ARTICLE IN PRESS

Alexandria Engineering Journal (2021) xxx, xxx–xxx

HOSTED BY

ELSEVIER

Alexandria University

Alexandria Engineering Journal

www.elsevier.com/locate/aej
www.sciencedirect.com

New Sequential and Parallel Support Vector Machine with Grey Wolf Optimizer for Breast Cancer Diagnosis

Elsayed Badr [a,d], Sultan Almotairi [b], Mustafa Abdul Salam [c], Hagar Ahmed [a,*]

[a] *Scientific Computing Department, Faculty of Computers and Artificial Intelligence, Benha University, Benha, Egypt*
[b] *Department of Natural and Applied Sciences, Community College, Majmaah University, Al-Majmaah 11952, Saudi Arabia*
[c] *Artificial Intelligence Department, Faculty of Computers and Artificial Intelligence, Benha University, Benha, Egypt*
[d] *Higher Technological Institute, 10th of Ramadan City, Egypt*

Received 8 December 2020; revised 31 May 2021; accepted 9 July 2021

KEYWORD

Machine learning;
Support vector machine;
Grey Wolf optimizer;
Scaling techniques;
Breast cancer;
Parallel processing

Abstract Breast cancer is one of the most common types of cancer worldwide. Early detection of cancer increases the probability of recovery. This work has three contributions. The first contribution is improving the performance of support vector machine (SVM) using a recent grey wolf optimizer (GWO) for diagnosis breast cancer with efficient scaling techniques. The second contribution is proposing three efficient scaling techniques against the classical normalization technique. The last contribution is using a parallel technique which applies task distribution to improve the efficiency of GWO. The proposed sequential model is applied on two different datasets, Wisconsin diagnosis breast cancer (WDBC) dataset and Electronic Health Records (EHR). Experimental results of WDBC show that the proposed hybrid GWO-SVM model achieves 98.60% with normalization scaling. Also, using the proposed scaling techniques with the proposed GWO-SVM model gives a fast convergence and achieves accuracy rate by 99.30%. The parallel version of the proposed model achieves a speedup by 3.9 on four CPU cores. On the other hand, Experimental results of EHR show that the proposed hybrid GWO-SVM model achieves 93.26% with normalization scaling against 82.05 for SVM.

1. Introduction

Breast cancer is the most common cancer of 11.6% among males and females of all ages. It is the most spreading in women worldwide, accounting 24.2% of the whole cases diagnosed in 2018 [1].

Enhancing the accuracy of detecting the breast cancer disease is very important task, and early and accurate detection

* Corresponding author.
E-mail addresses: almotairi@mu.edu.sa (S. Almotairi), hagar.abdelhalim@fci.bu.edu.eg (H. Ahmed).

Peer review under responsibility of Faculty of Engineering, Alexandria University.

https://doi.org/10.1016/j.aej.2021.07.024
1110-0168

Please cite this article in press as: E. Badr et al., New Sequential and Parallel Support Vector Machine with Grey Wolf Optimizer for Breast Cancer Diagnosis, Alexandria Eng. J. (2021), https://doi.org/10.1016/j.aej.2021.07.024

can save many lives. Defects in breast cancer diagnosis by experts can be avoided by expert systems and artificial intelligence techniques. These expert systems can examine the medical data in a short time and help junior physicians. Many artificial intelligence techniques were used to detect this type of cancers with high accuracies. C4.5 decision tree algorithm with accuracy 94.74% [2]. RIAC method achieved 94.99% accuracy with ten folds cross validation [3]. Linear discriminant analysis (LDA) method with 96.8% accuracy [4]. A diagnoses model with support vector machine (SVM), and the reported accuracy was 97.2% [5]. Neuro-fuzzy technique, the reported accuracy was 95.06% [6]. Performance of 97.36% was achieved by using fuzzy-GA method [7]. Three different methods, AIRS, big LVQ and optimized LVQ are proposed with 97.2%, 96.8% and 96.7% accuracies respectively [8]. The application of supervised fuzzy clustering method achieving performance 95.57% [9]. Accuracy 98.85% was obtained with the mixture expert's network structure for breast cancer detection [10]. A fuzzy-AIS and KNN (Fuzzy-AIS-KNN) are presented, with accuracy 99.14% [11]. A MLP-NN, four models, combined ANN, PNN, RNN with highest classification accuracy of 97.36% was achieved by SVM [12].

LS-SVM was used with 98.53% accuracy [13]. SVM-hybridized with F-score method was used with 99.51% accuracy [14]. Accuracy of 99.08% by ANFIS model for breast cancer [15]. The hybrid method integrating association rules and ANNs was introduced by with 97.4% accuracy [16]. SBS algorithm integrating with BPNN and LM, BPNN and PSO achieved 98.83%, 97.51% classification accuracy, respectively [17]. The Artificial Meta-plasticity MLP (AMMLP) algorithm was introduced with 99.26% accuracy [18]. RS-SVM classifier for breast cancer diagnosis with classification accuracy of 100% and 96.87% for the highest and the average respectively [19]. PSO-SVM with accuracy 99.3% [20]. An approach using GA-based on feature selection, and achieved 96.9% accuracy [21]. A comparison among six ML techniques: GRU-SVM, LR, MLP, KNN, Soft-max Regression, and SVM achieved 99.04% [22]. Accuracy of 98.24% using genetic programming and machine learning algorithms [23].

The scaling techniques can improve the classification accuracy and convergence speed. Ten efficient scaling techniques were proposed for optimizing SVM [24]. Shen et al. [25] studied the Support Vector Machine (SVM) algorithm with the Fruit-fly Optimization Algorithm (FOA) in various medical datasets such as Wisconsin breast cancer dataset, Pima Indians diabetes dataset, Parkinson's dataset, and thyroid disease diagnosis, got from UCI repository. The ML SVM technique is hybridized with Particle Swarm Optimization Algorithm based SVM (PSO-SVM), Genetic Algorithm-based SVM (GA-SVM), Bacterial Forging Optimization-based SVM (BFO-SVM), and Grid Search Technique-based SVM (Grid-SVM), and implemented with tools like MATLAB and LibSVM. 10-fold cross-validation technique was used. The SVM-FOA gives the highest accuracy as 96.9%, 77.46%, 77.46%, and 96.38% in Wisconsin dataset, Pima dataset, Parkinson dataset, and thyroid dataset respectively. A. Darwish et al. [26] presents a two-step system that first uses four different swarm algorithms namely; whale optimization algorithm (WOA), grey wolf optimizer (GWO), flower pollination algorithm (FPA), and moth flame optimization (MFO) for feature selection purpose. Then, several classifiers are applied including support vector machines, k-nearest neighbor, and decision tree. The experimental using WDBC and WPBC datasets outcomes positively that the proposed system was effective in undertaking breast cancer data classification and features selection tasks by accuracy 98.77% and 84.34% respectively. S. Kamel et al. [27] used data mining as a combination of feature selection method by Gray Wolf Optimization (GWO) and support vector machine (SVM), WDBC dataset were applied to evaluate the proposed method and assess the validity of the results in MATLAB. Application of the proposed method increased the improvement of the evaluated criteria, which increased the accuracy of diagnosis by 27.68%, compared to former works in the field. Kumar and Singh [28] proposed enhanced grey wolf optimization-support vector machine approach (EGWO-SVM) for breast cancer diagnosis. Their approach achieved 98.24% accuracy rate for Wisconsin Diagnostic Breast Cancer (WDBC) database. A. Rahmani et al. [29] presented a new method for minimizing the process of breast cancer diagnosis through the Grasshopper optimization algorithm to reduce the features then select the optimal features and improve the parameters using the SVM Classifier. The experiments in this study were performed on three datasets, namely WBC, WDBC and WPBC by accuracy 99.51, 98.83 and 91.38 respectively. Grey wolf optimizer (GWO) is one of the recently proposed swarm intelligence-based algorithms, which is developed by Mirjalili et al. [30] in 2014 GWO has been widely tailored for a wide variety of optimization problems due to its impressive characteristics over other swarm intelligence methods: it has very few parameters, and no derivation information is required in the initial search.. Faris et al. [31] proposed a review of recent variants and applications using GWO. They introduced several research publications using GWO have been overviewed and summarized. Al-Betar et al. [32] proposed six versions of GWO which are Greedy-based GWO (GGWO), Proportional-based GWO (PGWO), Tournament-based GWO (TGWO), Universal sampling-based GWO (UGWO), Linear rank-based GWO (LGWO), Random-based GWO (RGWO). The six versions are evaluated using 23 test functions circulated in the literature with different characteristics and complexity.

These scaling techniques are efficient for linear programming approach [33–43]. The scaling techniques that they applied with SVM on WDBC dataset are arithmetic mean, de Buchet for three cases (p = 1, 2″ and infinity“), de Buchet (p = ”infinity”), equilibration, geometric mean, IBM MPSX, Lp-norm for three cases (p = 1, 2″ and infinity”).

In [44] a parallel swarm technique applied for two-sided balancing problem was introduced. The parallel method was implemented to datasets for Massive passing which was introduced in [45]. In [46] parallel dynamic programming algorithms were introduced and discussed. A survey of many methods for algorithms' parallelization was discussed in [47]. In [48] parallel approach for constraint solving methods is introduced. Three techniques for enhancing the classical methods were introduced [49]. First technique is neighborhood search for reducing the solution space is used. Second technique is reducing calculations time by reducing the solutions space. The last technique is integrating of the two mentioned methods.

Optimizing SVM and extreme learning machine using swarm intelligence algorithms such as PSO, ABC, FPA, BA, and MFO, improve the performance of classical SVM model and overcome the local minima and overfitting problems easily [50–55].

In this paper, the performance of SVM was improved using GWO algorithm for efficient detection of breast cancer.

We used GWO algorithm because it has few number of parameters, it is simple to understand and implement as also as it results high classification accuracy, scalable, and it provides faster convergence by maintaining the right balance between the exploration and exploitation during the search. In addition to applying the parallel technique to accelerate the performance of GWO, the proposed sequential model was applied to WDBC breast cancer dataset that available from the Wisconsin from UCI with a total 569 instances and 33 attributes.

In this paper, three contributions are proposed. The first contribution is improving the performance of support vector machine (SVM) using a recent grey wolf optimizer (GWO) for diagnosis breast cancer. The second contribution is proposing three efficient scaling techniques against the classical normalization technique. The last contribution is using a parallel technique which applies task distribution to improve the efficiency of GWO. The proposed sequential model is applied on two different datasets, Wisconsin diagnosis breast cancer (WDBC) dataset and Electronic Health Records (EHR). Experimental results of WDBC show that the proposed hybrid GWO-SVM model achieves 98.60% with normalization scaling. Also, using the proposed scaling techniques with the proposed GWO-SVM model gives a fast convergence and achieves accuracy rate by 99.30%. The parallel version of the proposed model achieves a speedup by 3.9 on four CPU cores. On the other hand, Experimental results of EHR show that the proposed hybrid GWO-SVM model achieves 93.26% with normalization scaling against 82.05 for SVM.

The rest of this paper is organized as follows. The algorithms that are used in the study: SVM, GWO are described in Section 2. The proposed sequential and parallel hybrid models are introduced in Section 3. In Section 4, a detailed description of the new scaling techniques, equilibration, arithmetic mean and geometric mean are presented. In Section 5, Experimental Design is introduced. In Section 6, Experimental results and discussions are presented. Finally, conclusions and future works are introduced in Section 7.

2. Preliminaries

In this section, Support vector machine (SVM), and grew wolf optimizer (GWO) are presented and discussed.

2.1. Support vector machine (SVM)

Support vector machine (SVM), was introduced by Vapnik [56,57]. SVM is based on the Vapnik-Chervonenkis (VC) theory and structural risk minimization (SRM) principle. The goal of SVM is the specifying of a hyperplane in an N-dimensional space which can easily classify the available data vectors. SVM uses convex quadratic programming, which avoid local minima [58].

If we have a binary classification problem: having training dataset with class label: $(x_1, y_1), \ldots, (x_n, y_n), x_i \in R_d$ *and* $y_i \in (-1, +1)$, where x_i is an input or features vector and y_i the label of the class. The optimal hyperplane is:

$$wx^T + b = 0 \quad (1)$$

where w, x, and b are the weight, the input vector, the bias respectively. w And b satisfy:

$$wx_i^T + b \geq +1 \quad if y_i = 1 \quad (2)$$

$$wx_i^T + b \leq -1 \quad if y_i = -1 \quad (3)$$

where $y_i$is the class of each input vector

SVM model training objective is to specify w and b which maximizes the margin$\frac{1}{\|w\|^2}$.

Usually the problem is non-linearly separable. To convert the non– linear to linear, the input space is mapped into higher dimensional space. SVM uses kernels to model higher dimensional, non-linear models [59]. Kernel functions could be used to increase dimensions to the data and make it a linear problem. Linear and non-linear are shown in Fig. 1. On the other hand, the kernel functions could help in accelerating the calculations in high dimensional space. Such as, the linear kernel like the dot product of two features in extended feature space. RBF and polynomial kernels are the most common SVM kernels, which are respectively defined as:

$$K(x_i, x_j) = e^{-\gamma\|x_i - x_j\|^2} \quad (4)$$

$$K(x_i, x_j) = (1 + x_i^T x_j)^p \quad (5)$$

where $\|x_i - x_j\|^2$is recognized as the squared Euclidean distance and γ is the predefined parameter controlling the width of the Gaussian kernel and p is the polynomial order. It has been proved that proper model parameters setting can improve the SVM classification accuracy.

SVM parameters tuning is very sensitive part. SVM parameters as follows: (1) C parameter, which make the balance between the cost minimizing and model complexity; (2) gamma parameter which determines mapping from nonlinear to linear by raising the dimension; (3) the SVM kernel function, which responsible for building a non-linear decision hyperplane [60].

2.2. Grey wolf optimizer (GWO)

Grey Wolf Optimization (GWO) algorithm is a recent *meta*-heuristic algorithm [30]. GWO was proposed by Mirjalili et al. in 2014, which has a fast convergence to global optimum. GWO imitates the grey wolf's hunting mechanism. The grey

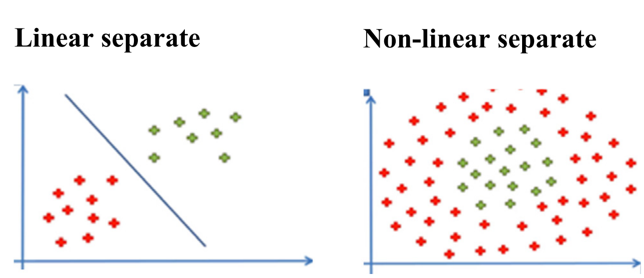

Fig. 1 Linear and nonlinear SVM.

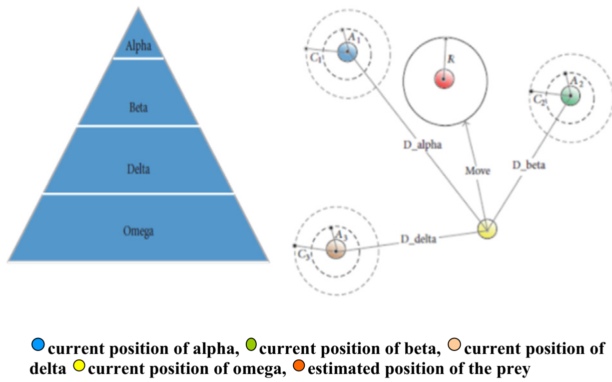

Fig. 2 GWO algorithm.

wolf population has four grades: α, β, δ, and ω. to mathematically model the grey wolves, suppose the best solution as *alpha*, and the second and third best solutions are *beta* and *delta*, respectively. The rest are named omega. The social hierarchy of GWs is shown in Fig. 2. GWs embracing prey during hunting. The embracing behavior of GWOs are defined as follow:

$$Dist = |Cg \cdot X_p(t) - X(t)| \tag{6}$$

$$X(t+1) = X_p(t) - A \times Dist \tag{7}$$

where $Dist$ the distance between the GW and the prey, t is the iteration number. X_p Is the prey position. X is the GW position vector. The vectors A and Cg are calculated as follows:

$$A = 2ar_1 - a \tag{8}$$

$$Cg = 2r_2 \tag{9}$$

where a is a vector where its values are linearly decreased from 2 to 0 during the course of run and among them r_1 and r_2 are both random vectors in the interval of [0, 1].

The GWO optimization is to score the prey location by α, β and δ wolves. Rest of wolves use the location as a reference and update their locations around the prey randomly. α, β, and δ wolves as shown in Fig. 2. GWs position updating is shown in Eqs. (10)–(16)

$$D_\alpha = |C_1 \cdot X_\alpha(t) - X(t)| \tag{10}$$

$$D_\beta = |C_2 \cdot X_\beta(t) - X(t)| \tag{11}$$

$$D_\delta = |C_3 \cdot X_\delta(t) - X(t)| \tag{12}$$

$$X_1 = X_\alpha(t) - A_1 \times D_\alpha \tag{13}$$

$$X_2 = X_\beta(t) - A_2 \times D_\beta \tag{14}$$

$$X_3 = X_\delta(t) - A_3 \times D_\delta \tag{15}$$

$$X(t+1) = (X_1 + X_2 + X_3) \tag{16}$$

The main objective of this research is to determine the optimal parameters of SVM by using GWO algorithm to classify the breast cancer data with high accuracy.

3. The proposed GWO-SVM classification model

3.1. The proposed sequential hybrid GWO-SVM model

GWO-SVM classification model for breast cancer diagnosis is proposed. The proposed classification model has two phases. In the first phase, SVM parameters are automatically adapted by GWO algorithm. In the second phase, the optimized SVM model performs the classification tasks. Ten-fold CV was applied to guarantee the optimal results and best accuracy.

GWO algorithm takes into consideration root mean square error (RMSE) as the fitness function was used to evaluate the best parameters of SVM. RMSE is defined as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} Predicted_i - Actual_i}{N}} \tag{17}$$

where N represent no. of observations in test dataset.

In GWO-SVM model for breast cancer, it initialize the number of population size with n which each grey wolf represent $X_i(i = 1, 2, \cdots., n)$, maximum number of iterations with Max_iter, number of features with dim which grey wolf search in it for prey, upper bound with ub, lower bound with lb, which are the boundary of positions, parameter a that linearly degrading from (2 to 0) over the iterations, A and Cg that are depend on a parameter value, $X_\alpha, X_\beta, X_\delta$are the positions of alpha, beta and delta wolves respectively where all wolves update their positions according the position of these three wolves and number of folds with k for cross validation then it generate the initial population ($n \times dim$) with random values and load the data and apply one of scaling techniques or normalization technique on it. In order to guarantee the effectiveness of the model then it apply k-folds cross validation and it do the following for each fold.

The model check if the number of iteration, does not reach toMax_iter, it pass each agent to specific two functions and set its output to parameters of SVM (C andγ) and train SVM and classify test set then it calculate the fitness function ($RMSE$) from Eq. (17) and update$X_\alpha, X_\beta, X_\delta$, where the least fitness is alpha value.

The model update the parameter a and for all dimensions in all agents, update A, Cg and the position of current agent according to$X_\alpha, X_\beta, X_\delta$, A and Cg values.

After that, it increase the number of iteration and check if the number of iteration, does not reach to Max_iter it go to Step 6. If Max_iter satisfies, it move to next fold and return to Step 4. If Max_iter and fold number k satisfy, it calculate the average $RMSE$ and average accuracy of k-folds. Finally, return mean $RMSE$ and mean classification accuracy. The algorithm in detail is explained in (Algorithm 1).

Algorithm1: The sequential hybrid GWO_SVM for Breast Cancer

Input: *SearchAgents_no* Number of search agents
Max_iter Number of Iterations
lb Lower Bound
ub Upper Bound
dim No. of features
Output : Average RMSE of SVM
Average classification accuracy rates of SVM over k testing set
1. Initialization:
- Agents size (n) and the GW population $X_i (i = 1, 2, ..., n)$.
- Parameter a, coefficient vector (A and Cg).
- *Max_iter*,*lb* and *ub*.
- The best search agent (X_α)
- The second best search agent (X_β)
- The third best search agent (X_δ)
- K-folds2. Generate the initial population randomly.

3. Load the data and apply one of scaling techniques on it.
4. For j = 1: k
Divide data randomly to train and test sets
5. while ($t < Max\_iter$)
6. for each search agent do
Pass it to specific functions and set its output to parameter of SVM (C, γ)
Train and test SVM model
7. Evaluate the search agent fitness by Eq. (17)
8. Update X_α, X_β, X_δ .based on fitness value
end for
9. Update a
for each search agent do
for each dimension do
Update A and C.
Update the position of the current search agent
end for
end for
10. t = t + 1
end while
t = 0
end for11. Calculate the average RMSE and accuracy of k-folds in each iteration.
12. Return RMSE and classification accuracy.

3.2. The parallel hybrid GWO_SVM model

Parallel algorithm can achieve high speed up and low execution time. Parallelism can be done by dividing the population into several groups, with multithreading.

Let that at the algorithm's start, the number of cores $N_c$is identified. An initial population contains n individuals is randomly initialized. The size of the group will be computed as follows:

$$n_g = \left[\frac{n}{N_c}\right] \tag{18}$$

The steps of the parallel hybrid GWO-SVM is shown in (Algorithm 2).

Algorithm 2: Parallel GWO-SVM

1: Begin,
2: Identifying N_c(no. of cores),
3: Randomly initializing the population,
4: compute n_g individuals with Eq. (18),
5: Make $N_c$groups
6: Divide the individuals on cores
7: Apply GWO-SVM Model on Each core
8: Select the optimal individuals from all threads,
9: Update the model's parameters and Position of individuals
10: Return Average classification accuracy for all folds
11: End.

4. Scaling techniques

Here, we introduce the mathematical notations of ten scaling techniques in addition to the normalization scaling techniques with ranges [0, 1] and [−1, 1]. First of all, we introduce the following mathematical preliminaries as shown in Table 1.

The scaled matrix is expressed as RAS, such that $R = diag(r_1, \cdots, r_m)$ and$S = diag(s_1, \cdots, s_n)$. All scaling techniques proposed in this section apply first rows scaling and after that columns scaling. Then, the matrix after full scaling (row and column) is given by:

$$A^R = RA; A^{RS} = A^R S \tag{19}$$

(1) Arithmetic Mean Scaling Technique

Arithmetic mean aims to decrease the variance between the nonzero elements in the coefficient matrix$A$. Eq. (20) represents the rows scaling such that each row (instance) is divided

Table 1 Mathematical preliminaries for scaling techniques.

Symbol	Description
$A(a_{ij})$	$m \times n$matrix (with m (observations) and n (attributes))
a_{ij}	The matrix element in $row_i$and$column_j$
r_i	The scaling agent of row$i$
s_j	The scaling agent of row$j$
R	Diagonal matrix such that$R = diag(r_1, \cdots, r_{m)}$
S	Diagonal matrix such that$S = diag(s, \cdots, s_{n)}$
N_i	$N_i = \{j \mid Aij \neq 0\}, such that 1 \le i \le m$
M_j	$M_j = \{i \mid Aij \neq 0\}, such that 1 \le j \le n$
n_i	The number of elements for the set N_i
m_j	The number of elements for the set M_j
a^R_{ij}	The matrix element in $row_i$and $column_j$ after row scaling
a^{RS}_{ij}	The matrix element in $row_i$and $column_j$ after row and column scaling
$A^R(a^R_{ij})$	The scaled matrix by row R scaling agent
$A^{RS}(a^{RS}_{ij})$	The final scaled matrix.

by the mean of the absolute value of the non-zero values (instance):

$$r_i = \frac{n_i}{\sum_{j \in N_i} a_{ij}} \tag{20}$$

Eq. (21) represents the columns scaling such that each column (attribute) is divided by the arithmetic mean of the absolute value of the nonzero elements in that column (attribute).

$$s_j = \frac{m_i}{\sum_{i \in M_j} a_{ij}{}^R} \tag{21}$$

(2) Equilibration Scaling Technique

The biggest value in absolute value is the corner stone for this scaling method. Each row of the matrix A is divided by the biggest value in absolute value in that row. Then, each column of the scaled matrix A by the row factor divided by the biggest value in absolute value in that column. The range of the final scaled matrix A is [−1, 1].

(3) Geometric Mean Scaling Technique

First, Eq. (22) represents the rows scaling such that each row (instance) is divided by the geometric mean of the absolute value of the non-zero elements in that row (instance).

$$r_i = \left(\max_{j \in N_i} a_{ij} \min_{j \in N_i} a_{ij} \right)^{-1/2} \tag{22}$$

Second, Eq. (23) represents the columns scaling such that each column (attribute) is divided by the geometric mean of the absolute value of the non-zero elements in that column (attribute).

$$s_j = \left(\max_{j \in M_j} a_{ij}^R \min_{j \in M_j} a_{ij}^R \right)^{-1/2} \tag{23}$$

(4) Normalization scaling technique

It is also known as min-max scaling or min-max normalization, is the simplest method and consists in rescaling the range of features to scale the specific range. Selecting the target range depends on the nature of the data. The general formula for a min-max of [0, 1] is given in Eq. (24).

$$x_{new=} \frac{x - x_{min}}{x_{max} - x_{min}} \tag{24}$$

where x is an original value, x_{new} is the normalized value, x_{min} and x_{max} are the minimum and maximum values of the dataset respectively. To rescale a range between an arbitrary set of values [a, b], the formula becomes as in Eq. (25).

$$x_{new} = a + \frac{(x - x_{min})(b - a)}{x_{max} - x_{min}} \tag{25}$$

5. Experimental design

In this section, we introduce data description, measure for performance evaluation and the comparative study.

5.1. Experimental setup

The proposed GWO-SVM detection model was developed by MATLAB. SVM, implementation was enhanced, which is originally developed by Chang and Lin [61]. Table 2 describes the experiments computing environment.

Salzberg [62] introduced the k-fold CV which is used to guarantee the valid results. In this paper, $k = 10$. The detail parameters for GWO-SVM are set as follows. The number of iterations, search agents, dimensions and k-fold are set to 1000 and 19, 25 and 10 respectively but the steady state occurs at about 317, 155, 551 and 146 iterations for normalization, arithmetic mean, Geometric mean, and Equilibration scaling techniques respectively. The lower bound and upper bound $[lb, ub]$ is set as $[-5, 5]$. Finally, the parameter a is calculated by:

Table 2 Description of the computing environment.

CPU	Intel (R) Core (TM) i5- 7200U CPU@ 2.70 GHz
RAM Size	4 GB RAM
MATLAB version	R2015a (8.5.0.197613)

Table 3 WDBC Dataset Description.

Attribute name	ID
Patient ID	A1
Outcome	B1
RADIUS1	C1
TEXTURE1	D1
PERIMETER1	E1
AREA1	F1
SMOOTHNESS1	G1
COMPACTNESS1	H1
CONCAVITY1	I1
CONCAVEPOINTS1	J1
SYMMETRY1	K1
FRACTALDIMENSION1	L1
RADIUS2	M1
TEXTURE2	N1
PREIMETER2	O1
AREA2	P1
SMOOTHNESS2	Q1
COMPACTNESS2	R1
CONCAVITY2	S1
CONCAVEPOINTS2	T1
SYMMETRY2	U1
FRACTALDIMENSION2	V1
RADIUS3	W1
TEXTURE3	X1
PREIMETER3	Y1
AREA3	Z1
SMOOTHNESS3	AA1
COMPACTNESS3	AB1
CONCAVITY3	AC1
CONCAVEPOINTS3	AD1
SYMMETRY3	AE1
FRACTALDIMENSION3	AF1

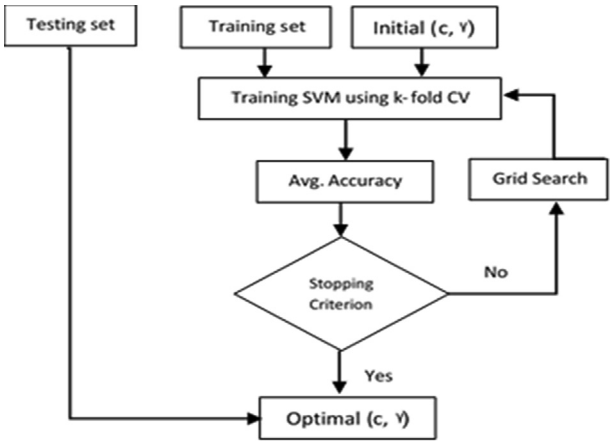

Fig. 3 SVM using grid search.

$$a = 2 - \left(t \times \frac{2}{Max_{iter}}\right) \quad (26)$$

where t is current iteration

5.2. Data description

In this work, we have run the proposed model on the Wisconsin diagnosis Breast Cancer (WDBC) dataset that available the UCI Machine Learning Repository [48]. The dataset consists of 569 instances divided into two classes.

The two classes malignant and benign have 357 and 212 cases respectively. Each record in the database has thirty-two attributes. The thirty two attributes listed in Table 3.

5.3. Measure for performance evaluation

In order to test the performance of the proposed GWO-SVM model, we use sensitivity, specificity, accuracy, precision, G-mean and F-score. According to the confusion matrix, sensitivity, specificity, accuracy, precision, G-mean and F-score are defined as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (27)$$

$$Sensitivity = \frac{TP}{TP + FN} \times 100 \quad (28)$$

$$Specificity = \frac{TN}{TN + FP} \times 100 \quad (29)$$

$$Precision = \frac{TP}{TP + FP} \times 100 \quad (30)$$

$$Gmean = \sqrt{Sensitivity \times Specificity} \quad (31)$$

$$Fmeasure = 2 \times \frac{Precision \times Sensitivity}{Precision + Sensitivity} \quad (32)$$

5.4. Comparative study

In this study, we compare the performance of the proposed GWO-SVM system and SVM using grid search technique. The best C and γ are computed by grid search. Fig. 3 shows steps for training SVM using grid search. The searching space of parameters C and γ are set to C = $\{2^{-5}, 2^{-3} \ldots 2^{15}\}$ and $\gamma = \{2^{-15}, 2^{-13}, \cdots, 2^{1}\}$, respectively.

6. Experimental results and discussions

In order to evaluate the efficiency of the proposed GWO-SVM model for breast cancer, we do experiments on the WDBC dataset. First of all, our results show the effectiveness of grid search techniques, the efficiency of the sequential proposed GWO-SVM model and the superiority of the newest scaling techniques which competitive the traditional normalization technique. Finally, the experimental results show that the parallel version of the proposed model achieves a speedup by 3.9 for four cores.

Table 4, Table 5 and Table 6 show a comparison among classification accuracies of SVM with normalization scaling [0, 1], without scaling, normalization scaling [−1, 1] arithmetic mean scaling, geometric mean scaling and equilibration scaling. It is apparent from Table 4 and Table 5 that the average accuracy rates achieved by SVM with geometric mean scaling technique (98.59%) is better than that obtained by SVM with normalization scaling techniques (96.49%) and (96.66%). On the other hand, the equilibration scaling technique overcomes

Table 4 Accuracy for WDBC database using SVM with grid search technique (Without scaling and Normalization scaling [0, 1]).

Fold	Without scaling (S0)			Normalization scaling [0,1] (S1)		
	C	γ	Acc.%	C	γ	Acc.%
1	2^{3}	2^{-13}	94.76	2^{13}	2^{-7}	100
2	2^{7}	2^{-15}	91.59	2^{15}	2^{-9}	98.25
3	2^{15}	2^{-13}	100	2^{15}	2^{1}	92.98
4	2^{5}	2^{-13}	97.18	2^{15}	2^{-1}	94.74
5	2^{1}	2^{-11}	96.23	2^{15}	2^{-1}	94.74
6	2^{-1}	2^{-9}	91.29	2^{15}	2^{1}	96.49
7	2^{11}	2^{-15}	97.59	2^{15}	2^{-3}	98.25
8	2^{9}	2^{-15}	98.6	2^{15}	2^{-13}	96.49
9	2^{9}	2^{-15}	97.59	2^{15}	2^{1}	94.74
10	2^{15}	2^{-9}	96.23	2^{15}	2^{1}	98.25
Avg.	6878	0.005	96.1	30310.4	0.91	96.49
Time	52.62167			14.448410		

Table 5 Accuracy for WDBC database using SVM with grid search technique (Normalization scaling [−1, 1] and Arithmetic mean scaling).

Fold	Normalization scaling [-1,1] (S2)			Arithmetic mean scaling (S3)		
	C	γ	Acc.%	C	γ	Acc.%
1	2^{11}	2^{1}	94.64	2^{3}	2^{-7}	100.00
2	2^{15}	2^{1}	92.98	2^{15}	2^{-9}	98.25
3	2^{13}	2^{1}	100	2^{9}	2^{-5}	96.49
4	2^{13}	2^{1}	98.25	2^{-1}	2^{-5}	96.49
5	2^{15}	2^{1}	96.49	2^{9}	2^{-9}	100.00
6	2^{15}	2^{-1}	96.49	2^{5}	2^{-5}	98.25
7	2^{13}	2^{1}	100	2^{7}	2^{-7}	98.25
8	2^{15}	2^{1}	96.49	2^{-1}	2^{-3}	98.25
9	2^{13}	2^{1}	94.74	2^{9}	2^{-9}	100.00
10	2^{13}	2^{-1}	96.49	2^{15}	2^{-9}	98.25
Avg.	17,408	1.7	96.66	6724	0.024	98.42
Time	19.208797			12.516496		

Table 6 Accuracy for WDBC database using SVM with grid search technique (Geometric mean scaling and Equilibration scaling).

Fold	Geometric mean scaling (S4)			Equilibration scaling (S5)		
	C	γ	Acc.%	C	γ	Acc.%
1	2^{1}	2^{-5}	100	2^{5}	2^{-1}	100.00
2	2^{9}	2^{-5}	98.25	2^{3}	2^{1}	98.25
3	2^{9}	2^{-5}	96.49	2^{5}	2^{-1}	100.00
4	2^{-1}	2^{-5}	96.49	2^{15}	2^{1}	98.25
5	2^{9}	2^{-9}	100	2^{1}	2^{-1}	100.00
6	2^{7}	2^{-5}	98.25	2^{9}	2^{-1}	98.25
7	2^{3}	2^{-3}	100.00	2^{15}	2^{1}	100.00
8	2^{15}	2^{-3}	98.25	2^{15}	2^{1}	100.00
9	2^{9}	2^{-9}	100	2^{3}	2^{1}	94.74
10	2^{5}	2^{-3}	98.25	2^{3}	2^{1}	100.00
Avg.	34,987	0.5352	98.59	9891	1.4	98.95
Time	15.143076			10.175330		

Table 7 Accuracy for WDBC database using SVM for all scaling techniques.

NO	Symbol	Scaling techniques	Accuracy	CPU time
1	(S5)	Equilibration	98.95	10.175330
2	(S4)	Geometric mean	98.60	15.143076
3	(S3)	Arithmetic mean	98.42	12.516496
4	(S2)	Normalization [−1, 1]	96.66	19.208797
5	(S1)	Normalization [0, 1]	96.49	14.558410
6	(S0)	Without scaling	96.10	52.62167

all other scaling techniques by (98.95%) accuracy that obtained by SVM.

Table 7 and Fig. 4 summarize the results of all scaling techniques that obtained by SVM according the accuracies and CPU times. It is apparent from Table 6 and Fig. 4 that the equilibration scaling technique overcomes all other scaling techniques according to the accuracy and CPU time. On the other hand, the geometric mean scaling technique outperforms the arithmetic mean scaling technique according to the accuracy but it is not like this about CPU time.

Tables 8, Table 9, Table 10, Table 11 and Table 12 show the superiority of the novel scaling techniques according to both accuracy rates and CPU time. It is clear that the average accuracy rates achieved by GWO-SVM with geometric mean scaling technique (99.12%) is better than that obtained by GWO-SVM with normalization scaling techniques (98.42%) and (98.60%). On the other hand, the equilibration scaling technique overcomes all other scaling techniques by (99.30%) accuracy that obtained by SVM. Finally, Table 13 summarizes all these results.

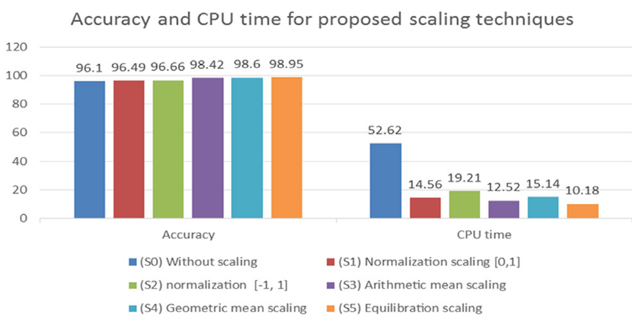

Fig. 4 Accuracy comparison among all proposed scaling techniques against the normalization [0, 1], normalization [−1, 1] and without scaling technique for SVM with grid search model.

Table 13 and Fig. 5, and Fig. 6 summarize the results of all scaling techniques that obtained by GWO-SVM according the accuracies and CPU times. It is apparent from Table 13 and Fig. 5 that the equilibration scaling technique overcomes all other scaling techniques according to the accuracy and CPU time. On the other hand, the arithmetic mean scaling technique outperforms the geometric mean scaling technique according to the accuracy and CPU time.

From Table 7-13 show the superiority of the proposed model GWO-SVM on the traditional SVM according to the accuracy rate. The accuracy rates of the proposed model GWO-SVM are 99.30, 99.12, 98.60, 98.60 and 98.42 for the scaling techniques Equilibration scaling, Geometric mean scaling, Arithmetic mean scaling, Normalization [−1, 1] and Normalization [0, 1] respectively. On the other hand, the accuracy rates of the traditional SVM are 98.95, 98.60, 98.42, 96.66 and 96.49 for the scaling techniques Equilibration scaling, Geometric mean scaling, Arithmetic mean scaling, Normalization [−1, 1] and Normalization [0, 1] respectively.

Table 8 Accuracy, sensitivity, specificity, precision, recall, F-score, G-mean and RMSE for WDBC database using the normalization scaling technique [0, 1].

Fold	GWO-SVM with Normalization Scaling between [0, 1] (S1)			
	Accuracy %	Sensitivity %	Specificity %	Precision %
1	100	100	100	100
2	100	100	100	100
3	96.49	90.91	100	100
4	100	100	100	100
5	96.49	90.48	100	100
6	98.25	95.24	100	100
7	98.25	100	97.22	95.45
8	98.25	100	97.22	95.45
9	98.25	100	97.22	95.45
10	98.25	95.24	100	100
Avg.	98.42	97.19	99.17	98.64
Time	2.28E + 04			
Fold	GWO-SVM with Normalization Scaling between [0, 1] (S1)			
	Recall %	F-score %	G-mean %	RMSE
1	100	100	100	0
2	100	100	100	0
3	90.91	95.24	95.35	0.1873
4	100	100	100	0
5	90.48	95	95.12	0.1873
6	95.24	97.56	97.59	0.1325
7	100	97.67	98.60	0.1325
8	100	97.67	98.60	0.1325
9	100	97.67	98.60	0.1325
10	95.24	97.56	97.59	0.1325
Avg.	97.19	97.84	98.14	0.1037
Time	2.28E + 04			

Table 9 Accuracy, sensitivity, specificity, precision, recall, F-score, G-mean and RMSE for WDBC database using tthe normalization scaling technique [−1, 1].

Fold	GWO-SVM with Normalization Scaling between [−1, 1] (S2)			
	Accuracy %	Sensitivity %	Specificity %	Precision %
%				
1	100	100	100	100
2	94.74	95.45	100	100
3	100	100	97.14	95.65
4	100	100	97.22	95.45
5	100	95.24	100	100
6	98.25	95.24	100	100
7	100	100	100	100
8	98.25	100	97.22	95.45
9	96.49	100	97.22	95.45
10	98.25	95.24	100	100
Avg.	98.60	98.12	98.88	98.20
Time	4.37E + 04			
Fold	GWO-SVM with Normalization Scaling between [-1, 1] (S2)			
	Recall %	F-score %	G-mean %	RMSE
1	100	100	100	0
2	95.45	97.67	97.70	0.1325
3	100	97.78	98.56	0.1325
4	100	97.67	98.60	0.1325
5	95.24	97.56	97.59	0.1325
6	95.24	97.56	97.59	0.1325
7	100	100	100	0
8	100	97.67	98.60	0.1325
9	100	97.67	98.60	0.1325
10	95.24	97.56	97.59	0.1325
Avg.	98.12	98.12	98.48	0.1060
Time	4.37E + 04			

In order to minimize CPU time of the proposed model GWO-SVM, we proposed the parallel version of the proposed model GWO-SVM that showed in Section 4. Table 14 and Fig. 7 show CPU times for all scaling techniques that obtained by GWO-SVM. On the other hand, Table15 and Fig. 8 show the speedup for all scaling techniques that obtained by GWO-SVM. The speedups for four cores are 3.50, 3.70, 3.61, 3.91 and 3.80 for the scaling techniques Equilibration scaling, Geometric mean scaling, Arithmetic mean scaling, Normalization [−1, 1] and Normalization [0, 1] respectively.

Table 16 shows the comparison between GWO-SVM and other approaches developed in the literature which shows the effectiveness of our approach. From the Table 16, it is clear that the proposed model outperforms other SVM classifiers on WDBC database. On the other hand, the proposed model doesn't outperform Kamel et al. [27] due to the last one used

Table 10 Accuracy, sensitivity, specificity, precision, recall, F-score, G-mean and RMSE for WDBC database using Arithmetic Mean Scaling technique

Fold	GWO-SVM with Arithmetic Mean Scaling technique (S3)			
	Accuracy %	Sensitivity %	Specificity %	Precision %
1	96.43	90.48	100	100
2	100	100	100	100
3	96.49	95.45	97.14	95.45
4	94.74	85.71	100	100
5	100	100	100	100
6	100	100	100	100
7	100	100	100	100
8	98.25	95.24	100	100
9	100	100	100	100
10	100	100	100	100
Avg.	98.60	96.69	99.71	99.55
Time	7.75E + 03			

Fold	GWO-SVM with Arithmetic Mean Scaling technique (S3)			
	Recall %	F-score %	G-mean %	RMSE
1	90.48	95	95.12	0.1890
2	100	100	100	0
3	95.45	95.45	96.30	0.1873
4	85.71	92.31	92.58	0.2294
5	100	100	100	0
6	100	100	100	0
7	100	100	100	0
8	95.24	97.56	97.59	0.1325
9	100	100	100	0
10	100	100	100	0
Avg.	96.69	98.03	98.16	0.0738
Time	7.75E + 03			

holdout validation but our model use k-fold validation which is more accurate technique.

Finally the strength of GWO-SVM algorithm that it results high classification accuracy especially when we used new scaling techniques and the shortcomings of it that it takes long running time but we were able to solve the problem relatively by the parallel version of GWO-SVM.

Table 11 Accuracy, sensitivity, specificity, precision, recall, F-score, G-mean and RMSE for WDBC database using Geometric Mean Scaling Technique.

Fold	GWO-SVM with Geometric Mean Scaling Technique (S4)			
	Accuracy %	Sensitivity %	Specificity %	Precision %
1	98.21	100	97.14	95.45
2	94.74	90.91	97.14	95.24
3	94.74	86.36	100	100
4	100	100	100	100
5	100	100	100	100
6	100	100	100	100
7	98.25	100	97.22	95.45
8	100	100	100	100
9	100	100	100	100
10	96.49	90.48	100	100
Avg.	98.24	96.77	99.15	98.61
Time	7.14E + 03			

Table 11 (*continued*)

Fold	GWO-SVM with Geometric Mean Scaling Technique (S4)			
	Accuracy %	Sensitivity %	Specificity %	Precision %
Fold	**GWO-SVM with Geometric Mean Scaling Technique (S4)**			
	Recall %	F-score %	G-mean %	RMSE
1	100	98.56	98.56	0.1336
2	90.91	93.97	93.97	0.2294
3	86.36	92.93	92.93	0.2294
4	100	100	100	0
5	100	100	100	0
6	100	100	100	0
7	100	98.60	98.60	0.1325
8	100	100	100	0
9	100	100	100	0
10	90.48	95.12	95.12	0.1873
Avg.	96.77	97.61	97.92	0.0912
Time	7.14E + 03			

Table 12 Accuracy, sensitivity, specificity, precision, recall, F-score, G-mean and RMSE for WDBC database using Equilibration Scaling technique.

Fold	GWO-SVM with Equilibration Scaling technique (S5)			
	Accuracy %	Sensitivity %	Specificity %	Precision %
1	98.21	100	97.14	95.45
2	100	100	100	100
3	100	100	100	100
4	100	100	100	100
5	96.49	90.48	100	100
6	100	100	100	100
7	100	100	100	100
8	98.25	95.24	100	100
9	100	100	100	100
10	100	100	100	100
Avg.	99.30	98.57	99.71	99.55
Time	3.13E + 03			

Fold	GWO-SVM with Equilibration Scaling technique (S5)			
	Recall %	F-score %	G-mean %	RMSE
1	100	97.67	98.56	0.1336
2	100	100	100	0
3	100	100	100	0
4	100	100	100	0
5	90.48	95	95.12	0.1873
6	100	100	100	0
7	100	100	100	0
8	95.24	97.56	97.59	0.1325
9	100	100	100	0
10	100	100	100	0
Avg.	98.57	99.02	99.13	0.0453
Time	3.13E + 03			

Table 13 Accuracy for WDBC database using GWOSVM for all scaling techniques.

NO	Symbol	Scaling techniques	Accuracy	CPU time
1	(S5)	Equilibration	**99.30**	3130
2	(S4)	Geometric mean	98.24	7140
3	(S3)	Arithmetic mean	98.60	7750
4	(S2)	Normalization [−1, 1]	98.60	43,700
5	(S1)	Normalization [0, 1]	98.42	22,800

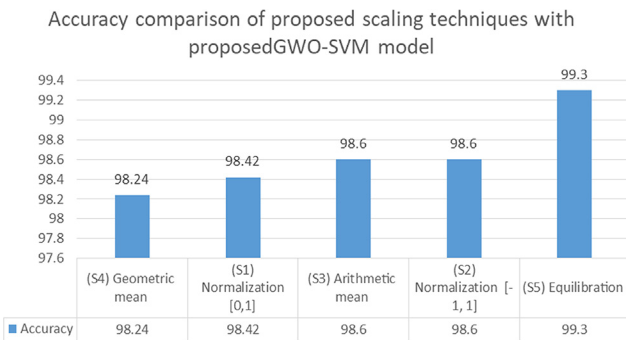

Fig. 5 Accuracy comparison among all proposed scaling techniques against the normalization [0, 1], normalization [−1, 1] for GWOSVM model.

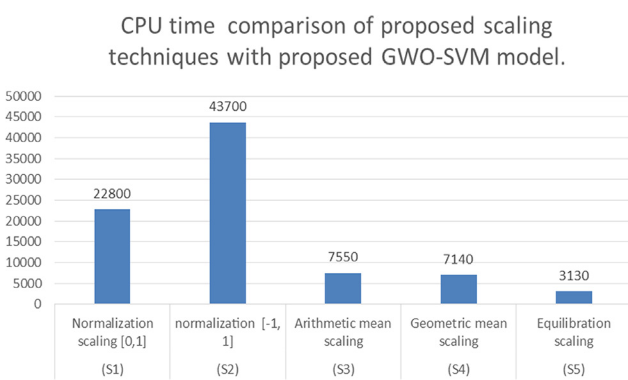

Fig. 6 CPU time comparison among all proposed scaling techniques against the normalization [0, 1], normalization [−1, 1] for GWOSVM model.

Table 14 CPU Time for WDBC database using parallel GWO-SVM for all scaling techniques

Scaling Techniques	GWO-SVM		
	Core1	Core2	Core4
Equilibration (S5)	3130	2019.35	894.29
Arithmetic mean (S4)	7550	4660.49	2091.41
Geometric mean (S3)	7140	4127.17	1929.73
Normalization [−1, 1] (S2)	43,700	22410.26	11176.47
Normalization [0, 1] (S1)	22,800	12459.02	6000

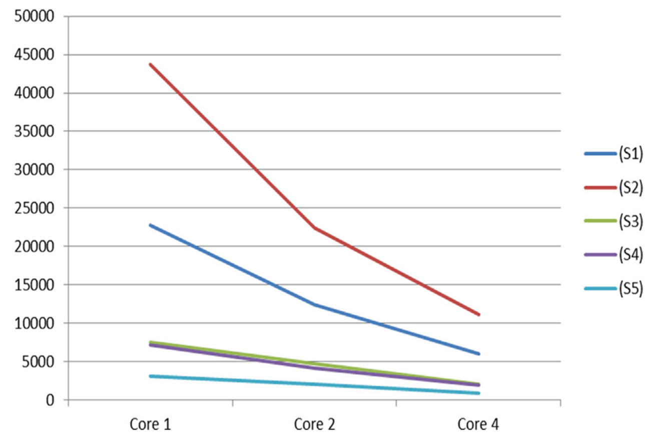

Fig. 7 CPU Time for WDBC database using GWOSVM for all scaling techniques.

Table 15 Speed up for WDBC database using parallel GWO-SVM for all scaling techniques

Scaling Techniques	GWO-SVM		
	Core1	Core2	Core4
Equilibration (S5)	1	1.55	3.5
Arithmetic mean (S4)	1	1.62	3.61
Geometric mean (S3)	1	1.73	3.70
Normalization [−1, 1] (S2)	1	1.95	**3.91**
Normalization [0, 1] (S1)	1	1.83	3.80

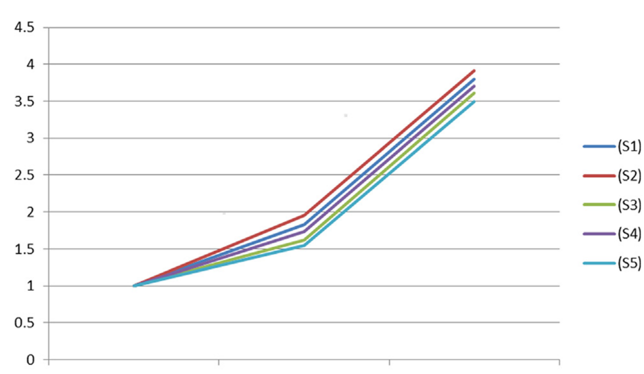

Fig. 8 Speedup for WDBC database using GWOSVM for all scaling techniques.

Table 16 Classification accuracies obtained with our method and other classifiers from literature for WDBC dataset

Study	Method	Accuracy (%)
Aalaei et al. (2016) [21]	GA-ANN	97.3 train: 80%-test:20%
S. Mandal (2017) [63]	Logistic Regression	97.9 train: 70%-test:30%
Agarap (2018) [22]	GRU-SVM	93.8 train: 70 %test:30%
Bharat et al. (2018) [64]	SVM	99.1 train: 80%-test:20%
Darwish et al. (2018) [26]	WOA-SVM	98.8 10xCV
Dhahri et al. (2019) [23]	GA-AB	98.23 10xCV
Kamel et al. (2019) [27]	GWO-SVM	100 train: 80%-test:20%
		95.7 train: 70%-test:30%
Rahmani et al. (2020) [29]	GOV-SVM	98.8 train: 70%-test:30%
Kumar et al. (2021) [28]	EGWO-SVM	98.24% 10xCV
Our study	GWO-SVM	99.3 10xCV

Table 17 Accuracy for EHR database using SVM for all scaling techniques.

Symbol	Scaling techniques	Accuracy	CPU time
(S5)	Equilibration	83.79	2.4498
(S4)	Geometric mean	80.15	2.9579
(S3)	Arithmetic mean	82.80	3.2539
(S2)	Normalization [−1, 1]	83.56	2.2924
(S1)	Normalization [0, 1]	82.47	1.9245
(S0)	Without scaling	75.99	2.2963

Table 18 Accuracy for EHR database using GWO-SVM for all scaling techniques.

NO	Symbol	Scaling techniques	Accuracy	CPU time
1	(S5)	Equilibration	92.20	2925
2	(S4)	Geometric mean	86.79	6950
3	(S3)	Arithmetic mean	86.29	7530
4	(S2)	Normalization [−1, 1]	92.35	41,521
5	(S1)	Normalization [0, 1]	93.26	21,200

7. Another dataset to validate GWO-SVM

In order to evaluate the proposed model we will use another data set Electronic Health Records EHR. The EHR was collected in routine blood analysis for 64 women with Breast cancer BC and 52 healthy volunteers after an overnight fasting. The EHR data for each participant comprises 9 parameters, namely, age, body mass index, glucose, insulin, HOMA, Leptin, Adiponectin, resisting and MCP-1. The utilized dataset EHR includes 116 samples [65].

Table 17 and Table 18 shows that the proposed model GWO-SVM overcomes the classic SVM by 93.26 against 82.47 for normalization [0, 1] scaling technique. On the other hand, for all other scaling technique the proposed model outperforms the traditional SVM. Finally, when we compare the proposed model with other reference [65], we note that the proposed model outperforms the reference [65] by 93.26 against 86.0 accuracy rate.

8. Conclusion and future work

This work has three contributions. The first contribution is improving the performance of support vector machine (SVM) using a recent grey wolf optimizer (GWO) for diagnosis breast cancer. The second contribution is proposing three efficient scaling techniques against the classical normalization technique. The last contribution is using a parallel technique which applies task distribution to improve the efficiency of GWO. The proposed sequential model is applied on two different datasets, Wisconsin diagnosis breast cancer (WDBC) dataset and Electronic Health Records (EHR). Experimental results of WDBC show that the proposed hybrid GWO-SVM model achieves 98.60% with normalization scaling. Also, using the proposed scaling techniques with the proposed GWO-SVM model gives a fast convergence and achieves accuracy rate by 99.30%. The parallel version of the proposed model achieves a speedup by 3.9 on four CPU cores. On the other hand, Experimental results of EHR show that the proposed hybrid GWO-SVM model achieves 93.26% with normalization scaling against 82.05 for SVM. The future work will give a lot of attention to evaluate the proposed model with other medical datasets. In addition to introducing more scaling techniques which will reduce the CPU time and they improve the performance of the accuracy of diagnostic system.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors extend their appreciation to the Deanship of Scientific Research at Majmaah University for funding this work under project number NO (RGP-2019-29).

References

[1] F. Bray, J. Ferlay, I. Soerjomataram, R.L. Siegel, L.A. Torre, A. Jemal, Global Cancer Statistics 2018, GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA Cancer J. Clin., Press (2018).

[2] J.R. Quinlan, Improved use of continuous attributes in C4. 5, J. Artif Intell. Res. 4 (1996).
[3] H.J. Hamilton, N. Shan, N. Cercone, RIAC: A rule induction algorithm based on approximate classification, Univ Regina, Tech Rep CS, 1996.
[4] B. Ster, A. Dobnikar, Neural Networks in Medical Diagnosis: Comparison with Other Methods, Proc Int Conf EANN96, 1996;1.
[5] K.P. Bennett, J.A. Blue, Support vector machine approach to decision trees, in: IEEE Int. Conf. Neural Networks - Conf. Proc., vol. 3, 1998. https://doi.org/10.1109/ijcnn.1998.687237.
[6] D. Nauck, R. Kruse, Obtaining interpretable fuzzy classification rules from medical data, Artif. Intell. Med. 16 (1999), https://doi.org/10.1016/S0933-3657(98)00070-0.
[7] C.A. Pe, M. Sipper, A fuzzy-genetic approach to breast cancer diagnosis, Artif. Intell. Med. 17 (1999), https://doi.org/10.1016/S0933-3657(99)00019-6.
[8] D.E. Goodman, L.C. Boggess, A.B. Watkins, Artificial immune system classification of multiple-class problems, Intell. Eng. Syst. Through Artif. Neural Networks, vol. 12, 2002.
[9] J. Abonyi, F. Szeifert, Supervised fuzzy clustering for the identification of fuzzy classifiers, Pattern Recogn. Lett. 24 (2003), https://doi.org/10.1016/S0167-8655(03)00047-3.
[10] E.D. Übeyli, A mixture of experts network structure for breast cancer diagnosis, J. Med. Syst. 29 (2005), https://doi.org/10.1007/s10916-005-6112-6.
[11] S. Şahan, K. Polat, H. Kodaz, S. Güneş, A new hybrid method based on fuzzy-artificial immune system and k-nn algorithm for breast cancer diagnosis, Comput. Biol. Med. 37 (2007), https://doi.org/10.1016/j.compbiomed.2006.05.003.
[12] E.D. Übeyli, Implementing automated diagnostic systems for breast cancer detection, Expert Syst. Appl. 33 (2007), https://doi.org/10.1016/j.eswa.2006.08.005.
[13] K. Polat, S. Güneş, Breast cancer diagnosis using least square support vector machine, Digit Signal Process A Rev. J. 17 (2007), https://doi.org/10.1016/j.dsp.2006.10.008.
[14] M.F. Akay, Support vector machines combined with feature selection for breast cancer diagnosis, Expert Syst. Appl. 36 (2009), https://doi.org/10.1016/j.eswa.2008.01.009.
[15] E.D. Übeyli, Adaptive neuro-fuzzy inference systems for automatic detection of breast cancer, J. Med. Syst. 33 (2009), https://doi.org/10.1007/s10916-008-9197-x.
[16] M. Karabatak, M.C. Ince, An expert system for detection of breast cancer based on association rules and neural network, Expert Syst. Appl. 36 (2009), https://doi.org/10.1016/j.eswa.2008.02.064.
[17] M.L. Huang, Y.H. Hung, W.Y. Chen, Neural network classifier with entropy based feature selection on breast cancer diagnosis, J. Med. Syst. 34 (2010), https://doi.org/10.1007/s10916-009-9301-x.
[18] A. Marcano-Cedeño, J. Quintanilla-Domínguez, D. Andina, WBCD breast cancer database classification applying artificial metaplasticity neural network, Expert Syst. Appl. 38 (2011), https://doi.org/10.1016/j.eswa.2011.01.167.
[19] H.L. Chen, B. Yang, J. Liu, D.Y. Liu, A support vector machine classifier with rough set-based feature selection for breast cancer diagnosis, Expert Syst. Appl. 38 (2011), https://doi.org/10.1016/j.eswa.2011.01.120.
[20] H.L. Chen, B. Yang, G. Wang, S.J. Wang, J. Liu, D.Y. Liu, Support vector machine based diagnostic system for breast cancer using swarm intelligence, J. Med. Syst. 36 (2012), https://doi.org/10.1007/s10916-011-9723-0.
[21] S. Aalaei, H. Shahraki, A. Rowhanimanesh, S. Eslami, Feature selection using genetic algorithm for breast cancer diagnosis: Experiment on three different datasets, Iran J. Basic Med. Sci. 19 (2016). https://doi.org/10.22038/ijbms.2016.6931.
[22] A.F.M. Agarap, On breast cancer detection: An application of machine learning algorithms on the Wisconsin diagnostic dataset, ACM Int. Conf. Proceeding Ser., 2018. https://doi.org/10.1145/3184066.3184080.
[23] H. Dhahri, E. Al Maghayreh, A. Mahmood, W. Elkilani, M. Faisal Nagi, Automated breast cancer diagnosis based on machine learning algorithms, J. Healthc Eng. 2019 (2019), https://doi.org/10.1155/2019/4253641.
[24] E. Badr, M. Abdulsalam, Almotairi Sultan, H. Ahmed, From linear programming approach to metaheuristic approach: scaling techniques. Complexity, Hindawi 2020.
[25] L. Shen, H. Chen, Z. Yu, W. Kang, B. Zhang, H. Li, et al, Evolving support vector machines using fruit fly optimization for medical data classification, Knowledge-Based Syst. 96 (2016), https://doi.org/10.1016/j.knosys.2016.01.002.
[26] A. Darwish, G. Sayed, A. Hassanien, Meta-heuristic optimization algorithms based feature selection for clinical breast cancer diagnosis, J. Egypt Math. Soc. 26 (2018). https://doi.org/10.21608/jomes.2018.2673.1023.
[27] S.R. Kamel, R. YaghoubZadeh, M. Kheirabadi, Improving the performance of support-vector machine by selecting the best features by Gray Wolf algorithm to increase the accuracy of diagnosis of breast cancer, J. Big Data 6 (2019), https://doi.org/10.1186/s40537-019-0247-7.
[28] S. Kumar, M. Singh, Breast cancer detection based on feature selection using enhanced Grey Wolf optimizer and support vector machine algorithms, Vietnam J. Comput. Sci. 08 (2021), https://doi.org/10.1142/s219688882150007x.
[29] A.E. Rahmani, M. Katouli, Breast cancer detection improvement by grasshopper optimization algorithm and classification SVM, Rev d'Intelligence Artif 34 (2020). https://doi.org/10.18280/ria.340210.
[30] S. Mirjalili, S.M. Mirjalili, A. Lewis, Grey Wolf Optimizer, Adv. Eng. Softw. 69 (2014), https://doi.org/10.1016/j.advengsoft.2013.12.007.
[31] H. Faris, I. Aljarah, M.A. Al-Betar, S. Mirjalili, Grey wolf optimizer: a review of recent variants and applications, Neural Comput. Appl. 30 (2018) 413–435, https://doi.org/10.1007/s00521-017-3272-5.
[32] M.A. Al-Betar, M.A. Awadallah, H. Faris, I. Aljarah, A.I. Hammouri, Natural selection methods for Grey Wolf Optimizer, Expert Syst. Appl. 113 (2018), https://doi.org/10.1016/j.eswa.2018.07.022.
[33] J.M. Elble, N.V. Sahinidis, Scaling linear optimization problems prior to application of the simplex method, Comput. Optim. Appl. 52 (2012), https://doi.org/10.1007/s10589-011-9420-4.
[34] N. Ploskas, N. Samaras, The impact of scaling on simplex type algorithms, in: ACM Int. Conf. Proceeding Ser., 2013. https://doi.org/10.1145/2490257.2490283.
[35] C. Triantafyllidis, N. Samaras, Three nearly scaling-invariant versions of an exterior point algorithm for linear programming, Optimization 64 (2015), https://doi.org/10.1080/02331934.2014.926356.
[36] N. Ploskas, N. Samaras, A computational comparison of scaling techniques for linear optimization problems on a graphical processing unit, Int. J. Comput. Math. 92 (2015), https://doi.org/10.1080/00207160.2014.890716.
[37] E.M. Badr, H.S. AlGendy, A hybrid water cycle particle swarm optimization for solving the fuzzy underground water confined steady flow, Indones J. Electr. Eng. Comput. Sci. 2020;19. https://doi.org/10.11591/ijeecs.v19.i1.pp492-504.
[38] E.M. Badr, M.I. Moussa, An upper bound of radio k-coloring problem and its integer linear programming model, Wirel Networks 26 (2020), https://doi.org/10.1007/s11276-019-01979-8.

[39] E.B. Aloufi, KA, A Robot's Response Acceleration Using the Metric Dimension Problem. Preprints 2019.
[40] E.S. Badr, K. Paparrizos, BT, GV, Some computational results on the efficiency of an exterior point algorithm, in: Proc. 18th Natl. Conf. Hell. Oper. Res. Soc. (HELORS), Rio, Greece, 2006, p. 1103–15.
[41] E.S. Badr, K. Paparrizos, N. Samaras, AS, On the Basis Inverse of the Exterior Point Simplex Algorithm, in: Proc. 17th Natl. Conf. Hell. Oper. Res. Soc. (HELORS), Rio, Greece, 2005, p. 677–87.
[42] E.S. Badr, M. Moussa, K. Paparrizos, N. Samaras, AS, Some computational results on MPI parallel implementation of dense simplex method, World Acad. Sci. Eng. Technol. 2008, 778–781.
[43] E. Badr, S. Almotairi, On a dual direct cosine simplex type algorithm and its computational behavior, Math. Probl. Eng. 2020 (2020), https://doi.org/10.1155/2020/7361092.
[44] Tapkan Pi, L. Özbakir, A. Baykasoɪlu, Bee algorithms for parallel two-sided assembly line balancing problem with walking times, Appl. Soft Comput. J. 39 (2016), https://doi.org/10.1016/j.asoc.2015.11.017.
[45] T. Tian, D. Gong, Test data generation for path coverage of message-passing parallel programs based on co-evolutionary genetic algorithms, Autom. Softw Eng. 23 (2016), https://doi.org/10.1007/s10515-014-0173-z.
[46] S. Maleki, M. Musuvathi, T. Mytkowicz, Efficient parallelization using rank convergence in dynamic programming algorithms, Commun. ACM 59 (2016), https://doi.org/10.1145/2983553.
[47] E.F. De Oliveira Sandes, A. Boukerche, A.C.M. Alves De Melo, Parallel optimal pairwise biological sequence comparison: Algorithms, platforms, and classification, ACM Comput. Surv. 48 (2016), https://doi.org/10.1145/2893488.
[48] C. Truchet, A. Arbelaez, F. Richoux, P. Codognet, Estimating parallel runtimes for randomized algorithms in constraint solving, J. Heuristics 22 (2016), https://doi.org/10.1007/s10732-015-9292-3.
[49] D. Połap, K. Kęsik, M. Woźniak, R. Damaševičius, Parallel technique for the metaheuristic algorithms using devoted local search and manipulating the solutions space, Appl. Sci. 8 (2018), https://doi.org/10.3390/app8020293.
[50] Osman Hegazy, Omar S. Soliman, MAS. A machine learning model for stock market prediction, Int. J. Comput. Sci. Telecommun. 4 (2013) 17–23.
[51] Osman Hegazy, Omar S. Soliman, MAS. LSSVM-ABC Algorithm for Stock Price Prediction, Int. J. Comput. Trends Technol. 7 (2014) 81–92.
[52] Osman Hegazy, Omar S. Soliman, MAS, Optimizing LS-SVM using Modified Cuckoo Search algorithm (MCS) for Stock Price Prediction, Int. J. Adv. Res. Comput. Sci. Manage. Stud. 3 (2015) 204–224.
[53] O. Hegazy, O.S. Soliman, M.A. Salam, Comparative Study between FPA, BA, MCS, ABC, and PSO Algorithms in Training and Optimizing of LS-SVM for Stock Market Prediction, Int. J. Adv. Comput. Res. 5 (2015).
[54] Osman Hegazy, Omar S. Soliman, MAS. FPA-ELM Model for Stock Market Prediction, Int. J. Adv. Res. Comput. Sci. Softw. Eng, 5 (2015) 1050–1063.
[55] E.M. Badr, M. Abdul, M. Ali, H. Ahmed, Social media sentiment analysis using machine learning and optimization techniques, Int. J. Comput. Appl. 178 (2019) 31–36, https://doi.org/10.5120/ijca2019919306.
[56] S.R. Sain, V.N. Vapnik, The nature of statistical learning theory, Technometrics 38 (1996), https://doi.org/10.2307/1271324.
[57] B.E. Boser, I.M. Guyon, V.N. Vapnik, Training algorithm for optimal margin classifiers, Proc. Fifth Annu. ACM Work. Comput. Learn. Theory, 1992. https://doi.org/10.1145/130385.130401.
[58] N. Cristianini, J. Shawe-Taylor, An Introduction to Support Vector Machines and Other Kernel-based Learning Methods, 2000. https://doi.org/10.1017/cbo9780511801389.
[59] M.A. Aizerman, E.A. Braverman, L. Rozonoer, Theoretical foundations of the potential function method in pattern recognition learning, Autom. Remote Control (1964).
[60] Chih-Wei Hsu, Chih-Jen Lin, C.-C. Chang, A Practical Guide to Support Vector Classification, BJU Int. 2008; 101.
[61] C.C. Chang, C.J. Lin, LIBSVM: A Library for support vector machines, ACM Trans. Intell. Syst. Technol. 2011; 2. https://doi.org/https://doi.org/10.1145/1961189.1961199.
[62] S.L. Salzberg, On comparing classifiers: Pitfalls to avoid and a recommended approach, Data Min. Knowl. Discov. 1 (1997), https://doi.org/10.1023/A:1009752403260.
[63] Mandal S. Kumar, Performance analysis of data mining algorithms for breast cancer cell detection using Naïve Bayes, logistic regression and decision tree, Int. J. Eng. Comput. Sci. 6 (2017).
[64] A. Bharat, N. Pooja, R.A. Reddy, Using Machine Learning algorithms for breast cancer risk prediction and diagnosis, in: 2018 IEEE 3rd Int. Conf. Circuits, Control. Commun. Comput. I4C 2018, 2018. https://doi.org/10.1109/CIMCA.2018.8739696.
[65] M. Patrício, J. Pereira, J. Crisóstomo, P. Matafome, M. Gomes, R. Seiça, et al, Using Resistin, glucose, age and BMI to predict the presence of breast cancer, BMC Cancer 18 (2018), https://doi.org/10.1186/s12885-017-3877-1.

Elsayed Badr is an Associate professor of computer science at Benha Faculty of Computers & Artificial Intelligence; Benha University in Egypt. He received his Ph.D. degree in Parallel Algorithms (mainly in parallel graph algorithms) in 2006 from the University of Macedonia; Greece. Dr. Badr holds a Certificate of Quality Assurance from the University of Benha, Egypt and M.Sc. in graph theory and graph algorithms applications, B.Sc. in Mathematics from Benha faculty of science in Egypt. In addition to over 8 years of teaching and academic experiences In Egypt and Greece, Dr. Badr has accumulated broad practical experiences and developed a solid set of skills in algorithms, Fuzzy theory, graph labeling, Wireless Networks, Distributed System, Parallel Programming and linear programming.

Sultan Almotairi received his B.Sc., M.Sc., and PhD. degrees in computer science from Florida Institute of Technology, Melbourne, USA, in 2010, 2012, and 2014 respectively. Currently, he is an assistant professor at Department of Natural and Applied Sciences in Community College of Majmaah University. He has acted as the dean of Community College at Majmaah University since June 2015. In 2016, he was elected as the Chairman of the Municipality Council of Majmaah. His research interests include neural networks, deep learning, pattern recognition, machine learning, image processing, and computer vision.

Mustafa Abdul Salam was born on November 1, 1981 in Sharkia, Egypt. He received the B. Sc. from Faculty of Computers and Informatics, Zagazig University, Egypt, and obtains master degree in information system from faculty of computers and information, Menoufia University, Egypt in 2009 specializing in Hybrid Machine Learning and Bio Inspired Optimization algorithms. He obtained his Ph.D. degree in information system from faculty of computers and information, Cairo University, Egypt in 2015. He is currently an assistant professor in Scientific Computing and Artificial Intelligence department, Faculty of Computers and Information, Benha University, Egypt. He has worked on a number of research topics. Mustafa has contributed more than +20 technical papers in the areas of neural networks, support vector machines, optimization, time series prediction, and extreme learning machine, hybrid CI models in international journals, international conferences, local journals and local conferences. His majors are Machine Learning, Big Data, Stream Data Mining, and Deep Learning.

Hagar Ahmed is a Teaching Assistant of Scientific Computing at Faculty of Computers & Artificial Intelligence; Benha University in Egypt. She received EHR Pre-Master in scientific computing in 2014 from the University of Benha, Egypt. She received the B.Sc. from Faculty of Computers and Informatics, Benha University, Egypt. Hagar is preparing master in early detection of cancer diseases by using data mining techniques and metaheuristic algorithms. In addition to 6 years of teaching and academic experiences In Egypt, Hagar has accumulated broad practical experiences and developed a solid set of skills in Parallel Programming, linear programming, data mining and metaheuristic algorithms.