Journal of AI and Data Mining
Vol 8, No 2, 2020, 237-245.
DOI: 10.22044/JADM.2020.8105.1945

H-BwoaSvm: A Hybrid Model for Classification and Feature Selection of Mammography Screening Behavior Data

E. Enayati[1*], Z. Hassani[2] and M. Moodi[3]

1. Computer Sciences Department, University of Bojnord, Bojnord, Iran.
2. Computer Sciences Department, Kosar University of Bojnord, Bojnord, Iran.
3. *Health Education and Health Promotion Department, Birjand University of Medical Sciences, Birjand, Iran.*

Received 19 February 2019; Revised 21 July 2019; Accepted 22 January 2020
**Corresponding author: e.enayati@ub.ac.ir (E. Enayati).*

Abstract
Breast cancer is one of the most common cancer in the world. An early detection of cancers causes a significant reduction in the morbidity rate and treatment costs. Mammography is a known effective diagnosis method of breast cancer. A way for mammography screening behavior identification is women's awareness evaluation for participating in the mammography screening programs. Today, intelligence systems could identify the main factors involved in a specific incident. These could help the experts a wide range of areas, specially health scopes such as prevention, diagnosis, and treatment. In this paper, we use a hybrid model called H-BwoaSvm BWOA is used for detecting the effective factors involved in the mammography screening behavior and SVM for classification. Our model is applied on a data-set which is collected from a segmental analytical descriptive study on 2256 women. The proposed model is operated on a data-set with 82.27% and 98.89% accuracy and selects the effective features on the mammography screening behavior.

Keywords: *Breast Cancer, Mammography Screening Behavior, Binary Whale Optimization Algorithm, SVM Algorithm, Feature Selection.*

1. Introduction

Breast cancer is one of the most common cancers among women throughout the world. in 2019, the breast cancer organization announced that1 in 8 US women buckled under the invasive breast cancer (about 12.4%) in their life-time. Based on these statistics, 268,600 new invasive breast cancer were distinguished in 2019, and the breast cancer fatality was 41,760 women in this year [1]. In 2018, the world Cancer Research Fund International (WCRF), has published a list of 25 countries with the most incidence rates of breast cancer. Belgium was the first that were Luxembourg, Netherlands, France, and New Caledonia Usually, the incidence rate of breast cancer is estimated to be n/100,000 cases. This rate in Belgium is 11,302/100,000 cases, in US is 84.9/100,000 [2], and in Iran is 33.1/100,000 [3]. The research works have shown that the breast cancer incidence is more popular in western countries than in the eastern countries. In despite, death rate is more in eastern countries because cancer is diagnosed ordinarily in the advanced stages [4].

Overall, women mortality has decreased significantly from 1989 up to now in the world. It maybe the result of progressive treatments or early detection mechanisms in the screening programs [1]. The average survival rate of breast cancer in the first 5-years of incidence is 90% [5]. Early detection of breast cancer is a survival factor. In this situation, patients can use various treatment services like inexpensive surgeries and chemotherapy with fewer serious side-effects [6]. There are two known methods for an early detection of breast cancer: clinical breast examination (CBE), known as mammography, and breast self-examination (BSE). A mammography test is a low-dose X-ray photography from a breast tissue. A mammogram test can figure out cancer 2 years junior. Mammography studies demonstrate mammogram tests, reducing the morbidity and mortality risks of breast cancer up to 20%. Moreover, new research works have shown that mammography screening programs in Europe and Canada have reduced this factor even up to 40%[6].

Incidence of mammogram test has been promoted with performing mammogram screening programs. This has been further than 50% between women older than 45 years, and between women 40-45 years old has been about 40% [6]. Mammogram screening programs can reduce death rate from breast cancer with growth of women participation rate [7]. Health researchers measure the breast cancer rate using the level of health services in various populations with some questionnaires. This measurement could reveal the people's preventive behavior in the society [8]. Breast cancer screening in Iran has been done by the Ministry of Health and Education from 2011. This program recommends that every women older than 40 should have a mammogram test every year [9]. Measuring women's awareness level of mammography screening reveals how much women are ready for participating in this program.

The transtheoretical model - also called the Stages of change model- was developed by Prochaska and DiClemente in the late 1970s. This model is a model of intentional change, and believes that the change in behavior is a process and occurs continuously through a series of stages.

Table 1. three-part questionnaire.

Features (questions)	
1.	Birth date
2.	Education level
3.	Marital status
4.	Occupation
5.	Marriage age
6.	First pregnancy age
7.	Children number
8.	Breastfeeding
9.	Breastfeeding duration
10.	Menopause
11.	Menopause age
12.	Income level
13.	Evaluation of income status
14.	Insurance status
15.	Breast problem background
16.	Breast problem type
17.	Breast cancer history in family
18.	Having mammography or sonography experience
19.	Having information about breast cancer and early detection methods (yes, no)
20.	Type of information resource about breast cancer screening (radio, journal, Tv, books, internet, family, friends, physician, responsive auxiliary phone, health center)
21.	Breast cancer screening methods (mammography, breast self-examination (BSE), clinical breast examination (CBE), sonography, pap smear)
22.	Target population for breast cancer screening (health women, suspected women, women with breast cancer)
23.	Target age group for mammography screening (20-40 years, > 40 years, any age group)
24.	Alternative time for BSE (monthly, every 6 months, yearly, every 3 years)
25.	Beginning age of BSE (>=20 years, 20-40 years, >40 years)
26.	Alternative time for mammography (monthly, every 6 months, yearly, every 3 years)
27.	Having mammography experience (yes/no)
28.	Mammography numbers
29.	Mammography status (precontemplation, contemplation, action, maintenance, relapse)
30.	Having BSE experience (yes/no)
31.	BSE status (precontemplation, contemplation, preparation, action, maintenance, relapse)
32.	Having CBE experience (yes/no)
33.	Alternative time for CBE
34.	CBE status (precontemplation, contemplation, action, maintenance, relapse)
35.	Having pap smear experience (yes/no)
36.	Pap smear numbers
37.	Pap smear status (precontemplation, contemplation, action, maintenance, relapse)
38.	Tendency to participate in educational class

This model assert that individuals move through stages of change in precontemplation, contemplation, preparation, action, maintenance, and relapse [10].

Now a days, a lot of digital data are collected which are related to the health fields such as prevention, treatment, and education areas. The intelligence algorithms used to classify, knowledge extraction and pattern discovery are effective and important steps for developing affiliate aims. These algorithms in the least possible time and resources gain the best results with an acceptable precision.

In this paper, we work on a data-set that was collected from 2256 questioners based on stages of change model. Women have filled out these in 2016. Our goals are data classification into 5 groups, as mentioned above, and extraction of important features that are effective in mammography screening behavior. We use Whale Optimization Algorithm (WOA) and Support-vector machine (SVM) algorithms.

The rest of this paper is organized as follows. In the next section, we introduce our data-set, basic concept of applied algorithms, and proposed method. In Section 3, the experimental results of our proposed method are shown. Finally, Section 4 concludes the research consequences.

2. Materials and methods

2.1. Data-set

Our data-set was collected from a cross-sectional analytic descriptive study. In this work, 2256 women personnel of governmental organizations were investigated in 2016-2017. The data-gathering tool was a three-part questionnaire: sociodemographic characteristics, knowledge about breast cancer screening methods and women's performance and stage of change regarding screening behaviors of mammography. This questionnaire is shown in table 1.
We used the data-set with 2256 records and 53 features, which were extracted from questions in questionnaire because some of the questions had more than one value. The extracted statistical information from data set is shown in table 2. Also, see the mammography screening behavior stage of statistical change in table 3. Each woman was grouped in one of five classes (precontemplation, contemplation, action, maintenance, relapse).

2-2. Whale Optimization Algorithm (WOA)

WOA was proposed by Seyedali Mirjalili to perform optimization. WOA mimics the social behavior of humpback whales. The algorithm simulates two phases of hunt behaviors of humpback whales, which are the exploitation phase and the exploration phase. The first phase is encircling a prey and spiral bubble-net attacking method (exploitation phase) and the second phase is searching randomly for a prey (exploration phase). In the following, we introduce the mentioned steps of algorithm in details from Mirjalili et al [11].

Table 2. Frequency Distribution of Demographic Variables.

Variables		Number	%
Education Level	Elementary	32	1.4
	High school	42	1.9
	Diploma	216	9.6
	University	1966	87.1
Marital Status	Married	1996	88.5
	Single	225	10
	Divorced& Widow	35	1.5
Menopause	Yes	125	5.5
	No	2131	94.5
Income Level	Good	528	23.4
	Moderate	1526	67.6
	Weak	2.2	9

Table 3. Frequency Distribution of breast cancer screening behavior stage of change.

Screening Behavior / Stage of change	Mammography N (%)
Precontemplation	1458 (64.6)
Contemplation	568 (25.2)
Action	154 (6.8)
Maintenance	58 (2.6)
Relapse	18 (0.8)
Total	2256 (100)

2.2.1. Encircling prey

Humpback whales can actually find the position of the prey and then hunt it in a shrinking circle way. Mathematical model of the movement of a whale around a prey is as follow:

$$\vec{D}(t) = |\vec{C}.\vec{X}^*(t) - \vec{X}(t)| \tag{1}$$

$$\vec{X}(t+1) = \vec{X}^*(t) - \vec{A}.\vec{D}(t) \tag{2}$$

where, t displays the current iteration, $\vec{D}(t)$ is a distance vector between $\vec{X}(t)$ and $\vec{X}^*(t)$ in the tth iteration $\vec{X}(t)$ is the current solution (current wale position), $\vec{X}^*(t)$ displays the best solution (prey position) acquired so far, | | is the absolute value, and • is an element-by-element multiplication. The solutions update their positions according to the position of the best-known solution using (2). $\vec{A}$ is a random value in the range of [−a, a], and $\vec{C}$ is a coefficient, obtained as follow:

$$\vec{A} = 2\vec{a}.\vec{r} - \vec{a} \tag{3}$$

$$\vec{C} = 2.\vec{r} \tag{4}$$

In these equations, $\vec{a}$ is a value in [0 2] that reduces linearly in each iteration and $\vec{r}$ is a random vector in the range of [0 1].

2.2.2. Bubble-net attacking

This phase consists of two approaches that update the position of whales, which are equally applied.

1- Shrinking encircling mechanism*:* this behavior is gained in (2,3) where $\vec{a}$ decreases linearly from 2 to 0 and $\vec{r}$ is a random vector in [0, 1]; the new position of search agent could be updated anywhere between the current position of the search agent and the position of the best search agent.

2- Spiral updating position: a spiral equation is then constructed between the positions of whale and prey to simulate the helix-shaped movement of the humpback whales.

$$\vec{X}(t+1) = \vec{D}'(t).e^{bl}.cos(2\pi l) + \vec{X}^*(t) \qquad (5)$$

where, $\vec{D}'(t) = |\vec{X}^*(t) - \vec{X}(t)|$ and displays the distance of the whale from the prey in the t[th] iteration, b is a constant for determining the shape of the logarithmic spiral; $b=1$ was used in our model, l is a random number in [−1, 1], and • is an element-by-element multiplication.

Humpback whale movement is around the prey within a shrinking circle and along a spiral-shaped path simultaneously. Therefore, we presume to update the position of whales, where there is 50% probability to choose between either the shrinking encircling or the spiral, as follows:

$$\vec{X}(t+1) = \begin{cases} \vec{X}^*(t) - \vec{A}.\vec{D}(t), & p < 0.5 \\ \vec{D}'(t).e^{bl}.cos(2\pi l) + \vec{X}^*(t), & p \geq 0.5 \end{cases} \qquad (6)$$

where, p is a random value in [0, 1].

2.2.3. Search for prey

This approach is *exploration phase that* is based on the variation in the $\vec{A}$ vector that can be utilized. The search process for humpback whales is randomly. Therefore, we used $\vec{A}$ with the random values greater than 1 or less than −1 to force search agent to move far away from a reference whale. This mathematical approach can be represented as follows:

$$\vec{D}'(t) = |\vec{C}.\vec{X}_{rand}(t) - \vec{X}(t)| \qquad (7)$$

$$\vec{X}(t+1) = \vec{X}_{rand}(t) - \vec{A}.\vec{D}(t) \qquad (8)$$

where, $\vec{X}_{rand}(t)$ is a random position vector for a random whale. In the following, figure 1 shows the pseudo-code of the WOA algorithm [11,12].

```
Initialize the whales population Xi (i = 1, 2, …, n) and number of iterations.
Calculate the fitness of each search agent.
 X*=the best search agent
while (t< maximum number of iterations)
    for each search agent
 Update a, A, C
    if1 (p < 0.5)
      if2 (| A |< 1)
          Update the position of the current search agent by Eq. 2
      else if2 (| A | >= 2)
                 Select a random search agent
                 Update the position of the current search agent by Eq. 8
          end if2
  else if1 (p >= 0.5)
      Update the position of the current search by Eq. 5
  end if1
        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*
```

Figure 1. pseudo-code of WOA[11]

2.3. Synthetic Minority Over-sampling Technique (SMOTE)

Sometimes, data-set class- member is not balanced. In this situation, mining algorithms could not work as well as possible. A lot of useful mechanisms are invented to amend this limitation. SMOTE is one of them. SMOTE adds new random samples to the minority class. This process benefits from KNN algorithm on the original data-set. New generated samples are created by attribute values that are in the data-set [15].

2.4. Support Vector Machine (SVM)

SVM was developed by Vladimir Vapnik who is known as a non-linear binary classifier. SVM is a supervised learning algorithm that is used for both classification and regression. SVM creates linear separating hyperplanes in high-dimensional vector spaces[16].

The best hyperplane has a maximum distance from the margin. Margin is the closest distance to the training data points. The instance in data space whose distance from the hyperplane is equal to the margin is named Support Vector (SV). Various forms of kernel function are [16]:

$$
\begin{aligned}
&Linearkernel: K(x_i, x_j) = x_i^T x_j \qquad (11)\\
&Polynomialkernel: K(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \gamma > 0\\
&RBFkernel: K(x_i, x_j) = exp(-\gamma \| x_i - x_j \|^2), \gamma > 0\\
&SigmoidKernel: K(x_i, x_j) = tanh(\gamma x_i^T x_j + r)
\end{aligned}
$$

$K(x_i, x_j)$ is the kernel function that is calculated from the inner product of the two instance vectors x_i and x_j. γ is a kernel parameter. The inner product of x_i map space is to the feature space ϕ and result of mapping is shown by $\phi(x_i)$ [16].

2.5. Features selection

On one hand, the experts in various areas, especially health experts, make effort to figure out dependency or independency of specific features in their research scopes. Thus, they have collected a lot of features in a certain subject, and afterwards, they find out the probable dependency or independency between features and their subject using different techniques like analytical techniques.

on the other hand, from the perspective of intelligence algorithms, dimension reduction and feature selection are approaches to reduce space problems and cause of time and complexity reduction. According to whether the grouped data are available or not, these problems are categorized into 3 types: supervised, unsupervised, and semi-supervised. The general aim is feature selection choice feature subsets with utmost effect on record classification [17].

2.6. Proposed method

In this section, we investigated and identified the effective factors in the mammography screening behavior with H-BwoaSvm model. We used BWOA for the feature selection and SVM for classification in the mammography screening behavior problem. This process was performed on the data-set as mentioned in Section 2-1.

Optimization binary algorithms are applied to the feature selection solutions. In these types of algorithms, attribute values are bounded to 0 and 1; 0 demonstrate an unselected attribute and 1 indicates a selected attribute. Our proposed method uses BWOA for feature selection, and also the SVM accuracy value is used as the evaluation cost in the fitness function.

First of all, normalization preprocess is employed on the data-set. After that, hybrid algorithm of BWOA and SVM are appraised on normalized data set. In the following, we explain the proposed method.

1. <u>Preprocess</u>:
   a) Normalization: Normalization process is defined as follows:

$$X = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \qquad (12)$$

   where, x is the feature value that is normalized to X. $x_{\min}$ and $x_{\max}$ are the minimum and maximum values of each feature.

   b) Balancing: If the data set is imbalanced, data-set is balanced by the SMOTE method.
   c) Changing normalized data to binary data: Normalized data set values are grouped with the 0 and 1 labels.

2- <u>Feature selection and classification</u>:
   BWOA is used for feature selection. In BWOA, a used SVM classification algorithm is used in the fitness function. SVM accuracy value is a criterion.

Figure 2 shows the model diagram. Preprocessing steps are prepared for feature selection. As shown in figure 3, H-BWoaSvm model operates on the preprocessed data-set (to be normalized, balanced, and binary). BWOA classifies the input data with association of SVM cost function.

3. Experiment and results:

In order to evaluate the performance of the proposed method, various evaluation metrics were used. They were accuracy, precision, and recall. We used evaluation metrics as:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}*100 \tag{13}$$

$$Precision = \frac{TP}{TP+FP}*100 \tag{14}$$

$$Recall = \frac{TP}{TP+FN}*100 \tag{15}$$

where, TP and FN are the number of items that are correctly classified. TN and FP are the number of items that are incorrectly classified.
The experiments were run on a system with an Intel Core i7-7700 HQ 2.80 GHz processor and 16 GB memory on Wndows10. The proposed method was coded by the Matlab R2017b software.

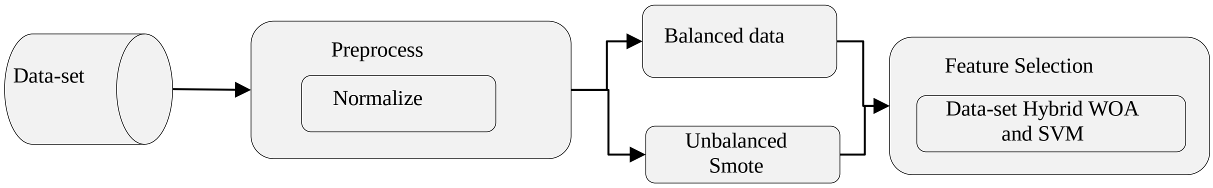

Figure 2. Proposed block diagram for feature selection algorithm.

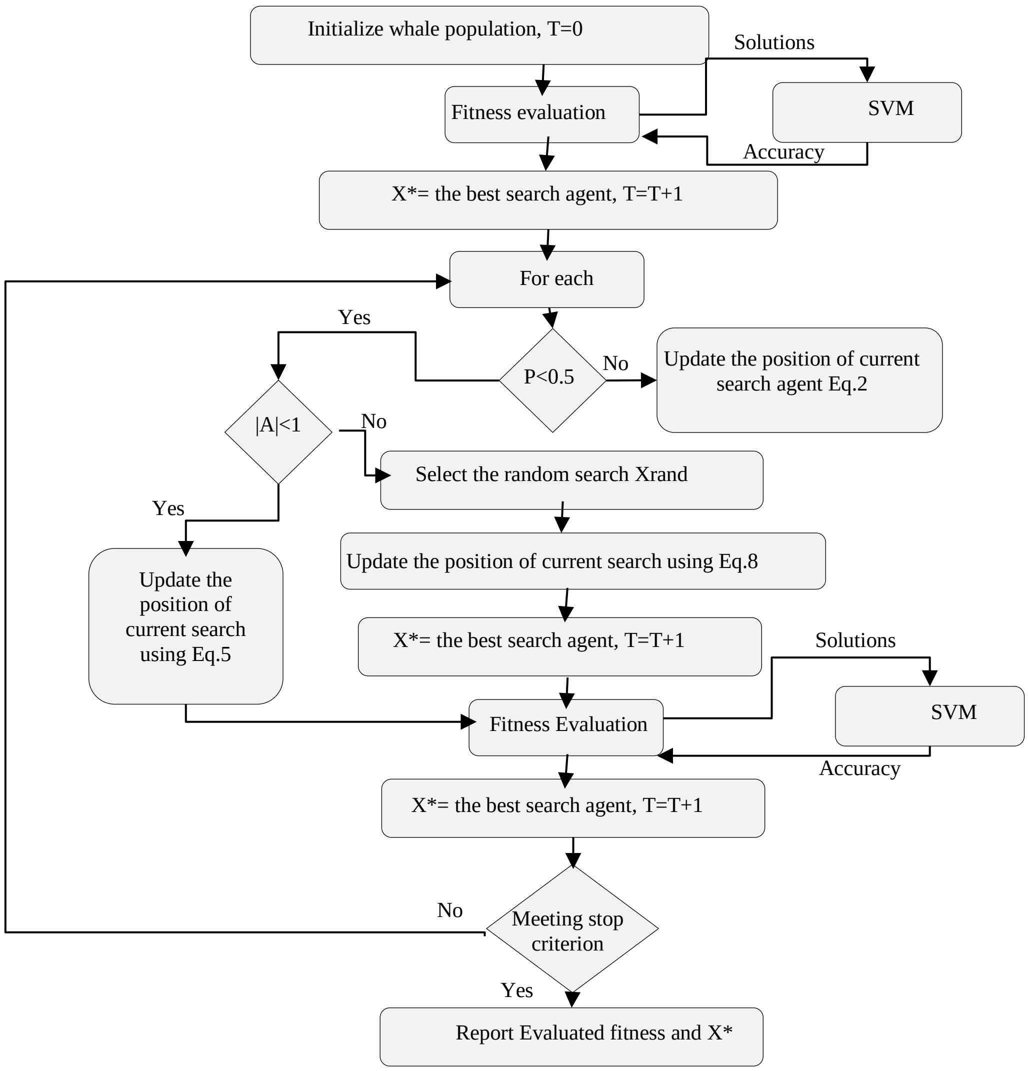

Figure 3. Flowchart of H-BWoaSvm algorithm for feature selection.

The values of parameters of H-BwoaSvm were set as follow: population size = 10, 15 and maximum iteration = 10, 20, 30. SVM algorithm was done in 10-fold cross-validation. As regard, we used 10-fold cross-validation, and 90% of data-set was used for train and 10% for the test data-set.
Two different types of binary data-set were made ready from the original data-set for algorithm normalized data-set that includes women who have not yet done a mammogram test (group 0) and who have done at least one mammogram test (group 1). Type 2 is a set of records that include women who have not decided to do a mammogram test next year (group 0) and women who have decided to do a mammogram test at next year (group 1).
Table 4 shows the simulation result of H-BwoaSvm on type 1 set. The best obtained accuracy was 82.27%. This accuracy value was obtained on all 3 different iterations with 3 different agents. The result of feature selection in our proposed method was selection of 11 features. In order, these features are type of occupation (q. 4), first pregnancy age (q. 6), income level (q. 12), breast problem background (q. 15), type of information resource about breast cancer screening from radio and friends (q.20-1, q. 20-7), having mammography experience (q. 18), mammography status (q. 29), having BSE experience in home (q. 30), having CBE experience (q. 32), and having experience of pap smear test (q. 35).

Table 4. Simulation result for type 1 data with H-BwoaSvm model.

Iteration	Wale	Accuracy	Precision	Recall
10	10	82.05	82.44	81.55
	15	82.27	82.71	**81.93**
20	10	82.09	**82.98**	81.33
	15	82.18	82.61	81.72
30	10	**82.27**	82.97	81.59
	15	82.27	82.98	81.68

Also you can see in table 5 the simulation result of H-BwoaSvm on type 2 set. Based on table 5, the best accuracy was 98.89 on simulation with 30 iterations.

Table 5. Simulation result for type 2 data with H-BwoaSvm model.

Iteration	Wale	Accuracy	Precision	Recall
10	10	98.77	96.15	100
	15	98.83	96.35	100
20	10	98.83	96.34	100
	15	98.89	96.54	100
30	10	98.83	96.35	100
	15	**98.89**	**96.54**	100

The selected features from these runs, in order, are: age group (q. 1), number of children (q. 7), income status (q. 12), breast problem (q. 15), acquiring mammography awareness from radio, books, journals, internet, and responsive auxiliary phone (q. 20-1, q. 20-4, q. 20-2, q. 20-5, q. 20-9), clinical examination (q. 21-3), time of BSE (q. 25), number of mammogram test (q.28), having clinical examination (q. 32), and status of clinical examination (q. 34).
As mentioned above, the features were selected by H-BwoaSvm model have a significant relationship with the characteristics. Since finding the main efficient factors is an important objective in many studies, we explored a lot of research works in this scope and could not find any work using the intelligent approach for extracting the principal features on mammography screening behavior. Mostly, experts assist in the analytical tools for analysis results. Patel et al. [18] have explored the factors influencing breast cancer screening in low-income African American women. They achieved some factors like marital status and having health insurance. In [19], Melvin et al. have predicted participation in mammography screening of None hispanic/hispanic white/black women just by analytical approaches. Another study in Turkey [20] has revealed women with familial breast cancer history had lower participation in mammography screening. They found the results by logistic regression analysis. As well, in many Iranian research works such as [21-23], analytical techniques have been used to discover mammography screening behavior of population or in other subjects. In this work, we found 11 effective factors on participation or not participation in mammography screening with 82.27% accuracy. Then we found 14 factors on mammography behavior repletion with 98.89% accuracy.

4. Conclusion

We have presented the H-BwoaSvm model for mammography screening behavior classification and feature selection. Our hybrid model was a combination of BWOA and SVM algorithms. We found the effective factors on participating/not participating in mammography screening test. The obtained results showed that our proposed model could detect intelligent patterns for doing educational interventions of breast cancer screening. These results were taken without ordinary statistical analysis. Therefore, it was a good approach for knowledge discovery in this scope. Participants in mammography screening test were screened for the risk of breast cancer for 5 years. They were traced for desiring to take part in screening test and their health status. Due to this fact, after the screening period and completing data gathering, we could use deep learning techniques for better mining. These techniques lead us to improve feature selection and discover the most important factors in the women mammography screening behavior.

References

[1] U.S. Breast Cancer Statistics (2019), Available:https://www.breastcancer.org/symptoms/understand_bc/statistics.

[2] Bray, F., Ferlay, J., Soerjomataram, I., Siegel R., Torre, L., & Jemal, A., (2018) Global Cancer Statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA Cancer J Clin, vol. 68, no. 6, pp. 394-424.

[3] Abachizadeh, K., Moradi Kouchi, A., Ghanbari Motlagh, A., Kousha, A., Shekarriz-Foumani, R., Erfani, A. (2018). Breast Cancer in Iran: levels, Variations and Correlates. Community Health, vol. 5, no. 1, pp. 11–21.

[4] Lee, M., Mariapun, S., Rajaram, N., Teo, S. H., & Yip, C. H., (2017). Performance of a subsidised mammographic screening programme in Malaysia, a middle-income Asian country, BMC Public Health, vol. 17, no. 1, pp. 1–7.

[5] Soerjomataram, I., Louwman, M. W. J., Ribot, J. G., Roukema, J. A., & Coebergh, J. W. W., (2008). An overview of prognostic factors for long-term survivors of breast cancer, Breast Cancer Research and Treatment, vol. 107, no. 3, pp. 309–330.

[6] American Cancer Society, (2017). Breast Cancer Facts & Figures 2017-2018, American Cancer Society, Inc.

[7] Tabar, L., Yen, M. F., Vitak, B., Chen, R. A. Smith, H. H. T. & Duffy, S. W., (2003). Mammography service screening and mortality in breast cancer patients: 20-year follow-up before and after introduction of screening, The Lancet, vol. 361, no. 9367, pp. 1405–1410.

[8] Diab, M. E. & Johnston, M. V., (2004). Relationships between level of disability and receipt of preventive health services, Archives of Physical Medicine and Rehabilitation, vol. 85, no. 5, pp. 749–757.

[9] Moodi, M., Rezaeian, M., Mostafavi, F., & Sharifirad, G.R. (2012). Determinants of mammography screening behavior in Iranian women: A population-based study, Journal of Research in Medical Sciences, vol. 17, no. 8, pp. 750–759.

[10] Prochaska, J. O. & Diclemente, C. C., (2005). The transtheoretical approach, in Handbook of psychotherapy integration, 2nd ed., New York: Oxford University Press, pp. 147–171.

[11] Mirjalili, S., & Lewis, A., (2016). The Whale Optimization Algorithm, Advances in Engineering Software, vol. 95, no. 3, pp. 51–67.

[12] Sai, L. & Huajing, F., (2017), A WOA-based algorithm for parameter optimization of support vector regression and its application to condition prognostics, in 36th Chinese Control Conference, Dalian,China, pp. 7345–7350.

[13] Amine L. M. & Nadjet, K., (2015). A Multi-objective Binary Bat Algorithm. Proceedings of the International Conference on Intelligent Information Processing, Security and Advanced Communication - IPAC '15, Batna, Algeria, pp. 1–5.

[14] Hussien, A. G., Houssein, E. H., & Hassanien, A. E., (2017). A binary whale optimization algorithm with hyperbolic tangent fitness function for feature selection. 2017 Eighth International Conference on Intelligent Computing and Information Systems (ICICIS), Cairo, pp. 166–172.

[15] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P., (2002). SMOTE: Synthetic Minority Over-sampling Technique, Journal of Artificial Intelligence Research, vol. 16, pp. 321–357.

[16] Aljarah, I., Al-Zoubi, A. M., Faris, H., Hassonah, M. A., Mirjalili, S., & Saadeh, H., (2018). Simultaneous Feature Selection and Support Vector Machine Optimization Using the Grasshopper Optimization Algorithm, Cognitive Computation, vol. 10, no. 3, pp. 478–495.

[17] Tang, J., Alelyani, S., & Liu, H., (2014). Feature Selection for Classification: A Review, Data Classification: Algorithms and Applications. pp. 37-64.

[18] Patel K., et al., (2014). Factors Influencing Breast Cancer Screening in Low-Income African Americans in Tennessee, J Community Health, vol. 39, no. 5, pp. 943–950.

[19] Melvin, C. L., Jefferson, M. S., Rice, L. J., Cartmell, K. B., & Halbert, C. H. (2016). Predictors of Participation in Mammography Screening among Non-Hispanic Black, Non-Hispanic White, and Hispanic Women, Front. Public Health, vol. 4, no. 15, pp.1-7.

[20] Kirca, N., Tuzcu, A., & Gozum, S., (2018). Breast Cancer Screening Behaviors of First Degree Relatives of Women Receiving Breast Cancer Treatment and the Affecting Factors, Eur J Breast Health, vol. 14, no. 1, pp. 23–28.

[21] Moodi, M., Rezaeian, M., Mostafavi, F., & Sharifirad G.R., (2012). Determinants of mammography screening behavior in Iranian women: A population-based study, Journal of Research in Medical Sciences, vol. 17, no. 8, pp. 750–759.

[22] Shiryazdi, S. M., Kholasehzadeh, G., Neamatzadeh, H., & Kargar, S., (2014). Health Beliefs and Breast Cancer Screening Behaviors among Iranian

Female Health Workers, Asian Pacific Journal of Cancer Prevention, vol. 15, no. 22, pp. 9817–9822.

[23] Imani, M., & Ghassemian, H. (2015). Feature reduction of hyperspectral images: Discriminant analysis and the first principal component, Journal of AI and Data Mining, vol. 3, no. 1, pp. 1-9.