Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

Hybrid Mahalanobis Taguchi System with Binary Whale Optimisation Feature Selection for the Wisconsin Breast Cancer Dataset

Chow Yong Huan[1], Wan Zuki Azman Wan Muhamad[1,2,*], Zainor Ridzuan Yahya[1,2], Nor Hizamiyani Abdul Azziz[1], Tan Li Mei[1], Tan Xiao Jian[3]

[1] Institute of Engineering Mathematics, University Malaysia Perlis, Pauh Putra Campus, 02600 Arau, Perlis, Malaysia
[2] Centre of Excellence for Advanced Computing (ADVCOMP), University Malaysia Perlis, Pauh Putra Campus, 02600 Arau, Perlis, Malaysia
[3] Centre for Multimodal Signal Processing, Department of Electrical and Electronic Engineering, Faculty of Engineering and Technology, Tunku Abdul Rahman University of Management and Technology, Jalan Genting Kelang, Setapak, 53300 Kuala Lumpur, Malaysia

ARTICLE INFO

Article history:
Received 25 April 2023
Received in revised form 16 July 2023
Accepted 22 July 2023
Available online 8 August 2023

Keywords:
Mahalanobis Taguchi System; Binary Whale Optimisation Algorithm; feature selection; Wisconsin Breast Cancer dataset

ABSTRACT

The Mahalanobis-Taguchi System (MTS) is a statistical approach used in breast cancer research to facilitate early detection and promote efficient treatment. The technique analyses mammogram images for significant features using a multivariate statistical analysis technique. It combines the Mahalanobis distance (MD) and Taguchi's method to determine the differences between benign and malignant samples. While orthogonal array (OA) has been widely used in MTS, it has been criticised for providing suboptimal results due to insufficient coverage of feature combinations during the feature optimisation process. To address this issue, the Binary Whale Optimisation Algorithm (BWOA) is proposed as an improved search algorithm for MTS. This paper aims to develop a novel hybrid method that enhances the efficiency of the Mahalanobis Taguchi System (MTS). The performance of feature selection ability due to different MTS hybrid algorithms were also compared. BWOA simulates the hunting behaviour of humpback whales and works by exploring new regions of the solution space, gradually narrowing the search space, and fine-tuning the solution. MTS-BWOA demonstrated its enhanced capability in feature optimisation compared to traditional MTS methods and has the potential to be applied in other medical imaging domains.

1. Introduction

In computational biology, feature selection is critical as it assists researchers in analyzing the collected data on patient conditions, enabling accurate treatment decisions [1]. In the medical treatment of breast cancer, the Mahalanobis Taguchi System (MTS) has emerged as a common and effective approach for feature selection [2–5]. MTS uses a multivariate statistical analysis technique that combines the Mahalanobis Distance (MD) and Taguchi's method to determine the significant features in the mammogram images. The technique involves creating a statistical model based on a database of mammogram images, which includes both benign and malignant samples.

* *Corresponding author.*
E-mail address: wanzuki@unimap.edu.my

https://doi.org/10.37934/araset.31.3.93105

The MTS method analyses feature of the mammogram images such as clump thickness, uniformity of cell size and shape, marginal adhesion, single epithelial cell size, bare nuclei, bland chromatin, normal nucleoli, mitoses, and tumor type to identify the differences between the benign and malignant samples [6]. The MD is used to measure the distance of the targeted observation from the center of the multivariate benign samples. Meanwhile, for Taguchi's method, orthogonal array (OA), signal-to-noise ratio (SNR) and gain are used to identify the critical features and reduce the impact of the non-critical features [7–11].

Despite prevalent applications of OA in MTS, they have been criticized for their inconsistency in producing optimal results [12–14]. This drawback is due to insufficient coverage of feature combinations, leading to missing pieces during the feature optimisation process. Orthogonal arrays are designed for simultaneous testing on multiple factors and levels with a minimum possible number of experimental runs. Although OA can reduce the time and cost of experimentation, it may not guarantee the best results. In some samples, conducting additional experiments or increasing the number of runs may be necessary to obtain more accurate results [15].

Another shortcoming of OA is its limited range of factors that can be tested, as they are typically designed to accommodate a fixed number of factors and levels [16,17]. In other words, if a new factor or level needs to be tested, the entire array may need to be redesigned, which can be time-consuming and expensive [18].

Hence, in order to boost performance, Taguchi *et al.,* [10] and Jugulum *et al.,* [19] suggested that the MTS methodology be improved with a better search algorithm. Woodall *et al.,* [11] also reviewed MTS and concluded that another feature selection technique could replace OA [20,21].

In the existing literature, artificial intelligence (AI) approaches which are commonly utilized for feature selection include genetic algorithms (GA), particle swarm optimisation (PSO), bee algorithms (BA) and many more [22]. These algorithms have been further hybridized with MTS, such as MTS-GA, MTS-PSO, Random Binary Search (RBS)-MTS, and MTS-BA [18,23,24]. Although these hybrid approaches have shown promise, they also possess inherent weaknesses. These algorithms and their hybrids suffer from the necessity of tuning multiple parameters to achieve optimal results. Determining the appropriate values for these parameters can be challenging and time-consuming, often necessitating extensive experimentation or domain expertise. Extra complexity and an increased likelihood of suboptimal settings will be incurred if the parameter-tuning process is not meticulously handled.

Moreover, these hybrid optimisation algorithms have been observed to suffer from the problem of becoming trapped in local optima. This issue can significantly cripple the algorithm's effectiveness in finding the optimal solution. When an algorithm gets stuck in a local optimum, it cannot explore the search space beyond that point, thereby missing out on better solutions that may exist elsewhere. As a result, the algorithm may converge to a suboptimal solution, limiting its potential effectiveness in certain scenarios. This drawback raises concerns about the algorithm's reliability and ability to guarantee the best solution for the given problem.

In this paper, a hybrid approach was proposed by combining the Binary Whale Optimization Algorithm (BWOA) with the Mahalanobis-Taguchi System (MTS). The main objective of this study is to develop a new approach: MTS-BWOA, to solve the OA issues mentioned earlier. The study focuses on comparing the abilities of different hybrid MTS algorithms in finding the best solution, evaluating the feature reduction rate, and analyzing the system's gain and variability range reduction. These comparisons demonstrated the effectiveness of feature reduction, highlighting the hybrid approach's potential in overcoming optimisation challenges and enhancing performance.

BWOA is a metaheuristic optimisation algorithm inspired by the hunting behavior of humpback whales [25]. It is a variant of the Whale Optimization Algorithm (WOA) specifically designed to solve optimisation problems with binary variables. BWOA has been successfully applied in a wide range of optimisation problems, such as feature selection, image segmentation, and pattern recognition [26–28].

BWOA works by mimicking the hunting behavior of humpback whales, which involves three main steps: searching for prey, encircling the prey, and attacking the prey. In BWOA, these steps are represented by three main operators: search operator, encircling operator, and attacking operator. The algorithm will explore new regions of the solution space, then gradually narrow the search space, fine-tune the solution, and improve accuracy [29,30]. BWOA also helps to introduce additional diversity into the population and prevent the algorithm from getting stuck in local optima by randomly flipping the binary values [31].

In a nutshell, Mahalanobis-Taguchi System Binary Whale Optimization Algorithm (MTS-BWOA) has demonstrated superior performance in the Wisconsin Breast Cancer case study by providing an improved feature optimisation compared to traditional MTS methods. The comparison between MTS and MTS-BWOA involved evaluating appearance time, feature reduction rate, number of selected features, computational time, gain, and variability range reduction. The promising results suggested that MTS-BWOA has the potential to be applied in other medical imaging domains, providing a useful tool for diagnosing a range of medical complications.

2. Methodology

2.1 Mahalanobis Taguchi System

MTS has been applied in fault condition identification, product analysis, risk prognosis, and assisting decision-making. It has been proven to be a successful method used for classification and feature selection in various studies [11,20,32,33]. The system starts with constructing Mahalanobis Space (MS) by determining MD for benign sample cases, followed by validation of malignant samples using mean, standard deviation, and correlation structure of features in MS. Eventually, the optimized feature can be identified using OA.

Step 1: Construct MS. The first step is to collect and remove any outliers and missing data. Subsequently, only benign samples are filtered out from the database to build the benign samples table. The benign samples $Data_{ij}$ have a j^{th} observation in a sample of size *n* with ith features in size *k* as described in Table 1 below where *i* equals 1, 2, 3, …., *k* while *j* equals 1, 2, 3, …, *n*.

Table 1
Benign samples table

	$Feature_i$	…	$Feature_k$
$Observation_j$	$Data_{ij}$	…	$Data_{kj}$
…		…	
…		…	
$Observation_n$	$Data_{in}$	…	$Data_{kn}$

Next, the benign dataset is transformed into a standardized dataset table by calculating the mean $\bar{x}_i$ using Eq. (1), the standard deviation S_i using Eq. (2) and the standardised data to Z_{ij} using Eq. (3).

$$\bar{x}_i = \frac{1}{n}\sum_{j=1}^{n} Data_{ij} \tag{1}$$

$$S_i = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(Data_{ij} - \bar{x}_i)^2} \tag{2}$$

$$Z_{ij} = \frac{Data_{ij} - \bar{x}_i}{S_i} \tag{3}$$

Once the standardized value Z_{ij} is calculated for all the data, a new standardized sample data table should resemble Table 2.

Table 2
Standardize data table

	$Feature_i$	…	$Feature_k$
$Observation_j$	Z_{ij}	…	Z_{kj}
…		…	
…		…	
$Observation_n$	Z_{in}	…	Z_{kn}

Before calculating *MD*, it is essential to calculate the correlation matrix beforehand. The correlation matrix of the dataset r_{ab} with *n* samples and *k* features where *a* and *b* are equal to 1, 2, 3, …, *k* can be referred to in Eq. (4) and the formula of the correlation matrix is listed in Eq. (5).

$$R = \begin{pmatrix} 1 & r_{12} & r_{13} & \cdots & r_{1k} \\ r_{21} & 1 & r_{23} & \cdots & r_{2k} \\ r_{31} & r_{32} & 1 & \cdots & r_{3k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{k1} & r_{k2} & r_{k3} & \ldots & 1 \end{pmatrix} \tag{4}$$

$$r_{ab} = \frac{\sum_{i=1}^{n}\left[\left(x_{ia} - \bar{x}_a\right)\left(x_{ib} - \bar{x}_b\right)\right]}{\left(\sqrt{\sum_{i=1}^{n}\left(x_{ia} - \bar{x}_a\right)^2}\right)\left(\sqrt{\sum_{i=1}^{n}\left(x_{ib} - \bar{x}_b\right)^2}\right)} \tag{5}$$

Finally, the *MD* of benign samples in the MS needs to be determined. The analysis involves inverting the correlation matrix *R* in Eq. (4) to R^{-1} and transposing the standardised data Z_{ij} to Z_{ij} '. Then, the formula in Eq. (6) can be used to calculate the *MD*. It is expected that the average *MD* of all samples in the MS will be equal to 1.

$$MD_j = \frac{Z_{ij} \times R^{-1} \times Z_{ij}^{'}}{k} \tag{6}$$

Step 2: Validate the measurement scale using malignant samples. To construct a new set of standardized malignant samples MZ_{ij}, the process outlined in Step 1 is replicated, with the exception that the mean and standardized data of the benign samples are employed in the calculation of Mz_{ij}

using Eq. (7). The MD values for the malignant samples *ZMD* is then calculated using MZ_{ij} and the correlation matrix of the benign samples can be described in Eq. (8).

$$MZ_{ij} = \frac{\left(MData_{ij} - \bar{x}_i\right)}{S_i} \tag{7}$$

$$ZMD_j = \frac{MZ_{ij} \times R^{-1} \times MZ_{ij}'}{k} \tag{8}$$

Step 3: Find significant features using OA, SNR and gain. Orthogonal arrays are used to generate a set of experiments that cover possible combinations of input variables systematically and efficiently. The number of experiments required depends on the number of input variables and their levels. Every experimental run-in designated OA will generate SNR value using MD from malignant cases. According to Taguchi, the larger-the-better type SNR exhibits superior results compared with the dynamic type SNR. Hence the larger-the-better type of SNR is applied, and it can be mathematically described as in Eq. (9).

$$SNR = -10\log\left(\frac{1}{t}\sum_{j=1}^{t}\frac{1}{MD_j}\right) \tag{9}$$

The gain acts as an indicator for variability improvement. The formula for improved variability range $VR_{improved}$ is presented in Eq. (10). The higher the SNR gain, the better the performance of reducing features in the system.

$$VR_{improved} = \left(\frac{1}{2}\right)^{\frac{Gain}{6}}\left(VR_{initial}\right) \tag{10}$$

The SNR mean of each feature is then calculated under two conditions: the presence condition and the absence condition. Both SNR mean in the presence condition $\bar{s}^{+}$ and SNR mean in the absence condition $\bar{s}^{-}$ are shown in Eq. (11) and Eq. (12), respectively.

$$\bar{s}^{+} = \frac{\sum \text{SNR value of feature}_i \text{ presence}}{\text{Total occurance of feature presence}} \tag{11}$$

$$\bar{s}^{-} = \frac{\sum \text{SNR value of feature}_i \text{ absence}}{\text{Total occurance of feature absence}} \tag{12}$$

Lastly, to identify the optimal combination of features, the gain for each feature must be calculated using Eq. (13). A positive result with the highest gain will be the optimal feature combination determined in OA methodology.

$$Gain_i = \bar{s}_i^{+} - \bar{s}_i^{-} \tag{13}$$

2.2 Mahalanobis Taguchi System-Binary Whale Optimization Algorithm (MTS-BWOA)

Mahalanobis-Taguchi System-Binary Whale Optimization Algorithm (MTS-BWOA) is a hybrid methodology that combines MTS and BWOA to improve the feature optimisation process. The method is useful for optimizing complex systems that have a large number of features and require a robust and efficient optimisation process [25,34,35]. The steps to run MTS-BWOA are similar to MTS, with the exception of substituting the OA approach with BWOA, which is used to optimize the selected significant features obtained from MTS.

i. Step 1: Determine MD values for both benign and malignant samples and validate the measurement scale. Repeat step 1 and step 2 in MTS to get MD values.
ii. Step 2: Initialization. BWOA begins by randomly generating an initial population of solutions, where each solution is represented as a binary string of 0s and 1s to show the feature that is absent or present in the solution.
iii. Step 3: Calculate fitness value using the objective function. In MTS-BWOA, the objective function used is the SNR value. Hence, each solution SNR value is calculated using the malignant sample's MD values by Eq. (9).
iv. Step 4: Find the best solution. The solution with the largest SNR will be selected as the global best solution.
v. Step 5: Entering BWOA main loop. The algorithm iteratively updates the candidate solutions until the maximum iteration is reached or a stopping criterion is met. A pseudocode to run the BWOA loop is shown in Figure 1.

```
While Current Iteration < Maximum Iteration
    For Each solution
        Calculate value a⃗, A⃗, C⃗;
        If (p<0.5)
            If (|A| <1)
                Update position of whale using Eq. (17)
            Else (|A|>=1)
                Generate random search agent.
                Update position of whale using Eq. (19)
            End if
        Else (p>=0.5)
            Update position of whale using Eq. (21)
        End if
    End
    Convert position into binary using Eq. (23) and Eq. (24)
    Calculate the SNR value for each feature combination.
    Find the best solution
End
```

Fig. 1. Pseudocode of BWOA

For each solution, coefficient $\vec{A}$ and coefficient $\vec{C}$ is a random number generated from Eq. (14) and Eq. (15), respectively, where $\vec{a}$ is a linearly decreasing number from 2 to 0 using Eq. (16), $rand_1$ and $rand_2$ are random numbers in the range of (0,1).

$$\vec{A} = 2\vec{a} \times rand_1 - \vec{a} \tag{14}$$

$$\vec{C} = 2 \times rand_2 \tag{15}$$

$$\vec{a} = 2 - \frac{2(\text{current\_iteration})}{(\text{maximum\_iteration})} \tag{16}$$

To model the humpback whales swimming around the prey within a shrinking circle and along a spiral-shaped path, a probability value p is randomly generated between 0 to 1 so that there is a 50% probability of selecting one of the mechanisms. For shrinking cases, when $\overrightarrow{|A|}$ is smaller than 1, the new solution of whales $X(t+1)$ will be updated using Eq. (17), where t is the current iteration, $X^*(t\ )$ is the best solution and $\vec{D}$ is the absolute distance calculated using Eq. (18).

$$X(t+1) = X^*(t) - \left(\vec{A} \times \vec{D}\right) \tag{17}$$

$$D = \left|\vec{C} \times X^*(t) - X(t)\right| \tag{18}$$

The same approach based on the variation of $\vec{A}$ can be utilised in the exploration phase. When $\overrightarrow{|A|}$ is larger than 1, a random search is conducted to generate new candidate solutions in the search space. The search is done by randomly selecting a whale and modifying its position based on a randomly generated vector and a predefined coefficient. The random position is generated and updated using Eq. (19) and Eq. (20).

$$X(t+1) = X_{random} - \left(\vec{A} \times \vec{D}\right) \tag{19}$$

$$D = \left|\vec{C} \times X_{random} - X(t)\right| \tag{20}$$

For the spiral updating position case with a random value p ≥ 0.5, a spiral equation, as in Eq. (21), is created between the current and best solutions to mimic the helix-shaped movement of humpback whales. The new solution $X(t+1)$ is calculated based on $\overrightarrow{D'}$, while the distance between the current solution $X(t\ )$ and the best solution $X^*(t)$ can be determined using Eq. (22). The random number *l* which lies in the range of (-1,1).

$$X(t+1) = \overrightarrow{D'} \times e^l \times \cos(2\pi l) + X^*(t) \tag{21}$$

$$\overrightarrow{D'} = \left|X^*(t) - X(t)\right| \tag{22}$$

Once all solutions have been updated, they must be converted back to binary form using the sigmoid transfer function, as shown in Eq. (23), before determining their new positions based on Eq. (24). Next, the SNR value for each solution is recalculated according to the absence or presence of features, using Eq. (9). The solution with the highest SNR value is then considered the best new solution. This iterative process continues until the preset maximum iteration is reached. At this point, the optimal feature combination is selected as the final best solution.

$$S\left(x(t+1)\right) = \frac{1}{1+e^{-x(t)}} \tag{23}$$

$$\text{New position} = \begin{cases} 1 & \text{if random value selector} < S\left(x\left(t+1\right)\right) \\ 0 & \text{otherwise} \end{cases} \tag{24}$$

vi. Step 6: System gain and variability range reduction are calculated to validate the performance of feature optimisation. Unlike in the OA approach, the highest SNR solution obtained in Step 5 is used as the optimal solution in MTS-BWOA. However, to ensure that the new system gain is larger than the gain achieved using MTS, a system gain value is computed by taking the difference between the original system SNR and the optimal system SNR, as shown in Eq. (25). Additionally, the variability range reduction is recalculated using Equation (10).

$$\text{System gain} = \text{Optimised system SNR} - \text{Original system SNR} \tag{25}$$

3. Experiments and Results

Breast Cancer Wisconsin (original), collected and made available by Dr William H. Wolberg from the University of Wisconsin Hospitals, is used in this experiment. The dataset contains a total of 699 samples, each with 10 attributes. The first 9 attributes are numerical and represent the characteristics of the cell nuclei, including clump thickness, uniformity of cell size and shape, marginal adhesion, single epithelial cell size, bare nuclei, bland chromatin, normal nucleoli and mitoses. Meanwhile, the last attribute is the class representing benign or malignant. However, since the dataset has missing attributes on 16 samples, these samples are removed and left only 683 samples, of which 444 are benign, and 239 are malignant.

In this MTS-BWOA methodology, the population and maximum iteration of BWOA were set to 10 and 50, respectively. This decision was based on the observation that BWOA was able to achieve the best possible solution within this iteration limit during the case study. The MTS-BWOA experiment was repeated 30 times to ensure the robustness and reliability of the results. The time of appearance for each feature was recorded in Table 3. Based on the recorded data, features that appeared more than 15 times (i.e., more than half of the total runs) were identified as significant features. This approach was adopted to ensure that the selected features substantially contributed to the classification accuracy.

The results obtained from Table 3 present the frequency of appearance for nine features using different variations of the MTS on the Wisconsin Breast Cancer dataset. The evaluated variations include MTS, MTS-BWOA, and RBS-MTS. In the MTS approach, all features, except for feature F7, were identified as significant. However, in contrast, the MTS-BWOA variation showed zero occurrences for features F1, F4, F7, and F9, and features F3 and F5 appeared fewer than 15 times. Similar to MTS-BWOA, RBS-MTS also had features F1, F3, F4, F5, F7, and F9, appearing less than 15 times. Both MTS-BWOA and RBS-MTS shared significant features F2, F6, and F8, with RBS-MTS having an additional significant feature, F9.

Table 3
Frequency of appearance for nine features using different variations of the MTS on the Wisconsin Breast Cancer dataset

Type of MTS	F1	F2	F3	F4	F5	F6	F7	F8	F9
MTS	Y	Y	Y	Y	Y	Y	N	Y	Y
MTS-BWOA	0	28	2	0	5	30	0	27	0
RBS-MTS	4	23	6	12	6	30	5	25	16

Table 4 lists the significant features selected from Table 3, along with each algorithm's feature reduction rate and computational time. Since this study focuses on implementing MTS and MTS-BWOA, the computational time is only available for these two algorithms. The times of appearance of each feature and features selected data for RBS-MTS are collected from previous literature [23,36]. From Table 4, it can be concluded that MTS-BWOA has the highest feature reduction rate of 66.67%, with only three significant features remaining. Features F2, F6, and F8 have shown their importance, as all other feature selection algorithms also include these features. However, MTS-BWOA has a longer computational time than MTS because it requires time to run through all the position updates compared to a fixed-designed orthogonal array structure.

Table 4
Result of feature reduction rate, selected features, and computational time between different algorithms

Algorithm	Feature Reduction Rate (%)	Significant Features	Computational Time
MTS	11.11	F1, F2, F3, F4, F5, F6, F8, F9	0.2042
MTS-BWOA	66.67	F2, F6, F8	3.388
RBS-MTS	55.56	F2, F6, F8, F9	Not available

Moreover, the effectiveness of feature reduction achieved by MTS-BWOA is supported by system gain and variability range reduction results, as presented in Table 5. Among all the algorithms, MTS-BWOA exhibits the highest system gain with 1.303 and a variability range reduction of 13.975%. The finding evidenced that MTS-BWOA can successfully reduce the number of features without sacrificing the ability to preserve the best combination of features. Meanwhile, RBS-MTS exhibits a lower system gain and variability range reduction when compared to MTS-BWOA. However, it is notable that all the hybrid MTS variations demonstrated significantly improved computational performance compared with the original MTS system.

Table 5
Result of gain and variability range reduction for MTS, MTS-BWOA and RBS-MTS

	MTS	MTS-BWOA	RBS-MTS
Optimized system SNR	11.142	12.210	-
Original system SNR	10.91	10.91	10.91
System Gain	0.235	1.303	0.9679
Variability Range Reduction	2.679%	13.975%	10.578%

By analysing the convergence of MTS-BWOA to the best solution depicted in Figure 2, it is evident that MTS-BWOA outperforms MTS. The initial SNR value obtained by MTS-BWOA is already greater than the best SNR obtained by MTS. Furthermore, MTS-BWOA continues to improve and achieve higher SNR values by updating the best solution. The algorithm attains the maximum SNR value of

12.21 on the 13th iteration. Therefore, the computational time is expected to be reduced as MTS-BWOA can achieve the maximum SNR value before the 50th iteration.

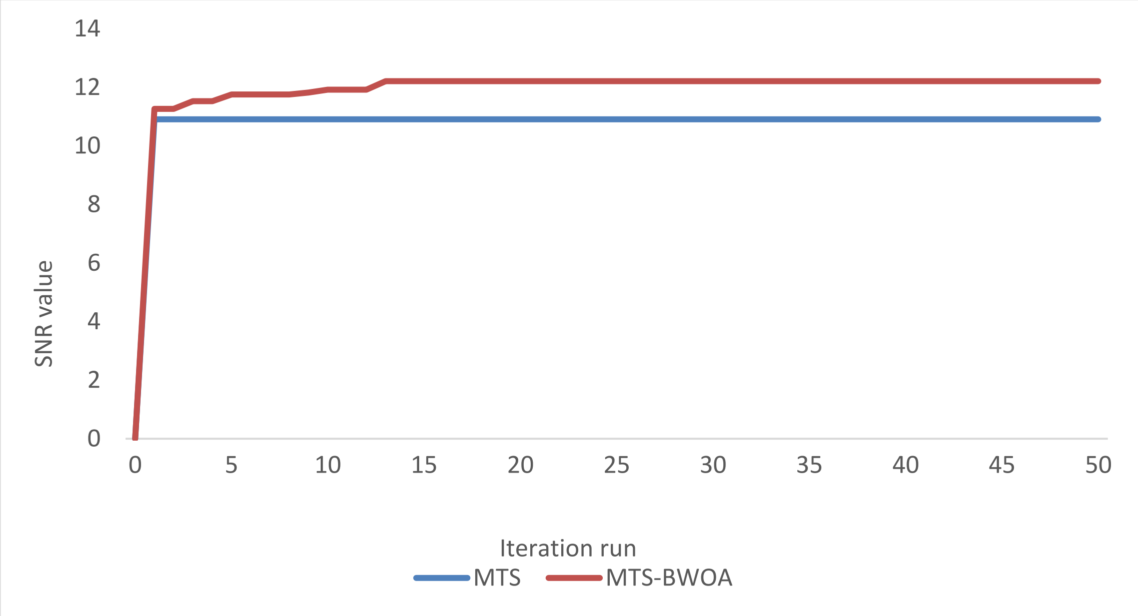

Fig. 2. Convergence of MTS and MTS-BWOA to best solution in the Wisconsin Breast Cancer dataset

The obtained results from Figure 3 and Figure 4 provide valuable insights into the performance of Mahalanobis distance distribution before and after optimisation. Figure 3, which represents the distribution prior to optimisation, elucidates that the Mahalanobis distances for healthy and unhealthy samples exhibit relatively closer values. The confusion indicates that, without feature selection or optimisation, the original dataset has limited discriminative power in distinguishing between healthy and unhealthy samples.

However, a significant improvement in the discriminant performance is observed upon applying the MTS-BWOA optimisation technique, as shown in Figure 4. The Mahalanobis distance distribution after optimisation demonstrates a clear distinction between healthy and unhealthy samples. The optimised MTS-BWOA approach effectively enhances the discriminative power, improving accuracy in classifying samples as healthy or unhealthy. These findings highlight the effectiveness of the proposed hybrid method in enhancing the discriminant performance of the MTS and potentially improving diagnostic accuracy in healthcare applications.

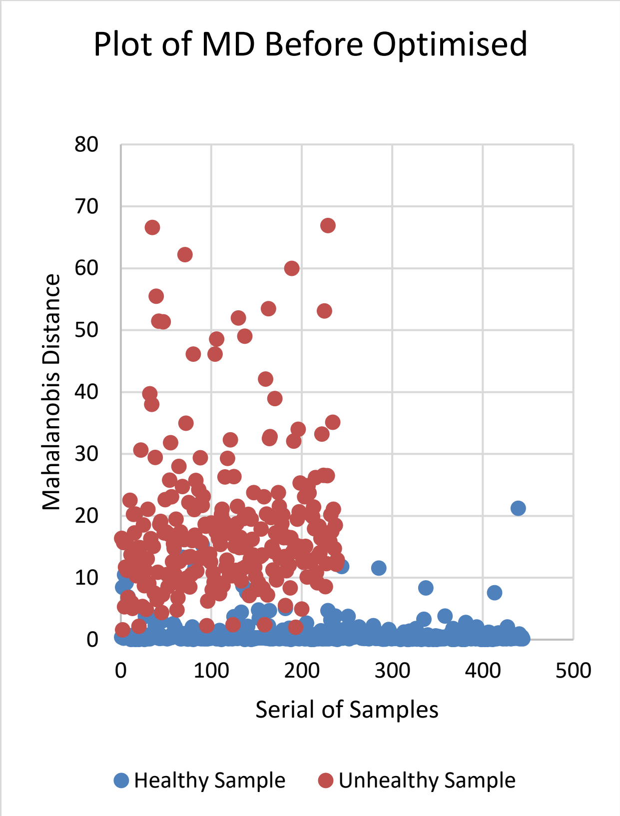

Fig. 3. MD distribution of healthy and unhealthy samples before optimized

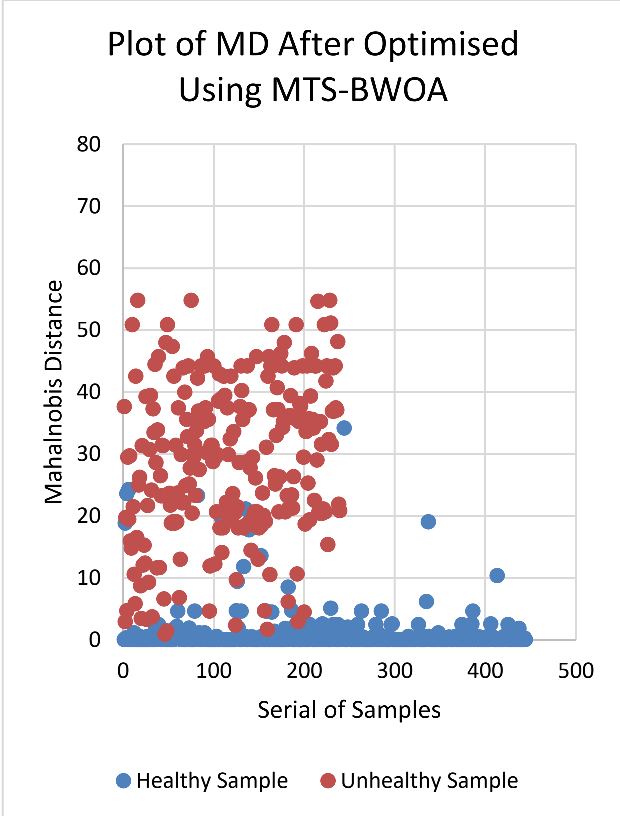

Fig. 4. MD distribution of healthy and unhealthy samples after optimized using MTS-BWOA

4. Conclusions

In this study, we applied the Mahalanobis-Taguchi System-Binary Whale Optimization Algorithm (MTS-BWOA) to optimise feature selection in breast cancer classification. Compared with the MTS approach, our results showed that MTS-BWOA could achieve greater performance in feature optimisation. MTS-BWOA selected fewer features, resulting in a higher system gain, higher variability range reduction and strong discriminant ability. Despite a longer computational time, the efficiency of MTS-BWOA outperformed MTS.

The success of MTS-BWOA in optimising feature selection in breast cancer classification suggests that this hybrid methodology can be used in other fields in the future. With its ability to efficiently identify significant features while reducing variability range and achieving a higher system gain, MTS-BWOA has the potential to be applied in various fields, such as image processing, natural language processing, and other medical diagnoses. Further research is required to examine the adaptability of MTS-BWOA and its potential to address various optimisation challenges across different fields.

Acknowledgement
The author would like to acknowledge the Fundamental Research Grant Scheme (FRGS) support under a grant number of FRGS/1/2017/STG06/UNIMAP/03/2 from the Ministry of Higher Education Malaysia.

References
[1] Ng, Win Son, Siew Chin Neoh, Kyaw Kyaw Htike, and Shir Li Wang. "Particle Swarm Feature selection for microarray Leukemia classification." *Progress in Energy and Environment* (2017): 1-8.
[2] Daniels, Benjamin, Steven M. Corns, and Elizabeth A. Cudney. "Introduction of R-LCS and comparative analysis with

FSC and mahalanobis-taguchi method for breast cancer classification." In *2012 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, pp. 283-289. IEEE, 2012. https://doi.org/10.1109/CIBCB.2012.6217242

[3] Muhamad, Wan Zuki Azman Wan, Khairur Rijal Jamaludin, Syafawati Ab Saad, Zainor Ridzuan Yahya, and Siti Aisyah Zakaria. "Random binary search algorithm based feature selection in Mahalanobis Taguchi system for breast cancer diagnosis." In *AIP Conference Proceedings*, vol. 1974, no. 1. AIP Publishing, 2018. https://doi.org/10.1063/1.5041558

[4] Hong, Jungeui, Rajesh Jugulum, Kioumars Paryani, Kenneth M. Ragsdell, Genichi Taguchi, and Elizabeth A. Cudney. "An evaluation of Mahalanobis-Taguchi system and neural network for multivariate pattern recognition." (2007).

[5] Kestle, Rodney S. "Classification application of mahalanobis-taguchi system for small datasets." In *IIE Annual Conference. Proceedings*, p. 1. Institute of Industrial and Systems Engineers (IISE), 2011.

[6] Wolberg, William, W. Street, and Olvi Mangasarian. "Breast cancer wisconsin (diagnostic)." *UCI Machine Learning Repository* (1995).

[7] Taguchi, Genichi, Subir Chowdhury, and Yuin Wu. "Taguchi's quality engineering handbook." *(No Title)* (2004). https://doi.org/10.1002/9780470258354

[8] Taguchi, Genichi, Yuin Wu, and Subir Chodhury. *Mahalanobis-Taguchi System*. McGraw-Hill Professional, 2000.

[9] Ghosh, Kushal Kanti, Ritam Guha, Suman Kumar Bera, Neeraj Kumar, and Ram Sarkar. "S-shaped versus V-shaped transfer functions for binary Manta ray foraging optimisation in feature selection problem." *Neural Computing and Applications* 33, no. 17 (2021): 11027-11041. https://doi.org/10.1007/s00521-020-05560-9

[10] Taguchi, Genichi, and J. Rajesh. "New trends in multivariate diagnosis." *Sankhyā: The Indian Journal of Statistics, Series B* (2000): 233-248.

[11] Woodall, William H., Rachelle Koudelik, Kwok-Leung Tsui, Seoung Bum Kim, Zachary G. Stoumbos, and Christos P. Carvounis. "A review and analysis of the Mahalanobis—Taguchi system." *Technometrics* 45, no. 1 (2003): 1-15. https://doi.org/10.1198/004017002188618626

[12] Abraham, B., and A. M. Variyath. "A review and analysis of the Mahalanobis-Taguchi System-Discussion." *Technometrics* 45, no. 1 (2003): 22-24. https://doi.org/10.1198/004017002188618644

[13] Hedayat, A., and J. Stufken. "On the maximum number of constraints in orthogonal arrays." *The Annals of Statistics* (1989): 448-451. https://doi.org/10.1214/aos/1176347030

[14] Hedayat, A. Samad, Neil James Alexander Sloane, and John Stufken. *Orthogonal arrays: theory and applications*. Springer Science & Business Media, 1999. https://doi.org/10.1007/978-1-4612-1478-6

[15] Hawkins, Douglas M. "Discussion," *Technometrics* 45, no. 1 (2003): 25–29. https://doi.org/10.1198/004017002188618653

[16] Iquebal, Ashif Sikandar, Avishek Pal, Darek Ceglarek, and Manoj Kumar Tiwari. "Enhancement of Mahalanobis–Taguchi system via rough sets based feature selection." *Expert Systems with Applications* 41, no. 17 (2014): 8003-8015. https://doi.org/10.1016/j.eswa.2014.06.019

[17] Iquebal, Ashif Sikandar, and Avishek Pal. "Artificial bee colony optimisation-based enhanced Mahalanobis Taguchi system for classification." *International Journal of Intelligent Engineering Informatics* 2, no. 2-3 (2014): 181-194. https://doi.org/10.1504/IJIEI.2014.066217

[18] Murata, Shinichi, and Hiroshi Morita. "Feature Analysis Using Mahalanobis-Taguchi Method and Genetic Algorithm for Recorded TV Data." *International Journal of Innovative Computing, Information and Control* 18, no. 1 (2022): 173-181.

[19] Jugulum, R., G. Taguchi, S. Taguchi, and J. O. Wilkins. "A review and analysis of the Mahalanobis-Taguchi system-Discussion." *Technometrics* 45, no. 1 (2003): 16-21. https://doi.org/10.1198/004017002188618635

[20] Ghasemi, Elham, Abdollah Aaghaie, and Elizabeth A. Cudney. "Mahalanobis Taguchi system: a review." *International Journal of Quality & Reliability Management* 32, no. 3 (2015): 291-307. https://doi.org/10.1108/IJQRM-02-2014-0024

[21] Tan, Li Mei, Wan Zuki Azman Wan Muhamad, Zainor Ridzuan Yahya, Ahmad Kadri Junoh, Nor Hizamiyani Abdul Azziz, Faizir Ramlie, Nolia Harudin, Mohd Yazid Abu, and Xiao Jian Tan. "A survey on improvement of Mahalanobis Taguchi system and its application." *Multimedia Tools and Applications* (2023): 1-17. https://doi.org/10.1007/s11042-023-15257-5

[22] Kek, Hong Yee, Huiyi Tan, Desmond Daniel Chin Vui Sheng, Yi Lee, Nur Dayana Ismail, Muhd Suhaimi Deris, Haslinda Mohamed Kamar, and Keng Yinn Wong. "A CFD assessment on ventilation strategies in mitigating healthcare-associated infection in single patient ward." *Progress in Energy and Environment* (2023): 35-45. https://doi.org/10.37934/progee.24.1.3545

[23] W. Z. A. W. Muhamad, "Hybrid Mahalanobis Taguchi System with Randomized Algorithms for Simultaneous Feature Selection and Binary Classification," University Teknologi Malaysia, 2016.

[24] Hancer, Emrah, Bing Xue, Dervis Karaboga, and Mengjie Zhang. "A binary ABC algorithm based on advanced

similarity scheme for feature selection." *Applied Soft Computing* 36 (2015): 334-348. https://doi.org/10.1016/j.asoc.2015.07.023

[25] Mirjalili, Seyedali, and Andrew Lewis. "The Whale Optimisation Algorithm," *Advances in Engineering Software* 95 (2016): 51–67. https://doi.org/10.1016/j.advengsoft.2016.01.008

[26] Hussien, Abdelazim G., Aboul Ella Hassanien, Essam H. Houssein, Mohamed Amin, and Ahmad Taher Azar. "New binary whale optimisation algorithm for discrete optimisation problems." *Engineering Optimization* 52, no. 6 (2020): 945-959. https://doi.org/10.1080/0305215X.2019.1624740

[27] Mafarja, Majdi, and Seyedali Mirjalili. "Whale optimisation approaches for wrapper feature selection." *Applied Soft Computing* 62 (2018): 441-453. https://doi.org/10.1016/j.asoc.2017.11.006

[28] Kumar, Vijay, and Dinesh Kumar. "Binary whale optimisation algorithm and its application to unit commitment problem." *Neural Computing and Applications* 32 (2020): 2095-2123. https://doi.org/10.1007/s00521-018-3796-3

[29] Soliman, Ghada MA, Tarek HM Abou-El-Enien, Eid Emary, and Motaz MH Khorshid. "A Hybrid Modified Whale Optimization Algorithm with Simulated Annealing for Terrorism Prediction." *Ingénierie des Systèmes d Inf.* 24, no. 3 (2019): 281-287. https://doi.org/10.18280/isi.240308

[30] Mafarja, Majdi M., and Seyedali Mirjalili. "Hybrid whale optimisation algorithm with simulated annealing for feature selection." *Neurocomputing* 260 (2017): 302-312. https://doi.org/10.1016/j.neucom.2017.04.053

[31] Chen, Huiling, Yueting Xu, Mingjing Wang, and Xuehua Zhao. "A balanced whale optimisation algorithm for constrained engineering design problems." *Applied Mathematical Modelling* 71 (2019): 45-59. https://doi.org/10.1016/j.apm.2019.02.004

[32] Chang, Z. P. "Research progress of Mahalanobis-Taguchi system." *Control. Decis* 34 (2019): 2505-2516.

[33] Muhamad, Wan Zuki Azman Wan, Khairur Rijal Jamaludin, Faizir Ramlie, Nolia Harudin, and Nur Najmiah Jaafar. "Criteria selection for an MBA programme based on the Mahalanobis Taguchi system and the kanri distance calculator." In *2017 IEEE 15th Student Conference on Research and Development (SCOReD)*, pp. 220-223. IEEE, 2017. https://doi.org/10.1109/SCORED.2017.8305390

[34] Eid, Heba F. "Binary whale optimisation: an effective swarm algorithm for feature selection." *International Journal of Metaheuristics* 7, no. 1 (2018): 67-79. https://doi.org/10.1504/IJMHEUR.2018.091880

[35] Hussien, Abdelazim G., Essam H. Houssein, and Aboul Ella Hassanien. "A binary whale optimisation algorithm with hyperbolic tangent fitness function for feature selection." In *2017 Eighth international conference on intelligent computing and information systems (ICICIS)*, pp. 166-172. IEEE, 2017. https://doi.org/10.1109/INTELCIS.2017.8260031

[36] Muhamad, W. Z. A. W., Khairur Rijal Jamaludin, Zainor Ridzuan Yahya, and Faizir Ramlie. "A hybrid methodology for the mahalanobis-taguchi system using random binary search-based feature selection." *Far East J. Math. Sci* 101, no. 2 (2017): 2663-2675. https://doi.org/10.17654/MS101122663