

3.2 Evaluation Metrics

The performances of the algorithms applied in this study for diagnosis of breast cancer are evaluated by well-known metrics sensitivity, specificity, precision and accuracy. These metrics are computed according to Eq. (4), (5), (6) and (7) respectively.

Sensitivity (True positive rate) = $\frac{TP}{TP+FN}$ (%) (4)

Specificity (True negative rate) = $\frac{TN}{TN+FP}$ (%) (5)

Precision (positive predictive value) = $\frac{TP}{TP+FP}$ (%) (6)

Accuracy (ACC) = $\frac{TP+TN}{TP+TN+FP+FN}$ (%) (7)

3.3 Experimental Setup

Swarm Intelligence algorithms applied in this study are implemented using MATLAB on an Intel Core-i5 CPU with 6 GB of RAM. To find the best algorithms with higher accuracy for diagnosis of breast cancer, each algorithm is evaluated five times by using the evaluation metrics described in Section 3.2. During each time, firstly, the datasets were randomly split into two sets of 70% and 30% as a training set and a test set respectively. Then, the multi-layer perceptron (MLP) network is used to evaluate the performance of each algorithm. The number of input and output neurons in MLP network is problem-dependent. In this study, the number of inputs is set with number of features and the number of hidden nodes is computed based on Kolmogorov theorem shown in Eq. (8):

$$\text{Hidden nodes} = 2 \times \text{Input neurons} + 1 \quad (8)$$

In addition, for all algorithms the maximum iteration is set to 15, the initial population size is set to 30. In each algorithm, different number of adjustable parameters is set. In IWO algorithm minimum and maximum number of seeds, variance reduction exponent, initial and final value of standard deviation are set to 0, 5, 2, 0.5 and 0.001 respectively. Phi in PSO algorithm is set to 2.05. Firefly algorithm is set with these parameters: 1 for light absorption coefficient, 2 for attraction coefficient base value, 0.2 for mutation and 0.98 for mutation coefficient damping ratio. Moreover, lower and upper bound of decision variables for PSO, IWO, ICA and FA are set to -10 and 10 respectively.

3.4 Experimental results

This section shows the results of experimental evaluation PSO, ICA, FA and IWO algorithms for diagnosis of breast cancer. We assess the behavior of each metrics during run of these algorithms, where Max, Mean, Min and STD indicate the maximum, average, minimum and standard deviation value respectively. As shown in Table2, the accuracy of FA algorithm with 98.54% is more than PSO, ICA, FA and IWO algorithms for diagnose of breast cancer. Furthermore, Table 3, 4 and 5 indicate the results obtained by sensitivity, specificity and precision metrics and the best values are shows in a bold face. The results show that the breast cancer can be diagnosed with high accuracy by swarm intelligence approaches. In the next experiment, the accuracy of swarm intelligence algorithms applied in this study is evaluated by different domain of features and the results are indicated in Fig. 1

Table 2. Comparison of the swarm intelligence algorithm in term of accuracy (%)

Accuracy	**PSO-ANN**	**ICA-ANN**	**FA-ANN**	**IWO-ANN**
Max	97.81	96.35	**98.54**	94.89
Mean	96.35	93.87	96.79	92.71
Min	95.62	91.97	95.62	91.24
STD	0.010	0.016	0.011	0.015

Table 3. Comparison of the swarm intelligence algorithm in term of Sensitivity (%)

Sensitivity	**PSO-ANN**	**ICA-ANN**	**FA-ANN**	**IWO-ANN**
Max	**100.00**	**100.00**	**100.00**	95.45
Mean	98.77	93.77	98.38	91.60
Min	97.87	87.75	93.88	85.71
STD	0.011	0.045	0.027	0.036

Table 4. Comparison of the swarm intelligence algorithm in term of Specificity (%)

Specificity	**PSO-ANN**	**ICA-ANN**	**FA-ANN**	**IWO-ANN**
Max	**100.00**	**100.00**	**100.00**	97.73
Mean	98.66	97.09	98.43	95.11
Min	97.87	94.19	96.51	93.33
STD	0.012	0.024	0.015	0.017

Table 5. Comparison of the swarm intelligence algorithm in term of Precision (%)

Precision	PSO-ANN	ICA-ANN	Firefly-ANN	IWO-ANN
Max	**100.00**	**100.00**	**100.00**	95.24
Mean	97.25	93.76	94.12	90.19
Min	95.35	88.23	96.98	87.23
STD	0.025	0.049	0.028	0.030

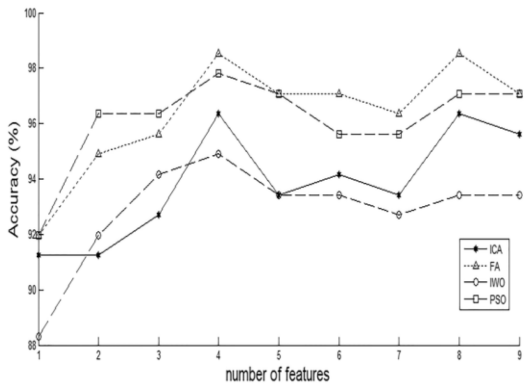

Fig 1. The accuracy of PSO, ICA, FA and IWO algorithms by different number of features

4. CONCLUSION AND FUTURE WORK

In diagnosis of disease base on Artificial intelligence approaches, the accuracy is very important particularly for breast cancer diagnosis. If this disease could be detected in the early stages with high accurate results, then the patient's chance of survival can be increased and also the time and cost associate with diagnosis of breast cancer can be decreased. Therefore, the present study focuses on the applying swarm intelligence for more accurate diagnosis of breast cancer. For this purpose PSO, ICA, FA and IWO were applied by the original Wisconsin breast cancer dataset. The experimental results show that FA algorithm with 98.54% accuracy can diagnose of breast cancer more accurate than other algorithms applied in this study. As a future work, the hybrid swarm intelligence approaches can be applied for diseases diagnosis.

5. REFERENCES

[1] http://www.nhs.uk/Conditions/Cancer-of-the-breast-female/Pages/Causes.aspx, Accessed 17th, April 2015.

[2] Blum, C., Li, X. 2008. Swarm intelligence in optimization, in: C. Blum, D. Merkle (Eds.), Swarm Intelligence: Introduction and Applications. Springer Verlag, Berlin, pp.43–86.

[3] Chang, P.-C., Lin, J.-J., and Liu, C.-H. 2012. An attribute weight assignment and particle swarm optimization algorithm for medical database classifications. Computer Methods and Programs in Biomedicine. vol. 107, pp. 382-392.

[4] Chen, H.-L., Yang, B., Wang, G., Wang, S.-J., Liu, J. and Liu, D.-Y. 2012. Support Vector Machine Based Diagnostic System for Breast Cancer Using Swarm Intelligence. Journal of Medical Systems. vol. 36, pp. 2505-2519.

[5] Ganji, M. F. and Abadeh, M. S. 2011. A fuzzy classification system based on Ant Colony Optimization for diabetes disease diagnosis. Expert Systems with Applications. vol. 38(12): pp.14650-14659.

[6] Chikh, M. A., Saidi, M. and Settouti, N. 2012. Diagnosis of Diabetes Diseases Using an Artificial Immune Recognition System2 (AIRS2) with Fuzzy K-nearest Neighbor. Journal of Medical Systems, vol. 36, pp. 2721-2729.

[7] Beloufa, F. and Chikh, M. A. 2013. Design of fuzzy classifier for diabetes disease using Modified Artificial Bee Colony algorithm. Computer Methods and Programs in Biomedicine. vol. 112, pp. 92-103.

[8] Quinlan, J.R. 1996. Improved use of continuous attributes in C4.5. Journal of Artificial Intelligence Research. Vol. 4, 77-90.

[9] Abonyi, J. and Szeifert, F. 2003. Supervised fuzzy clustering for the identification of fuzzy classifiers. Pattern Recognition Letters. vol. 24, pp. 2195-2207.

[10] Polat, K., Sahan, S., Kodaz, H. and Günes, S. 2005. A New Classification Method for Breast Cancer Diagnosis: Feature Selection Artificial Immune Recognition System (FS-AIRS). in Advances in Natural Computation: First International Conference. ICNC 2005, Changsha, China, August 27-29, 2005, Proceedings, Part II, L. Wang, K. Chen, and Y. S. Ong, Eds., ed Berlin, Heidelberg: Springer Berlin Heidelberg. pp. 830-838.

[11] Guijarro-Berdiñas, B., Fontenla-Romero, O., Pérez-Sánchez, B. and Fraguela, P. 2007. A Linear Learning Method for Multilayer Perceptrons Using Least-Squares. in Intelligent Data Engineering and Automated Learning-IDEAL 2007: 8th International Conference, Birmingham, UK, December 16-19, 2007. Proceedings, H. Yin, P. Tino, E. Corchado, W. Byrne, and X. Yao, Eds., ed Berlin, Heidelberg: Springer Berlin Heidelberg. pp. 365-374.

[12] Kennedy, J. and Eberhart, R. 1995. Particle swarm optimization. in Neural Networks. Proceedings., IEEE International Conference on. pp. 1942-1948 vol.4.

[13] Engelbrecht, A. P. 2007. Computational Intelligence: An Introduction: Wiley Publishing.

[14] Xing, B. and Gao, W.-J. 2014. Imperialist Competitive Algorithm. in Innovative Computational Intelligence: A Rough Guide to 134 Clever Algorithms, ed Cham: Springer International Publishing. pp. 203-209.

[15] Atashpaz-Gargari, E. and Lucas, C. 2007. Imperialist competitive algorithm: An algorithm for optimization inspired by imperialistic competition. in 2007 IEEE Congress on Evolutionary Computation. pp. 4661-4667.

[16] Yang, X.-S. 2010. Firefly Algorithm. in Engineering Optimization. ed: John Wiley & Sons, Inc., pp. 221-230.

[17] Fister, I. Fister Jr, I., Yang, X.-S. and Brest, J. 2006.A comprehensive review of firefly algorithms. Swarm and Evolutionary Computation, vol. 13, pp. 34-46.

[18] Mehrabian, A. R. and Lucas, C. 2006. A novel numerical optimization algorithm inspired from weed colonization. Ecological Informatics, vol. 1, pp. 355-366.

[19] Zwitter, M. and Soklic, M. (2015) University Medical Centre, Institute of Oncology, Ljubljana, Yugoslavia. Available: https://archive.ics.uci.edu/ml/datasets/Breast+Cancer.