Jundishapur J Nat Pharm Prod. 2023 November; 18(4):e142058. https://doi.org/10.5812/jjnpp-142058.

Published online 2023 November 27. **Research Article**

A Hybrid Intelligent Approach to Breast Cancer Diagnosis and Treatment Using Grey Wolf Optimization Algorithm

Mohammad Jafar Dehghan [1] and Amirabbas Azizi [2, *]

[1]Department of Computer Engineering, Technical and Vocational University (TVU), Tehran, Iran
[2]Department of Health Information Technology, School of Allied Medical Sciences, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran

[*] *Corresponding author*: Department of Health Information Technology, School of Allied Medical Sciences, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran. Email: Email: amirabbasazizi@gmail.com

Received 2023 October 14; **Revised** 2023 October 28; **Accepted** 2023 November 01.

Abstract

Background: Breast cancer is the second leading cause of death in women. The advent of machine learning (ML) has opened up a world of possibilities for the discovery and formulation of drugs. It is an exciting development that could revolutionize the pharmaceutical industry. By leveraging ML algorithms, researchers can now identify disease-related targets with greater accuracy. Additionally, ML techniques can be used to predict the toxicity and pharmacokinetics of potential drug candidates.
Objectives: The main purpose of ML techniques, such as feature selection (FS) and classification, is to develop a learning model based on datasets.
Methods: This paper proposed a hybrid intelligent approach using a Binary Grey Wolf Optimization Algorithm and a self-organizing fuzzy logic classifier (BGWO-SOF) for breast cancer diagnosis. The proposed FS approach can not only reduce the complexity of feature space but can also avoid overfitting and improve the learning process. The performance of this proposed approach was evaluated on the 10-fold cross-validation technique and the Wisconsin Diagnostic Breast Cancer dataset. Although the performance of breast cancer detection is highly dependent on classification accuracy, most good classification methods have an essential flaw in that they simply seek to maximize the accuracy of classification while ignoring the costs of misclassification among various categories. This is even more important in classification problems when the initial set of features is large. With such a large number of features, it is of special interest to search for a dependency between an optimal number of selected features and the accuracy of the classification model.
Results: In experiments, standard performance evaluation metrics, including accuracy, F-measure, precision, sensitivity, and specificity, were performed. The evaluation results demonstrated that the BGWO-SOF approach achieves 99.70% accuracy and 99.66% F-measure, which outperforms other state-of-the-art methods.
Conclusions: During the comparison of the results, it was observed that the proposed approach gives better or more competitive results than other state-of-the-art methods. By leveraging the power of ML algorithms and artificial intelligence (AI) and the findings of the current study, we can optimize the selection of natural pharmaceutical products for the treatment of breast cancer and maximize their efficacy.

Keywords: Natural Pharmaceutical Products, Breast Cancer Diagnosis, Self-Organizing Fuzzy Logic Classifier, Grey Wolf Optimization

1. Background

Breast cancer is the leading cause of death among women. Two types of abnormal cells are found in the breast: Benign and malignant. Due to the prevalence of thick and fatty tissue and the low ratio of malignant to benign cells, detecting malignant tumors is very challenging (1, 2). Classical methods for diagnosing breast cancer depend on human expertise, resulting in significant labor, time, and susceptibility to human error. Physicians use a standard system known as the breast imaging reporting and data system (BI-RADS) to communicate the findings and results of mammograms, categorizing results into groups numbered 0 through 6 (3).

Artificial intelligence (AI) algorithms can serve as appropriate models to assist physicians in diagnosing breast cancer and categorizing patients. The integration of AI in the field of medical science is of paramount importance for improving the accuracy and performance of disease diagnosis (4-6). Promising reports on machine learning (ML) methods and intelligent techniques in breast cancer diagnosis research indicate their potential to enhance diagnostic performance (7, 8). In recent years, various ML techniques have been applied to the diagnosis and classification of breast cancer to distinguish between malignant and benign cases.

Dutta et al. (9) introduced a classification approach for data mining in medicine. The approach applied an improved fireworks optimization algorithm with the best selection strategy (IFWABS) in the multilayer neural network. The IFWABS approach was tested on the Wisconsin Diagnostic Breast Cancer (WDBC) dataset, resulting in a testing accuracy of 96.98%.

Dora et al. (10) suggested a novel classification approach based on the Gauss-Newton method and sparse representation technique (GNRBA) for breast cancer diagnosis. The efficiency of the GNRBA approach was examined on the UCI WDBC dataset, resulting in a classification accuracy of 98.48%.

Saygili (11) conducted research on several classification methods for breast cancer, and the best accuracy was obtained by the multi-layer perceptron (MLP) neural network technique (98.41%). Jafari-Marandi et al. (12) offered a novel breast cancer diagnosis approach based on a life-sensitive method, a self-organizing map neural network, and an error-driven learning model (LS-SOED). This method employed a decision-oriented neural network based on hybrid supervised and unsupervised learning. The resulting classification accuracy of LS-SOED was 96.19% based on the WDBC dataset.

Wang et al. (13) suggested a novel breast cancer classification method applying the support vector machine based on weighted area under the curve and ensemble learning model (WAUCE). The simulation outcomes revealed that WAUCE achieved a 98.76% accuracy. Rao et al. (14) developed a new method based on the coherently integrated artificial bee colony optimization algorithm and gradient boosting decision tree technique (ABCODT) to select optimal feature subsets. Several UCI datasets, including the WDBC dataset, were subjected to the ABCODT approach. Tests were carried out using the WDBC dataset, resulting in a classification accuracy of 97.18%.

Liu et al. (15) employed a new hybrid technique for breast cancer detection. The information gain method, simulated annealing algorithm, and genetic wrapper approach (IGSAGAW) were used to select optimal features, and a support vector machine based on cost-sensitive (CSSVM) was used as a classifier in this hybrid technique (IGSAGAW-CSSVM). The results of experiments on the UCI-WDBC dataset demonstrated a classification accuracy of 95.7%.

Lu et al. (16) proposed a novel breast cancer classification method using a genetic optimization algorithm and an online gradient boosting model (GAOGB). On the WDBC dataset, the GAOGB method achieved a classification accuracy of 94.51%. Abdar et al. (17) presented a novel breast cancer diagnosis method by utilizing a two-layer nested ensemble model based on stacking and voting ensemble techniques and Naïve Bayes classifier (SV-Naïve-Bayes). The presented method achieved a classification accuracy of 98.07% when applied to the UCI WDBC dataset.

Dalwinder et al. (18) suggested a new breast cancer classification approach using a feature weighting technique based on the ant lion optimization algorithm and a multilayer neural network classifier (FW-BPNN). The approach was tested on the WDBC dataset, resulting in a classification accuracy of 99.3%. Kumar et al. (19) presented a novel medical data classification using genetic programming (GP) based on a new fitness function. The presented technique was applied to the WDBC dataset and achieved an accuracy of 97.69%.

Sahebi et al. (20) developed a new generalized wrapper feature selection (FS) technique based on a new parallel genetic approach (GeFeS). The performance of the GeFeS approach was evaluated using the k-nearest neighbor (KNN) classifier. Experiments conducted with the WDBC dataset demonstrated an accuracy of 98.51% for classification.

Khandezamin et al. (21) introduced a novel hybrid method for breast cancer detection. An FS algorithm based on logistic regression and a new deep neural network dubbed group method data handling (GMDH) were used in the hybrid technique. The experimental results demonstrated that the GMDH method achieved an

accuracy of 99.6% on the WDBC dataset.

Kalagotla et al. (22) presented a new method using FS based on correlation and AdaBoost techniques, along with a novel stacking technique with multi-layer perceptron, support vector machine, and logistic regression (Stacking). The method was tested on the WDBC dataset, and the classification accuracy was 97.4%. Wan et al. (23) developed a new hybrid FS approach considering feature interaction based on neighborhood rough set-based information theory (NCMI-IFS). The performance analysis was executed on the WDBC dataset and achieved a classification accuracy of 98.03%.

Although the performance of breast cancer detection highly depends on classification accuracy, most good classification methods have a fundamental flaw: They seek to maximize classification accuracy while ignoring the costs of misclassification among various categories. Based on our experience, the risk of missing a breast cancer case is unquestionably greater than the cost of mislabeling a benign case. Moreover, methods tend to overfit the data, resulting in poor generalization and high computational costs. Therefore, to increase classification performance and lower costs, it is crucial to explore a subset of efficient and optimal features while avoiding overfitting. The primary concept of FS is to choose a subset of variables that can significantly improve the time complexity and accuracy of a classification model. This is particularly important in classification problems when the initial set of features is large. With such a large number of features, it is of special interest to search for a dependency between the optimal number of selected features and the accuracy of the classification model. Therefore, it is critical for researchers to develop an effective algorithm, especially given the high cost of misdiagnosis in breast cancer diagnosis.

To carefully avoid overfitting and determine the relevance of features and class, this study proposes a novel hybrid approach that is more accurate and intelligent. The proposed approach significantly improves breast cancer classification performance and reduces the cost of misclassification. This approach employs a grey wolf optimization algorithm and a self-organizing fuzzy logic classifier (GWO-SOF) to accurately classify breast cancer data. The performance of the approach was evaluated on the WDBC dataset.

Grey wolf optimization, as a metaheuristic swarm intelligence algorithm, can provide the optimal trade-off between local search and global search. This algorithm has characteristics that are flexible, simple, and scalable, leading to favorable convergence. In addition, the SOF classifier is highly objective and non-parametric. This implies that data are not subjected to any generating model with parameters, and without any prior information and knowledge about the problems, all associated meta-parameters are immediately obtained from data. Therefore, the SOF is a highly adaptable classifier that has demonstrated excellent performance on a range of problems. On the other hand, the quality of the selection metrics and techniques has a significant impact on the performance of learning models that use the FS process. The GWO and the SOF have many inherent advantages that can be summarized as follows:

The GWO algorithm offers an optimal trade-off between local search and global search.

The GWO algorithm has the advantages of having fewer control parameters and favorable convergence.

The SOF classifier is highly objective and non-parametric.

The SOF classifier is a highly adaptable classifier.

This work hybridizes the BGWO algorithm and the SOF classifier to validate the supervised learning model effectively. The BGWO-SOF hybrid approach efficiently removes unimportant features from the feature space and generates an optimal feature subset while preserving the latent structure of the dataset.

2. Methods

In this section, the materials and methods applied in the proposed approach are described. First, the breast cancer dataset used in this work is introduced. Then, the GWO and SOF classifiers are explained. Finally, the proposed approach is presented.

2.1. Dataset

The WDBC dataset was obtained from the UCI Machine Learning Repository and utilized in this work (24). This dataset is frequently used in medical and breast cancer studies to evaluate ML approaches, including classification and FS. The tumor features in the WDBC dataset were extracted from digital images of fine needle aspirates (FNAs) of breast masses, describing characteristics of cell

Table 1. Description of the Wisconsin Diagnostic Breast Cancer (WDBC) Dataset

Attribute Number	Attributes	Comment	Attributes Range		
			Mean	Standard Error	Specificity
1	Radius	Mean of distances from the center to points on the perimeter	6.98 - 28.11	0.112 - 2.873	7.93 – 36.04
2	Texture	The standard deviation of gray-scale values	9.71 - 39.28	0.36 - 4.89	12.02 – 49.54
3	Perimeter		43.79 - 188.50	0.76 - 21.98	50.41 – 251.20
4	Area		143.50 – 2501.00	6.80 – 542.20	185.20 – 4254.00
5	Smoothness	Local variation in radius lengths	0.053 – 0.163	0.002 – 0.031	0.071 – 0.223
6	Compactness	Perimeter^2 / area - 1.0	0.019 – 0.345	0.002 – 0.135	0.027 – 1.058
7	Concavity	Severity of concave portions of the contour	0.000 – 0.427	0.000 – 0.396	0.000 – 1.252
8	Concave point	Number of concave portions of the contour	0.000 – 0.201	0.000 – 0.053	0.000 – 0.291
9	Symmetry		0.106 – 0.304	0.008 – 0.079	0.157 – 0.664
10	Fractal dimension	“Coastline approximation” - 1	0.050 – 0.097	0.001 – 0.030	0.055 – 0.208

nuclei present in the images. The WDBC dataset comprises 32 features collected from 569 individuals, including (a) an instance ID number, (b) 30 real-valued features computed for each cell nucleus, and (c) a class attribute indicating benign or malignant cases. Each instance has 10 cell nuclei properties measured, and various statistics of these 10 attributes are computed, such as mean, standard error, and maximum, resulting in a total of 30 features, as shown in Table 1. The dataset is distributed with 37.25% and 62.75% benign and malignant instances, respectively.

2.2. Feature Selection

Machine learning employs methods that allow the analysis of large amounts of data automatically (25). Classification algorithms are a type of supervised learning technique used to identify the category of new observations based on training data. In classification, a program learns from a given dataset and then classifies new observations into predefined classes. The primary objective of classification is to generalize from the training patterns to accurately categorize new patterns. The process of ML for classification begins with prerequisites, such as datasets, data cleaning procedures, FS techniques, and classification models (26-28).

Feature selection is a crucial aspect of ML. Feature selection techniques for classification problems are based on identifying significant features, and these techniques can enhance various standard ML methods (29, 30). Therefore, this process plays a vital role in the development of learning models. Selecting the appropriate and optimal feature subset can be challenging due to the complex and unpredictable interrelationships between features (31, 32). Since FS is a near-optimal (NP)-hard problem (33, 34), several optimization algorithms have been proposed to overcome its limitations. These algorithms include particle swarm optimization (PSO) (35, 36), ant colony optimization (ACO) (37, 38), whale optimization (WO) (39, 40), and GWO (41). Feature selection approaches based on optimization algorithms have the ability to efficiently explore large search spaces and often yield results that closely approximate the global solution. These approaches systematically eliminate unnecessary and redundant features, which can, in many cases, enhance the performance of learning models by reducing uncertainty and overfitting issues (42-44).

2.3. Grey Wolf Optimization Algorithm

Optimization algorithms refer to procedures for finding near-optimal solutions to multi-dimensional and complex optimization problems. One such optimization algorithm is the GWO algorithm. The GWO algorithm is a promising optimization technique based on swarm intelligence (SI) (45). This algorithm has garnered the attention of many researchers across various optimization domains (46-48). What sets the GWO algorithm apart from other evolutionary and swarm intelligence techniques are its distinctive characteristics. The GWO algorithm requires minimal parameter tuning, effectively balances global and local search, and demonstrates favorable convergence. Moreover, it is known for its simplicity of

implementation, adaptability, and scalability (49).

The GWO algorithm mimics the natural hierarchy of leadership and hunting mechanisms observed in grey wolves. It employs four types of grey wolves to simulate the hierarchy of leadership: Alpha (α), beta (β), delta (δ), and omega (ω). Grey wolves tend to live in herds and group environments, with group sizes ranging from a minimum of 5 to a maximum of 12. The hierarchy among grey wolves is highly structured, with α representing the best outcome in mathematical terms, followed by β and δ as the next level of preferred outcomes. The ω outcome is another consideration. The GWO algorithm posits that α, β, and δ wolves lead the hunt (optimization); however, ω wolves monitor these three leaders (45). The crucial phase of the hunt occurs when a wolf encircles its prey. Equations are employed to represent the encircling behavior of the GWO algorithm.

$$w\left(t+1\right)=p\left(t\right)-A.D_i \tag{1}$$

Where D_i is calculated in Equation the current iteration is defined as t; p and y are positions of prey and grey wolves, respectively.

$$D_i=\left|C.p\left(t\right)-w\left(t\right)\right| \tag{2}$$

Where A and C are coefficient vectors that are determined using Equations

$$A=2\ b.r_1-\ lb \tag{3}$$

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

Where r_1 and r_2 are random vectors in the range of [0,1], and the *lb* components are linearly lowered from 2 to 0. The hunting process is frequently directed by α. Grey wolves, both β and δ, sporadically attend the hunt. However, the exact position of the best solution (prey) in the problem's abstract search space is unknown. Therefore, it was believed that α would be the best potential solution for the simulation of wolf hunting behavior; nevertheless, β and δ would present a better understanding of where the prey could be found. As a result, the top three findings obtained were set reserved. Omegas and other search agents must change their positions to the best search agents' positions. Equation is utilized to update the positions of wolves.

$$w\left(t+1\right)=\frac{\left(W_1+\ W_2+W_3\right)}{3} \tag{5}$$

Where W_1, W_2, and W_3 are formulated in Equations to 8, accordingly.

$$W_1=\left|W_\alpha-A_1-D_{i\alpha}\right| \tag{6}$$

$$W_2=\left|W_\beta-A_2-D_{i\beta}\right| \tag{7}$$

$$W_3=\left|W_\delta-A_3-D_{i\delta}\right| \tag{8}$$

Where W_α, W_β, and W_δ are the GWO's first three most effective solutions at a certain iteration t, A1, A2, and A3 are defined in Equation and $D_{i\alpha}$, $D_{i\beta}$, and $D_{i\delta}$ are formulated in Equations to 11, accordingly.

$$D_{i\alpha}=\left|C_1.W_\alpha-W\right| \tag{9}$$

$$D_{i\beta}=\left|C2.W_\beta-W\right| \tag{10}$$

$$D_{i\delta}=\left|C3.W_\delta-W\right| \tag{11}$$

where C_1, C_2, and C_3 are formulated in Equation Finally, according to Equation the parameter lb was reduced (2 - 0) to emphasize global and local search.

$$lb=2-t\ \frac{2}{Maxit} \tag{12}$$

Where t denotes the current number of iterations, and Maxit denotes the maximum iterations permitted in the GWO.

2.4. Binary GWO

The binary GWO algorithm has been called BGWO, where each solution contains a combination of 0's and 1's. Emary et al. proposed a new BGWO algorithm applied for FS tasks (50). In this algorithm, the wolves that updated the equation represented a three-position vector function: W_α, W_β, and W_δ, which were in charge of inviting each of the wolves to the best three outcomes. The position of a specific wolf is included using the GWO principle while maintaining the binary constraint according to Equation The scholars employed the update on the GWO method, which is detailed in Equations to 23. The core update equation is expressed in Equation

$$W_i^{t+1}=\text{"}Crossover\text{"}\ \left(w_1.w_2.w_3\right) \tag{13}$$

Where w_1, w_2, and w_3 binary vectors reflect the influence of a wolf's movement toward α, β, and δ grey wolves in order. w_1, w_2, and w_3 vectors are computed by Equations and 20, respectively.

$$W_i^{t+1}=\begin{cases}1\ if\ \left(w_\alpha^d+stepb_\alpha^d\right)\geq 1\\ 0\ otherwise\end{cases} \tag{14}$$

where w_{α}^{d} is the α wolf position in dimension d, and $stepb_{\alpha}^{d}$ is a binary step in dimension d that can be determined in Equation

$$stepb_{\alpha}^{d} = \begin{cases} 1\, if\ stepb_{\alpha}^{d} \geq urand \\ 0\, otherwise \end{cases} \quad (15)$$

where urand is a random number generated from a standard uniform distribution in the range (0,1), and $stepb_{\alpha}^{d}$ is the dimension d's continuous-valued step size that can be computed using the sigmoidal function as in Equation

$$stepb_{\alpha}^{d} = \frac{1}{1+e^{-10\left(A_{1}^{d} Di_{\alpha}^{d}-0.05\right)}} \quad (16)$$

where A_{1}^{d} and Di_{α}^{d} are determined by Equations

$$w_{2}^{d} = \begin{cases} 1\, if\ \left(w_{\beta}^{d}+stepb_{\beta}^{d}\right) \geq 1 \\ 0\, otherwise \end{cases} \quad (17)$$

where w_{β}^{d} is the β wolf position vector in dimension d, and $stepb_{\beta}^{d}$ is a binary step in dimension d, which can be determined as in Equation

$$stepb_{\beta}^{d} = \begin{cases} 1\, if stepb_{\beta}^{d} \geq urand \\ 0\, otherwise \end{cases} \quad (18)$$

where urand is a random number generated from a standard uniform distribution in the range (0,1), and $stepb_{\beta}^{d}$ is the dimension d's continuous-valued step size that can be computed using the sigmoidal function as in Equation

$$stepb_{\beta}^{d} = \frac{1}{1+e^{-10\left(A_{2}^{d} Di_{\beta}^{d}-0.05\right)}} \quad (19)$$

where A_{2}^{d} and $Di_{\beta}^{d}$are determined by Equations

$$w_{3}^{d} = \begin{cases} 1\, if\ \left(w_{\delta}^{d}+stepb_{\delta}^{d}\right) \geq urand \\ 0\, otherwise \end{cases} \quad (20)$$

where w_{3}^{d} is the δ wolf position vector in the dimension d, and $stepb_{\delta}^{d}$ is a binary step in dimension d, which can be determined as in Equation

$$stepb_{\delta}^{d} = \begin{cases} 1\, if\ stepb_{\delta}^{d} \geq urand \\ 0\, otherwise \end{cases} \quad (21)$$

where urand is a random number generated from a standard uniform distribution in the range (0,1), and $stepb_{\delta}^{d}$ is the dimension d's continuous-valued step size that can be computed using the sigmoidal function as in Equation

$$stepb_{\delta}^{d} = \frac{1}{1+e^{\left(-10\left(A_{\delta}^{d} Di_{\delta}^{d}-0.05\right)\right.}} \quad (22)$$

where A_{δ}^{d} and Di_{δ}^{d} are determined by Equations A simple technique of stochastic crossover is used in each dimension to crossover w_1, w_2, and w_3 outputs by Equation

$$w_{d} = \begin{cases} w_{1}^{d}\, if\ rad < \frac{1}{3} \\ w_{2}^{d}\, if\ < \frac{1}{3}\, rad\ < \frac{2}{3} \\ w_{3}^{d}\, otherwise \end{cases} \quad (23)$$

2.5. Self-Organizing Fuzzy Logic

Gu et and Angelov introduced a novel classifier model based on SOF (51). The SOF classifier utilizes non-parametric statistical operators to objectively reveal essential data patterns, even in the absence of empirically acquired data samples. It identifies local peaks within the multi-modal data distribution to serve as prototypes. Additionally, the SOF classifier is highly objective and non-parametric. This means that it does not rely on a predefined model with parameters. Instead, it derives all associated meta-parameters directly from the data itself. Depending on the complexity of the problem and the availability of computational resources, the SOF classifier can address issues at various levels of granularity or detail. Furthermore, it supports both online and offline learning and can classify data using various dissimilarity/distance criteria. Therefore, the SOF is a versatile classifier known for its excellent performance across a range of problems. In this paper, the offline learning mode of the SOF classifier will be utilized.

The SOF classifier's offline method involves independently detecting prototypes for each class and constructing a zero-order fuzzy rule of the AnYa type based on the identified prototypes for each class (in the structure of Equation

The AnYa-type fuzzy rule-based scheme was introduced in (52) as an alternative approach to the commonly used fuzzy rule-based schemes, such as Takagi-Sugeno (53) or Mamdani (54) models. In comparison to the two previous

models, the pattern component (IF) in AnYa-type fuzzy rules is streamlined into a more concise, objective, and non-parametric vector structure without requiring the definition of ad-hoc membership functions, as needed in the two aforementioned predecessors. The following is the form of a zero-order fuzzy rule of the AnYa type:

$$If\ (xin \sim pro_1)\ OR\ (xin \sim pro_2)\ OR\ (xin \sim pro_N)\ THEN\ (class) \tag{24}$$

Where xin signifies vector of input and "~ " signifies similarity, which can also be considered a fuzzy degree of membership/satisfaction (55); pro_i (i=1,2,..., Np) represents the class's ith prototype; Np is the number of prototypes discovered from the data samples of this class. Different strategies, such as "fuzzily weighted average", might be used to determine the label for a specific data sample.

The fuzzy rule training procedures of separate classes will have no effect on each other. We will suppose for the remainder of this section that the training procedure is conducted using data samples from the cth class (c=1,2,..., C) indicated by $\{x\}_{k^c}^c = \{x_1^c, x_2^c, \ldots, x_{k^c}^c\}$ ($\{x\}_{k^c}^c \subset \{x\}_{k)}$, and the frequency of occurrence and the associated unique data sample set are denoted, respectively, by $\{u\}_{U_k^c}^c = \left\{u_1^c, u_2^c, \ldots, u_{U_k^c}^c\right\}$ and $\{f\}_{U_k^c}^c = \left\{f_1^c, f_2^c, \ldots, f_{U_k^c}^c\right\}$ where K^c is the number of data samples with $\{x\}_{k^c}^c$, U_k^c is the number of unique data samples of the cth class. Considering all the classes, we have $K = \sum_{c1}^{C} K^c = k\ and \sum_{c1}^{C} U_k^c = U_k$.

Prototypes are found using the densities and mutual distributions of data samples in the method. To begin with, multi-modal densities $D_{k^c}^{MM}(u_i^c) = f_i^c \frac{\sum_{l=1}^{k^c}\sum_{j=1}^{k^c} d^2(x_l^c, x_j^c)}{2k^c \sum_{j=1}^{k^c} d^2(x_i^c, x_j^c)}$ $(i = 1, 2, \ldots, U_k^c)$ (52, 56) at all the samples of unique data in $\{u\}_{U_k^c}^c$ are computed $D_k^{MM}(u_i) = f_i D_k(u_i) = f_i \frac{\sum_{i=1}^{k}\pi_k(x_i)}{2k\pi_k(u_i)}$ $i = 1, 2, \ldots, U_k)$. Then, in a list defined by {r}, the data samples are ordered according to multi-modal density values and their reciprocal distances.

By discovering the sample of data with the largest multi-modal density, $r_1 = \underset{i=1,2,\ldots,U_k^c}{argmax}(D_{k^c}^{MM}(u_i^c))$, the first element of list {r} is recognized (r_1). Then, the data sample was determined as the second element (r_2) that is closest to r_1 in terms of distance: $r_2 = \underset{i=1,2,\ldots,U_k^c-1}{argmin}(d(r_1, u_i^c)$. The minimal distance to r_2 is used to identify the third element of list {r}, indicated by r_3.

The entire list{r} is built by repeating the procedure until each of the data samples has been chosen, and based on the list{r}, the multi-modal densities of $\{u\}_{U_k^c}^c$ are ordered, indicated by $D_{k^c}^{MM}(r)$ (57, 58).

It is important to note that after a data sample is selected into list{r}, it cannot be selected for a second time.

Prototypes, indicated by $\{p\}_0$ are then

recognized as the local maximum of the ordered multi-modal densities, $D_{k^c}^{MM}(r)$, by condition one:

Condition 1:

$$IF\left(D_{k^c}^{MM}(r_i) > D_{k^c}^{MM}(r_{i+1})\ and\ D_{k^c}^{MM}(r_i) > D_{k^c}^{MM}(r_{i-1})\right)\ THEN\ (r_i \in \{p\}_0) \tag{25}$$

After all of the prototypes have been recognized with Equation some fewer representative ones might be found in $\{p\}_0$, thereby necessitating the use of a filtering process to eliminate them from P0.

Before beginning the filtering process, use the prototypes to attract close data samples to construct data clouds (55), similar to Voronoi tessellation (58):

$$Winning\ prototype = argmin\left(d(x_i, p);\ x_i \in \{x\}_{k^c}^c,\ p \in \{p\}_0\right) \tag{26}$$

After all the clouds of data are generated around the available prototypes $\{p\}_0$, one can acquire the data cloud centers Indicated by $\{\varphi\}_0$, and the multi-modal densities at the centers are computed by $D_k^{MM}(u_i) = f_i D_k(u_i) = f_i \frac{\sum_{i=1}^{k}\pi_k(x_i)}{2k\pi_k(u_i)}$ $i = 1, 2, \ldots, U_k)$ as $D_{k^c}^{MM}(\varphi_i) = S_i D_{k^c}(\varphi_i)$, where $\varphi_i \in \{\varphi\}_0$; S_i is the ith data cloud's support.

Following that, for each data cloud, supposing the ith one ($\varphi_i \in \{\varphi\}_0$), the set of the centers of its neighboring data clouds, indicated using $\varphi_i^{neighboring}$, are identified by Condition two:

Condition 2:

$$IF\left(d^2(\phi_i, \phi_j) \le G_{K^c}^{c,l}\right)\ THEN\left(\phi_i \in \{\phi\}_j^{neighboring}\right) \tag{27}$$

Where $(\varphi_j \in \{\varphi\}_{0,\ \varphi_i \neq \varphi_j}$; $G_{k^c}^{c,l}$ is described as the average radius of the local influential region surrounding, which corresponds to the Lth granularity level (L=1,2,3,...) and is produced from the cth class data based on an offline mode. Finally, the most representative prototypes of the cth class, indicated by $\{p\}^c$, are chosen from the centers of the available data clouds that fulfill Condition 3 (58):

Condition 3:

$$IF\left(D_{k^c}^{MM}(\phi_i) > \max_{\phi \in \left\{\{^{\phi}\}\right\}_j}{}_{neighboring} D_{k^c}^{MM}(\phi)\right) \quad (28)$$

$$THEN\ (\phi_i \in \{p\}^c)$$

In the end, the representative prototypes of the cth class $\{p\}^c$ are recognized, the fuzzy rule of AnYa type might be constructed as follows, where N^c denotes the number of prototypes in $\{p\}^c$:

$$If\ (x \sim p_1^c)\ OR\ (x \sim p_2^c)\ OR\ (x \sim p_{N^c}^c) \quad (29)$$

$$THEN\ (class\ c)$$

2.6. The Hybrid Intelligent Method

Taking into consideration the advantages of BGWO and SOF and the importance of breast cancer classification, this study proposes an intelligent approach for distinguishing benign from malignant breast cancers. In the proposed approach, BGWO acts as an FS technique to select the effective and optimal features; nevertheless, SOF functions as the classifier to evaluate the performance of these optimal features. The procedure of the proposed approach (BGWO-SOF) is described below.

The procedure begins with normalizing the values of the WDBC dataset and initializing the parameters for the BGWO algorithm. Then, the K-fold cross-validation technique is employed (with K = 10) to assess how effectively the classification approach can predict the tumor characteristics of an unknown instance. For each fold, the dataset is divided into 10 equally sized subsets. Consequently, in each fold, 9 subsets serve as the training data (90% of the dataset); however, 1 subset (10% of the dataset) is reserved for testing purposes.

In each fold, the algorithm executes the optimization process by generating an initial population of candidate solutions (i.e., individuals) within the search space. Each position of an individual is represented as a vector with N elements, where N is the number of features in the dataset. A 0 value indicates that a certain feature is not selected; nonetheless, a 1 value indicates that the related feature is selected. Each individual in the feature space constitutes a set of candidate features. The BGWO solution is depicted in Figure 1 (for example, N = 10).

Then, the fitness for each solution (Xi) is calculated. Because the approach's primary goal is to improve the performance of the classification, the quality of a solution is determined by two key criteria: The number of BGWO-selected features in the solution and the SOF classifier's error rate. Therefore, the optimal solution is a combination of features with the fewest number of selected features and the highest performance of classification. In this paper, the fitness function in Equation by the SOF classifier is applied to evaluate the quality of the features selected by the BGWO.

$$Fitness = \omega SOF_{Error} + \phi\frac{|f|}{|F|} \quad (30)$$

Where SOF_{Error} the SOF's error rate on the WDBC dataset applying optimal selected features. F is the number of the complete set of features in the cancer breast dataset, and f is the number of the selected features in the solution. $\omega \in (0,1)$ and φ =1-ω are balance factors between SOF_{Error} and the size of selected features to control the importance of feature space reduction and classification performance. The φ is equal to 0.05 in the present experiments.

At each iteration, the fitness or classification error rate of the new solution is compared to the fitness of the previous solution, and if it shows an improvement, the new solution is chosen. The process is repeated until the total number of iterations reaches a certain limit called Maxit. Finally, the sets of the selected features are used in the SOF learning model.

This procedure will be repeated until all of these subsets apply for both the training and testing phases. Finally, the evaluation metrics results from the 10 iterations are averaged to produce reliable statistical results. Figure 2 shows a flowchart of the proposed approach.

3. Results

This section presents the experimental setup, including the dataset and parameter settings, evaluation

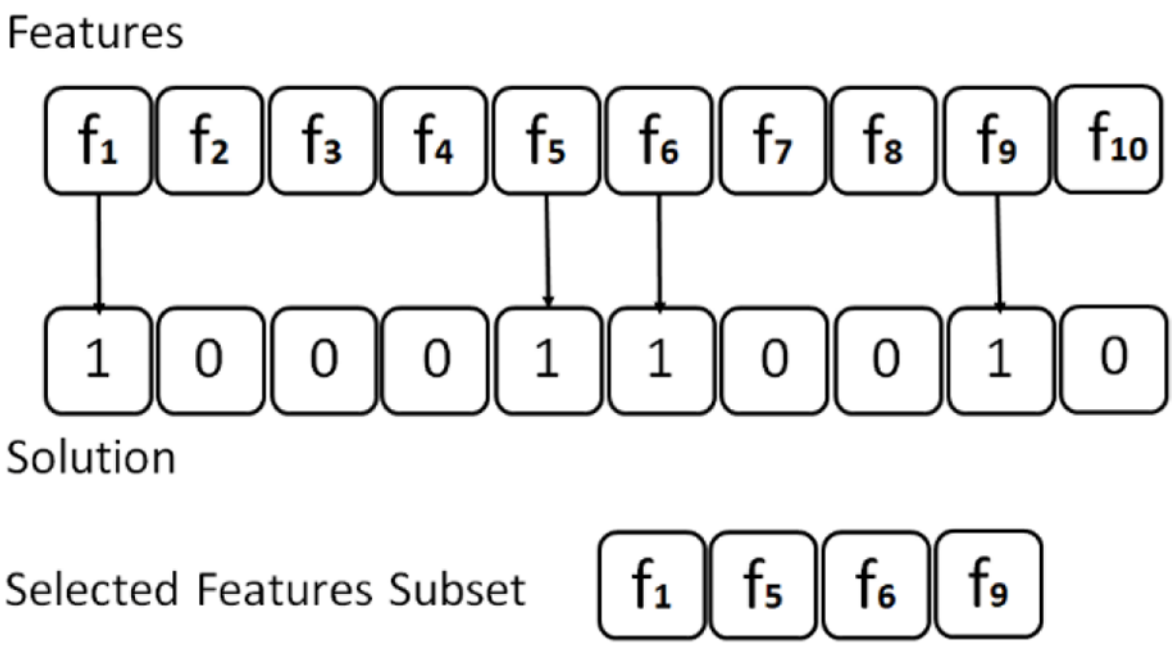

Figure 1. Solution representation of feature selection

metrics, and details of the experiment's execution and results analysis, to verify and examine the performance of the proposed approach.

3.1. Experimental and Parameter Settings

This study evaluated the performance of BGWO-SOF using the WDBC breast cancer dataset. For the experimental executions, this study utilized an Intel(R) Core (TM) i5 CPU 8250U 1.6GHz with 8 GB of RAM, running MATLAB 2018 on a Windows 10 (64-bit) operating system. The BGWO-SOF was executed with a population size of 20. The approach was formulated as an optimization problem, and BGWO was run for 50 iterations. In the SOF classifier, the Euclidean distance was employed as a dissimilarity metric with a granularity level equal to 12. To prepare for experimentation, the values from the WDBC dataset were normalized. Subsequently, a 10-fold cross-validation technique was employed to assess the performance of the proposed approach.

3.2 Evaluation Metrics

This work focused on a two-class classification problem, where a classifier produces two discrete results: Negative and positive. The experiments were conducted on the WDBC dataset, resulting in four possible outcomes in the confusion matrix, as shown in Figure 3. When a positive instance is correctly classified as positive, it is a true positive (TP). Conversely, a false negative (FN) occurs when a positive instance is incorrectly classified as negative. Similarly, when a negative instance is correctly classified as negative, it is a true negative (TN); however, an incorrect classification of a negative instance as positive is a false positive (FP).

In the WDBC dataset, patient instances with benign tumors are labeled as 0 (negatives); however, those with malignant tumors are labeled as 1 (positives). True negatives are instances where the actual class was negative, and the expected class was also negative. True positives are instances with a positive current class and a positive expected class. False negatives occur when a record's true class is positive but is expected to be negative; nevertheless, FPs occur when the record's actual class is negative; nonetheless, the expected class is positive. When dealing with cancer diagnosis results, minimizing FNs is crucial.

This study employed classification performance metrics, including accuracy, precision, recall (also known

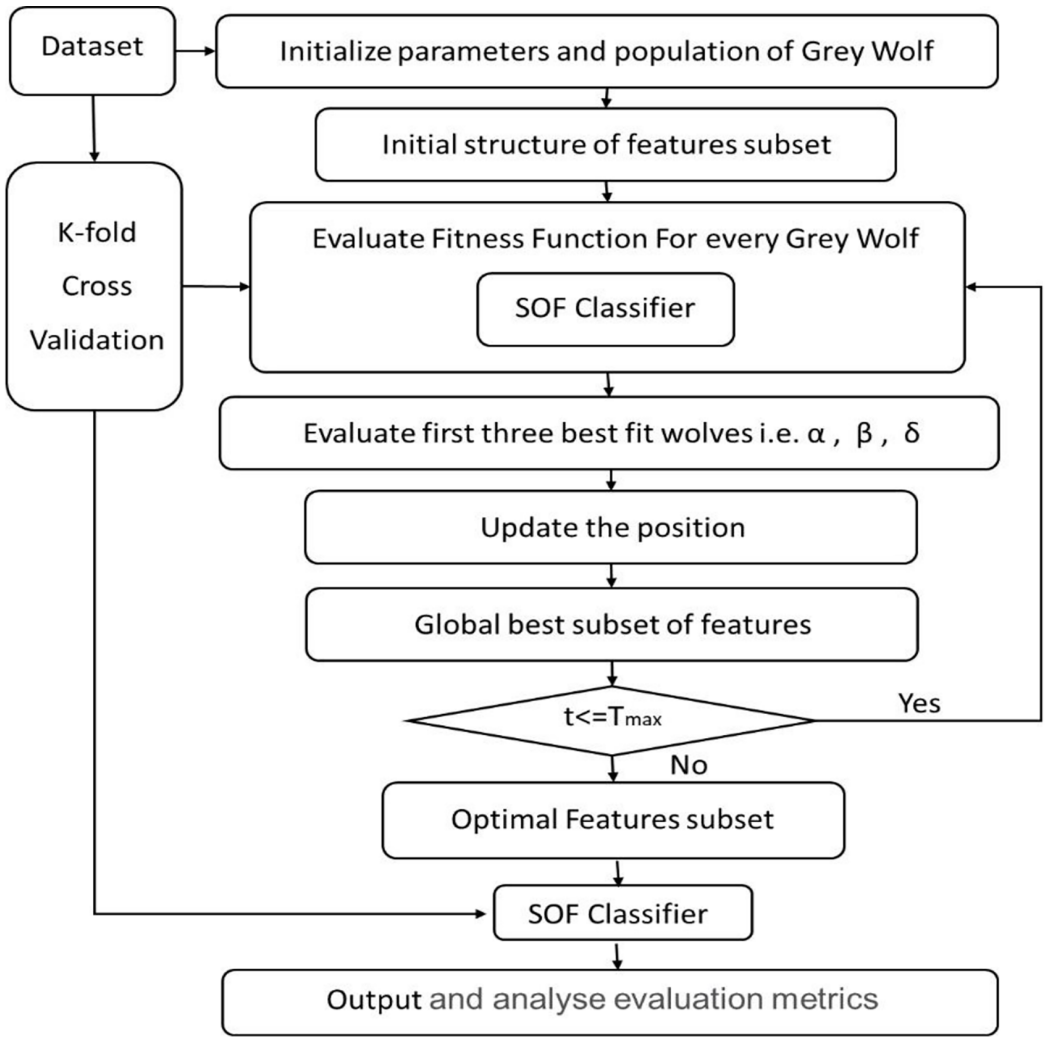

Figure 2. Flowchart of the proposed approach

as sensitivity), F-measure, Matthews correlation coefficient (MCC), and specificity, for the evaluation and comparison of the proposed approach on the WDBC dataset. Accuracy measures the percentage of times a classifier produces correct results. The precision determines how accurately a classifier predicts a positive pattern (yes). Recall, also known as sensitivity or true positive rate, indicates how well a classifier predicts a pattern's identity. Specificity measures the frequency at which a classifier predicts non-patterns (no). The F-Measure is the harmonic average of recall and precision. Matthews correlation coefficient assesses the quality of binary classifications and is a more reliable statistical rate than balanced accuracy. It produces a high score only if the prediction performs well in all four confusion matrix categories in proportion to both the size of positive and negative elements in the dataset. The following formulas are used to calculate the evaluation metrics:

The trade-off between the FP rate (1-specificity) and the TP rate (sensitivity) is depicted by a receiver operating characteristic curve (ROC curve). The ROC curve is a graphical diagram that illustrates how the diagnostic capacity of a binary classifier system changes with its discrimination threshold. Classifiers that yield curves closer to the top-left corner indicate better performance. The ROC curve does not depend on the class distribution,

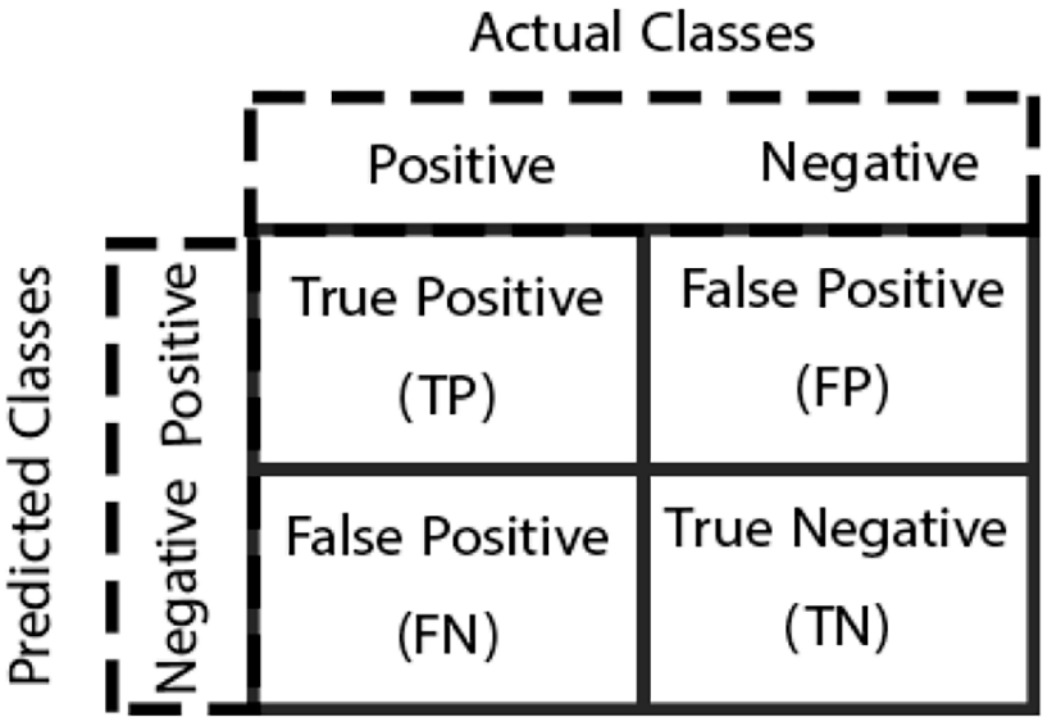

Figure 3. Confusion matrix in a two-class classification

making it useful for evaluating classifiers predicting rare events.

3.3. Experimental Results

The first experiment evaluates the performance of the present proposed approach (BGWO-SOF) and the SOF classifier without feature selection capability (SOF-WFS). The classification ability of the method is assessed using 10-fold cross-validation. Table 2 shows the experimental results as the average of the 10-fold cross-validation outcomes. Table 2 shows that by using BGWO-SOF, classification accuracy, and other metrics have improved dramatically, with the BGWO-SOF approach outperforming the SOF-WFS method. The average accuracy for the SOF method was 80.54%; however, the best accuracy for BGWO-SOF is 99.702%; nevertheless, the number of features was reduced by approximately 64%. Since the 10-fold cross-validation method was applied in this study, the ROC curve is generated from 10 folds. Figure 4 illustrates the ROC curve for the 10 folds, demonstrating the classifier's efficiency and its capability to distinguish between classes. Figure 5 presents the number of selected features by the proposed approach in each iteration. The proposed approach selects the best and optimal subset of features, reducing the feature count by 62% in the FS process. Figure 6 indicates the frequencies of the selected features using the BGWO-SOF approach on the WDBC dataset over 10 iterations.

4. Discussion

In this study, the BGWO algorithm was combined with the SOF classifier to validate the supervised learning model effectively. The BGWO-SOF hybrid approach can efficiently eliminate unimportant features from the feature space and generate an optimal feature subset while capturing the latent structure of the dataset. The quality of the selection metrics and techniques significantly impacts the performance of learning models utilizing the FS process. To demonstrate the effectiveness of the proposed approach, the BGWO-SOF findings (average metric values) were compared to several recently published approaches

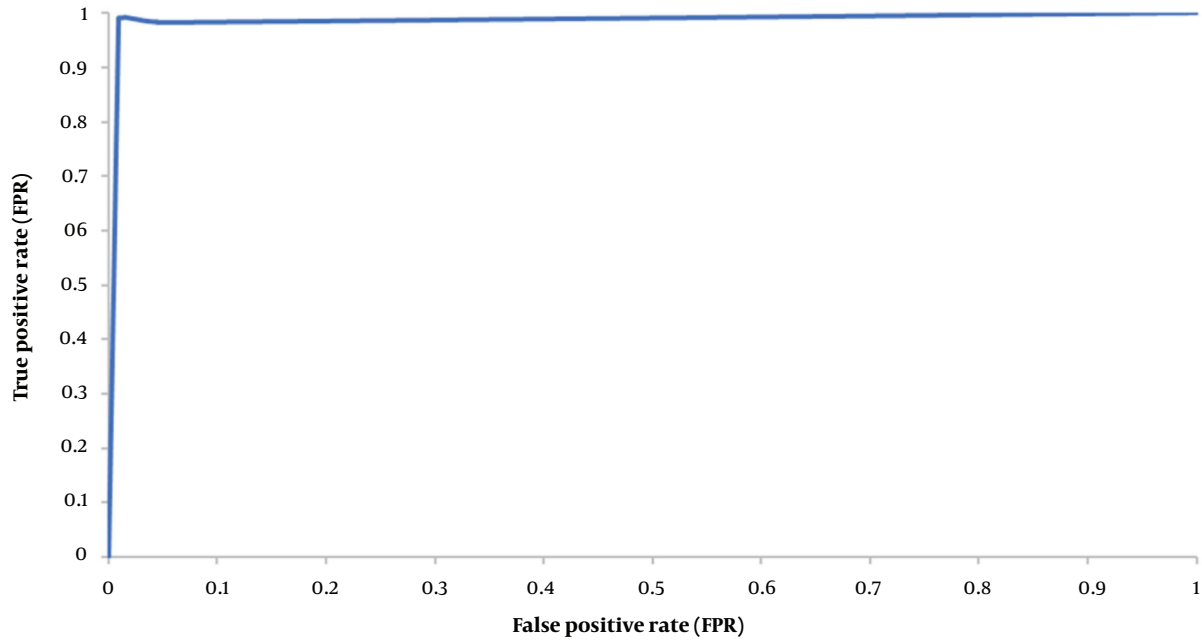

Figure 4. The receiver operating characteristic (ROC) curve

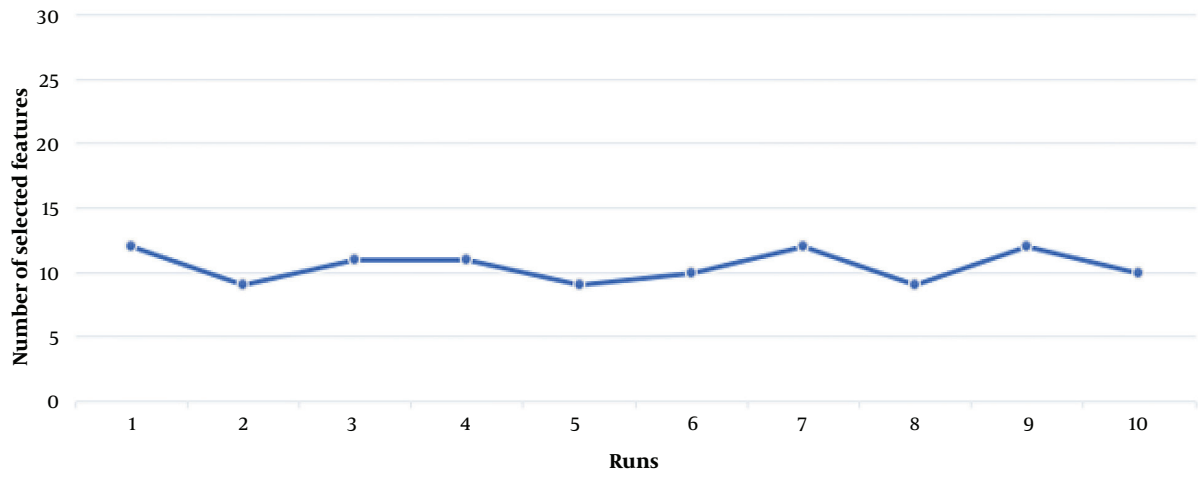

Figure 5. Number of selected features

Table 2. Experimental Results (%) of Binary Grey Wolf Optimization-Self-Organizing Fuzzy Logic Classifier (BGWO-SOF) and SOF Classifier Without Feature Selection (SOF-WFS) Methods

Method	Accuracy	Sensitivity	Specificity	Precision	F-measure	MCC
SOF without feature selection	80.54	85.33	72.34	80.27	82.60	78.13
BGWO-SOF (feature selection)	99.702	99.681	99.679	99.646	99.658	98.973

Abbreviations: MCC, matthews correlation coefficient; BGWO-SOF, binary grey wolf optimization-self-organizing fuzzy logic classifier

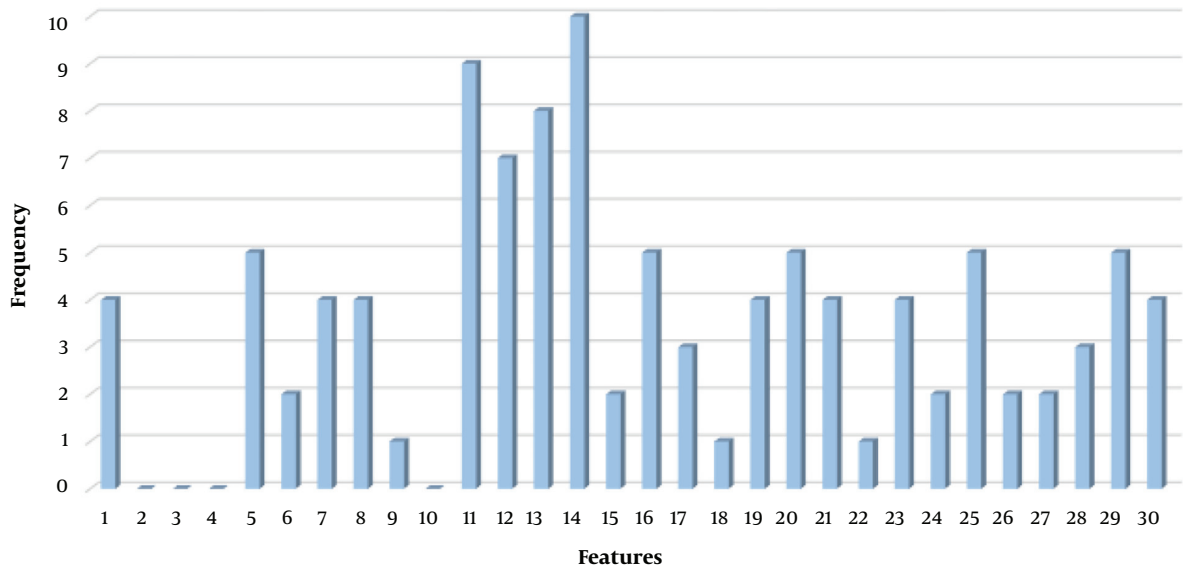

Figure 6. Frequency of selected features

in the field of FS and breast cancer classification. This study examined articles in the literature that utilized the same dataset (WDBC) and evaluation metrics (8, 12-15, 17-23, 59)

Table 3 shows the results of this comparison. In terms of classification accuracy, the results demonstrated that the proposed approach is more accurate and robust than state-of-the-art methods. Although classification accuracy is highly important for breast cancer detection, many good classification methods have significant drawbacks in that they aim solely to maximize classification accuracy, often ignoring the costs associated with misclassification across different categories. Evaluating a learning model's performance solely based on classification accuracy might not indicate its superiority. As shown in Table 3, to further assess the likelihood of class imbalance and its impact on accuracy, this study examined metrics such as recall, precision, F-measure, and specificity. The average F-measure value, close to 1 (0.99742), indicates that the proposed approach has a very low number of FP and FN, signifying excellent precision and recall percentages. Additionally, the average MCC value (close to 1) suggests that both classes are predicted effectively. Figure 7 depicts comparisons between the present study's approach's results and the findings of previous research. The results demonstrated that the proposed approach outperforms other state-of-the-art methods in all evaluation metrics.

In general, the results of the proposed approach are satisfactory for all the evaluation metrics. The evaluation results indicated that the proposed hybrid approach can not only effectively reduce the feature space dimensions but also ensure the efficiency and acceptable performance of the classification model. The experimental and comparison results for all the evaluation metrics demonstrate that the proposed approach outperforms other well-known methods in classification.

Since we used the standard WDBC dataset for this study, there was no access to the BI-RADS stages of cases in this dataset. Therefore, only two classification levels were

Table 3. Comparison of Results with State-of-the-art Methods

Variables	LS-SOED (12)	WAUCE (13)	ABCoDT (14)	IGSAGAW-CSSVM (15)	GAOGB (8)	SV-Naive-Bayes (17)	FW-BPNN (18)	GP (19)	GeFeS (20)	GMDH (21)	Stacking (22)	NCMI_IFS (23)	ESO-GSO (59)	BGWO-SOF
Accuracy	96.19	97.68	97.18	95.7	94.28	98.07	98.37	97.69	98.51	99.60	97.4	98.03	98.95	99.702
Recall		94.75	-	-	93.11	-	-	96.03	98.1	99.53	92.5	-	96.96	99.681
Specificity	-	99.49	-	-	93.20	-	-	97.02	-	-	-	-	-	99.679
Precision	-		-	-	-	-	-	95.03	99.3	99.53	96.2	-	100	99.646
F-measure	-	-	-	-	-	-	-	95.02	98.7	99.53	94.2	-	96.96	99.658
MCC	-	-	-	-	-	-	-	-	-	-	-	-	-	98.973

Abbreviations: Life-Sensitive Self-Organizing Error-Driven (LS-SOED); Weighted Area Under the Receiver Operating Characteristic Curve Ensemble (WAUCE); artificial bee colony and gradient boosting decision tree algorithm (ABCoDT); information gain directed simulated annealing genetic algorithm wrapper (IGSAGAW); cost sensitive support vector machine (CSSVM); genetic algorithm-based online gradient boosting (GAOGB); stacking and vote (SV); back-propagation neural networks (FW-BPNN); genetic programming (GP); generalized feature selection algorithm (GeFeS); group method data handling (GMDH); interaction feature selection algorithm based on neighborhood conditional mutual information (NCMI_IFS); eagle strategy optimization (ESO); gravitational search optimization (GSO); binary grey wolf optimization algorithm and a self-organizing fuzzy logic classifier (BGWO-SOF); matthews correlation coefficient (MCC).

considered: Benign and malignant. For future studies, it is recommended to use datasets of real patients with specific tumor grades (BI-RADS) to extract features for each BI-RADS category and evaluate the FS and classification performance for each of them. Furthermore, future studies can include more patient characteristics, such as demographic features, to create more accurate models.

It is important to note that, despite using the WDBC dataset from the UCI Machine Learning Repository, the proposed approach does not have any specific limitations for testing on real datasets. In practice, when using a real dataset, the following operations must be taken into account to ensure a clean dataset:

Features must be extracted from real samples.

Valid samples must be separated, and damaged or invalid samples must be removed.

An expert doctor must classify the data samples as benign or malignant.

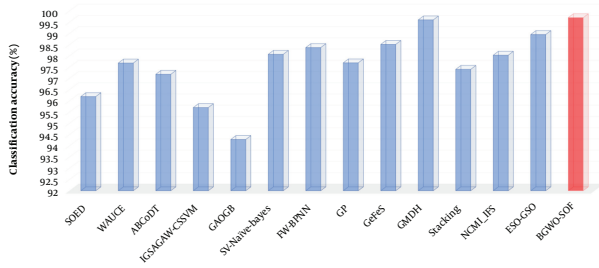

Figure 7. Accuracy comparison of the proposed approach to state-of-the-art methods (abbreviations: Weighted area under the receiver operating characteristic curve ensemble (WAUCE); artificial bee colony and gradient boosting decision tree algorithm (ABCoDT); information gain directed simulated annealing genetic algorithm wrapper (IGSAGAW); cost sensitive support vector machine (CSSVM); genetic algorithm-based online gradient boosting (GAOGB); Back-propagation neural networks (FW-BPNN); genetic programming (GP); generalized feature selection algorithm (GeFeS); group method data handling (GMDH); interaction feature selection algorithm based on neighborhood conditional mutual information (NCMI_IFS); eagle strategy optimization (ESO); gravitational search optimization (GSO); binary grey wolf optimization algorithm and a self-organizing fuzzy logic classifier (BGWO-SOF)

4.1. Conclusions

Breast cancer ranks among the leading causes of mortality worldwide, particularly among women. This pressing issue has prompted extensive research in the field of medicine. The primary objective of this study was to introduce an intelligent approach to breast cancer detection, with the aim of aiding clinical practitioners in making more informed decisions in the future.

In this paper, we proposed a novel hybrid intelligence approach that combines the GWO algorithm with the SOF classifier. The performance of this hybrid approach was assessed using the WDBC dataset and a stratified 10-fold cross-validation. Various standard performance evaluation metrics, including accuracy, F-measure, precision, sensitivity, and specificity, were employed in the experiments. Upon comparing the results, it was observed that the proposed approach consistently yielded superior or competitive results when compared to other state-of-the-art methods. In the future, it is planned to further develop and expand this approach by incorporating additional optimization algorithms and classification techniques.

Acknowledgments

This work has received support from the Research and Technology Deputy of Ahvaz Jundishapur University of Medical Sciences (AJUMS), Ahvaz, Iran, under grant number U-98094. The authors would like to express their appreciation for the support provided for this work.

Footnotes

Authors' Contribution: A. A. and M. D. served as the principal investigators and conceived the study. A. A. oversaw the overall administration of the grant. Data analysis expertise was contributed by A. A. and M. D. All the authors participated in the critical revision and protocol design phases of the study.

Conflict of Interests: The authors declare that there is no conflict of interest associated with this study.

Ethical Approval: This study has received ethical approval from the Research Ethics Committee of Ahvaz Jundishapur University of Medical Sciences, with ethics number IR.AJUMS.REC.1398.378.

Funding/Support: This work was financially supported by the Research and Technology Deputy of Ahvaz Jundishapur University of Medical Sciences (AJUMS) under grant number U-98094. The authors would like to express their gratitude for their support of this work.

References

1. Aličković E, Subasi A. Breast cancer diagnosis using GA feature selection and Rotation Forest. *Neural Computing applications*. 2017;**28**:753–63.
2. Chen HL, Yang B, Wang G, Wang SJ, Liu J, Liu DY. Support vector machine based diagnostic system for breast cancer using swarm intelligence. *J Med Syst*. 2012;**36**(4):2505–19. [PubMed ID: 21537848]. https://doi.org/10.1007/s10916-011-9723-0.
3. Swain M, Jeudy M. Breast Masses in Biological Females. *JAMA*. 2022;**328**(3):294–5. [PubMed ID: 35771591]. https://doi.org/10.1001/jama.2022.9554.
4. Kocher MR, Chamberlin J, Waltz J, Snoddy M, Stringer N, Stephenson J, et al. Tumor burden of lung metastases at initial staging in breast cancer patients detected by artificial intelligence as a prognostic tool for precision medicine. *Heliyon*. 2022;**8**(2). e08962. [PubMed ID: 35243082]. [PubMed Central ID: PMC8873537]. https://doi.org/10.1016/j.heliyon.2022.e08962.
5. Alshayeji MH, Ellethy H, Gupta R. Computer-aided detection of breast cancer on the Wisconsin dataset: An artificial neural networks approach. *Biomedical Signal Processing Control*. 2022;**71**:103141.
6. Huang PW, Ouyang H, Hsu BY, Chang YR, Lin YC, Chen YA, et al. Deep-learning based breast cancer detection for cross-staining histopathology images. *Heliyon*. 2023;**9**(2). e13171. [PubMed ID: 36755605]. [PubMed Central ID: PMC9900267]. https://doi.org/10.1016/j.heliyon.2023.e13171.
7. Sarkar JP, Saha I, Sarkar A, Maulik U. Machine learning integrated ensemble of feature selection methods followed by survival analysis for predicting breast cancer subtype specific miRNA biomarkers. *Comput Biol Med*. 2021;**131**:104244. [PubMed ID: 33550016]. https://doi.org/10.1016/j.compbiomed.2021.104244.
8. Lu H, Wang H, Yoon SW. A dynamic gradient boosting machine using genetic optimizer for practical breast cancer prognosis. *Expert Systems with Applications*. 2019;**116**:340–50.
9. Dutta RK, Karmakar NK, Si T. Artificial neural network training using fireworks algorithm in medical data mining. *Int. J. Comput. Appl.* 2016;**137**(1):1–5.
10. Dora L, Agrawal S, Panda R, Abraham A. Optimal breast cancer classification using Gauss–Newton representation based algorithm. *Expert Systems with Applications*. 2017;**85**:134–45.
11. Saygili A. Classification and diagnostic prediction of breast cancers via different classifiers. *International Sci Vocational Studies J*. 2018;**2**(2):48–56.
12. Jafari-Marandi R, Davarzani S, Gharibdousti MS, Smith BK. An optimum ANN-based breast cancer diagnosis: Bridging gaps between ANN learning and decision-making goals. *Applied Soft Computing*. 2018;**72**:108–20.
13. Wang H, Zheng B, Yoon SW, Ko HS. A support vector machine-based ensemble algorithm for breast cancer diagnosis. *European J Operational Res*. 2018;**267**(2):687–99.
14. Rao H, Shi X, Rodrigue AK, Feng J, Xia Y, Elhoseny M, et al. Feature selection based on artificial bee colony and gradient boosting decision tree. *Applied Soft Computing*. 2019;**74**:634–42.
15. Liu N, Qi E, Xu M, Gao B, Liu G. A novel intelligent classification model for breast cancer diagnosis. *Information Processing & Management*. 2019;**56**(3):609–23.
16. Lu SY, Wang SH, Zhang YD. SAFNet: A deep spatial attention network with classifier fusion for breast cancer detection. *Comput Biol Med*. 2022;**148**:105812. [PubMed ID: 35834967]. https://doi.org/10.1016/j.compbiomed.2022.105812.
17. Abdar M, Zomorodi-Moghadam M, Zhou X, Gururajan R, Tao X, Barua PD, et al. A new nested ensemble technique for automated diagnosis of breast cancer. *Pattern Recognition Letters*. 2020;**132**:123–31.
18. Dalwinder S, Birmohan S, Manpreet K. Simultaneous feature weighting and parameter determination of neural networks using ant lion optimization for the classification of breast cancer. *Biocybernetics Biomedical Engineering*. 2020;**40**(1):337–51.
19. Kumar A, Sinha N, Bhardwaj A. A novel fitness function in genetic programming for medical data classification. *J Biomed Inform*. 2020;**112**:103623. [PubMed ID: 33197613]. https://doi.org/10.1016/j.jbi.2020.103623.
20. Sahebi G, Movahedi P, Ebrahimi M, Pahikkala T, Plosila J, Tenhunen H. GeFeS: A generalized wrapper feature selection approach for optimizing classification performance. *Comput Biol Med*. 2020;**125**:103974. [PubMed ID: 32890978]. https://doi.org/10.1016/j.compbiomed.2020.103974.
21. Khandezamin Z, Naderan M, Rashti MJ. Detection and classification of breast cancer using logistic regression feature selection and GMDH classifier. *J Biomed Inform*. 2020;**111**:103591. [PubMed ID: 33039588]. https://doi.org/10.1016/j.jbi.2020.103591.
22. Kalagotla SK, Gangashetty SV, Giridhar K. A novel stacking technique for prediction of diabetes. *Comput Biol Med*. 2021;**135**:104554. [PubMed ID: 34139440]. https://doi.org/10.1016/j.compbiomed.2021.104554.
23. Wan J, Chen H, Yuan Z, Li T, Yang X, Sang B. A novel hybrid feature selection method considering feature interaction in neighborhood rough set. *Knowledge-Based Systems*. 2021;**227**:107167.
24. Wolberg W, Mangasarian O, Street N, Street W. Breast Cancer Wisconsin (Diagnostic). *UCI Machine Learning Repository*. 1995.
25. Shokrzade A, Ramezani M, Tab FA, Mohammad MA. A novel extreme learning machine based kNN classification method for dealing with big data. *Expert Systems with Applications*. 2021;**183**:115293.

26. Hosni M, Abnane I, Idri A, Carrillo de Gea JM, Fernandez Aleman JL. Reviewing ensemble classification methods in breast cancer. *Comput Methods Programs Biomed*. 2019;**177**:89–112. [PubMed ID: 31319964]. https://doi.org/10.1016/j.cmpb.2019.05.019.
27. Momeni Z, Hassanzadeh E, Saniee Abadeh M, Bellazzi R. A survey on single and multi omics data mining methods in cancer data classification. *J Biomed Inform*. 2020;**107**:103466. [PubMed ID: 32525020]. https://doi.org/10.1016/j.jbi.2020.103466.
28. Aurna NF, Yousuf MA, Taher KA, Azad AKM, Moni MA. A classification of MRI brain tumor based on two stage feature level ensemble of deep CNN models. *Comput Biol Med*. 2022;**146**:105539. [PubMed ID: 35483227]. https://doi.org/10.1016/j.compbiomed.2022.105539.
29. Dokeroglu T, Deniz A, Kiziloz HE. A comprehensive survey on recent metaheuristics for feature selection. *Neurocomputing*. 2022;**494**:269–96.
30. Zheng Y, Ma Y, Cammon J, Zhang S, Zhang J, Zhang Y. A new feature selection approach for driving fatigue EEG detection with a modified machine learning algorithm. *Comput Biol Med*. 2022;**147**:105718. [PubMed ID: 35716435]. https://doi.org/10.1016/j.compbiomed.2022.105718.
31. Cai J, Luo J, Wang S, Yang S. Feature selection in machine learning: A new perspective. *Neurocomputing*. 2018;**300**:70–9.
32. Mahendran N, P M. A deep learning framework with an embedded-based feature selection approach for the early detection of the Alzheimer's disease. *Comput Biol Med*. 2022;**141**:105056. [PubMed ID: 34839903]. https://doi.org/10.1016/j.compbiomed.2021.105056.
33. Kılıç F, Kaya Y, Yildirim S. A novel multi population based particle swarm optimization for feature selection. *Knowledge-Based Systems*. 2021;**219**:106894.
34. Deep K. A random walk Grey wolf optimizer based on dispersion factor for feature selection on chronic disease prediction. *Expert Systems with Applications*. 2022;**206**:117864.
35. BinSaeedan W, Alramlawi S. CS-BPSO: Hybrid feature selection based on chi-square and binary PSO algorithm for Arabic email authorship analysis. *Knowledge-Based Systems*. 2021;**227**:107224.
36. Hu P, Pan J, Chu S, Sun C. Multi-surrogate assisted binary particle swarm optimization algorithm and its application for feature selection. *Applied soft computing*. 2022;**121**:108736.
37. Ma W, Zhou X, Zhu H, Li L, Jiao L. A two-stage hybrid ant colony optimization for high-dimensional feature selection. *Pattern Recognition*. 2021;**116**:107933.
38. Wang Z, Gao S, Zhang Y, Guo L. Symmetric uncertainty-incorporated probabilistic sequence-based ant colony optimization for feature selection in classification. *Knowledge-Based Systems*. 2022;**256**:109874.
39. Got A, Moussaoui A, Zouache D. Hybrid filter-wrapper feature selection using whale optimization algorithm: A multi-objective approach. *Expert Systems with Applications*. 2021;**183**:115312.
40. Kundu R, Chattopadhyay S, Cuevas E, Sarkar R. AltWOA: Altruistic Whale Optimization Algorithm for feature selection on microarray datasets. *Comput Biol Med*. 2022;**144**:105349. [PubMed ID: 35303580]. https://doi.org/10.1016/j.compbiomed.2022.105349.
41. Rajammal RR, Mirjalili S, Ekambaram G, Palanisamy N. Binary grey wolf optimizer with mutation and adaptive k-nearest neighbour for feature selection in Parkinson's disease diagnosis. *Knowledge-Based Systems*. 2022;**246**:108701.
42. Shrivastava P, Shukla A, Vepakomma P, Bhansali N, Verma K. A survey of nature-inspired algorithms for feature selection to identify Parkinson's disease. *Comput Methods Programs Biomed*. 2017;**139**:171–9. [PubMed ID: 28187888]. https://doi.org/10.1016/j.cmpb.2016.07.029.
43. Abualigah LM, Khader AT, Hanandeh ES. A new feature selection method to improve the document clustering using particle swarm optimization algorithm. *Journal of Computational Science*. 2018;**25**:456–66.
44. Nguyen BH, Xue B, Zhang M. A survey on swarm intelligence approaches to feature selection in data mining. *Swarm and Evolutionary Computation*. 2020;**54**:100663.
45. Mirjalili S, Mirjalili SM, Lewis A. Grey wolf optimizer. *Advances in engineering software*. 2014;**69**:46–61.
46. Liu M, Luo K, Zhang J, Chen S. A stock selection algorithm hybridizing grey wolf optimizer and support vector regression. *Expert Systems with Applications*. 2021;**179**:115078.
47. Nadimi-Shahraki MH, Taghian S, Mirjalili S. An improved grey wolf optimizer for solving engineering problems. *Expert Systems with Applications*. 2021;**166**:113917.
48. Achom A, Das R, Pakray P. An improved Fuzzy based GWO algorithm for predicting the potential host receptor of COVID-19 infection. *Computers in Biology and Medicine*. 2022;**151**:106050.
49. Faris H, Aljarah I, Al-Betar MA, Mirjalili S. Grey wolf optimizer: a review of recent variants and applications. *Neural computing and applications*. 2018;**30**:413–35.
50. Emary E, Zawbaa HM, Hassanien AE. Binary grey wolf optimization approaches for feature selection. *Neurocomputing*. 2016;**172**:371–81.
51. Gu X, Angelov PP. Self-organising fuzzy logic classifier. *Information Sciences*. 2018;**447**:36–51.
52. Angelov P, Gu X, Kangin D. Empirical data analytics. *International Journal of Intelligent Systems*. 2017;**32**(12):1261–84.
53. Takagi T, Sugeno M. Fuzzy identification of systems and its applications to modeling and control. *IEEE transactions on systems, man, and cybernetics*. 1985;(1):116–32.
54. Mamdani EH, Assilian S. An experiment in linguistic synthesis with a fuzzy logic controller. *International journal of man-machine studies*. 1975;**7**(1):1–13.
55. Angelov P, Yager R. A new type of simplified fuzzy rule-based system. *International Journal of General Systems*. 2012;**41**(2):163–85.
56. Angelov PP, Xiaowei G, Principe JC. A Generalized Methodology for Data Analysis. *IEEE Trans Cybern*. 2018;**48**(10):2981–93. [PubMed ID: 29035234]. https://doi.org/10.1109/TCYB.2017.2753880.
57. Gu X, Angelov PP, Príncipe JC. A method for autonomous data partitioning. *Information sciences*. 2018;**460**:65–82.
58. Okabe A, Boots B, Sugihara K, Chiu SN. *Spatial tessellations: concepts and applications of Voronoi diagrams*. John Wiley & Sons; 2009.
59. Singh LK, Khanna M, Singh R. Artificial intelligence based medical decision support system for early and accurate breast cancer prediction. *Advances in Engineering Software*. 2023;**175**:103338.