Neural Processing Letters (2023) 55:3631–3654
https://doi.org/10.1007/s11063-021-10700-w

Breast Cancer Detection Based on Modified Harris Hawks Optimization and Extreme Learning Machine Embedded with Feature Weighting

Feng Jiang[1] · Qiannan Zhu[1] · Tianhai Tian[2]

Accepted: 17 November 2021 / Published online: 23 January 2022

Abstract
Computer-aided diagnosis (CAD) can assist doctors with clinical diagnosis and improve diagnosis accuracy and efficiency further. It is significative and valuable for cancer detection by using machine learning. In this paper, a hybrid model based on optimization algorithm and machine learning with feature weighting is carried out to detect breast cancer. Firstly, to surmount the limitation of nonlinear and imbalanced data distribution, we apply feature weighting (FW) based on K-Means to make benign and malignant samples more separate. Then Particle Swarm Optimization (PSO) is used to enhance searching ability of Harris Hawks Optimization (HHO). Moreover, the HHO optimized by PSO (PHHO) is employed to optimize Extreme Learning Machine (ELM). Finally, in order to verify the availability of our proposed FW-PHHO-ELM model, experiments are implemented on Wisconsin Diagnosis Breast Cancer (WDBC) data set. The results indicate that the proposed model can achieve 98.76%, 97.37% and 99.46% on accuracy, sensitivity and specificity respectively. The comparison results demonstrate that the proposed model outperformed reported benchmark models and existing models on accuracy and sensitivity. Besides, the proposed model could also balance sensitivity and specificity well.

Keywords Breast cancer detection · Feature weighting · Extreme learning machine · Harris hawks optimization · Particle swarm optimization

1 Introduction

According to the report issued by WHO in 2018, cancer caused 9.6 million deaths, out of which 627,000 were caused by breast cancer [1]. As one of the three most common cancers reported worldwide [2], breast cancer seriously treats human lives, with a high mortality rate and difficulty to cure [3]. Therefore, early diagnosis of breast cancer is

✉ Feng Jiang
fjiang@zuel.edu.cn

1 School of Statistics and Mathematics, Zhongnan University of Economics and Law, Wuhan 430073, China

2 School of Mathematical Science, Monash University, Melbourne, Clayton 3800, Australia

important in improving patients' survival rate and quality of life effectively. Histopathological is considered as the gold standard of breast cancer diagnosis in clinical right now [4]. However, the examination conducted under a microscope costs numerous time and energy. Therefore, new technologies and methodologies are in need urgently to improve accuracy and efficiency for breast cancer detection.

At present, artificial intelligence (AI) and machine learning (ML) methods have been widely used in various fields such as price prediction [5–7], air quality prediction [8, 9], facial expression recognition [10], portrait parsing [11] and classification [12, 13] et al. Meanwhile, it is also popular in terms of medical science [14, 15]. For example, Extreme Learning Machine (ELM) has been utilized in lung cancer diagnosis [16], disease detection [17, 18] et al. Machine learning methods are data-driven and can be trained end-to-end, which are more time-saving and more accurate for breast cancer detection compared with manual methods. Therefore, in clinic, it is vital to build and develop an efficient computer-aided diagnosis (CAD) model to improve diagnosis performance.

In recent years, researchers have presented many detection models for breast cancer, which can be grouped into two categories: machine learning method [19–23] and optimization algorithm [24–26]. For example, Patricio et al. [27] applied machine learning models, including random forest (RF), logistic regression (LR) and support vector machine (SVM) into detection of breast cancer. Rahman et al. [28] developed a RBF-based SVM model for early detection of breast cancer, they got an accuracy of 93.9%. To reduce detection variances of individual detection models and increase accuracy, Wang et al. [29] proposed an ensemble model based on SVM of different kernel function for breast cancer detection. To improve the interpretability of ML-based detection model, Wang et al. [30] proposed an improved random forest-based rule extraction (IRFRE) model for breast cancer detection. They obtained 95.09% accuracy on Wisconsin Diagnostic Breast Cancer (WDBC) data set. However, machine learning models are not enough to meet demands for high accuracy because of the imbalance between bias and variances, which are often induced by imbalanced data, parameter settings of algorithm and the indetermination of training set. To surmount the limitation of machine learning models and improve the performance, the hybrid model attracts more attention. In which, feature engineering such as feature weighting (FW) [31] and feature selection (FS) [32] are utilized to overcome the shortcomings of noisy data and imbalanced data. Optimization algorithms, including Particle Swarm Optimization (PSO) [33], Artificial Bee Colony (ABC) [34], Grey Wolf Optimization (GWO) [35], Harris Hawks Optimization (HHO) [36] are applied to increase the stability and accuracy of basic classifier models. For example, Singh et al. [37] proposed a hybrid model based on Ant Lion Optimization (ALO) and Back Propagation Neural Network (BPNN), in which, BPNN was the basic classifier, and ALO was used for feature weighting and parameter determination simultaneously. They finally got 98.37% of accuracy on the WDBC data set.

To the best of our knowledge, HHO and ELM have not been applied to breast cancer detection in the framework of hybrid model. In this paper, a hybrid model, namely FW-PHHO-ELM, is proposed for breast cancer detection based on feature weighting, HHO enhanced by PSO (PHHO) and ELM. In order to surmount the limitation of nonlinear and imbalanced data distribution, we first apply K-Means-based feature weighting to make benign and malignant samples more separate. Continuously, we use PSO to enhance the global searching ability of HHO. Then, in order to increase the detection accuracy and stability, we employ PHHO to optimize ELM. Finally, the optimized ELM is used to output the detection results.

The rest of this paper is organized as follows. Section 2 gives the main theory of methodologies used in this paper, including K-Means feature weighting, HHO, ELM and evaluation

criteria. In Sect. 3, FW-PHHO-ELM learning algorithm is presented, including PHHO, optimized ELM and learning framework. Empirical analysis for breast cancer detection and comparisons are discussed in Sect. 4. Finally, conclusions and future works are given.

2 Methodologies

2.1 Feature Weighting Based on K-Means

In the fields of classification and pattern recognition, the contributions of each feature are different in general. That is to say, some features are important while others are unimportant. Generally speaking, feature selection and feature weighting are two common ways used to improve classification accuracy by preprocessing the original data set. Feature selection is a dimensionality reduction method since it selects higher relevant features out and abandons lower relevant features, then a subset of original data set will be obtained after feature selection. However, the idea of feature weighting is different from feature selection, which gives higher weights to important features and lower weights to less important features, then all the features will be retained after feature weighting. Therefore, feature weighting contains more information than feature selection, for the latter one contains partial features only [37].

For nonlinear and imbalanced data in classification problem, K-Means based feature weighting is an effective method to make samples more separable by reducing variance value within the class and increasing discrimination among classes simultaneously [38]. The processes of feature weighting based on K-Means are as follows. The pseudo code is given as Algorithm 1.

Step 1 Set the number of cluster centers equal to the number of actual classes, then K-Means is utilized to find out the cluster center for each feature.

Step 2 Assigned L2-norm as distance and calculate distances between each sample and corresponding cluster center for all features respectively. Then, figure out the mean value of distances in each class for all features.

Step 3 Divide the mean distance M by corresponding cluster center C to get relative crowding degree. A lower value of relative crowding degree means a lower variance value within class, then a higher weight should be given. Therefore, weighting coefficient η could be designed as reciprocal of relative crowding degree, which is calculated as Eq. (1).

$$\eta = \frac{C}{M} \tag{1}$$

Step 4 Multiply weighting coefficient η by the corresponding feature values in the original data set to obtain a new weighted data set.

Algorithm 1 Feature Weighting Based on K-Means

```
Load original data set A = m×n (m samples, n features)
Assign the number of cluster centers (K) equals the number of actual classes
for each feature in the original data set
    using K-Means to figure out K cluster_centers C(K)
    for i in 1 to n
        for k in 1 to K
            for j in 1 to m
                while sample in class k
                    distance (j, i) = norm (A (j, i) – cluster_centers (k, i))
                end while
            end for
        end for
    end for
    for k in 1 to K
        sum_distance = 0
        while sample in class k
            sum_distance = sum_distance + distance
        end while
        M (k) = sum_distance / corresponding sample size
    end for
    calculate weighting coefficient η by Eq. (1)
    multiply η by the corresponding features to get a weighted data set
end for
Return weighting coefficient η and weighted data set.
```

2.2 Harris Hawks Optimization

Harris Hawks Optimization (HHO) is a nature-inspired optimization algorithm proposed by Heidari et al. in 2019 [39]. The main inspiration of HHO is the smart hunting behavior of hawks. In HHO, we regard rabbits as the prey of hawks. The location vectors of hawks denote all the candidate solutions to a problem, and the location vector of rabbit is considered as the best solution. Three phases during hunting are mimicked: exploration phase, transition from exploration to exploitation phase and exploitation phase.

(a) exploration phase.

In the exploration phase, hawks randomly rest on some locations based on two strategies as shown in Eq. (2) and wait to detect a rabbit.

$$L(t+1) = \begin{cases} L_r(t) - \gamma_1 |L_r(t) - 2\gamma_2 L(t)|, q \geq 0.5 \\ (L_{best}(t) - L_m(t)) - \gamma_3 (LB + \gamma_4 (UB - LB)), q < 0.5 \end{cases} \quad (2)$$

where t is the number of iterations. $\gamma_1, \gamma_2, \gamma_3, \gamma_4$ and q are generated between [0,1] randomly, which are updated in each iteration, q is utilized to decide which strategy will be implemented.$L(t)$ is the current location of hawks. $L(t+1)$ is the hawks' location in the next iteration. $L_r(t)$ is one of the hawks' locations selected from the current population randomly. $L_{best}(t)$ is rabbit location, which represents the best solution. $L_m(t)$ is the average position of the current population, which can be calculated as Eq. (3)

$$L_m(t) = \frac{1}{N}\sum_{i=1}^{N} L_i(t) \tag{3}$$

where N is population size. $L_i(t)$ is the location of i^{th} hawks in t^{th} iteration.

(b) transition from exploration to exploitation phase.

In this phase, assign rabbit energy as E, which can be defined as Eq. (4). When $|E| \geq 1$, exploration phase will be implemented, when $|E| < 1$, turn to exploitation phase.

$$E = 2E_0\left(1 - \frac{t}{T}\right) \tag{4}$$

where E is rabbit energy of current state, E_0 indicates the initial energy of rabbit, t and T represent current and maximum number of iterations respectively.

(c) exploitation phase.

Assign c as the probability of the rabbit escaping successfully. According to the behavior of hawks in chasing and hunting a prey, four strategies are presented to mimic the exploitation phase: soft besiege, hard besiege, soft besiege with progressive rapid dives, hard besiege with progressive rapid dives.When $c \geq 0.5$ and $0.5 \leq |E| < 1$, soft besiege will be implemented. Hunting behavior can be modeled as Eq. (5)-Eq. (7).

$$L(t+1) = \Delta L(t) - E \times |\xi \times L_{best}(t) - L(t)| \tag{5}$$

$$\Delta L(t) = L_{best}(t) - L(t) \tag{6}$$

$$\xi = 2 \times (1 - \gamma_5) \tag{7}$$

where $\Delta L(t)$ indicates the distance between rabbit location and current location of hawks, ξ is a random jump strength of rabbit in the process of escaping, γ_5 is generated between [0,1] randomly.When $c \geq 0.5$ and $|E| < 0.5$, hard besiege works. Hunting behavior can be modeled as Eq. (8).

$$L(t+1) = L_{best}(t) - E \times |\Delta L(t)| \tag{8}$$

When $c < 0.5$ and $0.5 \leq |E| < 1$, turn to soft besiege with progressive rapid dives. Hunting behavior can be modeled as Eq. (9)-Eq. (11).

$$L(t+1) = \begin{cases} H, F(H) < F(L(t)) \\ G, F(G) < F(L(t)) \end{cases} \tag{9}$$

$$H = L_{best}(t) - E \times |\xi \times L_{best}(t) - L(t)| \tag{10}$$

$$G = H + S \times LF(D) \tag{11}$$

where $F(\cdot)$ is fitness function, $LF(\cdot)$ is Levy flight function, D indicates the dimensions of the solution space, S is a random vector of size $1 \times D$.When $c < 0.5$ and $|E| < 0.5$, hard besiege with progressive rapid dives will be implemented. In this strategy, update Eq. (10) into Eq. (12), and hunting behavior can be modeled as Eq. (9), Eq. (11), Eq. (12).

$$H = L_{best}(t) - E \times |\xi \times L_{best}(t) - L_m(t)| \tag{12}$$

2.3 Extreme Learning Machine

Extreme Learning Machine is a single-hidden layer feedforward neural network [40]. It contains an input layer, an output layer and a hidden layer. The connection weights between the input layer and hidden layer are defined as ω, the connection weights between output layer and hidden layer are defined as β, and the thresholds of hidden layer are defined as b. ELM can randomly generate initial weights and thresholds to train a model, and obtain the single optimal solution without parameters' adjustment during training. Therefore, ELM has better learning rate. Specifically, the following steps are contained during training a ELM.

Step 1 Initialize the connection weights ω and thresholds b randomly, as Eq. (13)-Eq. (14).

$$\omega = \begin{bmatrix} \omega_{11} & \omega_{12} & \cdots & \omega_{1n} \\ \omega_{21} & \omega_{22} & \cdots & \omega_{2n} \\ & \vdots & & \\ \omega_{l1} & \omega_{l2} & \cdots & \omega_{ln} \end{bmatrix} \tag{13}$$

$$b = [b_1, b_2, \cdots, b_l]^{'} \tag{14}$$

Step 2 Calculate the output matrix h of hidden layer through Eq. (15), and then ELM can be described as Eq. (16).

$$h(\omega_1, \cdots, \omega_l, b_1, \cdots b_l, x_1, \cdots x_Q) = \begin{bmatrix} g(\omega_1 x_1 + b_1) & g(\omega_2 x_1 + b_2) & \cdots & g(\omega_l x_1 + b_l) \\ g(\omega_1 x_2 + b_1) & g(\omega_2 x_2 + b_2) & \cdots & g(\omega_l x_2 + b_l) \\ & \vdots & & \\ g(\omega_1 x_Q + b_1) & g(\omega_2 x_Q + b_2) & \cdots & g(\omega_l x_Q + b_l) \end{bmatrix} \tag{15}$$

$$h\beta = T^{'} \tag{16}$$

where x is input matrix, $g(x)$ is an active function of hidden layer, T is the label of network. $T^{'}$ denotes the transposition of T.

Step 3 To train a network with minimal error, objective function can be designed as Eq. (17). Then the estimated value of connection weights $\hat{\beta}$ between output layer and hidden layer can be obtained by Eq. (18).

$$\min \left\| h\beta - T^{'} \right\| \tag{17}$$

$$\hat{\beta} = h^{+}T^{'} \tag{18}$$

where h^{+} is the Moore–Penrose inverse of h.

Step 4 Final output can be obtained by Eq. (19).

$$T_{sim} = (h\hat{\beta})^{'} \tag{19}$$

2.4 Evaluation Criteria

In classification models, confusion matrix is a helpful approach to measure performance. In this paper, two classes are contained in the data set: benign and malignant, in which malignant is defined as positive class and benign is defined as negative class. Table 1 illustrates the confusion matrix.

There are four indexes in Table 1, where, a indicates the number of actual malignant samples which are correctly classified, b means the number of actual benign samples which are wrongly classified into malignant class, c shows the number of actual malignant samples which are wrongly classified into benign class, and d represents the number of actual benign samples which are rightly classified. Based on Table 1, other evaluation criteria can be designed as follows.

$$Accuracy(Acc) = \frac{a+d}{a+b+c+d} \times 100\% \tag{20}$$

$$Sensitivity(Sen) = TPR = \frac{a}{a+c} \times 100\% \tag{21}$$

$$Specificity(Spe) = TNR = \frac{d}{b+d} \times 100\% \tag{22}$$

$$FNR = 1 - Sen = \frac{c}{a+c} \times 100\% \tag{23}$$

$$FPR = 1 - Spe = \frac{b}{b+d} \times 100\% \tag{24}$$

Table 1 Confusion matrix

Detection class	Actual class		Total
	Malignant (Positive)	Benign (Negative)	
Malignant (Positive)	a (True Positive)	b (False Positive)	$a+b$
Benign (Negative)	c (False Negative)	d (True Negative)	$c+d$
Total	$a+c$	$b+d$	$a+b+c+d$

$$\begin{aligned} Youden's\,index(\gamma) &= Sen + Spe - 1 \\ &= 1 - (FNR + FPR) \end{aligned} \tag{25}$$

where accuracy is a comprehensive index, which inflects the percentage of all the samples detected correctly. In medical statistics, there are two other important indexes: sensitivity and specificity. Here, sensitivity represents the percentage of malignant samples detected correctly, which is related to missed diagnosis rate (*FNR*). According to Eq. (23), a higher sensitivity indicates a lower missed diagnosis rate. In clinic, missed diagnosis means that the malignant samples are diagnosed as benign samples, which will lead to missing the best time for treatment. In terms of specificity, it refers to the percentage of benign samples which are detected correctly, which is related to misdiagnosis rate (*FPR*). According to Eq. (24), a higher specificity indicates a lower misdiagnosis rate. In clinic, misdiagnosis means that the benign samples are diagnosed as malignant samples, which will lead to needless anxiety. However, sensitivity and specificity are two conflict indexes. In clinical, the improvement of the specificity will be based on the sacrifice of sensitivity. In order to evaluate the model's ability of balancing sensitivity and specificity, we introduce Youden's index, which is a comprehensive index constructed based on sensitivity and specificity as Eq. (25). It can well reflect the model's capacity to distinguish benign and malignant samples.

3 FW-PHHO-ELM Learning Algorithm

3.1 HHO Modified by PSO

In majority swarm intelligence optimization algorithms, the balance between global and local searching ability is essential. In HHO, benefit from four strategies mimicked in stage of exploitation, the capability of finding the best solution in a limited area is promising, which makes the local searching ability excellent. To enhance the global searching ability and improve the overall optimization ability of HHO, inspired by improved pigeon-inspired optimization (IPIO) [8] and memory-based hybrid dragonfly algorithm (MHDA) [41], PSO is employed to optimize the initial population of HHO. A two-stage optimization algorithm, HHO modified by PSO (PHHO), is proposed. In PHHO, PSO is first employed in global search, then based on the best N individuals obtained from PSO, exploration and exploitation are implemented respectively in HHO. PHHO combines the global search ability of PSO in the first stage and local search ability of HHO in the second stage to achieve global optimal solutions. Due to the balance between global searching offered by PSO and local searching offered by HHO, PHHO performs better than parent algorithms. The pseudo code of PHHO is given as Algorithm 2.

Algorithm 2 PHHO algorithm

```
Load parameters of HHO and PSO, and define fitness function
Generate the initial population of PSO
Calculate the fitness of the initial population and assign the best location
t1 = 0
while t1 < T1
    for i = 1 to 2N
        Update velocity vector and location of individual
        Calculate new fitness of the updated individual
        if new fitness is better than the previous one
            Update the best location
        end if
    end for
    t1 = t1 + 1
end while
Initialize population of HHO with the best N individuals in PSO
t2 = 0
while t2 < T2
    for i=1 to N
        Update initial energy E0
        Update escaping energy of rabbit by Eq. (4)
        if c≥0.5 and 0.5≤|E|<1
            Update the location of Harris Hawks according to Eq. (5)- Eq. (7)
        else if c≥0.5 and |E|<0.5
            Update the location of Harris Hawks according to Eq. (8)
        else if c<0.5 and 0.5≤|E|<1
            Update the location of Harris Hawks according to Eq. (9)- Eq. (11)
        else if c<0.5 and |E|<0.5
            Update the location of Harris Hawks according to Eq. (9), Eq. (11), Eq. (12)
        end if
        Calculate new fitness of updated population
        if new fitness is better than the previous one
            Update the best location
        end if
    end for
    t2 = t2 + 1
end while
Return best location and fitness value
```

3.2 Optimized ELM

As a single-hidden layer feedforward neural network (SLFN), ELM is an effective algorithm with fast training speed. In ELM, initial weights and thresholds, as two essential parameters,

are generated randomly. Hence, stability and generalization performance will be influenced. Therefore, PHHO is utilized to reduce randomness and enhance the performance of ELM by optimizing the initial weights and thresholds. Specific steps are as follows.

Step 1 Assign the fitness function of PHHO as Eq. (24).

$$fitness = \frac{1}{2} \times (sensitivity + specificity) \tag{24}$$

Step 2 Generate the initial weights and thresholds of ELM randomly. Train a ELM based on random parameters. Calculate the evaluation criteria of ELM by Eq. (20)-Eq. (22), and obtain the fitness of PHHO by Eq. (24).

Step 3 Optimize ELM by PHHO. Then, train a new ELM based on the optimized weights and thresholds. Further, calculate new fitness by Eq. (24). If new fitness is better than the previous one, an updated ELM is got.

Step 4 Repeat step 3 until maximum iterations. An optimized ELM is obtained.

3.3 Learning Framework

In this paper, the key objective is to construct an efficient breast cancer detection model based on feature weighting, optimization algorithm and machine learning. The proposed novel hybrid model is utilized to increase the accuracy, sensitivity, specificity of the detection of breast cancer. Two main phases are considered in the proposed hybrid model: data preprocessing based on K-Means feature weighting and detection of breast cancer based on PHHO and ELM. Figure 1 illustrates the architecture of the proposed model. Specific processes of each phase are as follows.

Phase 1 Data preprocessing based on feature weighting.

Feature weighting is an important approach in decreasing prediction error and increasing accuracy. High quality feature engineering can reduce computation complexity, risk of

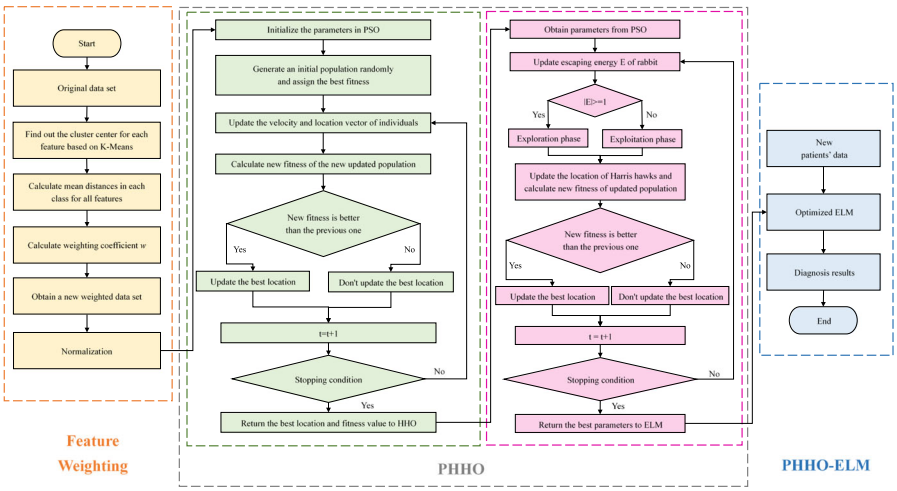

Fig. 1 Learning framework of the proposed FW-PHHO-ELM: Feature weighting based on K-means is applied for data preprocessing. PHHO is a novel optimization algorithm proposed in this paper, which is utilized to optimize ELM. PHHO-ELM is the final optimized model, which is adopted to output detection results

over-fitting, and enhance model's applicability effectively. In traditional CAD system, feature engineering mainly depends on medical workers and experienced engineers, which is subjective and time costing. With the development of machine learning and big data technology, the data-driven approach becomes common. In this phase, K-Means-based feature weighting, an efficient approach in data preprocessing and feature engineering, is applied to make samples more separate.

Phase 2 Detection of breast cancer based on PHHO and ELM.

In this phase, machine learning and optimization algorithms are utilized to train a breast cancer detection model. ELM is a single-hidden layer feedforward neural network. In this paper, ELM is used for the detection of breast cancer. Since the parameters of ELM are generated randomly in each run, which will lead to unstable results. HHO improved by PSO (PHHO) is utilized to optimize the weights and thresholds of ELM, which will increase accuracy of detection and make diagnostic results stable.

4 Empirical Analysis for Breast Cancer Detection

4.1 Data Description

To evaluate the performance of our proposed model for breast cancer detection, several experiments are conducted on WDBC data set (https://archive.ics.uci.edu/ml/datasets). This data set contains 569 samples (357 benign samples and 212 malignant samples) and 32 variables, including ID, diagnosis result and 30 variables describing tumor cell nucleus. These 30 variables describe 10 features of cell nucleus which are tightly associated with the morphological characteristics of the cell nucleus. These 10 features include radius, texture, perimeter, area, smoothness, compactness, concavity, concave points, symmetry and fractal dimension, while each feature has 3 corresponding variables, including mean values, standard deviations and worst values. These 30 variables are numbered as variable 1 to 30. Variables 1 to 10 are the mean values of the 10 features of cell nucleus, one variable representing one feature. Similarly, variables 11 to 20 are the standard deviations of the 10 features and variables 21 to 30 are the worst values of the 10 features. The average and standard deviation of each variable among benign, malignant and all samples are calculated respectively. Table 2 summaries these variables and calculation results.

4.2 Parameter Settings

In stage of detection of breast cancer, we apply optimization algorithm and machine learning to distinguish cancer between benign and malignant. To evaluate the performance of the proposed FW-PHHO-ELM in this paper, several optimization algorithms and machine learning models are chosen as comparison models. In optimization algorithm, HHO, PHHO, Whale Optimization Algorithm (WOA), Artificial Fish Swarm Algorithm (AFSA) are set as comparison methods. In terms of machine learning, ELM, BPNN and Generalized Regression Neural Network (GRNN) are compared. Parameter settings of all the methods used in this paper are shown in Table 3.

In order to evaluate the proposed model synthetically, we take two categories comparison models into consider: single machine learning model and optimization-based hybrid model. First, to verify each module in the hybrid model is effective, some models are compared, including BPNN, GRNN, ELM, FW-ELM, FW-HHO-ELM. Besides, other models based on

Table 2 Descriptive statistics of tumor cell nucleus features in 569 breast tumor patients

Variables ID	Benign samples			Malignant samples			All samples		
	Average	±	Std	Average	±	Std	Average	±	Std
1	12.15	±	1.78	17.46	±	3.20	14.13	±	3.52
2	17.91	±	3.99	21.60	±	3.77	19.29	±	4.30
3	78.08	±	11.79	115.37	±	21.80	91.97	±	24.28
4	462.79	±	134.10	978.38	±	367.07	654.89	±	351.60
5	0.09	±	0.01	0.10	±	0.01	0.10	±	0.01
6	0.08	±	0.03	0.15	±	0.05	0.10	±	0.05
7	0.05	±	0.04	0.16	±	0.07	0.09	±	0.08
8	0.03	±	0.02	0.09	±	0.03	0.05	±	0.04
9	0.17	±	0.02	0.19	±	0.03	0.18	±	0.03
10	0.06	±	0.01	0.06	±	0.01	0.06	±	0.01
11	0.28	±	0.11	0.61	±	0.34	0.41	±	0.28
12	1.22	±	0.59	1.21	±	0.48	1.22	±	0.55
13	2.00	±	0.77	4.32	±	2.56	2.87	±	2.02
14	21.14	±	8.83	72.67	±	61.21	40.34	±	45.45
15	0.01	±	0.00	0.01	±	0.00	0.01	±	0.00
16	0.02	±	0.02	0.03	±	0.02	0.03	±	0.02
17	0.03	±	0.03	0.04	±	0.02	0.03	±	0.03
18	0.01	±	0.01	0.02	±	0.01	0.01	±	0.01
19	0.02	±	0.01	0.02	±	0.01	0.02	±	0.01
20	0.00	±	0.00	0.00	±	0.00	0.00	±	0.00
21	13.38	±	1.98	21.13	±	4.27	16.27	±	4.83
22	23.52	±	5.49	29.32	±	5.42	25.68	±	6.14
23	87.01	±	13.51	141.37	±	29.39	107.26	±	33.57
24	558.90	±	163.37	1422.29	±	596.56	880.58	±	568.86
25	0.12	±	0.02	0.14	±	0.02	0.13	±	0.02
26	0.18	±	0.09	0.37	±	0.17	0.25	±	0.16
27	0.17	±	0.14	0.45	±	0.18	0.27	±	0.21
28	0.07	±	0.04	0.18	±	0.05	0.11	±	0.07
29	0.27	±	0.04	0.32	±	0.07	0.29	±	0.06
30	0.08	±	0.01	0.09	±	0.02	0.08	±	0.02

different optimization algorithms are compared, including FW-WOA-ELM and FW-AFSA-ELM.

4.3 Results Analysis and Comparisons

In this section, we first discuss how feature weighting in the proposed FW-PHHO-ELM benefits the detection results. Then, analyze the superiorities of the proposed FW-PHHO-ELM by comparing it with some benchmark models and existing models.

Table 3 Parameter settings

Processes	Methods	Parameters	Values
Feature weighting	K-Means	number of cluster center	2
Optimization algorithm	HHO/WOA	maximum iterations	100
		size of population	30
	PSO	maximum iterations	100
		size of population	60
		acceleration factors c1	2
		acceleration factors c2	2
		weight coefficient	0.9
	AFSA	maximum iterations	100
		size of population	30
		maximum number of try	10
		visual distance	1
		crowding factor	0.618
		step	0.1
Machine learning	BPNN/GRNN/ELM	number of hidden layer neurons	60

4.3.1 Feature Weighting for Breast Cancer Data

In this paper, variance within the class and discrimination among classes are used to measure the effect of clustering. In order to show the effect of feature weighting based on K-Means, the class distributions of original and weighted data set on WDBC have been given in Fig. 2 to demonstrate and prove how the proposed feature weighting method improves performance. In Fig. 2, the class distributions for both original and weighted data set are according to the first three features: radius, texture, perimeter. Where, Fig. 2a, c, e are corresponding to original data set, Fig. 2b, d, f are corresponding to weighted data set.

According to Fig. 2, in the original data sets, there are many overlaps between benign and malignant data points. After applying the weighting method to the data sets, the class distribution seems to be closer to each other in class and more separate between classes. That is to say, the discrimination between classes has increased and the distance within the class has decreased. Thus high classification performance has been obtained.

4.3.2 Detection Results and Comparisons

To validate the performance of the proposed model, two types of models are selected as benchmark models: machine learning models and optimization models. Machine learning models without optimization algorithms, namely BPNN, GRNN, ELM, FW-ELM, and optimization models, including FW-HHO-ELM, FW-WOA-ELM, FW-AFSA-ELM and FW-PHHO-ELM are compared. After 10 independent experiments on WDBC dataset for each model, the comparison results on accuracy, sensitivity and specificity are given in Table 4. In which, lowest, highest, average and standard deviation values are summarized on three evaluation criteria respectively. Further, the average values with 10 independent experiments on three indicators for each model are given in Fig. 3.

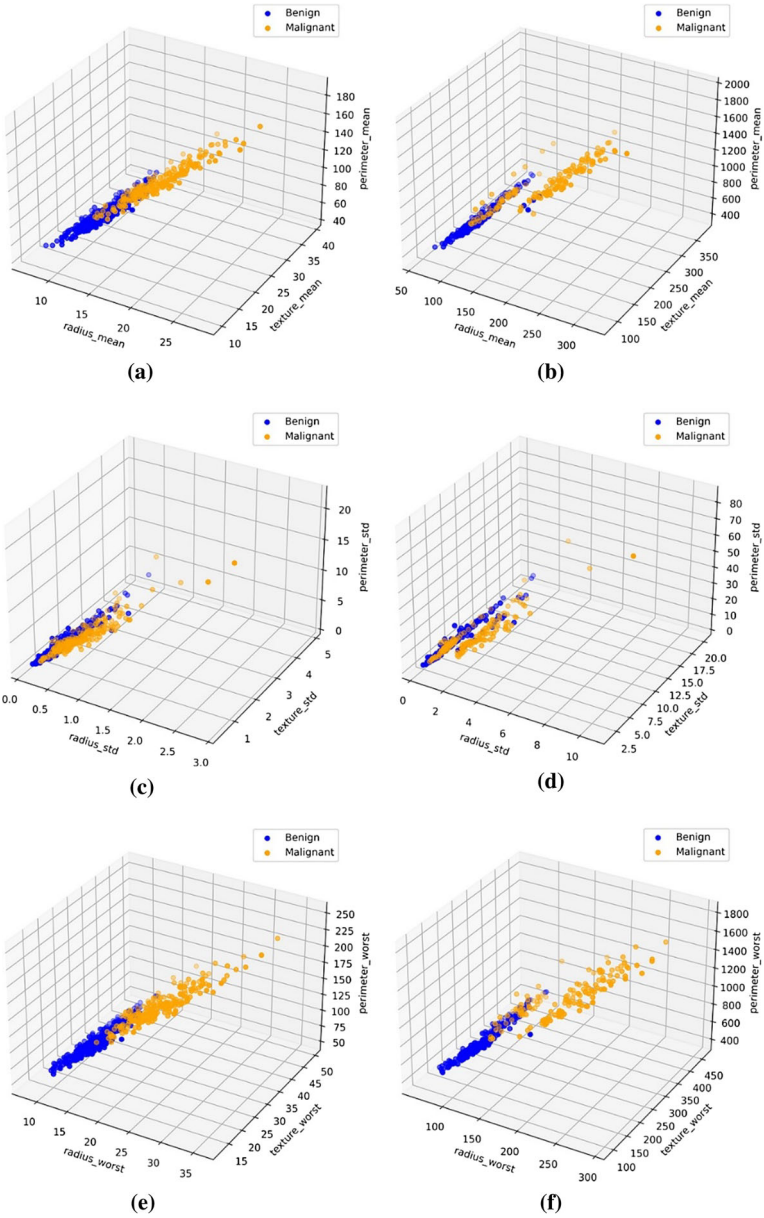

Fig. 2 Classification of both original and weighted data set according to the first three features: radius, texture, perimeter. **a** original data set for mean value, **b** weighted data set for mean value, **c** original data set for standard deviation, **d** weighted data set for standard deviation, **e** original data set for worst value, **f** weighted data set for worst value

Table 4 Comparisons of proposed model with benchmark models

	Models	Lowest (%)	Highest (%)	Average (%)	Standard deviation
accuracy	BPNN	84.0237	94.0828	91.0651	3.0108
	GRNN	89.9408	89.9408	89.9408	0.0000
	ELM	92.8994	95.8580	94.3195	0.9243
	FW-ELM	94.0828	96.4497	95.0888	0.7026
	FW-HHO-ELM	97.0414	98.8166	97.7515	0.4428
	FW-WOA-ELM	96.4497	98.8166	97.8107	0.6509
	FW-AFSA-ELM	95.8580	98.2249	97.0414	0.7001
	FW-PHHO-ELM	98.2249	99.4083	98.7574	0.3187
sensitivity	BPNN	70.4918	95.0820	82.9508	7.5196
	GRNN	86.8852	86.8852	86.8852	0.0000
	ELM	85.2459	90.1639	87.8689	1.8255
	FW-ELM	85.9649	91.2281	88.7719	1.4035
	FW-HHO-ELM	94.7368	98.2456	96.3158	1.4573
	FW-WOA-ELM	94.7368	96.4912	96.1404	0.7018
	FW-AFSA-ELM	92.9825	98.2456	95.9649	1.3702
	FW-PHHO-ELM	96.4912	98.2456	97.3684	0.8772
specificity	BPNN	89.8148	99.0741	95.6481	3.0160
	GRNN	91.6667	91.6667	91.6667	0.0000
	ELM	96.2963	100.0000	97.9630	1.1565
	FW-ELM	96.4286	100.0000	98.3036	1.1607
	FW-HHO-ELM	97.3214	100.0000	98.4821	0.8036
	FW-WOA-ELM	96.4286	100.0000	98.6607	1.1469
	FW-AFSA-ELM	95.5357	99.1071	97.5893	1.2012
	FW-PHHO-ELM	99.1071	100.0000	99.4643	0.4374

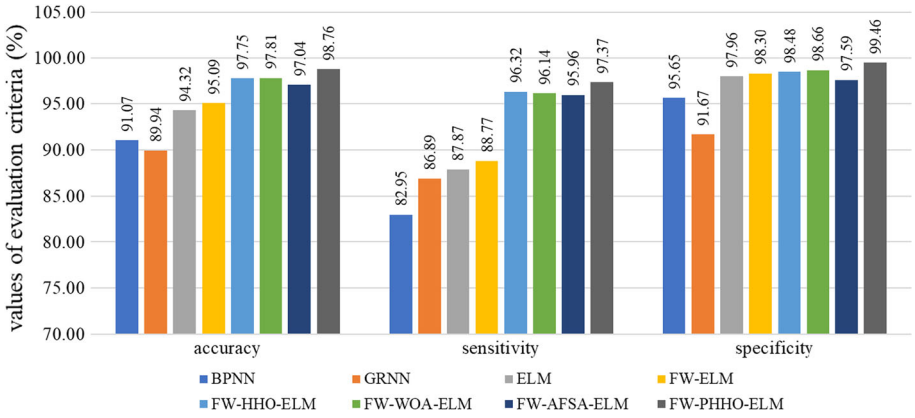

Fig. 3 Average values of three indicators on proposed model and benchmark models

According to Table 4 and Fig. 3, we can draw the following conclusions. First of all, the proposed FW-PHHO-ELM has better classification ability than all the other benchmark models on three indicators. From Table 4, the lowest, highest and average accuracy on the FW-PHHO-ELM model is 98.2249, 99.4083, 98.7574% respectively, which is higher than all the other benchmark models. In particular, the proposed FW-PHHO-ELM can improve up to 9.80% on average accuracy rate in comparison to GRNN (89.9408%). In terms of sensitivity, the lowest and average values on the FW-PHHO-ELM model are higher than other models. For instance, in comparison to BPNN, the proposed FW-PHHO-ELM improves lowest sensitivity from 70.4918 to 96.4912%, and average sensitivity from 82.9508 to 97.3684%. The corresponding improvement rates are 36.88 and 17.38%. For the highest sensitivities of the proposed FW-PHHO-ELM, FW-HHO-ELM and FW-AFSA-ELM can achieve 98.2456%, and the equal values indicate that these three models can detect the same number of malignant samples among the truly malignant samples. Similarly, the lowest and average values of proposed FW-PHHO-ELM are better than other benchmark models. The highest specificity of the proposed FW-PHHO-ELM, ELM, FW-ELM, FW-HHO-ELM and FW-WOA-ELM can achieve 100.0000%. These same values mean that these five models can detect all the benign samples correctly in the best cases. To sum up, the proposed FW-PHHO-ELM has better performance on breast cancer detection compared with other benchmark models.

In addition, feature weighting and optimization algorithms can effectively enhance the performance of detection models. According to Table 4, FW-ELM achieves 95.0888, 88.7719, 98.3036% on average accuracy, sensitivity, specificity respectively, which is better than that of ELM (94.3195, 87.8689, 97.9630%). It indicates that feature weighting is an efficient approach to improving the performance of breast cancer detection. From Fig. 3, all the hybrid models with optimization algorithm have better performance than the machine learning models without optimization algorithm. Especially, in terms of sensitivity, the latter four models with optimization algorithm obtain more than 95% on sensitivity while the former four models have not reached 89%. Particularly, in different optimization models, FW-PHHO-ELM achieves the best performance in terms of accuracy, sensitivity and specificity. These results show that optimization algorithms can improve the performance of breast cancer detection, and PHHO is a more effective optimizer than other reported optimizers.

Moreover, the stability of the proposed model is superior to most benchmark models. From Table 4, GRNN achieves the best stability for the standard deviation equal 0.0000 in 10 runs, that is to say, the classification results are the same for every experiment. Except GRNN, FW-PHHO-ELM achieves the lowest standard deviation in terms of accuracy and specificity, which is 0.3187 and 0.4374 specifically. In the aspect of sensitivity, the standard deviation of FW-PHHO-ELM (0.8772) ranks third only to GRNN (0.0000) and FW-WOA-ELM (0.7018). Therefore, FW-PHHO-ELM performs better than most benchmark models in terms of standard deviation, and achieves superior stability.

Further, to verify the statistical significance between the proposed FW-PHHO-ELM and comparison models, independent samples T-test is employed. The null hypothesis of T-test in this case is that there is no significant difference in classification ability between the benchmark model and the tested model, and the alternative hypothesis is that the difference of classification ability between the benchmark model and the tested model is significant. In aspect of accuracy, specificity and sensitivity, T-test results on benchmark models and tested models are given in Tables 5, 6 and 7 respectively.

In Tables 5, 6 and 7, the values in [] are standard deviation of corresponding evaluation indicators and the numbers above [] are average values. Similarly, the symbols in () indicate the levels of significance and the values above () are p-values of T-test.

Table 5 Results of T-test for accuracy on proposed model and benchmark models

Benchmark models	Tested models							
	BPNN	GRNN	ELM	FW-ELM	FW-HHO-ELM	FW-WOA-ELM	FW-AFSA-ELM	FW-PHHO-ELM
BPNN	91.0651%	0.2773	0.0062	0.0010	0.0000	0.0000	0.0000	0.0000
	[3.0108]	(·)	(***)	(***)	(***)	(***)	(***)	(***)
GRNN		89.9408%	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
		[0.0000]	(***)	(***)	(***)	(***)	(***)	(***)
ELM			94.3195%	0.0623	0.0000	0.0000	0.0000	0.0000
			[0.9243]	(*)	(***)	(***)	(***)	(***)
FW-ELM				95.0888%	0.0000	0.0000	0.0000	0.0000
				[0.7026]	(***)	(***)	(***)	(***)
FW-HHO-ELM					97.7515%	0.8241	0.0192	0.0000
					[0.4428]	(·)	(**)	(***)
FW-WOA-ELM						97.8107%	0.0266	0.0010
						[0.6509]	(**)	(***)
FW-AFSA-ELM							97.0414%	0.0000
							[0.7001]	(***)
FW-PHHO-ELM								98.7574%
								[0.3187]

Remark: ***, **, * and · represent a significant level of 0.01, 0.05, 0.1 and not significant respectively

Table 6 Results of T-test for specificity on proposed model and benchmark models

Benchmark models	Tested models							
	BPNN	GRNN	ELM	FW-ELM	FW-HHO-ELM	FW-WOA-ELM	FW-AFSA-ELM	FW-PHHO-ELM
BPNN	95.6481% [3.0160]	0.0009 (***)	0.0454 (**)	0.0240 (**)	0.0139 (**)	0.0118 (**)	0.0897 (*)	0.0014 (***)
GRNN		91.6667% [0.0000]	0.0000 (***)	0.0000 (***)	0.0000 (***)	0.0000 (***)	0.0000 (***)	0.0000 (***)
ELM			97.9630% [1.1565]	0.5407 (.)	0.2833 (.)	0.2150 (.)	0.5099 (.)	0.0019 (***)
FW-ELM				98.3036% [1.1607]	0.7088 (.)	0.5197 (.)	0.2158 (.)	0.0117 (**)
FW-HHO-ELM					98.4821% [0.8036]	0.7065 (.)	0.0803 (*)	0.0047 (***)
FW-WOA-ELM						98.6607% [1.1469]	0.0688 (*)	0.0652 (*)
FW-AFSA-ELM							97.5893% [1.2012]	0.0003 (***)
FW-PHHO-ELM								99.4643% [0.4374]

Table 7 Results of T-test for sensitivity on proposed model and benchmark models

Benchmark models	Tested models							
	BPNN	GRNN	ELM	FW-ELM	FW-HHO-ELM	FW-WOA-ELM	FW-AFSA-ELM	FW-PHHO-ELM
BPNN	82.9508%	0.1339	0.0726	0.0348	0.0001	0.0001	0.0001	0.0000
	[7.5196]	(.)	(*)	(**)	(***)	(***)	(***)	(***)
GRNN		86.8852%	0.1234	0.0008	0.0000	0.0000	0.0000	0.0000
		[0.0000]	(.)	(***)	(***)	(***)	(***)	(***)
ELM			87.8689%	0.2547	0.0000	0.0000	0.0000	0.0000
			[1.8255]	(.)	(***)	(***)	(***)	(***)
FW-ELM				88.7719%	0.0000	0.0000	0.0000	0.0000
				[1.4035]	(***)	(***)	(***)	(***)
FW-HHO-ELM					96.3158%	0.7486	0.6051	0.0798
					[1.4573]	(.)	(.)	(*)
FW-WOA-ELM						96.1404%	0.7364	0.0042
						[0.7018]	(.)	(***)
FW-AFSA-ELM							95.9649%	0.0186
							[1.3702]	(**)
FW-PHHO-ELM								97.3684%
								[0.8772]

According to Tables 5, 6 and 7, several conclusions can be summarized as follows. First of all, each module of the proposed hybrid model is indispensable. ELM outperforms GRNN and BPNN on accuracy and specificity at a significant level of 0.05. As for sensitivity, it is 82.9508, 86.8852 and 87.8689% on BPNN, GRNN and ELM respectively. *P*-values are 0.1234 between ELM and GRNN, and 0.0726 between ELM and BPNN, which are not significant at level of 0.05 because these two values are greater than 0.05. That is to say ELM is better or close than GRNN and BPNN on sensitivity, while the standard deviation is smaller than BPNN. Thus, we can say that it is suitable to select ELM as a basic classifier for optimization. In addition, all the optimized models are significantly superior to single machine learning models without optimization on accuracy and sensitivity. Interestingly, as for specificity, we find that there is no significant difference between FW-HHO-ELM and ELM, FW-HHO-ELM and FW-ELM, FW-WOA-ELM and ELM, FW-WOA-ELM and FW-ELM, FW-AFSA-ELM and ELM, FW-AFSA-ELM and FW-ELM. One possible reason is that the specificity of basic models has a high value already. As can be seen from Table 6, all the specificities have achieved more than 97.5% on ELM and FW-ELM, thus, it is tough to improve further. In summary, we can say that hybrid model with optimization algorithm outperforms single machine learning models without optimization in most cases and optimization algorithm is an effective tool for improving performance.

From Table 5, when FW-PHHO-ELM is set as a tested model, all the p-values between test model and benchmark models are less than 0.05 on accuracy. The results show that FW-PHHO-ELM statistically outperforms all the benchmark models utilized for breast cancer detection. Furthermore, the accuracy of FW-PHHO-ELM is 98.7574%, the classification accuracy has been increased by 8.82% compared with the worst model GRNN (89.9408%). It is worth noticing that accuracy is a comprehensive indicator for classification model. In summary, FW-PHHO-ELM significantly increases the classification accuracy and it is an effective model for breast cancer detection.

In terms of specificity. All the p-values are smaller than 0.05 between FW-PHHO-ELM and benchmark models except FW-WOA-ELM. From Table 6, *p*-value of *T*-test between FW-PHHO-ELM and FW-WOA-ELM is 0.0652, which is greater than 0.05 but smaller than 0.1. In fact, a significant level of 0.1 is also acceptable in practice. The specificity of FW-PHHO-ELM and FW-WOA-ELM are 99.4643 and 98.6607% respectively. Thus, we can say that FW-PHHO-ELM achieves higher or close specificity compared to FW-WOA-ELM, and achieves significantly higher specificity compared to other benchmark models.

Similar conclusions can be drawn in the aspect of sensitivity. According to Table 7, all the p-values are smaller than 0.05 between FW-PHHO-ELM and benchmark models except FW-HHO-ELM. *P*-value of *T*-test between FW-PHHO-ELM and FW-HHO-ELM is 0.0798, which is greater than 0.05 but smaller than 0.1. The sensitivity of FW-PHHO-ELM and FW-HHO-ELM are 97.3684 and 96.3158% respectively. Therefore, we can draw conclusions that the sensitivity of FW-PHHO-ELM is similar to that of FW-HHO-ELM and significantly higher to those of other benchmark models. It is worth to noticing that, in clinical practice, it is vital to increase sensitivity for the purpose of making sure the patients with malignant tumors could be detected correctly as much as possible. Thus, the increase of sensitivity is important in clinical practice.

4.3.3 Discussions

To further evaluate the performance of the proposed FW-PHHO-ELM, some works proposed in [29–32, 34, 35, 37] are compared with FW-PHHO-ELM on accuracy, sensitivity, specificity and Youden's index.

According to Table 8, it can be seen that our proposed method outperforms existing methods in accuracy. In clinical diagnosis tests, accuracy is a comprehensive evaluation index, it indicates the degree to which the diagnosis class is consistent with the actual class. Our method achieves 98.76% on accuracy which is higher than the highest accuracy 98.37% achieved in reference [37], that is to say, our model achieves better comprehensive performance.

As for sensitivity, which is also called true positive rate (*TPR*), it refers to the percentage of malignant samples which are detected correctly. According to Eq. (23), a higher sensitivity indicates a lower missed diagnosis rate (*FNR*). In clinic, missed diagnosis means that the malignant samples are diagnosed as benign samples, which will lead to missing the best time for treatment. Thus, a high sensitivity is essential. As can be seen from Table 8, our model achieves 97.37% on sensitivity which is higher than the highest sensitivity 96.43% achieved in reference [37]. Therefore, our model achieves a lower missed diagnosis rate compared with reported models.

In terms of specificity, which is also called true negative rate (*TNR*), it refers to the percentage of benign samples which are detected correctly. According to Eq. (24), a higher specificity indicates a lower misdiagnosis rate (*FPR*). In clinic, misdiagnosis means that the benign samples are diagnosed as malignant samples, which will lead to needless anxiety. According to Table 8, our model is superior to references [30, 32], but lower than references [29, 31, 37] on specificity. It should be noticed that these references with high specificities have attained lower sensitivities, especially the model in reference [31]. It achieves 100% on specificity but 84.00% on sensitivity. The gap between these two indexes is too large. Because of the conflict of sensitivity and specificity, we can say that the specificity in references [31] is high with a sacrifice of sensitivity. However, an efficient model utilized in clinical practice should have the ability to balance sensitivity and specificity on data set well. Youden's index

Table 8 Comparisons of proposed model with existing models in references

References	Models	Accuracy (%)	Sensitivity (%)	Specificity (%)	Youden's index (%)
Our proposed model	FW-PHHO-ELM	98.76	97.37	99.46	96.83
Pratheep et al. (2021) [31]	FW-BOA-RDF	–	84.00	100	84.00
Singh et al. (2020) [37]	FW-ALO-BPNN	98.37	96.43	99.52	95.95
Wang et al. (2020) [30]	IRFRE	95.09	93.40	96.09	89.49
Abdel-Basset et al. (2020) [35]	FS-TMGWO-KNN	94.82	–	–	–
Naik et al. (2019) [32]	FS-BBA-OGCNN	93.54	90.76	95.24	86.00
Rao et al. (2019) [34]	FS-ABC-GBDT	92.80	–	–	–
Wang et al. (2018) [29]	SVM-Ensemble	97.68	94.75	99.49	94.24

Remark: Lacking data in original literature is represented by –

is a comprehensive indicator to evaluate the ability to balance sensitivity and specificity, which can infect the capacity to distinguish benign and malignant samples. According to Table 8, we can see that our model achieves a higher Youden's index than that of references [29, 31, 37]. In addition, the specificity of references [29, 31, 37] and our model all reaches more than 99.45%, which is a high level. Overall, our method performs superior to other existing models and can well balance sensitivity and specificity. Thus, it is suitable to be employed in breast cancer detection.

5 Conclusions

CAD has attracted researchers' attention for application in breast cancer diagnosis. However, the accuracy and efficiency of the reported models still need to be improved. In this paper, a two-phase hybrid method, FW-PHHO-ELM, based on feature weighting, optimization and machine learning was proposed for breast cancer detection. In the first phase, K-Means based feature weighting, an efficient approach in data preprocessing and feature engineering, was utilized to make benign and malignant samples more separate by reducing variance within the class and increasing discrimination between classes simultaneously. In the second phase, a machine learning model was constructed with optimization algorithm. The initial parameters of HHO were optimized by PSO firstly, and then the HHO optimized by PSO (PHHO) was utilized to optimize the initial weights and thresholds of ELM, and finally the optimized ELM was employed for breast cancer detection and output the detection results. The empirical results on WDBC data set showed that the proposed model could achieve better stability and performance on breast cancer detection compared to benchmark models and existing models, and could balance sensitivity and specificity well. Therefore, we conclude that the proposed model is suitable for breast cancer detection.

In the future, multi-objective optimization algorithms can be considered to balance the conflict of sensitivity and specificity well. Further, it is feasible to employ this hybrid model into multi-class classification, and also in other data sets or other kinds of cancer detection. In addition, different forms of data such as images and X-ray signals could be input into other machine learning models to detect cancer directly. Besides, early detection of cancer and prediction of cancer staging could be considered in future work.

Funding National Natural Science Foundation of China, 61773401; Science and Technology Department of Jiangxi Province, 20202BABL202028, Feng Jiang.

References

1. Desai M, Shah M (2021) An anatomization on breast cancer detection and diagnosis employing multi-layer perceptron neural network (MLP) and convolutional neural network (CNN). Clinic eHealth 4:1–11
2. Harbeck N, Gnant M (2017) Breast cancer. Lancet 389:1134–1150
3. Al-Azzam N, Shatnawi I (2021) Comparing supervised and semi-supervised machine learning models on diagnosing breast cancer. Ann Med Surg (Lond) 62:53–64
4. Zhang Z, Chen B, Xu S et al (2021) A novel voting convergent difference neural network for diagnosing breast cancer. Neurocomputing 437:339–350
5. Bisoi R, Dash PK, Das PP (2018) Short-term electricity price forecasting and classification in smart grids using optimized multikernel extreme learning machine. Neural Comput Appl 32:1457–1480
6. Jiang F, He J, Zeng Z (2019) Pigeon-inspired optimization and extreme learning machine via wavelet packet analysis for predicting bulk commodity futures prices. Sci China Inf Sci. 62:070204

7. Shrivastava NA, Panigrahi BK, Lim M-H (2014) Electricity price classification using extreme learning machines. Neural Comput Appl 27:9–18
8. Jiang F, He J, Tian T (2019) A clustering-based ensemble approach with improved pigeon-inspired optimization and extreme learning machine for air quality prediction. Appl Soft Comput 85:105827
9. Jiang F, Qiao Y, Jiang X et al (2021) MultiStep ahead forecasting for hourly PM10 and PM2.5 based on two-stage decomposition embedded sample entropy and group teacher optimization algorithm. Atmosphere 12:64
10. Li S, Li W, Wen S et al (2021) Auto-FERNet: a facial expression recognition network with architecture search. IEEE Trans Netw Sci Eng. https://doi.org/10.1109/TNSE.2021.3083739
11. Lyu B, Wen S, Shi K et al (2021) Multi-objective reinforcement learning based neural architecture search for efficient portrait parsing. IEEE Trans Cybern. https://doi.org/10.1109/TCYB.2021.3104866
12. Wang H, Jing W, Li Y et al (2020) Fault diagnosis of fuel system based on improved extreme learning machine. Neural Process Lett 53:2553–2565
13. Yahia S, Said S, Zaied M (2020) A novel classification approach based on extreme learning machine and wavelet neural networks. Multimed Tools Appl 79:13869–13890
14. Shanthi S, Rajkumar N (2020) Lung cancer prediction using stochastic diffusion search (SDS) based feature selection and machine learning methods. Neural Process Lett 53:2617–2630
15. Turkson RE, Qu H, Mawuli CB et al (2021) Classification of Alzheimer's disease using deep convolutional spiking neural network. Neural Process Lett 53:2649–2663
16. Daliri MR (2012) A hybrid automatic system for the diagnosis of lung cancer based on genetic algorithm and fuzzy extreme learning machines. J Med Syst 36:1001–1005
17. Bi X, Zhao X, Huang H et al (2019) Functional brain network classification for Alzheimer's disease detection with deep features and extreme learning machine. Cognit Comput 12:513–527
18. Peng X, Lin P, Zhang T et al (2013) Extreme learning machine-based classification of ADHD using brain structural MRI data. PLoS ONE 8:e79476
19. Ghiasi MM, Zendehboudi S (2021) Application of decision tree-based ensemble learning in the classification of breast cancer. Comput Biol Med 128:104089
20. Gopal VN, Al-Turjman F, Kumar R et al (2021) Feature selection and classification in breast cancer prediction using IoT and machine learning. Measurement 178:109442
21. Hajiabadi H, Babaiyan V, Zabihzadeh D et al (2020) Combination of loss functions for robust breast cancer prediction. Comput Electr Eng 84:106624
22. Hsieh SL, Hsieh SH, Cheng PH et al (2012) Design ensemble machine learning model for breast cancer diagnosis. J Med Syst 36:2841–2847
23. Okagbue HI, Adamu PI, Oguntunde PE et al (2021) Machine learning prediction of breast cancer survival using age, sex, length of stay, mode of diagnosis and location of cancer. Health Technol 11:887–893
24. Huang H, Feng X, Zhou S et al (2019) A new fruit fly optimization algorithm enhanced support vector machine for diagnosis of breast cancer based on high-level features. BMC Bioinformatics 20:290
25. Punitha S, Amuthan A, Joseph KS (2019) Enhanced Monarchy Butterfly Optimization Technique for effective breast cancer diagnosis. J Med Syst 43:206
26. Sayed GI, Darwish A, Hassanien AE (2019) Binary whale optimization algorithm and binary moth flame optimization with clustering algorithms for clinical breast cancer diagnoses. J Classif 37:66–96
27. Patricio M, Pereira J, Crisostomo J et al (2018) Using Resistin, glucose, age and BMI to predict the presence of breast cancer. BMC Cancer 18:29
28. Rahman MM, Ghasemi Y, Suley E et al (2021) Machine learning based computer aided diagnosis of breast cancer utilizing anthropometric and clinical features. IRBM 42:215–226
29. Wang H, Zheng B, Yoon SW et al (2018) A support vector machine-based ensemble algorithm for breast cancer diagnosis. Eur J Oper Res 267:687–699
30. Wang S, Wang Y, Wang D et al (2020) An improved random forest-based rule extraction method for breast cancer diagnosis. Appl Soft Comput 86:105941
31. Pratheep K, Mary A, Nair G (2021) An efficient classification framework for breast cancer using hyper parameter tuned Random Decision Forest Classifier and Bayesian Optimization. Biomed Signal Process Control 68:102682
32. Naik AK, Kuppili V, Edla DR (2019) Efficient feature selection using one-pass generalized classifier neural network and binary bat algorithm with a novel fitness function. Soft Comput 24:4575–4587
33. Jain I, Jain VK, Jain R (2018) Correlation feature selection based improved-binary particle swarm optimization for gene selection and cancer classification. Appl Soft Comput 62:203–215
34. Rao H, Shi X, Rodrigue AK et al (2019) Feature selection based on artificial bee colony and gradient boosting decision tree. Appl Soft Comput 74:634–642
35. Abdel-Basset M, El-Shahat D, El-Henawy I et al (2020) A new fusion of grey wolf optimizer algorithm with a two-phase mutation for feature selection. Expert Syst Appl 139:112824

36. Abdel-Basset M, Ding W, El-Shahat D (2020) A hybrid Harris Hawks optimization algorithm with simulated annealing for feature selection. Artif Intell Rev 54:593–637
37. Singh D, Singh B, Kaur M (2020) Simultaneous feature weighting and parameter determination of Neural Networks using Ant Lion optimization for the classification of breast cancer. Biocybern Biomed Eng 40:337–351
38. Polat K (2018) Similarity-based attribute weighting methods via clustering algorithms in the classification of imbalanced medical datasets. Neural Comput Appl 30:987–1013
39. Heidari AA, Mirjalili S, Faris H et al (2019) Harris hawks optimization: algorithm and applications. Futur Gener Comp Syst 97:849–872
40. Huang G, Zhu Q, Siew CK (2006) Extreme learning machine: theory and applications. Neurocomputing 70:489–501
41. Sree Ranjini KS, Murugan S (2017) Memory based Hybrid Dragonfly Algorithm for numerical optimization problems. Expert Syst Appl 83:63–78