Mammograms Classification Using ELM Based on Improved Sunflower Optimization Algorithm

Yeheng Sun*
School of Business, University of Shanghai for Science and Technology, Shanghai 200093, China
Email: syh54222@163.com

Abstract. To assist specialists in detecting breast cancer on mammograms with better accuracy and less time consuming, this paper proposes an approach based on improved sunflower optimization algorithm (ISFO) and extreme learning machine (ELM). Firstly, features were extracted by using lifting scheme and gray-level co-occurrence matrix (GLCM). Then, the parameters of ELM were optimized by (ISFO) to obtain the final classification results. Finally, in order to avoid overfitting, the proposed model’s performance was evaluated with k-fold random stratified cross validation, and the experiments compared the model with other models on MIAS datasets. The experimental results show that the proposed model has higher classification accuracy, shorter learning time and stronger robustness on mammograms classification task. Thus, this method could be a promising application in bio-medical and provide a basis for the early diagnosis of breast cancer.

Keywords: mammograms; classification model; improved sunflower optimization algorithm ((ISFO)); extreme learning machine (ELM); lifting wavelet decomposition (LWD); ray-level co-occurrence matrix (GLCM); random stratified cross validation (RSCV)

1. Introduction
Breast cancer ranks first in the incidence and second in the mortality among women according to the global cancer report released by the World Health Organization (WHO) in 2018 [1]. However, early diagnosis and treatment are of importance to reduce the mortality rate and improve the cure rate of breast cancer [2].

Screening examinations are the main method to detect breast cancer, and mammography method was preferred for early detection of breast cancer due to its non-invasive, effective, economic [3]. At present, Computer Aid Detection and Diagnosis (CAD) has been increasingly used in the detection and diagnosis of breast cancer [4]. However, the traditional models have some limitations such as low accuracy and long training time [5].

Texture features and morphological features are the main features used for classification on mammograms, and Lifting Wavelet Decomposition (LWD) is a widely used method to extract morphological features of mammograms with the advantages of multi-scale, faster calculation speed and less memory [6-7]. Another effective method for tumor detection from mammograms is texture features obtained by gray-level co-occurrence Matrix (GLCM), which is a two-dimensional feature extraction method. In this method, texture features are acquired by inherent coherence and relationship among near gray pixels of images [8-9].

Generally, the gradient-descent learning algorithms such as Back Propagation (BP) were used in most conventional CAD models to train neural networks, and then different classifiers such as Support
Vector Machine (SVM), Naive Bayes (NB) were utilized for classification. However, the gradient-descent learning algorithms get local optimal solution too easily and converge slowly because all the parameters need to be adjusted. Extreme Learning Machine (ELM) can greatly improve learning speed and reduce training time by randomly initializing parameters [10]. However, setting the parameters randomly also limits the performance of the model. To tackle this problem, an intelligent optimization algorithm Sunflower Optimization Algorithm (SFO) has been used to obtain the best parameters of ELM [11-12].

In this paper, a reliable and efficient automatic diagnosis model on mammograms was proposed, which adopted LWD and GLCM for feature extraction and using ELM as classifier by the improved SFO which optimized the initial input weight and bias of ELM. The paper is organized as follows: Section 1 illustrates the problem and the approaches to counter it. Section 2 contains the details of related methodology such as Region of Interest (ROI) extraction, feature extraction and the architecture of the proposed model. Section 3 evaluates the experimental results, which are discussed in Section 4.

2. Methods

2.1. ROIs Extraction

ROIs contains cancerous information such as neoplasm and nodules which are the key indicators for breast cancer. Therefore, it is necessary to extract the ROIs for better classification result. In this paper, a method of cropping by pixel matrix is adopted to extract ROIs. For the normal images, ROIs are extracted by random, as for the abnormal images, ROIs extraction is based on the radius and center value given by specialists [13-14]. The ROIs from images are shown in figure 1.

![Figure 1](image.png)

**Figure 1.** Raw images and the ROIs of MIAS datasets. (a) and (c) are the original mammogram images, (b) and (d) are the ROIs extracted from (a) and (c), respectively.

2.2. Feature Extraction

2.2.1. Lifting Wavelet Decomposition. Lifting wavelet transform (LWD) adopts lifting scheme to construct the wavelet and provides temporal and frequency conversion and representation for the mammogram image [15].

In general, LWT contains three fundamental steps, which are split, predict and update.

**Split:** The signal is split into even samples $S_E(i, j)$ and old samples $S_O(i, j)$ defined as Eq. (3) and Eq. (4), respectively.

$$S_E(i, j) = S(i, 2j)$$  \hspace{1cm} (3)
Predict: Here, the value of a specific sample can be predicted from the nearest sample. The resultant high-pass coefficient $h(i, j)$ is calculated by Eq. (5).

$$h(i, j) = S_0(i, j) - P(S_L(i, j))$$

Where $P(\cdot)$ represents the prediction operator.

Update: Adopting the updating operation can effectively avoid signal deviation, which is based on the prediction of low-pass coefficient. The low-frequency coefficient is calculated by Eq. (6):

$$l(i, j) = S_L(i, j) - U(P(i, j))$$

Where $U(\cdot)$ is the updating operator.

In this paper, a 2-D LWD includes Low-Low (LL), Low-Low (LL), High-High (HH), High-Low (HL) and Low-High (LH) sub-bands. The high frequency and low frequency components based on their scales are calculated by Eq. (7). The level-4 decomposition of mammograms is shown in figure 2.

$$P_{LW}(s) = \begin{cases} D_{M,N} = \sum F(s) h \ominus M(s - 2MN) \\ D_{M,N} = \sum F(s) l \ominus M(s - 2MN) \end{cases}$$

Where, $\ominus$ is the convolution operator, the terms $M$ and $N$ represent the scale of the wavelet and the translation factors, respectively, $D_{M,N}$ describes the component feature of the signal $F(s)$ in accordance with wavelet function, and $h$ and $l$ are the high-pass and the low-pass filter coefficients, respectively.

2.2.2. Gray-Level Co-Occurrence Matrix. GLCM extracts the features from the image intensities histogram [16]. In this study, the LL and HL sub-bands were used to extract GLCM features, and the five features including Angular Second Moment (ASM), Inverse Different Moment (IDM), contrast, energy and entropy are calculated by Eqs. (8)-(12).

The first feature, ASM is calculated by Eq. (8).

$$ASM = \sum_{i=1}^{m} \sum_{j=1}^{n} f(i, j)^2 \cdot (i - u_i) \cdot (j - u_j)$$

where $f(i, j)$ represents the gray-level co-occurrence matrix after normalization.

IDM reflects the degree of change between different regions of texture features which is achieved by Eq. (9).

$$IDM = \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} f(i, j) \cdot d \cdot \theta$$

Where $d$ and $\theta$ represent the distance and the angle between two pixels, respectively.

The next feature is Contrast ($Con$) measuring the intensity of pixels of the image and is formulated by Eq. (10).

$$Con = \sum_{i=1}^{m} \sum_{j=1}^{n} f(i, j) \cdot (i - j)^2$$

Energy ($Eng$) can be calculated by Eq. (11).

$$Eng = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} f^2(i, j)}$$

Entropy ($Ent$) is a measure of complexity of grayscale distribution evaluated by Eq. (12).

$$Ent = - \sum_{i=1}^{m} \sum_{j=1}^{n} f(i, j) \cdot \log f(i, j)$$
2.3. Hybrid Classification Algorithm Based on Elm and Improved SFO

2.3.1. Extreme Learning Machine. ELM is composed of the input layer, the hidden layer and the output layer and all the nodes are fully connected [17]. Taken \( N \) arbitrary distinct samples \((x_j, t_j)\), where \( x_j = [x_{j1}, x_{j2}, ..., x_{jn}]^{T} \in R^n \) and \( t_j = [t_{j1}, t_{j2}, ..., t_{jm}]^{T} \in R^m \). Here \( n \) is the number of samples and \( m \) is the number of classes. \( N \) represents the total number of hidden neuron and \( \eta(\cdot) \) is the activation function. Consider the weights between the input layer and hidden layer as \( \beta_i = [\beta_{i1}, \beta_{i2}, ..., \beta_{in}]^{T} \). The weight vector of \( i \)th hidden neuron is \( w_i = [w_{i1}, w_{i2}, ..., w_{in}]^{T} \) and the bias of \( i \)th hidden neuron is \( b_i \). \( T = [t_1, t_2, ..., t_N]_{m \times N} \) represents the output of the network and \( H \) is the input-hidden layer matrix which are obtained by Eq. (13) and Eq. (14), respectively.

\[
H = \begin{bmatrix}
\eta(w_1x_1 + b_1) & \ldots & \eta(w_1x_n + b_1) \\
\vdots & \ddots & \vdots \\
\eta(w_nx_1 + b_1) & \ldots & \eta(w_nx_n + b_1)
\end{bmatrix}_{N \times l}
\]

(13)

\[
t_j = \begin{bmatrix}
\sum_{i=1}^{l} \beta_{i1}\eta(w_{i}x_j + b_i) \\
\vdots \\
\sum_{i=1}^{l} \beta_{im}\eta(w_{i}x_j + b_i)
\end{bmatrix}_{m \times 1}
\]

(14)

Therefore, \( T' \) can be formulated by Eq. (15).

\[
H\beta = T'
\]

(15)

Where \( T' \) represents inverse of matrix \( T \).

2.3.2. SFO Algorithm

2.3.2.1 The Standard SFO Algorithm. The Sun Flower Optimization algorithm (SFO) is one of the modern optimization algorithms and is generally used in combinatorial optimization problem [18-19]. In the proposed method, each plant is considered as a feasible solution in the solution space and the pollination is provided at random with the minimal distance between the plant \( i \) and the plant \( i+1 \), which is approximate to the process of neighborhood search. The absorbed radiation of each flower is shown as Eq. (16).

\[
Q_i = \frac{P}{4\pi d_i^2}
\]

(16)

Where, \( P \) represents the source power and \( d_i \) is the distance between the flower and the sun.

The sunflowers’ orientation to the sun is equated by Eq. (17).

\[
\bar{O}_i = \frac{x - X_i}{\|x - X_i\|}
\]

(17)

Where, \( X^* \) and \( X_i \) are the current plantation and the \( i \)th plantation, respectively.

The step on the specific direction is evaluated by Eq. (18).

\[
D_i = \gamma \times P_i (\|X_i + X_{i-1}\|) \times \|X_i + X_{i-1}\|
\]

(18)

Where, \( \gamma \) is a constant value that describe the inertial displacement of the sunflowers, \( P_i \) describes the pollination probability.

To avoid getting into the local optimum, the algorithm restricts the maximum given step which is formulated by Eq. (19).
\[ S_{\text{max}} = \frac{\|X_{\text{max}} - X_{\text{min}}\|}{2 \times N_{\text{pop}}} \]  

(19)

Where, \( N_{\text{pop}} \) is the total number of plants. \( X_{\text{max}} \) and \( X_{\text{min}} \) are the minimum and maximum limitation, respectively.

The new plantation is obtained by Eq. (20).

\[ \bar{X}_{i+1} = \bar{X}_i + D_i \times \hat{O}_i \]  

(20)

2.3.2.2. The Improved SFO Algorithm. To increase population diversity and enhance the local searching ability of SFO, Gauss mutation operator is introduced to the algorithm by selecting \( X_i \) randomly from the population then producing \( M \) mutation plantations according to Eq. (21) [20].

\[ \bar{X}_i = X_i \times e \]  

(21)

Where, \( e \sim \mathcal{N}(0,1) \).

To prevent the new plantations produced after pollination exceeding the boundary, \( \bar{X}_i \) will be mapped to a new location by Eq. (22), when it crosses the boundary.

\[ \bar{X}_i = X_{\text{max}} + |\bar{X}_i| \% (X_{\text{max}} - X_{\text{min}}) \]  

(22)

In this algorithm, the \( \text{Lévy Flight (LF)} \) mechanism is used to improve the early convergence problem, which adopts a random walk behavior to deal with the local search position by Eqs. (23)-(26). The \( \text{Lévy distribution} \) can be approximated as:

\[ L(S) \approx |S|^{-1-\gamma} \]  

(23)

Where,

\[ S = \frac{u}{|v|^{1/\gamma}} \]  

(24)

\[ u \sim \mathcal{N}(0, \sigma_u^2), v \sim \mathcal{N}(0, \sigma_v^2) \]  

(25)

\[ \sigma_u = \sqrt{\frac{\Gamma(1+\gamma)\sin(\pi\gamma/2)}{\Gamma([1+(\gamma/2)]\sin(\pi(\gamma/2)/2))}}, \sigma_v = 1 \]  

(30)

Where \( \gamma \in [0,2] \) (here, \( \gamma = 1.5 \)), \( \Gamma(\cdot) \) is Gamma function, \( S \) describes the step size.

By considering the Gauss mutation operator and the LF mechanism, the new plantation is obtained by Eq. (31).

\[ \bar{X}_{i+1} = (\bar{X}_i + D_i \times \hat{O}_i) \times L(S) \]  

(26)

The pseudo-code of Improved Sunflower Optimization (ISFO) is expressed as follows:

**Algorithm 1: ISFO**

1. Initial a uniform/random population of \( n \) flowers
2. Find the sun (best solution \( S^{*} \)) in the initial population
3. Orient all plants toward the sun
4. **while** (\( k < \text{MaxDays} \))
   5. obtain the Gauss mutation plants
   6. Calculate the orientation vector for each plant
   7. Remove \( m\% \) plants further away from the sun
   8. Calculate the step scale for each plant
   9. Best plants will pollinate around the sun
6

10. Evaluate the new individuals
11. if a new individual is a global best, update the sun
12. end while
13. Best solution found

3. Results

3.1. Data
The proposed model has experimented on MIAS datasets, which is consist of 322 images, out of which 206 images are normal, 116 images are abnormal and among the abnormal images, classified as 68 benign and 48 malignant. The MIAS dataset contains mammogram images with a size of $1024 \times 1024$ pixels after padding, which are held as 8-bit gray level scale of 256 different gray levels (0-255).

To avoid the over-fitting problem, a random stratified cross-validation (RSCV) method has been applied for validating the proposed ISFO-ELM model, the distributed samples for MIAS are shown in figure 2.

![Figure 2](image)

**Figure 2.** Distribution of samples for each trial by 5-fold RSCV.

3.2. Experimental Results
In this paper, all the experiments were implemented by the GPU version of V100 in Ubuntu 14.04 system with Python3.6.4. In order to improve performance evaluation reliability and test the generalization ability of the proposed method, the runtime environment and experimental steps for all models are the same, and the experiment for each model is repeated five times with 5-fold RSCV to calculate the mean of statistical indicators for evaluation.

The numbers of parameter such as Confusion Matrix, Accuracy (Acc), precision, sensitivity, specificity, $F_1$-score, and Area Under ROC Curve (AUC).

The average accuracy for the proposed model of five times with 5-fold RSCV is shown in table 1.

| Dataset | Benign vs. malignant |
|---------|---------------------|
| Run 1F  | 2F  | 3F  | 4F  | 5F  | $A_{avg}$ |
| MIAS    | 98.78 | 100 | 97.65 | 93.44 | 100 | 97.97 |
| 2       | 100 | 95.33 | 96.34 | 94.34 | 95.45 | 96.29 |
| 3       | 96.53 | 98.32 | 97.54 | 94.13 | 93.24 | 95.95 |
| 4       | 100 | 95.68 | 96.23 | 92.99 | 100 | 96.98 |
| 5       | 99.87 | 94.45 | 98.32 | 95.11 | 98.44 | 97.24 |

$A_{avg}(\%)$: Average accuracy in percentage (%) over 5-fold RSCV.
The experimental results including Acc, sensitivity, specificity, confusion matrix, F1-Score and AUC of the proposed model, PSO-ELM, BPNN and SVM on MIAS dataset are illustrated in figures 3-4.

![Figure 3. Results of ISFO-ELM, PSO-ELM, BPNN and SVM.](image)

![Figure 4. The optimal confusion matrix for ISFO-ELM, PSO-ELM, BPNN and SVM on MIAS dataset.](image)

**4. Discussion**

The experimental results shown in figures 3-4 indicated that the proposed classification model in this paper earns better performance than other models in the context of classification accuracy, sensitivity and specificity on MIAS dataset, which prove that LWD and GLCM method are more effective stable in capturing approximate features from mammograms and greatly improving the capability of classification on mammogram images than other classifier. In the future, three-class classification problems for breast cancer detection will be studied.

**Contribution**

Y. S. was involved in the data acquisition, data analysis, modeling, manuscript write-up and revision. J. H. proposed the early conception of the study. All authors have provided final approval of the current version to be published and agree to be accountable for all aspects of the work.

**References**

[1] Bray F, Ferlay J, Soerjomataram I, Siegel R L, Torre L A and Jemal A 2018 Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries CA: A Cancer Journal for Clinicians 68(6): 394-424.

[2] Gupta N P, Malik P K, Ram B S 2020 A Review on Methods and Systems for Early Breast Cancer Detection Proceeding of the 2020 International Conference on Computation, Automation and Knowledge Management (ICCAKM) IEEE: 42-46.

[3] Herskind C, Steil V, Kraus-Tiefenbacher U, et al. 2005 Radiobiological aspects of intraoperative radiotherapy (IORT) with isotropic low-energy X rays for early-stage breast cancer [J] Radiation Research 163(2): 208-215.

[4] Lee R S, Gimenez F, Hoogi A, et al. 2017 A curated mammography data set for use in computer-aided detection and diagnosis research [J] Scientific Sata 4: 170-177.
[5] Saxena S, Gyanchandani M 2020 Machine Learning Methods for Computer-Aided Breast Cancer Diagnosis Using Histopathology: A Narrative Review [J] Journal of Medical Imaging and Radiation Sciences 51(1): 182-193.

[6] Liu W, Zhang R, Ling Y, et al. 2020 Automatic recognition of breast invasive ductal carcinoma based on terahertz spectroscopy with wavelet packet transform and machine learning [J] Biomedical Optics Express 11(2): 971-981.

[7] Antonini M, Barlau M, Mathieu P, and Daubechies I 1992 Image coding using wavelet transform[J] IEEE Transactions on image processing 1(2): 205-220.

[8] Beura S, Majhi B, Dash R 2015 Mammogram classification using two-dimensional discrete wavelet transform and gray-level co-occurrence matrix for detection of breast cancer [J] Neurocomputing 154: 1-14.

[9] Gardezi S J S, Faye I, Adjad F, et al. 2016 Mammogram classification using curvelet GLCM texture features and GIST features Proceeding of the International Conference on Advanced Intelligent Systems and Informatics Springer Cham: 705-713.

[10] Dhillon A, Singh A 2020 eBreCaP: extreme learning-based model for breast cancer survival prediction [J] IET Systems Biology 14(3): 160-169.

[11] Zhu B, Feng Y, Gong D, et al. 2020 Hybrid particle swarm optimization with extreme learning machine for daily reference evapotranspiration prediction from limited climatic data [J]. Computers and Electronics in Agriculture 173: 105430.

[12] El-Sehiemy R A, Hamida M A, Mesbahi T 2020 Parameter identification and state-of-charge estimation for lithium-polymer battery cells using enhanced sunflower optimization algorithm [J] International Journal of Hydrogen Energy 45(15): 8833-8842.

[13] Wong M T, He X, Nguyen H, et al. 2012 Particle swarm optimization based feature selection in mammogram mass classification Proceeding of the 2012 International Conference on Computerized Healthcare (ICCH) IEEE: 152-157.

[14] Mohanty F, Rup S, Dash B, et al. 2019 Mammogram classification using contourlet features with forest optimization-based feature selection approach [J] Multimedia Tools and Applications 78(10): 12805-12834.

[15] Sweldens W 1998 The lifting scheme: A construction of second generation wavelets [J] SIAM Journal on Mathematical Analysis 29(2): 511-546.

[16] Haralick R M, Shanmugam K, Dinstein I H 1973 Textural features for image classification [J] IEEE Transactions on Systems, Man, and Cybernetics (6): 610-621.

[17] Pal S S 2018 Grey wolf optimization trained feed forward neural network for breast cancer classification [J] International Journal of Applied Industrial Engineering (IJAIE) 5(2): 21-29.

[18] Muduli D, Dash R, Majhi B 2020 Automated breast cancer detection in digital mammograms: A moth flame optimization based ELM approach [J] Biomedical Signal Processing and Control 59: 101912.

[19] Gomes G F, da Cunha S S, Ancelotti A C 2019 A sunflower optimization (SFO) algorithm applied to damage identification on laminated composite plates [J] Engineering with Computers 35(2): 619-626.

[20] Shen L, He M, Shen N, et al. 2020 Optimal breast tumor diagnosis using discrete wavelet transform and deep belief network based on improved sunflower optimization method [J] Biomedical Signal Processing and Control 60: 101953.