

Mohammad Moradi and Abdalhossein Rezai*

High-performance breast cancer diagnosis method using hybrid feature selection method

<https://doi.org/10.1515/bmt-2024-0185>

Received April 14, 2024; accepted December 2, 2024;

published online December 23, 2024

Abstract

Objectives: One of the primary causes of the women death is breast cancer. Accurate and early breast cancer diagnosis plays an essential role in its treatment. Computer Aided Diagnosis (CAD) system can be used to help doctors in the diagnosis process. This study presents an efficient method to performance improvement of the breast cancer diagnosis CAD system using thermal images.

Methods: The research strategy in the proposed CAD system is using efficient algorithms in feature extraction and classification phases, and new efficient feature selection algorithm. In the feature extraction phase, the Segmentation Fractal Texture Analysis (SFTA) algorithm that is a texture analysis algorithm is used. This algorithm utilizes two-threshold binary decomposition. In the feature selection phase, the developed feature selection algorithm, which is hybrid of binary grey wolf optimization algorithm and firefly optimization algorithm, is applied to extracted features. Then, the kNN, SVM, and DTree classification techniques are applied to check whether the selected features are efficiently discriminated the group successfully with minimal misclassifications.

Results: The DMR database is utilized for performance evaluation of the proposed method. The results indicate that the obtained accuracy, specificity, sensitivity, and MCC are 97, 96, 98, and 94.17 %, respectively.

Conclusions: The developed breast cancer diagnosis method has advantages compared to other breast cancer diagnosis using thermal images.

Keywords: Computer Aided Diagnosis (CAD) system; breast cancer; feature selection; feature extraction; thermography

Introduction

One of the primary causes of the death among women is breast cancer. Accurate and early breast cancer diagnosis plays an essential role in its treatment. It is because it can help in avoiding surgeries. In addition, it can increase the patient's survival probability [1, 2]. There is more metabolic activity in cancerous tissues. So, the blood flow and thereby the temperature in the cancerous tissues are increased. On the other hand, the infrared breast thermography is a screening tool that measures the distribution of the temperature in breast tissue. As a result, it is an important juncture tool for mammography screening tool in early breast cancer detection [3, 4].

By using Computer Aided Diagnosis (CAD) system for analysis, the breast thermographic images have emerged as an efficient tool for supporting the radiologists for cancer diagnosis [5]. The system has several phases including pre-processing, feature extraction, feature selection and classification [6]. Among them, the feature selection is an important phase. It is because selecting the best features can help to decrease the processing time and increase the accuracy of diagnosis [7].

On the other hand, Swarm Intelligence (SI) algorithms are one of the most important branches of metaheuristics based on natural processes. In nature, groups of animals such as bats, hawks, bees, and whales show certain SI behaviors that are greater than each individual intelligence. The researchers have utilized these SI behaviors to predict the solution to impossible problems [8]. The Grey Wolf Optimization (GWO) algorithm, Binary Grey Wolf Optimizer (BGWO) algorithm and Firefly Algorithm (FA) are instances of the SI algorithms that are utilized for optimization applications [9, 10].

Recently, there are several attempts [4, 11–20] to present the CAD systems for diagnosis of breast cancer using SI algorithms and thermography. Although, these CAD systems have suitable performance, but the performance of the CAD systems can be improved for diagnosis of the breast cancer.

This study presents an efficient method to improve the performance of the CAD systems that are used for diagnosis of the breast cancer using thermographic images. The main contributions of this study are as follows:

- It extracts the features with Segmentation Fractal Texture Analysis (SFTA) technique.

*Corresponding author: Abdalhossein Rezai, Department of Electrical Engineering, University of Science and Culture, Tehran, Iran, E-mail: rezai@usc.ac.ir

Mohammad Moradi, ACECR Institute of Higher Education, Isfahan Branch, Isfahan, Iran, E-mail: Mohamadmoradi0078@gmail.com

- It selects the best features by applying the BGWO algorithm, firefly algorithm and the proposed hybrid of these algorithms to extracted features.
- It applies k-Nearest Neighbors (kNN), DTree and Support Vector Machine (SVM) classification algorithms to classify these three selected feature groups.

The Database for Mastology Research (DMR) is used for verification. Figures 1 and 2 depict examples of healthy and cancerous tissues of this database.

The comparison indicate that the developed method has advantages compared to other breast cancer diagnosis methods.

The rest of this study is organized as follows. Section “Related works” demonstrates the related works. Section “Methodology” demonstrates the developed method. Section “Algorithms” presents the utilized algorithms. Section “Experimental results and discussion” demonstrates the experimental results. Section “Comparative analysis” discusses the results, and Section “Conclusion” concludes this work.

Related works

There are several attempts [4, 11–20] to present the CAD system for the breast cancer diagnosis using thermography. Mishra and Rath [11] have offered a comprehensive study of the effect of feature selection methods for the abnormalities diagnosis in thermography images. The utilized feature selection method is principal component analysis and

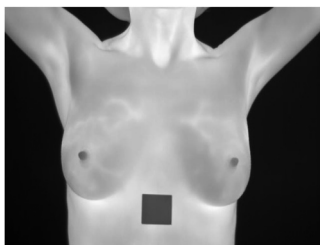


Figure 1: Healthy image.

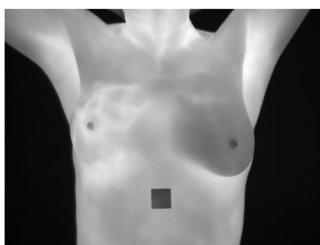


Figure 2: Cancerous image.

autoencoder technique. They evaluate the proposed CAD system using the DMR database and kNN algorithm for classification. The obtained accuracy, specificity and sensitivity in [11] are 95.45, 88.07, and 99.17 %, respectively.

Rajinikanth et al. [12] have used marine predators algorithm for diagnosis of abnormalities in thermography images. In [12], the Gray-Level Co-occurrence Matrix (GLCM) is utilized for feature extraction. They utilized 200 images for training the algorithm and 100 images for validation. The obtained accuracy, specificity, and sensitivity in [12] are 93.5, 93 and 94 %, respectively.

Sathish et al. [13] have proposed wavelet transform method. Statistical feature extraction is performed after applying Discrete Wavelet Transform (DWT) algorithm. They evaluate the developed CAD system using DMR database. The obtained accuracy, sensitivity and specificity in [13] are 87.23, 94.34 and 91 %, respectively.

Gogoi et al. [14] have offered a comprehensive study of the effect of feature selection methods for diagnosis of abnormalities in thermography images. They used statistical and texture features. They evaluate the developed CAD system using DMR database and SVM, kNN, DTree and Artificial Neural Network (ANN) algorithms for classification. The best obtained accuracy, specificity and sensitivity for ANN in [14] are 87.5, 93.3, and 80 %, respectively.

Fikadu [16] has proposed Binary Particle Swarm Optimization (BPSO) method for breast cancer diagnosis. Feature selection is performed by applying BPSO algorithm. The evaluation is performed using the DMR database and SVM algorithm for classification. The obtained accuracy in [16] is 96.22 %.

Resmini et al. [17] have developed a computational method that is hybrid of dynamic and static infrared thermography for cancer diagnosis using supervised and unsupervised techniques. They have offered Genetic Algorithm (GA) for feature selection. They applied their method on thermography images of the DMR database. They also used SVM for classification. The obtained accuracy, specificity and sensitivity in [17] are 95, 95 and 95 %, respectively.

Zarei et al. [4] have offered a CAD system for breast cancer diagnosis. They utilize a new segmentation method for the thermal images that is based on the Gaussian mean shift algorithm. The obtained sensitivity, specificity, and accuracy in [4] are 91.81, 84.86, and 87.4 %, respectively.

Darabi et al. [20] have used feature selection algorithms for diagnosis of the breast cancer in thermal images. In this CAD system, the hybrid of minimum Redundancy Maximum Relevance (mRMR) algorithm and GA with Random Subset Feature Selection (RSFS) algorithm are utilized for feature selection. In addition, the SVM and kNN algorithms are utilized as classifier algorithm. The best obtained sensitivity

and accuracy for kNN and hybrid of GA and RSFS in [20] are 96 and 83.87 %, respectively.

Salimian et al. [19] have focused on effective features in thermal images for breast cancer diagnosis. They tested several features such as energy, entropy, correlation, contrast and mean.

Kalita et al. [21] have utilized mammography images of breast tissue. They performed feature extraction using the Local Binary Pattern (LBP) algorithm. They further processed the features using the 2-way thresholding technique to find the lower bound and upper bound for cancer detection. Furthermore, the SVM and kNN algorithms are employed as classifier algorithms. The achieved metrics in [21] include specificity of 96.2 % and accuracy of 99 %.

Khafaga et al. [22] have used thermal images for breast cancer diagnosis by extracting features using the VGG-16, VGG-19, ResNet-50, GoogLeNet, and AlexNet. The best obtained sensitivity, specificity, and accuracy in [22] are 98.36, 62.51, and 96.15 %, respectively.

Moayedi et al. [23] have offered an intelligent method to improve the performance of the CAD system using thermal images. In the developed method, the features in the LBP and GLCM are extracted from thermal images. Then, the features are selected using firefly feature selection algorithm. They used the kNN, SVM, and DTree classifiers for detection of malignancy in the breast. The achieved metrics in [23] include specificity of 98.2 %, accuracy of 98.8 %, and sensitivity of 99 %, where using SVM as classifier.

Although these methods have suitable performance, but they suffer from accurate diagnosis. So, we try to improve the performance of the breast cancer diagnosis method in this study.

Methodology

Figure 3 displays the research methodology in this paper.

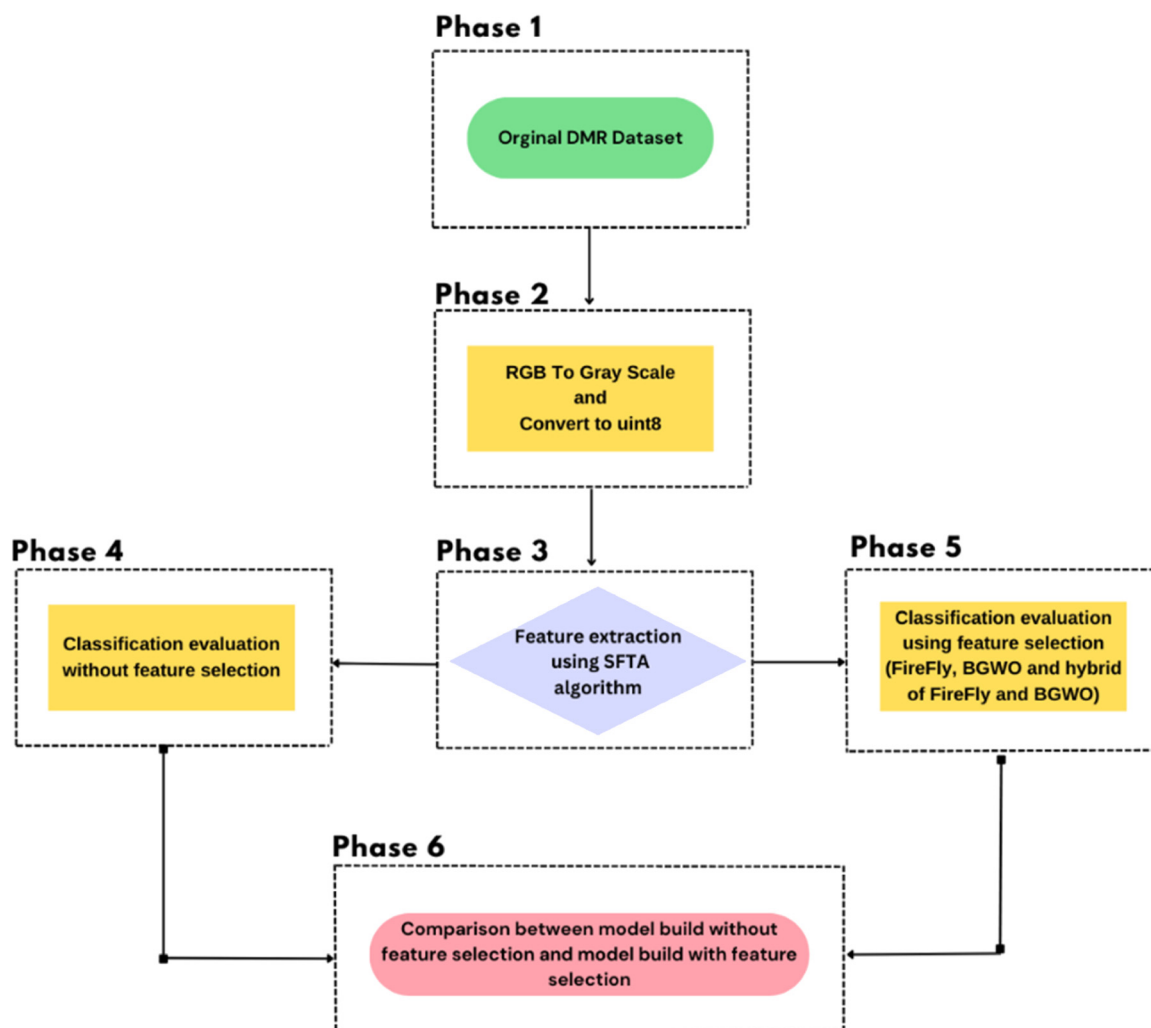


Figure 3: Research methodology.

Our research methodology has 6 phases. The data acquisition phase is the first stage of this methodology, in which we collect the necessary data for the study. The second stage is data preprocessing, which involves RGB to gray scale, convert to unit8, and transforming the collected data so that it is suitable for classification prediction. In phase three, we extract features. The results are then carried over to phase 4 for classification prediction without feature selection. In phase 5, the BGWO, firefly, and hybrid algorithms are designed, trained, and tested on data for feature selection. Then, the results are sent for classification prediction. In the final stage of the research, we conduct a comparison of models with and without feature selection. The details of these 5 phases are described in next section. The phase 6 is described in Section “Experimental results and discussion”.

Phase 1 – data acquisition

The required data is gathered from the DMR public database [24] in this phase. A set of 200 thermograms of healthy and cancerous patients are taken in this work.

Phase 2 – data preprocessing

At first, image format changed from RGB to black and white, then converted to unit8. At this phase, the brightness levels of the images are divided into a range of 0–255 levels.

Phase 3 – feature extraction

We use SFTA algorithm for feature extraction. In this phase, 576 features are extracted from each image. All 115,200 extracted features are useful.

Phase 4 – classification evaluation without feature selection

We have done classification without feature selection for comparison in this phase. In addition, DTree, SVM and kNN algorithms are used for classification. All 115,200 extracted features are useful even could detect cancer, but it is a time-consuming process.

Phase 5 – classification evaluation with feature selection

Using all features may reduce the accuracy due to high processing. So, using the feature selection not only can reduce the

processing time, but also can increase the accuracy. We propose hybrid of BGWO and firefly algorithms for feature selection. However, the BGWO algorithm, firefly algorithm, and hybrid of BGWO and firefly algorithms are applied to extracted features for feature selection in this phase.

Algorithms

Feature extraction algorithm

The SFTA that is a texture analysis algorithm, is used for feature extraction. This algorithm is divided into two stages. In the first stage, by using Two Threshold Binary Decomposition (TTBD), a grayscale image is decomposed into a set of binary images. The mean gray level and pixels count are then calculated from each binary image. In addition, the regions boundaries are utilized to calculate the fractal dimensions [25].

Feature selection algorithm

Feature selection has vital role in high-dimensional data mining and analysis, and machine learning. In this study, an efficient hybrid feature selection algorithm is presented based on the FA and BGWO algorithm.

GWO algorithm

The pyramidal hierarchy in the grey wolves group is depicted in Figure 4 [26].

In this figure, the first three leaders are denoted by Alpha (α), beta (β) and delta (δ). It should be noted that omega (ω) wolves are at the bottom of this pyramidal hierarchy that follow the above mentioned leaders [9]. The GWO algorithm is depicted in Algorithm 1.

Algorithm 1: The GWO algorithm

```

Initialize the related parameters of GWO
Randomly generate the positions of the wolves
Compute the fitness of each wolf
Find  $Y_\alpha$ ,  $Y_\beta$  and  $Y_\delta$ 
for  $i=1$ :  $MAX\_ITERATION$  do
    Update  $k$ ,  $P$  and  $U$  by Eqs. (1), (2) and (6)
    Calculate the position of each wolf by Eqs. (3)–(5)
    Compute the fitness of each wolf
    Update  $Y_\alpha$ ,  $Y_\beta$  and  $Y_\delta$ 
End for
Output  $Y_\alpha$ 

```

In this algorithm, the locations of ω wolves are updated based on the locations of α , β and δ wolves. A distance of wolf

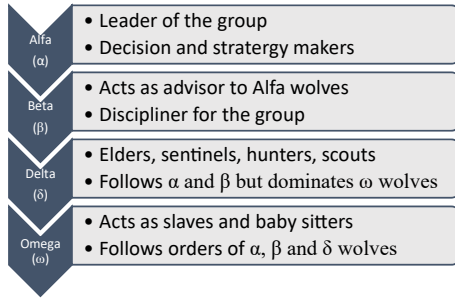


Figure 4: The pyramidal hierarchy in the grey wolves group [26].

(i) is calculated using Eqs. (1)–(4) and then its position is updated using Eq. (5) [9].

$$\vec{P} = 2k \cdot \vec{g}_1 - k \quad (1)$$

$$\vec{U} = 2\vec{g}_2 \quad (2)$$

$$\vec{U}_\alpha = \left| \vec{Q} \cdot \vec{Y}_\alpha - \vec{Y}_i \right|, \vec{U}_\beta = \left| \vec{Q} \cdot \vec{Y}_\beta - \vec{Y}_i \right|, \vec{U}_\delta = \left| \vec{Q} \cdot \vec{Y}_\delta - \vec{Y}_i \right| \quad (3)$$

$$\vec{Y}_1 = \vec{Y}_\alpha - \vec{P}, \vec{U}_\alpha, \vec{Y}_2 = \vec{Y}_\beta - \vec{P}, \vec{U}_\beta, \vec{Y}_3 = \vec{Y}_\delta - \vec{P}, \vec{U}_\delta \quad (4)$$

$$\vec{Y}_i(nt) = \frac{\vec{Y}_1 + \vec{Y}_2 + \vec{Y}_3}{3} \quad (5)$$

Where $\vec{Y}_i, \vec{Y}_\alpha, \vec{Y}_\beta$ and $\vec{Y}_\delta$ denote the position vectors for i, α, β and δ wolves, respectively. In addition, $\vec{U}_\alpha, \vec{U}_\beta$ and $\vec{U}_\delta$ denote the distance vectors between α, β, δ and i wolves, respectively. The nt indicates the next iteration. Moreover, $\vec{g}_1$ and $\vec{g}_2$ are random vectors between $[0, 1]$; $\vec{P}$ and $\vec{Q}$ are coefficient vectors. The parameter k is computed as follows:

$$k = 2 - 2 * a/b \quad (6)$$

Where b and a indicate the maximum iterations count and the current iteration, respectively.

BGWO algorithm

Positions in the GWO algorithm can be found in the continuous space at any point. As a result, the implementation of the utilized equations for updating are simple, but the wolves' positions are not always continuous. The BGWO algorithm can be used in these cases in which wolves are only found in values of 1 or 0. Thus, the same utilized updating equations for GWO cannot be used for BGWO [9], but the same strategy can be used to obtain α, β and δ . In addition, Eq. (3) is used to compute $\vec{U}_\alpha, \vec{U}_\beta$ and $\vec{U}_\delta$. Then, the sigmoid function is utilized to compute v_1, v_2 and v_3 as follows.

$$v_1^c = 1 / (e^{-10(P^c \cdot U_\alpha^c - 0.5)} + 1) \quad (7)$$

$$v_2^c = 1 / (e^{-10(P^c \cdot U_\beta^c - 0.5)} + 1) \quad (8)$$

$$v_3^c = 1 / (e^{-10(P^c \cdot U_\delta^c - 0.5)} + 1) \quad (9)$$

Where the c th dimension of a wolf is indicated by c . The Eqs. (10)–(12) are utilized to compute the $bstep_1, bstep_2$ and $bstep_3$. Then, the result has binary value. It then utilizes Eqs. (7)–(9) as transfer function [9].

$$bstep_1^c = \begin{cases} 1 & \text{if } (v_1^c \geq rnd) \\ 0 & \text{else} \end{cases} \quad (10)$$

$$bstep_2^c = \begin{cases} 1 & \text{if } (v_2^c \geq rnd) \\ 0 & \text{else} \end{cases} \quad (11)$$

$$bstep_3^c = \begin{cases} 1 & \text{if } (v_3^c \geq rnd) \\ 0 & \text{else} \end{cases} \quad (12)$$

Where rnd is a random number between $[0, 1]$, and $bstep_3, bstep_2$ and $bstep_1$ are the distances that i will move relative to δ, β and α . Next, $\vec{Y}_1, \vec{Y}_2$, and $\vec{Y}_3$ are computed as follows.

$$Y_1^c = \begin{cases} 1 & \text{if } (Y_\alpha^c + bstep_1^c \geq 1) \\ 0 & \text{else} \end{cases} \quad (13)$$

$$Y_2^c = \begin{cases} 1 & \text{if } (Y_\beta^c + bstep_2^c \geq 1) \\ 0 & \text{else} \end{cases} \quad (14)$$

$$Y_3^c = \begin{cases} 1 & \text{if } (Y_\delta^c + bstep_3^c \geq 1) \\ 0 & \text{else} \end{cases} \quad (15)$$

Finally, as illustrated in Eq. (16), a simple stochastic crossover is utilized for updating Y_i position in the next iteration. The BGWO algorithm is shown in Algorithm 2.

$$Y_i^c(nt) = \begin{cases} Y_1^c & \text{if } (rnd < \frac{1}{3}) \\ Y_2^c & \text{elseif } (\frac{1}{3} \leq rnd < \frac{2}{3}) \\ Y_3^c & \text{else} \end{cases} \quad (16)$$

Algorithm 2: The BGWO algorithm

```

Initialize the related parameters of BGWO
Randomly generate the positions of the wolves
Compute the fitness of each wolf
Find  $Y_\alpha, Y_\beta$  and  $Y_\delta$ 
for  $i=1: b$  Do
    Update  $k, P$  and  $U$  by Eqs. (1), (2), and (6)
    Calculate the position of each wolf by Eqs. (10)–(16)
    Compute the fitness of each wolf
    Update  $Y_\alpha, Y_\beta$  and  $Y_\delta$ 
End for
Output  $Y_\alpha$ 

```

The utilized parameters for the BGWO algorithm in this study are [0 1], 10, 500, and 1,000 for the search domain, No. of Search agents, Problem dimension, and No. of iterations, respectively.

Firefly algorithm

This algorithm was first presented by Yang in [27]. Firefly algorithm is based on a physical law, which states that light intensity decreases as square of distance increases. It should be noted that by increasing the distance between the light observer and the light source, the light became weak due to light absorption. Yang in [27] has present the firefly optimization algorithm based on this phenomena. Thus, the firefly algorithm can be expressed as shown in Algorithm 3.

Algorithm 3: The firefly algorithm

```
i=0; z*=∅; μ=1.0; //initialize: Best solution, attractiveness
K(0)=Init FA(); //initialize a population
while (i<MF) do
  α(i)=ANew(); //determine a new value of α
  EvalFA(K(i),f(z));
  OrdFA(K(i),f(z));
  Z*=FinBesFA(K(i),f(z)); //determine the best solution
  K(i+1)=MovFA(K(i)); //vary the attractiveness accordingly
  i=i + 1;
End while
```

The following firefly flashing characteristics are utilized to formulate:

- All fireflies are gender neutral.
- The landscape of the fitness function influences or determines the light intensity of a firefly.
- Their attractiveness is proportional to the intensity of their light.

The 'Init FA' function initializes randomly the firefly population. The firefly search process is contained within the while loop. The while loop includes the following steps.

- (a) The 'ANew' function can be utilized for changing the initial value of α . This step is optional.
- (b) The 'EvalFA' function evaluates the solution's quality. This contains the implementation of a fitness function $f(z)$.
- (c) The 'OrdFA' function sorts the firefly population based on fitness values.
- (d) The 'FinBesFA' function chooses the best individual from the population.
- (e) The positions of the fireflies are moved using 'MovFA' function.

It should be noted that the fireflies are drawn to the more attractive individuals. The firefly search procedure adheres to the maximum number of fitness function evaluations, MF. In fact, each FA implementation can fall somewhere between these two asymptotic states. The utilized parameters for firefly in this study are 10, 50, and 100 for No. Selection, No. Fireflies, and No. Max Iteration, respectively.

Hybrid of the firefly and BGWO algorithms

The proposed algorithm is a hybrid feature selection algorithm, which is a combination of the BGWO and firefly feature selection algorithms. The developed algorithm is displayed in Algorithm 4.

Algorithm 4: The developed feature selection algorithm

```
Step 1: Set parameter of firefly algorithm:
  Number of selections=10
  Number of fireflies=50
  Maximum iteration=100
Step 2: Start feature selection using firefly algorithm
Step 3: Save firefly algorithm output in array
Step 4: Set parameter of BGWO algorithm:
  Data validation=20 %
  Maximum iteration=1,000
Step 5: Start feature selection using BGWO algorithm
Step 6: Save BGWO algorithm output in array
```

In the first step, the parameters of the firefly algorithm are set. Then, the feature selection is started using the firefly algorithm in the second step. In step 3, the results are saved. Then, the parameters of the BGWO algorithm are sets in step 4. The feature selection is started using BGWO algorithm in step 5 by applying the BGWO algorithm to results of the firefly algorithm. Finally, the results of the BGWO algorithm are saved as the output of the proposed algorithm.

Classification algorithm

The DTree, kNN and SVM algorithms [28, 29] are applied to the selected/extracted features and the performance is analyzed. To faire comparison, the BGWO and firefly feature selection algorithms are also applied to extracted features. Then, the results are classified using DTree, kNN and SVM. In the following, a brief explanation of each algorithm is provided.

The SVM algorithm

The SVM algorithm is an important classification algorithm, which develops a relationship between input and output

based on labeled cases. In this classification algorithm, the output of algorithm is the categorized inputs. This classification algorithm is worked according to nonlinear kernel functions, which are utilized for performing a conversion of initial dataset to a higher-dimensional feature case [30].

The kNN algorithm

The kNN algorithm is a widely used algorithm in image classification. This is a supervised classification algorithm in which the k nearest neighbors of a point is selected using the Euclidean or Mahalanobis distance. For determining the class of an unlabeled instance, this algorithm calculates the distance of this instance, labeled them and identifies its k nearest neighbors and respective labels. Then, this unlabeled instance is classified by a weighted majority or majority voting [31].

The DTree algorithm

The decision tree builds classification models using tree structure. It simultaneously creates a related decision tree and incrementally splits a dataset into smaller and smaller subsets. The outcome is a tree with decision nodes and leaf nodes. A decision node has at least two branches. A leaf node is utilized to represent a classification. In a decision tree, the root node is the topmost node that corresponds to the best predictor. Decision trees can handle both categorical and numerical data [32].

Experimental results and discussion

The developed method is evaluated with MATLAB 2019a. The evaluation is performed using DMR database. The effectiveness measures are classification sensitivity, accuracy,

specificity, Matthews Correlation Coefficient (MCC), and Kappa. Eqs. (17)–(21) display these classification measures.

$$\text{Accuracy} = \frac{TP + TN}{TN + FN + FP + TP} \quad (17)$$

$$\text{Sensitivity} = \frac{TP}{FN + TP} \quad (18)$$

$$\text{Specifity} = \frac{TN}{TN + FP} \quad (19)$$

$$\text{MCC} = \frac{(TP * TN) - (FN * FP)}{\sqrt{(TP + FP) * (TP + FN) * (TN + FP) * (TN + FN)}} \quad (20)$$

$$\text{Kappa} = \frac{2 * ((TP * TN) - (FN * FP))}{(TP + FP) * (FP + TN) + (TP + FN) * (FN + TN)} \quad (21)$$

The True Positive (TP) in these equations indicates a patient who is predicted to have breast cancer and who has actually breast cancer. The True Negative (TN) indicates a patient who is predicted to be healthy and who is actually healthy. The False Positive (FP) indicates a patient who is predicted to have breast cancer, but who is actually healthy, and the False Negative (FN) indicates a patient who is predicted to be healthy, but who has actually breast cancer.

Classification evaluation without feature selection phase

In this section, we done classification without feature selection using the DMR dataset for the DTree, SVM and kNN classifiers. The experimental results are displayed in Figure 5 with regard to specificity, accuracy, and sensitivity.

The experimental results without feature selection show that the DTree classification algorithm provides better

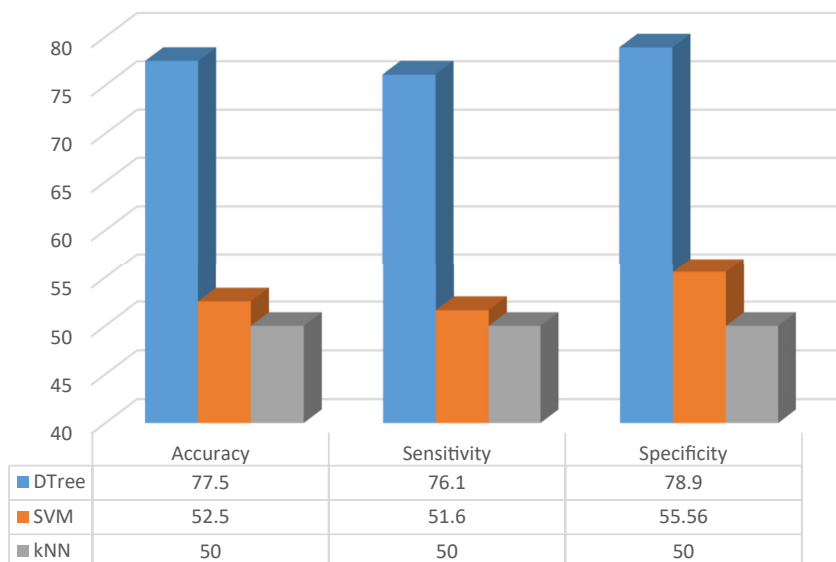


Figure 5: The results without feature selection.

performance with regard to sensitivity, accuracy, and specificity. The kNN classification algorithm has lowest performance, it is because the kNN algorithm does not perform well with high-dimensional data input.

Classification evaluation with feature selection phase

In this section, we have done classification with firefly, BGWO and hybrid of these feature selection algorithms. The experimental results are displayed in Figure 6.

The experimental results using firefly feature selection algorithm show that the DTree and SVM classification algorithms provide better performance with regard to specificity and sensitivity, respectively. In addition, the kNN classification algorithm provides better performance with regard to accuracy, MCC, and kappa.

The experimental results using BGWO feature selection algorithm show that the DTree classification algorithm provide better performance with regard to specificity, sensitivity, and accuracy. In addition, the SVM classification algorithm provides better performance with regard to MCC and kappa in these results.

The experimental results using hybrid of the BGWO and firefly feature selections show that the DTree classification algorithm provide better performance in terms of specificity, sensitivity, accuracy, MCC, and kappa.

Comparative analysis

This section includes a discussion of the comparison results of the performance evaluation metrics based on the previous section's experiments. Figure 7 displays the experimental results for various classifications.

According to Figure 7, the best obtained accuracy, specificity and sensitivity are 97, 96, and 98 %, respectively.

In terms of accuracy, the results show that when feature selection algorithms are used, all of the classifiers outperform their counterparts without feature selection. The results of the accuracy for breast cancer are 97, 96.5, and 84 % for DTree, SVM, and kNN, respectively. In addition, the results of the sensitivity (specificity) for breast cancer are 98 % (96 %), 96.9 % (95.9 %), 82.6 % (85.4 %) for DTree, SVM, and kNN classifier, respectively. It should be noted that the performance of the proposed CAD system is improved by using feature selection algorithms. It is because by using the feature selection algorithms not only the effective features are selected but also the irrelevant and redundant features are removed. In addition, by using

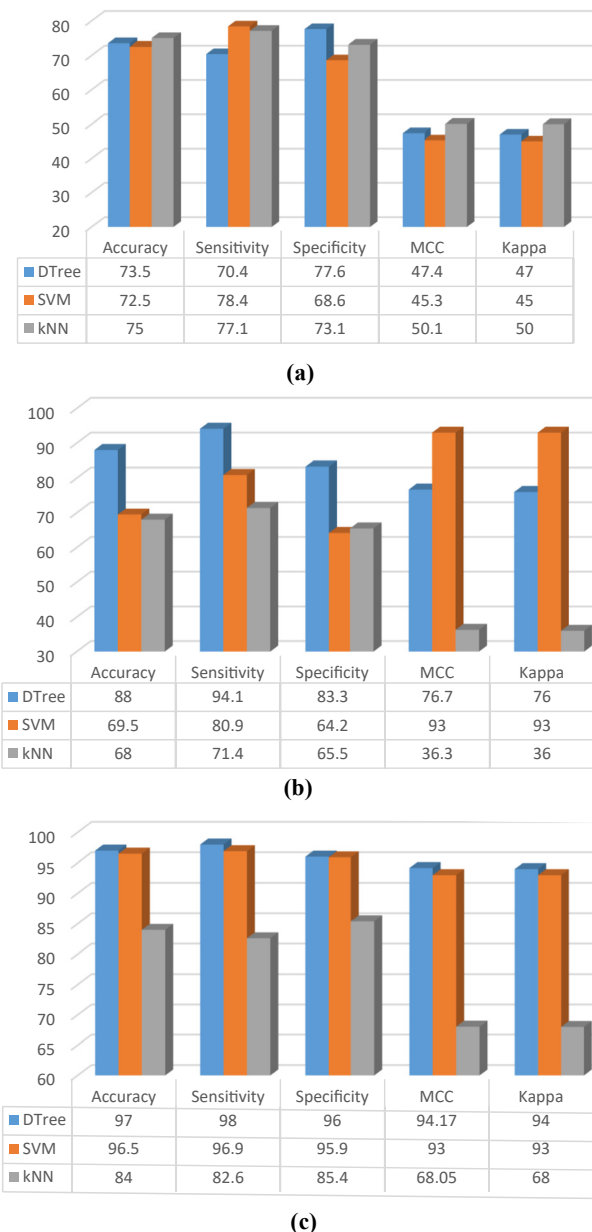


Figure 6: The experimental results using a) firefly feature selection algorithm, b) BGWO feature selection algorithm, c) hybrid of BGWO and firefly feature selection algorithms.

the hybrid feature selection algorithm, the CAD system performance is further improved. It is because the most effective features are selected. Since the kNN has dimensional sensitive, the performance of the CAD system is considerably improved by reducing the dimensions in feature selection phase compared to classification phase without feature selection.

Table 1 summarizes the experimental results with regard to accuracy, specificity and sensitivity compared to other CAD systems.

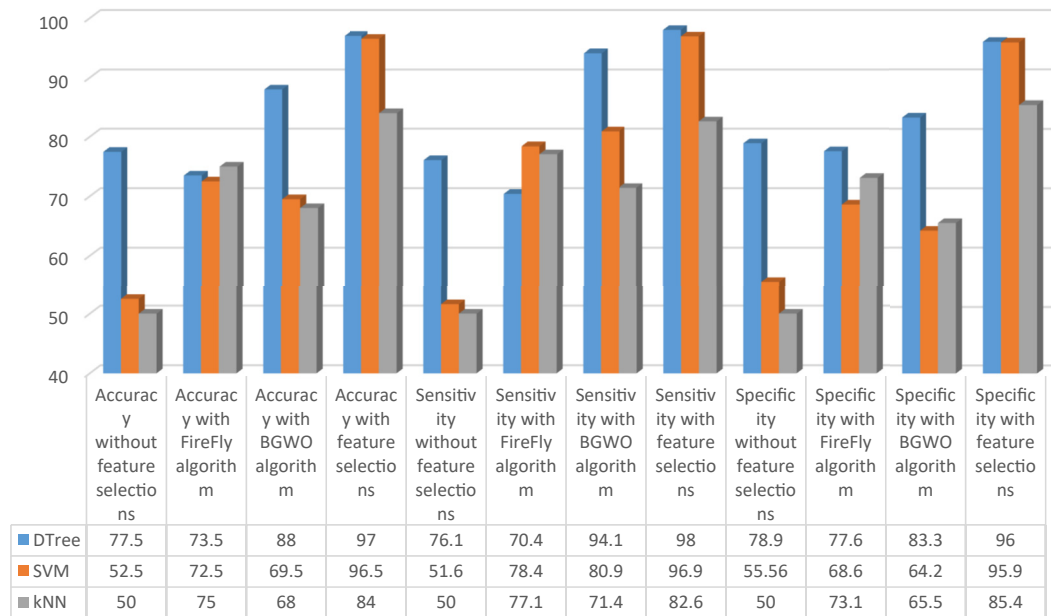


Figure 7: The experimental results of classifications.

Table 1: Results comparison.

Reference	Year	Database	Feature selection method	Classifier	Classifier & performance rate, %
[11]	2021	DMR	GLRLM	kNN	Accuracy=95.45 % Specificity=88.07 % Sensitivity=99.17 %
[26]	2019	Immunofluorescence (IIF)	BGWO	kNN	Accuracy=91.6 %
[14]	2017	DMR	Statistical features	ANN	Accuracy=87.5 % Specificity=93.3 % Sensitivity=80 %
[16]	2021	DMR	BPSO	SVM	Accuracy=96.22 %
[17]	2021	DMR	GA	SVM	Accuracy=95 % Specificity=95 % Sensitivity=95 %
[20]	2021	121 breast thermography images	Hybrid of GA and RSFS	kNN	Accuracy=83.87 % Sensitivity=96 %
[33]	2022	DMR	Grunwald-Ietnikov-aided dragonfly	SVM	Accuracy=97 %
[34]	2023	DMR	ACO and PSO	SVM	Accuracy=97.4 %
[22]	2022	DMR	CNN		Accuracy=96.15 % specificity=62.51 % Sensitivity=98.36 %
This paper	2025	DMR	Hybrid of firefly and BGWO	DTree	Accuracy=97 % Specificity=96 % Sensitivity=98 % MCC=94.17 % Kappa=94 %
				SVM	Accuracy=96.5 % Specificity=95.9 % Sensitivity=96.9 % MCC=93 % Kappa=93 %
				kNN	Accuracy=84 % Specificity=85.4 % Sensitivity=82.6 % MCC=68.05 % Kappa=68 %

Based on these results, the highest accuracy, specificity, sensitivity, MCC, and Kappa obtained by the proposed method for the CAD system are 97, 96, 98, 94.17, and 94 %, respectively, which are acceptable results in comparison with other works for the breast cancer detection using thermography.

Conclusions

Thermography has benefits including painlessness, non-invasiveness, and affordability. It has the potential to provide a significant chance for patient treatment. As a result, it has been identified as a diagnosing breast cancer method. However, feature selection has a vital role in the accurate diagnosis of breast cancer. Using infrared thermal images, this paper presented a new method for improving breast cancer diagnosis. A hybrid feature selection algorithm was used in this method. The developed feature selection algorithm was a hybrid algorithm that utilized the firefly algorithm and BGWO algorithm. Evaluation results indicate that the accuracy, specificity, sensitivity, MCC and kappa in the hybrid feature selection algorithm are 97 , 96, 98, 94.17 and 94 %, respectively. The results of comparing show that the developed CAD system has advantages over other CAD systems.

Research ethics: The local Institutional Review Board deemed the study exempt from review.

Informed consent: Informed consent was obtained from all individuals included in this study.

Author contributions: All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Use of Large Language Models, AI and Machine Learning Tools: None declared.

Conflict of interests: Authors state no conflict of interest.

Research funding: There is no fund.

Data availability: Not applicable.

References

- Kozegar E, Soryani M. A cost-sensitive Bayesian combiner for reducing false positives in mammographic mass detection. *Biomed Eng/Biomed Tech* 2019;64:39–52.
- Cengiz E, Kelek MM, Oğuz Y, Yılmaz C. Classification of breast cancer with deep learning from noisy images using wavelet transform. *Biomed Eng/Biomed Tech* 2022;67:143–50.
- Hakim AS, Awale RN. Extraction of hottest blood vessels from breast thermograms using state-of-the-art image segmentation methods. *Quant InfraRed Thermogr J* 2022;19:347–65.
- Zarei M, Rezai A, Falahieh Hamidpour SS. Breast cancer segmentation based on modified Gaussian mean shift algorithm for infrared thermal images. *Comput Methods Biomech Biomed Eng: Imag & Visualiz* 2021; 9:574–80.
- Singh VP, Srivastava S, Srivastava R. Effective mammogram classification based on center symmetric-LBP features in wavelet domain using random forests. *Technol Health Care* 2017;25:709–27.
- Singh VP, Srivastava R. Content-based mammogram retrieval using wavelet based complete-LBP and K-means clustering for the diagnosis of breast cancer. *Int J Hybrid Intell Syst* 2017;14:31–9.
- Bezdan T, Cvetnic D, Gajic L, Zivkovic M, Strumberger I, Bacanin N. Feature selection by firefly algorithm with improved initialization strategy. 7th Conference on the engineering of computer based systems. Novi Sad, Serbia: Association for Computing Machinery; 2021. Article 8.
- Bezdan T, Cvetnic D, Gajic L, Zivkovic M, Strumberger I, Bacanin N. Feature selection by firefly algorithm with improved initialization strategy. *ECBS 2021: 7th Conference on the Engineering of Computer Based Systems* 2021;1–8. <https://doi.org/10.1145/3459960.3459974>.
- Hu P, Pan J-S, Chu S-C. Improved binary grey wolf optimizer and its application for feature selection. *Knowl Base Syst* 2020;195. <https://doi.org/10.1016/j.knsys.2020.105746>.
- Tilahun SL, Ong HC. Modified firefly algorithm. *J Appl Math* 2012;2012. <https://doi.org/10.1155/2012/467631>.
- Mishra V, Rath SK. Detection of breast cancer tumours based on feature reduction and classification of thermograms. *Quant InfraRed Thermogr J* 2021;18:300–13.
- Rajinikanth V, Kadry S, Taniar D, Damaševičius R, Rauf HT, editors. Breast-cancer detection using thermal images with marine-predators-algorithm selected features. 2021 Seventh International conference on Bio Signals, Images, and Instrumentation (ICBSII) Chennai, India: IEEE; 2021. 25–27 March 2021.
- Sathish D, Kamath S, Prasad K, Kadavigere R. Role of normalization of breast thermogram images and automatic classification of breast cancer. *Vis Comput* 2019;35:57–70.
- Gogoi UR, Bhowmik MK, Ghosh AK, Bhattacharjee D, Majumdar G, editors. Discriminative feature selection for breast abnormality detection and accurate classification of thermograms. 2017 International Conference on Innovations in Electronics, Signal Processing and Communication (IESC) Shillong, India: IEEE; 2017. 6–7 April 2017.
- Silva TAE, Silva LF, Muchaluat-Saade DC, Conci AJS. A computational method to assist the diagnosis of breast disease using dynamic thermography. *Sensors* 2020;20:3866.
- Fikadu M. Thermography breast cancer detection using appropriate segmentation and feature selection approach [Ph.D. thesis]. Bahir Dar: Institute of Technology; 2021.
- Resmini R, Faria da Silva L, Medeiros PRT, Araujo AS, Muchaluat-Saade DC, Conci A. A hybrid methodology for breast screening and cancer diagnosis using thermography. *Comput Biol Med* 2021;135. <https://doi.org/10.1016/j.combiomed.2021.104553>.
- Saadizadeh S, editor. Breast cancer detection in thermal images using GLRLM algorithm. 2021 International Conference on Electrical, Communication, and Computer Engineering (ICEECE); 2021 12–13.
- Salimian M, Rezai A, Hamidpour S, Khajeh-Khalili F. Effective features in thermal images for breast cancer. In: 2nd National Conference on New Technologies in Electrical and Computer Engineering, SID. Isfahan, Iran; 2019.
- Darabi N, Rezai A, Hamidpour SSF. Breast cancer detection USING RSFS-based feature selection algorithms in thermal images. *Biomed Eng: Appl, Basis and Commun* 2021;33:2150020.

21. Kalita D, Singh V, Kumar V. Two way threshold based intelligent water drops feature selection algorithm for accurate detection of breast cancer. *Soft Comput* 2022;26:2277–305.
22. Khafaga DS, Alhussan AA, El-kenawy E-SM, Takieldeens AE, Hassan TM, Hegazy EA, et al. Meta-heuristics for feature selection and classification in diagnostic breast cancer. *Comput Mater & Continua* 2022;73.
23. Moayedi SMZ, Rezai A, Hamidpour SSF. Toward effective breast cancer detection in thermal images using efficient feature selection algorithm and feature extraction methods. *Biomedical Eng: Appl, Basis and Commun* 2024;2450007. <https://doi.org/10.4015/s1016237224500078>.
24. Database D. <https://visual.ic.uff.br/dmi/>.
25. Casagrande L, Macarini LAB, Bitencourt D, Fröhlich AA, de Araujo GM. A new feature extraction process based on SFTA and DWT to enhance classification of ceramic tiles quality. *Mach Vision and Appl* 2020;31: 71.
26. Devanathan K, Ganapathy N, Swaminathan R. editors. Binary grey wolf optimizer based feature selection for nucleolar and centromere staining pattern classification in indirect immunofluorescence images. 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2019. 23–27 July 2019.
27. Yang X-S, editor. Firefly algorithms for multimodal optimization. In: *Stochastic Algorithms: Foundations and Applications*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2009.
28. Kalita DJ, Singh VP, Kumar V. Detection of breast cancer through mammogram using wavelet-based LBP features and IWD feature selection technique. *SN Comput Sci* 2022;3. <https://doi.org/10.1007/s42979-022-01071-7>.
29. Singh VP, Srivastava S, Srivastava R. Automated and effective content-based image retrieval for digital mammography. *J X-ray Sci Technol* 2018;26:29–49.
30. Yue S, Li P, Hao P. SVM classification: Its contents and challenges. *Appl Mathematic-A J Chin Univers* 2003;18:332–42.
31. Guo G, Wang H, Bell D, Bi Y, Greer K, editors. KNN Model-Based Approach in Classification. *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2003.
32. Yang Y. An evaluation of statistical approaches to text categorization. *Inform Retrieval* 1999;1:69–90.
33. Chatterjee S, Biswas S, Majee A, Sen S, Oliva D, Sarkar R. Breast cancer detection from thermal images using a Grunwald-Letnikov-aided Dragonfly algorithm-based deep feature selection method. *Comput Biol Med* 2022;141. <https://doi.org/10.1016/j.compbiomed.2021.105027>.
34. Kalaiyarasi V, Meenakshi JS, Jain S, Umapriya R, Sarala R. Bio-inspired optimization technique for feature selection to enhance accuracy of BC detection. 2023 International Conference on Inventive Computation Technologies (ICICT); 2023. 26–28 April 2023.