Breast Cancer Prediction using a novel Modified Whale Optimization, Grey Wolf Optimization and Support Vector Machine Algorithms

Mr.K.Madhavan, Dr. R. Manicka chezian

Ph.D., Research Scholar (Part-Time), Department of Computer Science,
Bharathiar University, Coimbatore , Tamilnadu-641046
Associate Professor P.G. Department of Computer Science,
Nallamuthu Gounder Mahalingam College, Pollachi**.**

madankmadan3@gmail.com

Abstract

Globally, breast cancer has been identified as one of the deadliest disease and top causes of cancer related mortality in females. 16% of malignant lesions that are diagnosed in the world are linked to consequence of breast cancer. Because of this, it is of the utmost importance to make a diagnosis of malignant tumours at the earliest possible stage in order to give oneself the best possible chance of surviving them. Thus, an accurate and timely diagnosis of the condition ensures the patient's long survival. Breast cancer diagnosis depends on the ability to recognise benign and malignant tumours at the appropriate time. Traditional approaches in the diagnosis of breast cancers have several drawbacks including human errors related discrepancies, inaccurate diagnosis, and time factor. Recently, machine learning algorithms together with different hyperparameter tuning optimization techniques were proved as a viable option that support the early detection of cancer.

In this work, an attempt is made to create a novel prediction system that utilises a modified Whale Optimization Algorithm (WOA) and Grey Wolf Optimization algorithm (GWO) methods to fully exploit the capabilities of a support vector machine (SVM). The Modified Whale Optimization Algorithm (WOA), Grey Wolf Optimization algorithm (GWO) and Support Vector Machine (SVM) technique (MWOA+GWO+SVM) meta-heuristic algorithm has been created to extract useful information from popular breast cancer datasets like the Wisconsin Diagnostic Breast Cancer Database (WDBC) for the early-stage detection of the disease. The proposed method was compared with other ensemble and popular classification algorithms to evaluate its performance.

Keywords: *Breast cancer detection, Whale Optimization Algorithm, Grey wolf optimization, SVM algorithm*

1. Introduction

Breast cancer is, undoubtedly, the most common cause of mortality from cancer among the female population in every region of the globe. It is estimated that over 2.1 million women are diagnosed with breast cancer each year. Breast cancer is the most common kind of cancer among women (1). The national average of cancer cases for 2022 is 100.4 per 100,000, with a large number of women (105.4 per 100,000) being diagnosed with breast cancer (Priyanka, 2020).

In 2018, breast cancer was responsible for the deaths of roughly 627,000 women, which represents approximately 15% of all cancer-related fatalities among women. Therefore, finding breast cancer at an early stage is essential if one want to enhance their chances of surviving the disease. The correct categorization of a breast cancer tumour as either benign or malignant is necessary for appropriate diagnosis of breast cancer (Ades et al. 2014). The diagnosis of breast cancer is becoming more interested in the use of soft computing approaches (Dubey et al. 2015; Arya & Tiwari, 2016). Researchers from all around the world have been developing different strategies and techniques in the hopes of further improving the categorization capabilities of their breast cancer diagnostic system. However, there is still a significant opportunity to create a system that is more effective for the categorization of breast cancer.

The manual detection of breast cancer takes a large amount of time and presents challenges to the medical professional in terms of classification. As a result, it is very important to look for signs of cancer using a range of different automated diagnostic techniques. In an attempt to counteract the steadily growing number of cases of cancer, several research studies on the prevention of illness have focused substantially on the development of methods for the early identification of cancer. Periodic mammography (Ades et al., 2014), gene identification (Jazaeri et al. 2002), clinical diagnosis (Sotiriou and al., 2003), and other similar procedures are examples of pre-diagnostic treatments that are often applied. In addition, as a result of the development of biomedical and information technology over the course of the years, a large number of prognostic factors pertaining to breast cancer have been revealed. This has enabled a large number of researchers to develop more complex early detection models based on a variety of data-driven prediction methodologies, including support vector machines (SVMs), logistic regression (LR), multilayer perceptrons (MLPs), and decision trees (DTs).

2. Background

Rojas-Domnguez et al. (2017) have evaluated a variety of algorithms, such as the Boltzmann-UMDA algorithm, the Firefly method, the Fruit-fly optimisation algorithm, the particle Swarm optimisation algorithm, and the Bat algorithm, to find which ones have the best overall performance based on criteria such as their effectiveness, generality, efficiency, and complexity. The estimate of distribution algorithms (EDAs) have produced the best results possible, as shown by the outcomes of 15 separate experiments on medical diagnostic concerns. The process of optimisation was directed by a one-of-a-kind performance indicator that has the potential to boost the generalizability of the solutions while simultaneously maintaining their effectiveness.

Support vector machines (SVM) are the foundation of the ensemble learning methodology that Wang et al. (2018) have presented as a method for the diagnosis of breast cancer. The suggested strategy has reduced the diagnostic variance while simultaneously improving the accuracy of diagnosis in order to circumvent the limitations imposed by individual model performance. Twelve unique SVMs were hybridised using the Weighted Area Under the Receiver Operating Characteristic Curve Ensemble (WAUCE) approach that was provided. It has been determined whether or not the suggested model is effective by analysing the breast cancer datasets from the Surveillance, Epidemiology, and End Results (SEER) programme; the Wisconsin Breast Cancer dataset; the Wisconsin Diagnostic Breast Cancer dataset; and the Wisconsin Breast Cancer dataset. The results of the experiments have shown that the WAUCE model works better than five distinct ensemble processes and two commonly used ensemble models, namely adaptive boosting and bagging classification tree, when it comes to the diagnosis of breast cancer in terms of accuracy and variance. When measured against the best single SVM model on the SEER dataset, the suggested WAUCE model's accuracy is found to be 33.34 percent better, while the best single SVM model's variance has been reduced by 97.89 percent.

Khuriwal and Mishra (2018) have used the Wisconsin Breast Cancer database to apply an adaptive ensemble voting strategy to the process of identifying breast cancer. In the research, it was shown that ANN and logistic algorithms performed better than ensemble machine learning approaches when it came to diagnosing breast cancer. This was true even when the number of variables was decreased. When compared to one of the other machine learning algorithms, the findings showed that the ANN approach that made use of the logistic algorithm obtained 98.50 percent accuracy.

Yin et al. (2019) have introduced a novel technique for SVM parameter optimisation that is based on the advanced whale optimisation algorithm (AWOA). This algorithm is an improved version of the whale of algorithm (WOA) that incorporates an external archiving strategy. A whole new framework for SVM parameter optimisation was built, and it was based on AWOA. In order to demonstrate that the technique that was suggested is effective, six sets of data that are typical of the whole were chosen to analyse the impact of the SVM classification issue. The results of experiments have shown that the AWOA method of parameter optimisation may achieve greater accuracy and better convergence than the three standard techniques to parameter optimisation (WOA, PSO, and DE).

Mallika and Selvamuthukumaran (2021) have developed an effective method for detecting diabetes (SVM) by making use of a Support Vector Machine that is based on hybrid optimisation. In order to make full use of the potential offered by SVM in the diabetes detection system, the suggested hybrid optimisation strategy brought together two different optimisation strategies: the Crowd Search algorithm (CSA) and the Binary Grey Wolf Optimizer (BGWO). The effectiveness of the hybrid optimization-based SVM approach that was suggested (henceforth CS-BGWO-SVM) was fully examined utilising real-world datasets, such as the UCIPima Indian standard dataset and the diabetes type dataset that was retrieved from the Data World repository. The findings of the empirical investigation indicated that the CS-BGWO-SVM classification approach is a more effective one that also has an exceptionally high degree of accuracy.

Using a learning model that is based on SVM has been suggested as a method for the detection of breast cancer by Wang et al. (2015). The model had a total of six distinct kernel functions and SVM structures, one of which was an a-SVM, in addition to a C-SVM. WAUCE is an acronym that stands for Weighted Area Under the Receiver Operating Characteristic Curve Ensemble. This approach was developed for use in the hybridization of models with various base classifiers. The databases known as Wisconsin Breast Cancer (WBC), Wisconsin Diagnostic Breast Cancer (WDBC), and Surveillance, Epidemiology and End Results (SEER) were among those that were examined as part of this study. When the newly developed model is compared to earlier experiments that were just based on a single SVM, a significant improvement in diagnosis accuracy is seen. The showy nature of the system, as well as the inadequate amount of training time, are both determined to be negatives.

Support vector machine (SVM) is a pattern classification approach whose classification performance is significantly affected by the kernel parameter setting as well as feature selection. Wang and Chen (2020) have developed an improved whale optimisation algorithm (CMWOA) that combines chaotic and multi-swarm techniques in order to concurrently conduct parameter optimisation and feature selection for SVM. Comparing the proposed SVM model, which was given the name CMWOAFS-SVM, to multiple competitive SVM models based on other optimisation algorithms, such as the original algorithm, particle swarm optimisation, bacterial foraging optimisation, and genetic algorithms, was done by using several well-known medical diagnosis problems, such as breast cancer, diabetes, and erythemato-squamous. The results of the experiments show that CMWOAFS-SVM performs much better than any of the other rivals when it comes to classification performance as well as the size of the feature subset.

The process of building an architecture for a machine learning model can take several different forms, and it is important to investigate all of these different forms before deciding on the best one. The process of determining the best possible values for the model parameters is referred to as hyperparameter tuning, and it is accomplished through the use of machine learning. This step is necessary for the selection of the best possible model architecture. The hyper parameters have an effect on the performance and behaviour of the model, which helps to increase the efficiency of the model (Wu et al., 2019). Several optimization algorithms have been developed for tuning the parameters of the machine learning models in different applications like Bayesian optimization algorithm (BOA), Whale Optimization Algorithm (WOA), Particle Swarm Optimization algorithm (PSO), Genetic algorithm (GA), Grey Wolf Optimization Algorithm (GWO), Artificial Bee Colony Optimization Algorithm (ABC), etc.

3. Grey Wolf Optimization algorithm (GWO)

Mirjalili et al (2014) have proposed Grey Wolf Optimization algorithm (GWO) based on the concept of grey wolf hunting mechanisms and social hierarchy. In GWO algorithm, grey wolves are classified into four categories viz. alpha (α) which is the leader, beta (β) which assists the leader, delta (δ) which follows both prior wolves, and omega (ω) (Mirjalili et al 2014).

There is a social hierarchy in the operation of Grey Wolf Optimization algorithm. The alpha (a) is considered to be the best solution, followed by the beta (b), and the delta (d) is considered to be the third-best option. The candidate's solution that is still in left is omega (w). These wolves (w) are below the three wolves that came before them in the hierarchy. Figure 1 depicts the hierarchy of grey wolves.

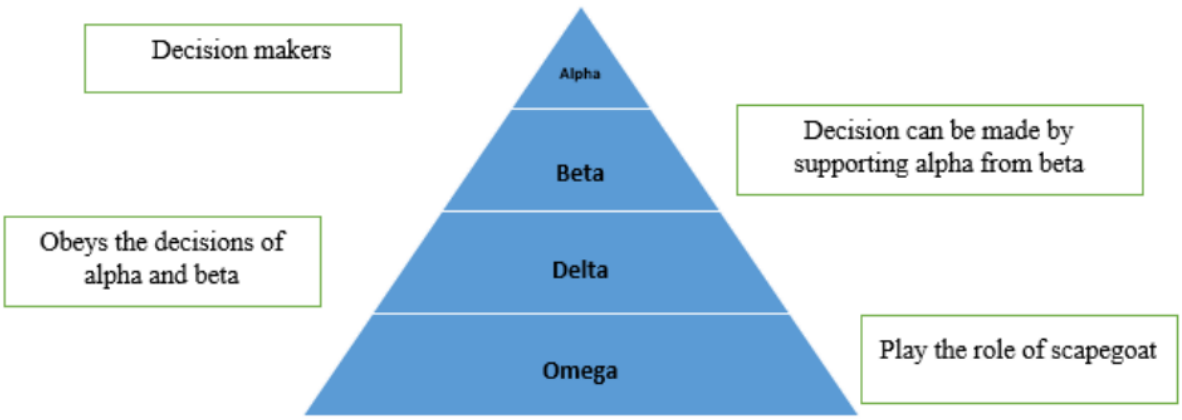

Figure 1 Hierarchy of Grey Wolves

In general, Grey wolves encircle their prey for hunting by using based on the equations (1) and (2).

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

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

Where, D is the distance between the grey wolf and its prey. “t” represents the number of iteration. X_p represents the positional location of the prey and X represents positional location of the grey wolf.

$$A = 2ar_1 - a \tag{3}$$

$$C = 2r_2 \tag{4}$$

“a” represents a vector whose values decline linearly from 2 to 0 over the length of run, and r1 and r2 are random vectors within the interval [0, 1]. The values of A and C vectors determine how close the wolves will be to the prey.

3.1 Hunting Process

After the encirclement process, a grey wolf begins searching for the optimal option. Even if the optimal answer needs optimization, alpha wolf saves the optimal solution in each iteration and updates it if it is improved. Beta and delta are able to pinpoint the position of the prey. Consequently, each subspecies of grey wolf stores the optimal solution and uses the following equations to update the position of grey wolves. The position updation of grey wolves is represented in the equations (5) to (11).

$$D_a = |C_1.X_a(t) - X(t)| \tag{5}$$

$$D_b = |C_2.X_b(t) - X(t)| \tag{6}$$

$$D_d = |C_3.X_d(t) - X(t)| \tag{7}$$

$$X_1 = X_a(t) - A_1.D_a \tag{8}$$

$$X_2 = X_b(t) - A_2.D_b \tag{9}$$

$$X_3 = X_d(t) - A_3.D_d \tag{10}$$

$$X(t+1) = X_1 + X_2 + X_3 \tag{11}$$

Here, $X_a(t)$, $X_b(t)$, $X_d(t)$ are the position vectors of three best solutions at a given iteration t. The coefficient vectors like A_1, A_2, A_3, are essential to the operation of GWO. The exploitation portion is favoured by the coefficient values of A that lie between -1 and 1, which forces the search agents to converge (attack) in the direction of the prey. The search agents are forced to deviate from the prey in quest of a better or more effective solution when the coefficient values are greater than one or less than equal to one favouring the exploration phase. Coincident vector C is another GWO control parameter that encourages exploration. This parameter's value always falls between 0 and 2. It controls the role that the prey plays in determining the next location.

3.2 Exploitation Phase - Attacking Prey

In this part of the hunting process, a grey wolf attempts to halt the movement of its prey in order to attack it. This technique is implemented by lowering the value of a. The value of A is also decreased by the value a and lies between -1 and 1. The grey wolf can attack the victim if A is larger than -1 and less than 1. However, GWO suffers from stagnation in the local optimum, and researchers are attempting to identify several strategies to address this issue (Mirjalili et al., 2014).

3.3 Exploration Phase – Searching for Prey

Alpha, beta, and delta have an effect on the search method. These three groups are distinct from one another. Consequently, they require a mathematical equation in order to converge and assault prey. If the value of A is larger than 1 or less than -1, the search agents are driven to diverge from the prey. In addition, if A is larger than 1, the search agent attempts to locate superior prey. C is an additional component element that affects the exploration phase in GWO (Mirjalili et al., 2014). When C > 1, the solution gravitates more toward the prey, and this contribution is hence considerable.

In summary, the GWO algorithm generates the random population. Alpha, beta, and delta assume the prey's position. The gap between possible solutions is then modified. After then, an is decreased from 2 to 0 to achieve equilibrium between the two phases. If A is greater than one, the search agents stop attacking the victim. If A is less than 1, then they pursue the prey. The GWO has achieved a successful conclusion and is now ended. Algorithm 2 covers the GWO algorithm in depth (Mirjalili et al., 2014).

Table 1 Pseudocode for GWO Algorithm

```
Initialize the grey wolfs (search agents)
Initialize a, A and C
Calculate the fitness value using eqn. 1
Compute the values for search agents X_α, X_β, X_δ
    While (iter<iter_max)
        For each search agent
        Update the location of the present search agent by (11)
        End for
    Update a, A and C
    Search agents evaluated by a fitness
    Update X_α, X_β, X_δ
    iter = iter +1
    End While
Display X_α and fitness value
```

4. Whale Optimisation Algorithm (WOA)

The Whale Optimisation Algorithm (WOA) is a recently created swarm-based meta-heuristic algorithm that depends on the bubble-net hunting manoeuvre strategy used by humpback whales to solve complicated optimisation problems. The programme was inspired by the humpback whales' ability to capture prey using their bubble nets. Swarm intelligence is a technique that has gained widespread acceptance in a variety of engineering domains due to its straightforward structure, reduced need for human operators, lightning-fast convergence speed, and enhanced capacity for striking a better balance between the exploration and exploitation phases. In recent years, the applications of the algorithm have seen widespread use across a wide variety of sectors due to the algorithm's outstandingly high levels of performance and efficiency in its use of resources.

The whale optimisation algorithm, also known as the WOA, is a relatively new form of the metaheuristic algorithm that was initially presented by Mirjalili and Lewis (40). WOA begins, in a manner analogous to that of other metaheuristic optimisation algorithms, by producing a random population of possible solutions for the issue. From this population, the global optimal solution, which may be either the maximum or the minimum, is searched for and located. The algorithm will continue to make the answer better and provide updates based on its structure until it reaches the value that is optimal for the situation.

4.1 Encircling Prey and Attacking-Bubble Net Mechanism

Humpback whales have the ability to pinpoint the position of their prey and then surround it. The WOA method operates on the assumption that the current best candidate solution is either the target prey or is very near to the optimum. This is due to the fact that the location of the optimal design inside the search space is unknown a priori. After the best search agent has been identified, the other search agents will work to improve their standings in relation to the top search agent in order to remain competitive. The following equations provide a representation of this pattern of behaviour.

$$\vec{X}(t+1) = \vec{X}^*(t) - \vec{A}\,.\vec{D} \tag{12}$$

$$\vec{D} = \left|\vec{C}.\vec{X}^*(t) - \vec{X}(t)\right| \tag{13}$$

where "t" represents the current iteration, "A" and "C" are coefficient vectors, "X*" is the position vector of the best solution achieved so far, "X" is the position vector, "| |" represents the absolute value, and "." represents an element-by-element multiplication. It is important to note that "X*" has to be modified after each iteration if there is a better answer. This is something that should be done. The vectors $\vec{A}$ and $\vec{C}$ are computed as follows:

$$\vec{A} = 2.\vec{a}.\vec{r} - \vec{a} \tag{14}$$

$$\vec{C} = 2.\vec{r} \tag{15}$$

where "r" is a random vector in the range [0,1] and "a" is reduced linearly from 2 to 0 throughout the length of iterations (in both the exploration and exploitation phases).

The humpback whales approach their prey in a spiralling pattern while swimming in a circle that is gradually becoming smaller. A chance of fifty percent that one of these processes will be selected, and spiralling motion of the humpback whale can be modelled using a mathematical formulation as shown below:

$$\vec{X}(t+1) = \begin{cases} \vec{C}.e^{bl}.cos(2\pi l) - \vec{X}(t) & \text{if p} > 0.5 \\ \vec{X}(t) - \vec{A}.\vec{D} & \text{if p} < 0.5 \end{cases} \tag{16}$$

where p is a random integer that falls between 0 and 1. In addition to searching for prey using bubble nets, humpback whales also randomly look for prey.

4.2 Searching for prey

The humpback whales often search randomly for prey. The vector A is used to search for prey when j j A > 1. This approach can be mathematically represented as follows:

$$\vec{D} = \left|\vec{C}.\vec{X}rand - \vec{X}\right| \tag{17}$$

$$\vec{X}(t+1) = \vec{X}rand - \vec{A}.\vec{D} \tag{18}$$

The way in which the rules of WOA are used to enhance and update the outcome is the primary distinction that can be drawn between it and other types of metaheuristic algorithms. The WOA was conceived after seeing how whales hunt their prey, which they do by swimming in a spiralling pattern around the target, enclosing it in a trap, and then attacking it. This behaviour served as the inspiration for the WOA. This type of feeding behaviour is referred known as bubble-net feeding.

The illustration demonstrates that prior to attacking its victim, the humpback whale makes bubbles by moving in a circular pattern around the target. This feeding behaviour served as the inspiration for the primary architecture of the WOA.

Pseudocodes of WOA

```
// Initialization
Initialize a population of whales randomly within the search space
Evaluate the fitness of each whale in the population
Set the best whale as the one with the highest fitness
X* shows the best solution in the current iteration
// Main loop
Repeat until a termination condition is met:
   // Update exploration and exploitation rates
   Calculate the current iteration as a fraction of the maximum number of iterations
   Calculate the a parameter to control the linearly decreasing exploration rate
   // Update position and encircling mechanism
   For each whale in the population:
      Generate random vectors r1 and r2
      Calculate the distance to the best whale (D)
      // Update the position of the current whale
      If D > 0:
         Update the position of the whale towards the best whale using equation (1)
      Else:
         Perform the encircling mechanism around the best whale using equation (2)

      // Update search agents' position
      Apply a search operator (e.g., random walk, spiral update) to improve exploration
      Clip the positions of the whales to ensure they remain within the search space
      // Evaluate fitness
      Evaluate the fitness of the updated whale
      // Update the best whale
      If the fitness of the updated whale is better than the best whale's fitness, update the best
whale
   // Update a and A values
   Calculate a decreasing linearly from 2 to 0 over the iterations
   Calculate A to control the spiral updating mechanism
// Termination
Return the best whale found
// Equation (A): Update the position of the whale towards the best whale
D = abs(C * best_whale_position - current_whale_position)  // Calculate distance
new_position = best_whale_position - (D * exp(-a * current_iteration) * cos(2 * pi *
current_iteration)) // Update position
// Equation (B): Encircling mechanism around the best whale
new_position = best_whale_position - (A * abs(C * best_whale_position -
current_whale_position) * exp(-a * current_iteration) * cos(2 * pi * current_iteration))
// Update position
```

In this pseudocode, equations (A) and (B) are included to update the position of the whales. Equation (A) is used when the distance to the best whale (D) is greater than zero, and equation (B) is used when the distance is zero or negative. The variables C, a, and A control the movement of the whales, with a decreasing linearly from 2 to 0 over the iterations. The positions of the whales are clipped to ensure they remain within the search space.

5. Modified WOA+GWO Algorithm

In this work, a novel algorithm is proposed as a means to enhance the functionality of the WOA algorithm. The technique makes use of the leadership hierarchy of the GWO to apply to the bubble-net assaulting tactic used by the WOA. During the phase of exploitation, the proposed algorithm chose the three best candidate solutions from among the whole search agents (the first level alpha (a), and the second and third level in the group is beta (b) and delta (d)). The remaining search agents modified their positions in accordance with the positions of the best search agents in order to improve the WOA algorithm's overall performance.

The humpback whales have two different processes that they use while swimming around their prey. The following is an example of how the suggested mathematical model for updating the location of whales during optimisation by employing the leadership hierarchy of GWO may be expressed mathematically. The following is one possible formulation for the process of updating the position of whales by employing the hierarchical leadership of GWO:

5.5 Bubble-net attacking strategy

The shrinking encircling process is the method that humpback whales use to adjust their position using the equation (11).

5.6 Spiral updating position

The process of updating the location of humpback whales along a route curved like a spiral and may be expressed as follows:

$$\overrightarrow{D_\alpha}' = |X_\alpha(t) - \vec{X}(t)\,| \tag{19}$$

$$\overrightarrow{D_\beta}' = |X_\beta(t) - \vec{X}(t)\,| \tag{20}$$

$$\overrightarrow{D_\gamma}' = |X_\gamma(t) - \vec{X}(t)\,| \tag{21}$$

$$\overrightarrow{X_1}(t) = \vec{X}_\alpha(t) + \overrightarrow{D_\alpha}'.e^{bl}.cos(2\pi l) \tag{22}$$

$$\overrightarrow{X_2}(t) = \vec{X}_\beta(t) + \overrightarrow{D_\beta}'.e^{bl}.cos(2\pi l) \tag{23}$$

$$\overrightarrow{X_3}(t) = \vec{X}_\gamma(t) + \overrightarrow{D_\gamma}'.e^{bl}.cos(2\pi l) \tag{24}$$

$$\vec{X}(t+1) = \frac{\overrightarrow{X_1} + \overrightarrow{X_2} + \overrightarrow{X_3}}{3} \tag{25}$$

The pseudocode for the proposed modified WOA+GWO algorithm can be written in steps as follows:

Generate initial population of search agent.
Evaluate the objective function value for each search agent.
X_α is best candidate solution.
X_β is the second-best candidate solution.
X_γ is the third-best candidate solution.
While (n < Max number of iterations
for i=1 to number of each search agent
Update control parameter (A, C, a, l, and p).
If$_1$ (p < 0.5)
If$_2$ (|A| < 1)

Update the position of the search agent by (18)
else If_2 ((|A|≥1)
Select a random search agent (X_{rand}).
Update the position of the search agent by (21)
end If_2
else If_1 (p≥0.5)
Update the position of the search agent by (25).
end If_1
end for
Check if any search agent goes outside the search space.
Evaluate the objective function value for each search agent.
Update the position of X_α, X_β, and X_δ.
n=n+1
end while
Return X_α

6. SVM Algorithm

Boser, Guyon, and Vapnik (1992) have proposed the SVM classification method. The SVM method is commonly utilised in bioinformatics because to its high accuracy and capacity to manage data with vast dimensions (Cristianini & Shawe-Taylor 2000). SVM aims to maximise the margin by discovering a hyper-plane between two distinct data categories. The hyper-plane linear model is described by the following equation:

$$f(x) = sign(W^T X + \mathrm{b}) \quad (26)$$

Where, w = weight vector, b = bias, term x = input vector.

7. Experimental Design

7.1 Dataset Description

Wisconsin Diagnostic Breast Cancer (WDBC) database retrieved from the UCI Machine Learning Repository was used to validate the performance of the proposed Modified WOA+GWO+SVM technique. The WDBC dataset consists of 569 records, of these, 357 patients have been classified as having benign breast cancer, while the remaining patients have been classified as having malignant breast cancer. 32 features make up each record which comprises of a patient ID, a diagnosis, and 30 real-valued attributes. These parameters define the features of the cell nuclei that the digital picture of the FNA of the breast mass captured. The ten distinct characteristics of each cell nucleus are represented by the 30 real valued qualities, which are the radius, texture, perimeter, area, smoothness, compactness, concavity, concave point, symmetry, and fractional dimension. The mean value, standard error, and maximum value for each characteristic have all been recorded. The WDBC dataset's ten feature categories are summarised in Table 2.

Table 2 Summary of WDBC Dataset

Attributes	Measurement range		
	Mean	Standard error	Maximum
Radius	6.99–28.12	0.121–2.923	7.95–37.01
Texture	9.80–40.02	0.37–4.90	112.10–50.01
Perimeter	44.02–189.09	0.80–22.01	50.48–252.03
Area	144.04–2503.01	6.90–543.10	186.01–4255.00
Smoothness	0.054–0.164	0.003–0.035	0.072–1.102
Compactness	0.020–0.350	0.002–0.138	0.030–1.060
Concavity	0.001–0.501	0.000–0.400	0.000–1.255
Concave points	0.0001–0.202	0.000–0.055	0.000–1.296
Symmetry	0.108–0.305	0.009–0.080	0.158–0.668
Fractal dimension	0.051–0.098	0.001–0.031	0.057–0.210

8. Results and Discussion

The suggested model's accuracy in classification was evaluated based on the results of a number of different performance measurements. Accuracy, sensitivity, specificity, precision, recall, and F-measure are some of the parameters that are taken into consideration while evaluating the performance of the proposed WOA+GWO+SVM technique. Matlab R2020a and LIBSVM (Version 3.3) were used in the development of the method that was proposed by Chang (2011).

By decreasing the number of parameters and making use of the modified WOA+GWO+SVM technique, the major objective of this study was to enhance classification performance and raise the level of diagnostic precision for breast cancer. In order to conduct a more precise analysis of the method that was presented, the following three instances were taken into consideration:
Case 1: Training accounts for 60% of the data, whereas testing accounts for 40% of the data
Case 2: Training accounts for 70% of the data, whereas testing accounts for 30% of the data
Case 3: Training accounts for 80% of the data, whereas testing accounts for 20% of the data

At first, the modified WOA+GWO+SVM method was tested out using the WDBC Cancer dataset with ten attributes. Table 3 displays the results of the algorithm for the evaluation parameters like that Accuracy, Sensitivity, Specificity, Precision and F-measure. The performance of the proposed modified WOA+GWO+SVM was compared with GWO-SVM suggested by Singh et al. (2020) and the results are presented here.

Table 3 Classification Performance of modified WOA+GWO+SVM Method

Parameter	Case 1		Case 2		Case 3	
	GWO -SVM	MWOA+ GWO	GWO -SVM	MWOA+ GWO	GWO-SVM	MWOA+ GWO
Accuracy	97.28	98.64	96.24	98.79	95.82	97.97
Sensitivity	98.29	99.91	95.50	96.89	96.00	98.11
Specificity	95.10	96.69	97.51	98.80	93.12	94.82
Precision	98.74	99.64	97.20	98.29	96.84	98.10
F-measure	98.43	99.47	96.33	97.63	96.84	98.20

From the above table it is inferred that the proposed WOA+GWO+SVM approach has recorded better results across the performance parameters like accuracy, Sensitivity, Specificity, Precision and F-measure when compared with GWO-SVM (Singh et al. 2020). Similarly, Case 1 scenario where the training data (60%) was higher than the testing data (40%) has recorded the best results.

9. Comparison of Results with Previous Studies

The findings obtained through the proposed method were compared to the results of different recently developed models. The purpose of this comparison was to illustrate the robustness of the proposed model. The outcomes of the other recently constructed models that made use of the Wisconsin Breast Cancer Dataset (WDBC) are shown and discussed in Table 4. On the WDBC Dataset, Asri et al. (2016) used different classifiers based on SVM, NB, K-NN, DT and generated an accuracy of 97.1%, 96.0%, 95.3% and 95.1% respectively.

The Adjusted BAT Algorithm for the WDBC Dataset (Tube et al. 2016) has achieved a success rate of 96.88% in terms of accuracy. The accuracy of the Support Vector Machine Parameters Optimisation by Enhanced Fireworks Algorithm (EFWA-SVM) algorithm that was proposed by Rube et al. (2016) was measured as 96.31%. The accuracy of 92.98%. was recorded by the improved model developed by applying the Particle Swarm Optimisation (PSO) method for optimizing SVM parameters in WDBC dataset (Alhakbani & al-Rifaie, 2017). Similarly, Wang and Chen (2020) improved the SVM parameters by utilising the Whale Optimizer Algorithm (WOA), and they achieved an accuracy of 96.65 with WBCD dataset. Table 4 makes it abundantly clear that the proposed Modified WOA+GWO+SVM algorithm has provided outstanding accuracy and exceeded the other methods already existing for diagnosing breast cancer.

Table 4 Performance Comparisons

Paper Refence	Classification Algorithm	Accuracy	Dataset
Asri et al. (2016)	SVM, NB, K-NN, DT	97.1%, 96.0%, 95.3%, 95.1%	WBCD
Alhakbani & al-Rifaie(2017)	PSO-SVM	92.98%	WBCD
Rube et al. (2016)	EFWA+SVM	96.31%.	WBCD
Tuba et al. (2016)	ABA+SVM	96.88%	WBCD
Wang & Chen (2020)	WOA+SVM	96.65%	WBCD
Proposed Work		99.78%	WBCD

10. Conclusion

When multidimensional data are combined with a variety of classification, feature selection, and dimensionality reduction algorithms, an effective model could be generated for inference in the field of medical diagnosis. A large number of actual datasets are used for training purposes in the clinical diagnosis system's data classification, which enables qualified medical experts to unearth hidden patterns and draw useful insights from data samples. As a result, it is essential for accurate prediction, analysis, and management of breast cancer in a timely manner when utilising an efficient, reliable, and effective breast cancer diagnosis model. A model of this kind will prove to be extremely helpful to medical experts in the accurate prediction and treatment of cancer patients. If breast cancer is discovered at an early stage, when it is more treatable, the prognosis will be far more favourable.

WOA, GWO and SVM were used in this paper to improve the accuracy of breast cancer detection by picking the most relevant features and data classification. MATLAB was used during the development and implementation of the method, and the dataset from UCI was utilised during the experimental testing of the algorithm. This body of work has successfully demonstrated the implementation of Whale Optimization Algorithm (WOA) and in the process of improving the performance of Support Vector Machines (SVM) for applications such as the diagnosis of breast cancer.

The best results might be achieved by combining the Whale Optimization algorithm with Grey Wolf Optimization algorithm and Support Vector Machine algorithm in order to determine the subset of useful features. The suggested method exhibited noticeably higher levels of accuracy, sensitivity, and specificity when measured against earlier algorithmic approaches. In upcoming medical investigations, this point of view may prove useful in the detection of cardiovascular disease, diabetes, as well as other disorders. In subsequent research, a particular emphasis will be placed on making comparisons between the suggested model and new medical datasets. This work is a component of a broader study, and future improvements will concentrate on incorporating feature selection strategies into the proposed algorithm and evaluating the classification performance in relation to that of other renowned models.

References

- Ades, F., Zardavas, D., Bozovic-Spasojevic, I., Pugliano, L., Fumagalli, D., De Azambuja, E., Viale, G., Sotiriou, C. and Piccart, M., 2014. Luminal B breast cancer: molecular characterization, clinical management, and future perspectives. *Journal of clinical oncology*, *32*(25), pp.2794-2803.
- Alhakbani, H.A. and al-Rifaie, M.M., 2017, September. Optimising SVM to classify imbalanced data using dispersive flies optimisation. In *2017 Federated Conference on Computer Science and Information Systems (FedCSIS)* (pp. 399-402). IEEE.
- Ali, A.H. and Abdullah, M.Z., 2020. An efficient model for data classification based on SVM grid parameter optimization and PSO feature weight selection. *International Journal of Integrated Engineering*, *12*(1), pp.1-12.
- Arya, C. and Tiwari, R., 2016, January. Expert system for breast cancer diagnosis: A survey. In *2016 international conference on computer communication and informatics (ICCCI)* (pp. 1-9). IEEE.
- Asri, H., Mousannif, H., Al Moatassime, H. and Noel, T., 2016. Using machine learning algorithms for breast cancer risk prediction and diagnosis. *Procedia Computer Science*, *83*, pp.1064-1069.
- Basir, M. A., & Hussin, M. S. (2021, September). Exploitation of Meta-Heuristic Search Methods with Bio-Inspired Algorithms for Optimal Feature Selection. In *2021 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)* (pp. 1-5). IEEE.
- Breast Cancer - WHO [Online], Available at https://www.who.int/cancer/prevention/diagnosis-screening/breast-cancer/en/.
- Chang, C.C., 2011. " LIBSVM: a library for support vector machines," ACM Transactions on Intelligent Systems and Technology, 2: 27: 1--27: 27, 2011. *http://www. csie. ntu. edu. tw/~ cjlin/libsvm*, *2*.
- Cristianini, N. and Shawe-Taylor, J., 2000. *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press.
- Dubey, A.K., Gupta, U. and Jain, S., 2015. A survey on breast cancer scenario and prediction strategy. In *Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2014: Volume 1* (pp. 367-375). Springer International Publishing.
- Guo, Z., Xu, L. and Ali Asgharzadeholiaee, N., 2022. A Homogeneous Ensemble Classifier for Breast Cancer Detection Using Parameters Tuning of MLP Neural Network. *Applied Artificial Intelligence*, *36*(1), p.2031820.

- Jazaeri, A.A., Yee, C.J., Sotiriou, C., Brantley, K.R., Boyd, J. and Liu, E.T., 2002. Gene expression profiles of BRCA1-linked, BRCA2-linked, and sporadic ovarian cancers. *Journal of the National Cancer Institute*, *94*(13), pp.990-1000.
- Kamel, S.R., YaghoubZadeh, R. and Kheirabadi, M., 2019. 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. *Journal of Big Data*, *6*(1), pp.1-15.
- Karthik, S., Srinivasa Perumal, R. and Chandra Mouli, P.V.S.S.R., 2018. Breast cancer classification using deep neural networks. *Knowledge Computing and Its Applications: Knowledge Manipulation and Processing Techniques: Volume 1*, pp.227-241.
- Khuriwal, N. and Mishra, N., 2018, March. Breast cancer diagnosis using adaptive voting ensemble machine learning algorithm. In *2018 IEEMA engineer infinite conference (eTechNxT)* (pp. 1-5). IEEE.
- Kumar, P. and Nair, G.G., 2021. An efficient classification framework for breast cancer using hyper parameter tuned Random Decision Forest Classifier and Bayesian Optimization. *Biomedical Signal Processing and Control*, *68*, p.102682.
- Maglogiannis, I., Zafiropoulos, E. and Anagnostopoulos, I., 2009. An intelligent system for automated breast cancer diagnosis and prognosis using SVM based classifiers. *Applied intelligence*, *30*(1), pp.24-36.
- Mallika, C. and Selvamuthukumaran, S., 2021. A hybrid crow search and grey wolf optimization technique for enhanced medical data classification in diabetes diagnosis system. *International Journal of Computational Intelligence Systems*, *14*(1), pp.1-18.
- Mirjalili, S., Mirjalili, S.M. and Lewis, A., 2014. Grey wolf optimizer. *Advances in engineering software*, *69*, pp.46-61.
- Nassif, A.B. and Al-Chikh Omar, A., 2022, October. A Comprehensive Study on Machine Learning in Breast Cancer Detection and Classification. In *2022 The 6th International Conference on Advances in Artificial Intelligence* (pp. 81-87).
- Priyanka, S. 2020. ICMR data shows unequal toll of cancer on women. National Cancer Registry Programme (NCRP) data for 2020-2022. Retrieved from https://www.livemint.com/news/india/icmr-data-shows-unequal-toll-of-cancer-on-women-11670349329355.html.
- Rojas-Domínguez, A., Padierna, L.C., Valadez, J.M.C., Puga-Soberanes, H.J. and Fraire, H.J., 2017. Optimal hyper-parameter tuning of SVM classifiers with application to medical diagnosis. *Ieee Access*, *6*, pp.7164-7176.
- Singh, I., Bansal, R., Gupta, A. and Singh, A., 2020, November. A Hybrid Grey Wolf-Whale Optimization Algorithm for Optimizing SVM in Breast Cancer Diagnosis. In *2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC)* (pp. 286-290). IEEE.
- Sotiriou, C., Neo, S.Y., McShane, L.M., Korn, E.L., Long, P.M., Jazaeri, A., Martiat, P., Fox, S.B., Harris, A.L. and Liu, E.T., 2003. Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proceedings of the National Academy of Sciences*, *100*(18), pp.10393-10398.
- Stephan, P., Stephan, T., Kannan, R. and Abraham, A., 2021. A hybrid artificial bee colony with whale optimization algorithm for improved breast cancer diagnosis. *Neural Computing and Applications*, *33*(20), pp.13667-13691.
- Tuba, E., Tuba, M. and Simian, D., 2016, July. Adjusted bat algorithm for tuning of support vector machine parameters. In *2016 IEEE congress on evolutionary computation (CEC)* (pp. 2225-2232). IEEE.

- UCI. Breast Cancer Wisconsin Dataset. Available online: https://archive.ics.uci.edu/ml/datasets/Breast+ Cancer+Wisconsin+(Diagnostic) (accessed on 9 May 20229).
- Wang, H., Zheng, B., Yoon, S.W. and Ko, H.S., 2018. A support vector machine-based ensemble algorithm for breast cancer diagnosis. *European Journal of Operational Research*, *267*(2), pp.687-699.
- Wang, H.; Yoon, S.W. Breast cancer prediction using data mining method. In Proceedings of the IIE Annual Conference Expo 2015, Nashville, TN, USA, 30 May–2 June 2015; pp. 818–828.
- Wang, M. and Chen, H., 2020. Chaotic multi-swarm whale optimizer boosted support vector machine for medical diagnosis. *Applied Soft Computing*, *88*, p.105946.
- Wu, J., Chen, X. Y., Zhang, H., Xiong, L. D., Lei, H., AND Deng, S. H., 2019. Hyperparameter optimization for machine learning models based on Bayesian optimization. *Journal of Electronic Science and Technology*, *17*(1), pp.26-40.
- Yin, X., Hou, Y., Yin, J. and Li, C., 2019, June. A novel SVM parameter tuning method based on advanced whale optimization algorithm. In *Journal of Physics: Conference Series* (Vol. 1237, No. 2, p. 022140). IOP Publishing.