Global Journal of Pure and Applied Mathematics.
ISSN 0973-1768 Volume 12, Number 3 (2016), pp. 2651–2671

http://www.ripublication.com/gjpam.htm

Optimal Feature Selection of Taguchi Character Recognition in the Mahalanobis-Taguchi System using Bees Algorithm

Faizir Ramlie, Khairur Rijal Jamaludin, and Rozzeta Dolah
UTM Razak School of Engineering and Advanced Technology, UTM International Campus Kuala Lumpur, Jalan Sultan Yahya Petra, 54100 Kuala Lumpur, Malaysia.

Wan Zuki Azman Wan Muhamad
UTM Razak School of Engineering and Advanced Technology, UTM International Campus Kuala Lumpur, Jalan Sultan Yahya Petra, 54100 Kuala Lumpur, Malaysia.
Institute of Engineering Mathematics, Universiti Malaysia Perlis, Kampus Pauh Putra, 02600 Arau, Perlis, Malaysia.

Abstract

The Mahalanobis-Taguchi System (MTS) is a data mining method employing Mahalanobis distance (MD) and Taguchi′s Robust Engineering philosophy to explore and exploit data in a multidimensional system. The MD calculation provides a measurement scale to discriminate sample data and gives an approach of measuring the level of severity among them. One unique feature of MTS lies its robustness to assess variability among all levels of samples (noise) and ability to evaluate significant and insignificant factors which contributed to the system (optimization) by means of simplistic yet robust technique via orthogonal array (OA) and signal to noise ratio (SNR). The optimized system obtained is considered robust, since the SNR identifies the useful variables that are most insensitive to variation, and cost efficient, as it constitutes a smaller number of attributes with better system performance. In this paper, a novel useful variable selection (feature selection) approach using Bees Algorithm (BA) replacing conventional OA technique is presented. BA is a heuristic search technique that finds optimal (or near optimal) result which falls under the Swarm Intelligence field. The solution search strategy mimics social behaviour of animals or insects (bee colony in particular). MD is used as the result assessment metric while the larger-the-better type of SNR is deployed as the algorithm objective function. Character recognition based on Taguchi concepts (exploiting variation and abundance items) is used as the case study on which the comparison between BA and OA performances is made. The results show a promising discriminant power of the optimized system via BA as compared to OA, however, the OA approach outperforms BA in terms of optimization speed to a great extent.

AMS subject classification:
Keywords: Mahalanobis–Taguchi System, Orthogonal Array, Bees Algorithm, Taguchi Character Recognition.

1. Introduction

The Mahalanobis-Taguchi System (MTS) is a pattern information technology that aids quantitative decision making process by constructing a multivariate measurement scale using data analytic methods [1]. It was developed by the renowned Japanese Quality guru Dr. Genichi Taguchi. The MTS methodology started with the theory of Mahalanobis distance (MD) formulated by the famous Indian statistician, Dr. P.C. Mahalanobis in 1936 inspired from his determination to examine if the Indian people who married European people came from specific caste levels [2] . The formulation of MD was then extended by Dr. Taguchi whom integrated the MD formulation with his robust engineering concepts to enhance the MD methodology to become a popular application tool for diagnosis and forecasting technique in multidimensional systems [3].

One unique feature of MTS lies in its robustness to assess variability among all levels of samples (noise) and the ability to evaluate significant and insignificant factors which contribute to the system (optimization) by means of a simplistic yet robust approach which utilizes orthogonal array (OA) and signal to noise ratio (SNR) techniques. The optimized system obtained is considered robust because the SNR identifies the useful variables that are most insensitive to variation [4] and cost efficient as it constitutes a smaller number of attributes with better system performance [5].

However, the deployment and the operational aspect of OA as a mean of optimizing MTS system are critically debated in the literature. The optimization procedure of deploying OA for dimensionality reduction and search procedure is claimed inadequate and produces sub-optimal solution [6], [7], [8], [9]. [10] agree and recommend a better search algorithm should be incorporated into the MTS methodology to improve MTS performance. Several attempts to replace the OA with other techniques have been reported in the literature which are discussed in the following sub-sections.

2. Alternative optimization schemes replacing OA

[11] introduced a method based on principal component transformation and multi-modal overlap methods called the Principal component Feature overlap Measure (PFM) method for product inspection in a manufacturing system. The PFM theory is based on machine vision concepts to enhance statistical classification by improving the methods for selecting optimum features for pattern recognition [11]. The Mahalanobis distance (MD) metric was used as the statistical measurement of the relationships in a dataset,; however, the selection of significant features was based on the degree of overlap of the features probability mass functions (PFM approach) instead of using an OA and SNR in conventional MTS methods. In an application, the PFM achieved a significantly higher signal to noise ratio (+80 dB) with equal or better performance than MTS.

In another study, [12] provided adaptations to the conventional MTS method called Modified MTS (MMTS) to monitor and diagnose a faulty chemical process. Instead of using an OA, they proposed Multiple Regression Analysis (MRA) as the tool for variable selection which is claimed to be more statistical sound approach [12]. The performance of MMTS was evaluated through an application of a continuous- stirred-tank-reactor problem and was shown to yield a promising result for on-line fault detection in the chemical process application.

[13] promoted the usage of MTS in an unsupervised data mining environment dubbed as Unsupervised Mahalanobis Distance Classifier (UNMDC). Their work attempted to enhance MTS functional capability in an unsupervised data mining environment against its present supervised data mining nature. The UNMDC algorithm based on MD values was developed to perform the classification and feature selection replacing the use of an OA and SNR. The performance of this algorithm was studied thoroughly on three different types of steel products on the basis of their composition and processing parameters. Performance in future diagnosis based on useful features by the new scheme was found to be quite satisfactory.

[14] proposed an omni-optimizer method to replace the OA and SNR to solve an optimization problem according to the purpose and characteristics of the data classification problem and optimization model. Test results on several training datasets revealed an effective outcome not only for classification but also for optimum feature selection. The proposed model called Mahalanobis-Taguchi System Optimizer (MTSO) was practically validated by performing quality inspections for notebook computers. Implementation results showed a significant reduction in the number of inspections attributed with high inspection accuracy in the inspection process which reduced production costs and enhanced productivity.

3. Fusion with Swarm Intelligence (SI)

Several studies adopted superior heuristic search algorithms to replace OA as suggested by [10] and [8]. Swarm Intelligence (SI) with its exploratory search approach is one option. The emergence of optimization studies adopting SI has gained great interest

Table 1: Fusion studies of MTS with other state of the art approaches replacing the orthogonal array

Author	MTS with	Search Method	Selection Criteria Objective Function
[15]	ACO	Binary Ant	TWFM
[16]	PSO	Gompertz PSO	TWFM
[14]	OOM	OOM	OOM
[13]	UNMDC	UNMDC	Partial F-Statistics
[17]	PSO	Binary PSO	TWFM
[12]	MRA	Binary PSO	TWFM
[11]	PCA	PCA based Wavelet	Probability Mass Function

in the literature due to the advantages of collective solution strategies offered by this technique.

[9] formulated a mathematical model for a feature selection scheme based on sound operation research principles and solved using binary particle swarm optimization (PSO). This was a novel attempt using a swarm heuristic approach replacing the OA and SNR in conventional MTS methodology. The performance of the proposed method was tested using data gathered from an Indian foundry shop. Results based on the comparison with conventional MTS reveals controversial findings on the validity of SNR practices [9]. [15][16] attempted to replace the OA with Binary Ant Colony Optimization (BACO) and Gompertz Binary Particle Swarm Optimization (GBPSO); however, the outcome of the studies were only to compare the outcomes of [9]. No critical evaluation was made by comparing their approaches with conventional MTS.

Table 11 summarizes the works found in the literature which attempt to replace the OA as the search and optimization strategy in MTS methodology. However, most of the assessment metrics presented in selecting important features or variables are found to be lacking in the use of SNR and leaning to conventional statistical approaches. The authors believe that the deployment of the SNR in MTS methodology should remain as it is. SNR is considered an important measurement since it is the core essence of the Taguchi philosophy [17]. Further, attempts to fuse SI into the MTS optimization procedure are found to be scarce. Hence, possible strategies offered by other SI techniques are worthy of study to evaluate their potential in solving the MTS optimization problem.

Therefore, in this work, another recent SI optimization technique called Bees Algorithm (BA) is introduced to replace the OA in MTS optimization procedure. In this strategy, the SNR measurement metrics will be retained as the measurement characteristic for selecting the important features of the system.

The paper is presented as follows, A theoretical overview on fundamental concept of

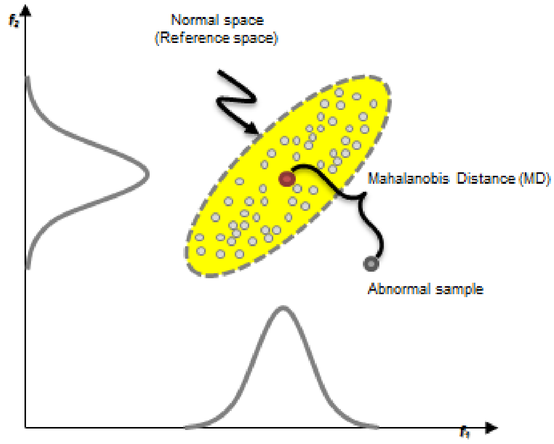


Figure 1: Mahalanobis Distance based on two characteristics/variables

MD and MTS are explained in Section 4. A brief concept of BA and the fusion strategies between the MTS and BA are discussed in Section 5. Section 7 explains the approaches to investigate the performance of the algorithm using a case study based on character recognition proposed by [18]. The discussion on the result presented in Section 8 while Section 9 concludes the key findings and future research.

4. The Concept of Mahalanobis Distance (MD)

MD is a dimensionless distance measure based on correlation between variables and pattern differences that can be analysed with respect to a reference population [19], as shown in Figure 1. This reference population is called as the normal space. The distance measure termed as the Mahalanobis Scale (MS) and aids the discriminant analysis approach by assessing the level of abnormality of datasets against the normal space.

MD has an elliptical shape (see Figure 1) due to the correlation effect between the variables. If there is no correlation, the MD is the same as the Euclidean Distance (ED) that has a circular shape. MD is different from Euclidean Distance since the later does not consider the correlation among the variables of the data points.

4.1. Mahalanobis Distance Formulation

MD is defined as in Equation 4.1

$$MD_j = D_j^2 = Z_{ij}^T C^{-1} Z_{ij} \quad \text{with} \quad Z_{ij} = \frac{x_{ij} - m_i}{s_i} \tag{4.1}$$

where;
k = the total numbers of variables;
i = the number of variables ($i = 1, 2, \dots, k$);

j = the number of samples ($j = 1, 2, \ldots, n$);
Z_{ij} = the standardized vector of normalized characteristics of x_{ij};
x_{ij} = the value of the ith characteristics in th jth observation;
m_i = the mean of the ith characteristics;
s_i = the standard deviation of the ith characteristics;
T = the transpose of the vector;
C^{-1} = the inverse of the correlation coefficient matrix.

MD has been well deployed in a broad array of applications [20], [21] mainly because it is very effective in tracking intervariable correlations in data.

4.2. MTS Procedures

Taguchi extended the MD methodology with his robust engineering concepts to become an efficient and effective strategy for prediction and forecasting in multidimensional systems. In the MTS methodology, the formulation of MD is scaled where the existing MD formulation stated in Equation 4.1 is divided by a term k that denotes the number of variables or features of a recognition system. Therefore, the equation for calculating the scaled MD in the MTS methodology becomes:

$$MD_j = D_j^2 = \frac{1}{k} Z_{ij}^T C^{-1} Z_{ij} \tag{4.2}$$

From this point onwards, the MD computation will be based on Equation 4.2. The MD offers a statistical measure to diagnose unknown sample conditions with known samples and provides information to make future prediction. The fundamental steps in the MTS methodology are explained in the next section.

STAGE 1: Construction of measurement scale

To construct a measurement scale, a homogeneous data set from normal observations needs to be collected to build a reference group called the normal group. It is used as a base or reference point in the scale. The collected normal datasets need to be standardized to obtain a dimensionless unit vector followed by the MD computation. Practically, the MD for unknown data is interpreted as the nearness to the mean of the normal group. As a countercheck, the average value of the MDs for the normal group must always be close to unity; therefore they are called the normal space or Mahalanobis Space (MS) [22].

The steps for the construction of the MS are outlined below:

Step 1: Calculate the mean characteristic in the normal data set as:

$$\bar{x}_i = \frac{\sum_{j=1}^{n} X_{ij}}{n} \tag{4.3}$$

Step 2: Then, calculate the standard deviation for each characteristic:

$$\bar{x}_i = \sqrt{\frac{\sum_{j=1}^{n}(X_{ij} - x_i)^2}{n-1}} \tag{4.4}$$

Step 3: Next, standardise each characteristic to form the normalized data matrix (Z_{ij}) and its transpose (Z_{ij}^T):

$$Z_{ij} = \frac{X_{ij} - x_i}{s_i} \tag{4.5}$$

Step 4: Then, verify that the mean of the normalized data is zero:

$$\bar{z_i} = \frac{\sum_{j=1}^{n} z_{ij}}{n} = 0 \tag{4.6}$$

Step 5: Then, verify that the mean of the normalized data is zero:

$$s_i = \sqrt{\frac{\sum_{j=1}^{n}(z_{ij} - z_i)^2}{n-1}} = 1 \tag{4.7}$$

Step 6: Form the correlation coefficient matrix (C) of the normalized data. The element matrix (c_{ij}) is calculated as follows

$$c_{ij} = \frac{\sum_{m=1}^{n}(z_m z_{jm})}{n-1} \tag{4.8}$$

Step 7: Compute inverse correlation coefficient matrix (C^{-1})

$$C_{ij} = \frac{Cov(X, Y)}{V(X)V(Y)} \tag{4.9}$$

where

$$Cov(X, Y) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) \tag{4.10}$$

n is the number of samples
X and Y are two different variables being correlated.
$\bar{X}$ and $\bar{Y}$ are the averages among the data in each variable, and
$V(X)$ and $V(Y)$ are the variances of X and Y.

Step 8: Finally, calculate the MD_j using Equation 4.2.

STAGE 2: Assessment of the measurement scale

To evaluate the measurement scale, observations outside the MS or abnormal datasets are used. The same mathematical calculation is repeated to calculate the same goal (MD

Table 2: An example of an OA structure of type $L_8(2^7)$ array.

Run	1	2	3	4	5	6	7
1	1	1	1	1	1	1	1
2	1	1	1	2	2	2	2
3	1	2	2	1	1	2	2
4	1	2	2	2	2	1	1
5	2	1	2	1	2	1	2
6	2	1	2	2	1	2	1
7	2	2	1	1	2	2	1
8	2	2	1	2	1	1	1

value) using the abnormal sample data. However, the abnormal data is normalized based on the mean, standard deviation and correlation matrix of the normal group. The normal MDs and abnormal MDs are then compared. An acceptable measurement scale should demonstrate significant discrimination between the normal and abnormal MD values.

STAGE 3: Identify significant variables

In the third stage, the system is optimized by means of selecting only the features that are known to be significant or useful for the system. This is where the OA and SNR are utilized. The variables are assigned to an orthogonal array experimental run of two-level, in which used is signified as level 1 and not used as level 2. The MD for each experiment run for all used variables from each abnormal sample is calculated. The calculated MD values are recorded according to the experimental run. The SNR based on the MD values for all samples is then computed.

The role of the orthogonal array (OA) in MTS

Orthogonal array (OA) is a type of fractional factorial design of experiment introduced by [23]. It is different from the traditional fractional factorial DOE in the sense that it tries to balance the combination or interaction of factors equally with the minimum number of experimental run. In MTS, the orthogonal array structure is represented by Latin symbology as $L_a(b^c)$ where L is the Latin Square, a is the number of runs, b is the number of factor levels and c is the number of main factors. Table 2 illustrates an example of an OA structure for 7 factors with eight runs and two factor levels.

The name orthogonal is suggested not because of the perpendicular attribute of the structure but rather it is defined as any pairs of columns with the same repetition number of combination of factors [23]. To illustrate further, using the OA in Table 2 as an example, take a pair between column 1 and column 2, the repetition number of each level of combinations in this column pair is the same (which is twice in this case). The

Table 3: The number of repetition of level combinations

Combination	Col 1 & Col 2	Col 1 & Col 3	Col 1 & Col 7	Col 3 & Col 6
1 1	2	2	2	2
1 2	2	2	2	2
2 2	2	2	2	2
2 1	2	2	2	2

same number of repetition should be obtained for the rest of the column pairs thus the $L_8(2^7)$ array depicted by Table 2 can be said to be orthogonal. Table 3 illustrates the number of repetitions in level combination for another three more column pairs.

The same repetition number of levels (twice) of all two column pairs of this OA structure is obtained; therefore the $L_8(2^7)$ array, as depicted by Table 3, can be said to be orthogonal.

In MTS, OAs are used to select the variables of importance by minimizing the different combinations of the original set of variables. The variables are assigned to the different columns of array. Since the variables have only two levels, a two-level array is used in MTS as illustrated in Tabletableoare. For each run of an OA, MDs corresponding to the known abnormal conditions are computed. The importance of variables is judged based on their ability to measure the degree of abnormality on the measurement scale [1]. This is where the signal to noise ratio metric is deployed. Further discussion on OA concepts can be found from [1], [24], [25], [26].

The role of the SNR in MTS

The signal to noise ratio (SNR) concept which can be considered as the core essence of Taguchi philosophy, is developed by Taguchi who was inspired when he was practicing the engineering profession in a Japanese telecommunication company in the 1950s. In telecommunication context, the SNR captures the magnitude of true information (i.e. signals) after making some adjustment for uncontrollable variation (i.e. noise) [1]. In Taguchi's robust engineering concept, the SNR is defined as the measure of the functionality of the system, which exploits the interaction between control factors and noise factors. A gain in the SNR value denotes a reduction in the variability, hence a reduction in the cost associated with the overall significant factors of interest. [26] and [25] provide a detailed description of SNR concepts and its origin of formulation.

In the context of MTS, the SNR is defined as the measure of accuracy of the measurement scale for predicting abnormal conditions [1]. In MTS, a higher value of SNR, expressed in decibels (dB), means a lower prediction error. SNR is used as a metric to assess how significant each variable in the system contributes to the ability to discriminate between the normal and abnormal observations. It could also be used to assess the overall performance of a given MTS system and the degree of improvement after optimization.

Table 4: An example of useful feature selection using OA ($L_8[2^7]$) and SNR

Run	1	2	3	4	5	6	7	MD Computation				SNR
1	1	1	1	1	1	1	1	MD_1	MD_2	MD_3	MD_4	SNR_1
2	1	1	1	2	2	2	2	MD_1	MD_2	MD_3	MD_4	SNR_2
3	1	2	2	1	1	2	2	MD_1	MD_2	MD_3	MD_4	SNR_3
4	1	2	2	2	2	1	1	MD_1	MD_2	MD_3	MD_4	SNR_4
5	2	1	2	1	2	1	2	MD_1	MD_2	MD_3	MD_4	SNR_5
6	2	1	2	2	1	2	1	MD_1	MD_2	MD_3	MD_4	SNR_6
7	2	2	1	1	2	2	1	MD_1	MD_2	MD_3	MD_4	SNR_7
8	2	2	1	2	1	1	1	MD_1	MD_2	MD_3	MD_4	SNR_8
Level 1	SNR	SNR	SNR	SNR	SNR	SNR	SNR					
Level 2	SNR	SNR	SNR	SNR	SNR	SNR	SNR					
Gain	(+/−)	(+/−)	(+/−)	(+/−)	(+/−)	(+/−)	(+/−)					

The three most commonly used types of SNR in MTS are larger-the-better (LTB), nominal-the-best (NTB), and dynamic [1], [5], [**?**], [27]. In this study, the larger-the-better SNR will be deployed.

Larger-the-better SNR

LTB is formulated as in Equation 4.11 below, where t are the abnormal conditions and $D_1^2, D_2^2, \ldots, D_t^2$ are the MDs corresponding to the abnormal situations. The SNR (for the larger-the-better criterion) corresponding to qth run of OA is given as:

$$SNR = \eta_q = -10\log_{10}\left[\frac{1}{t}\sum_{i=1}^{t}\left(\frac{1}{D_i^2}\right)\right] \tag{4.11}$$

For each variable X_i, SNR^1 represents the average SNR of level 1 for X_i while SNR^2 represents the average SNR of level 2 for X_i throughout the vertical columns of the OA. Thus, positive gains from Equation 4.12 constitute useful variables while negative gains constitute otherwise. Table 4 illustrates the assessment made using the SNR to evaluate significant factors of the L_8 OA structure.

$$Gain = SNR^1 - SNR^2 \tag{4.12}$$

STAGE 4: Future deployment with significant variables

The optimized system is then re-evaluated with the abnormal samples to validate the effectiveness of assessing the discriminant power. Once confirmed, the optimized system is used for future application in diagnosis, classification, or forecasting purposes.

5. Methodology

This section presents an overview of Bees Algorithm (BA) methodology developed by [28] followed by the fusion strategies between MTS and BA.

5.1. Bees Algorithm

Bees Algorithm was proposed by D.T. Pham and his colleagues in 2005 [29]. BA is an SI technique that provides an optimization solution based on the collective interaction between bee agents within the colony. This collective intelligent mechanism serves as the backbone of the technique at which faster convergence (or divergence) towards the most promising solution is made possible.

BA gets inspiration from the food foraging behavior of bees to search for the most promising solution to a given optimization problem. Each potential solution in the search space is treated as a food source for the bees. Therefore, each bee carries one possible solution for the problem.

In BA, a population of 'scout'bees randomly search the solutions (food) in the sample space and evaluates the quality of each solution (food) based on a predefined fitness function of the optimization problem. All found solutions by the 'scout bees'are then ranked in either ascending or descending order (depending on the fitness function objective). In a real bee colony, the food quality is evaluated via a 'waggle dance'in which information about the food found (the direction of the food source, its distance from the hive and quality of the food) is choreographed [29].

The highest ranked solution is chosen as a potential optimum solution and more bees will be recruited to exploit further solution (if any) around the neighborhood area. The neighborhood of a solution is called a 'flower patch'in bee colony term. In BA, the neighborhood is the search landscape area near the best solution chosen during recruitment. Other lower ranked solutions will also be exploited selectively under a similar strategy.

Despite the exploitation being performed in the neighborhood area, the remaining scout bees will be assigned for global random searches to explore and locate (if any) other promising solutions better than the one that has been exploited before. This explorative search strategy is crucial to avoid the scout bees population being trapped at a local optimum during the search. Finally, a new set of 'scout bee'population is formed comprising the bees with the best solutions among all exploited sites found so far. A similar process (random search, neighborhood exploitation, and global exploration) is repeated until a global best with a promising solution is found or the search limiting criterion is met. Figure 2 shows the basic flow chart of the Bees Algorithm proposed by [29]. It requires a number of parameters to be set beforehand as depicted by Table 5.

6. MTS-BA fusion strategy

BA is deployed replacing the conventional OA as the optimization technique in MTS. Figure 3 illustrates the conceptual idea behind the strategy.

Table 5: Basic Bees Algorithm Parameters

Parameter	Description
ns	number of scout bees
ne	number of elite sites
nb	number of best sites
nrb	recruited bees for remaining sites
ngh	initial size of the neighborhood
$stlim$	limit or stopping criterion

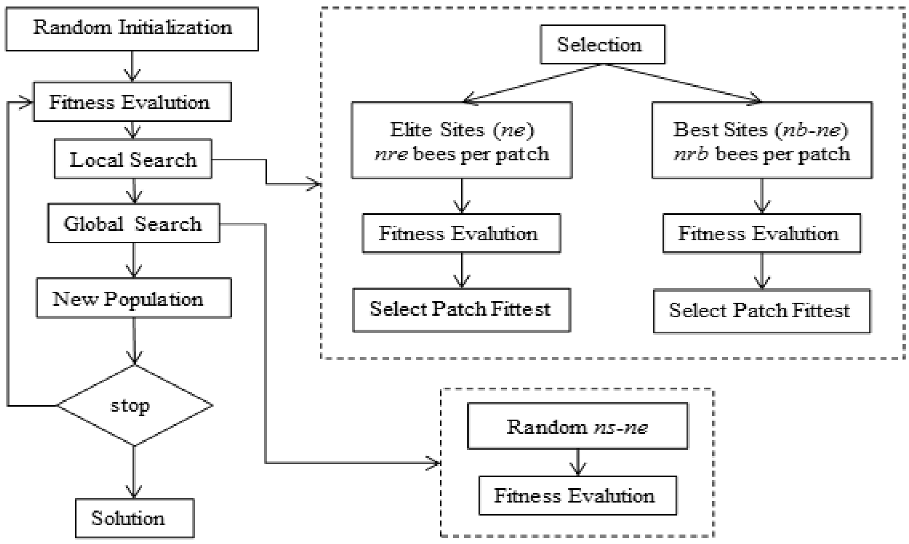


Figure 2: Flowchart of the basic Bees Algorithm.

The scope of the work in this paper concentrates on an optimization study in Stage 3 of the conventional MTS methodology. Normalized unit data and normalized signal data are assumed to be calculated beforehand.

In this fusion strategy, the conventional scaled MD is used as the metric to build the baseline multivariate space using Equation 4.2 while the larger-the-better (LTB) SNR is used as the objective function to be solved as denoted in Equation 4.11. Hence, the objective of BA algorithm in the fusion strategy is to determine the maximum SNR value among the different sets of variables using Equation 4.11. The set of variables with the largest SNR value is the potential solution.

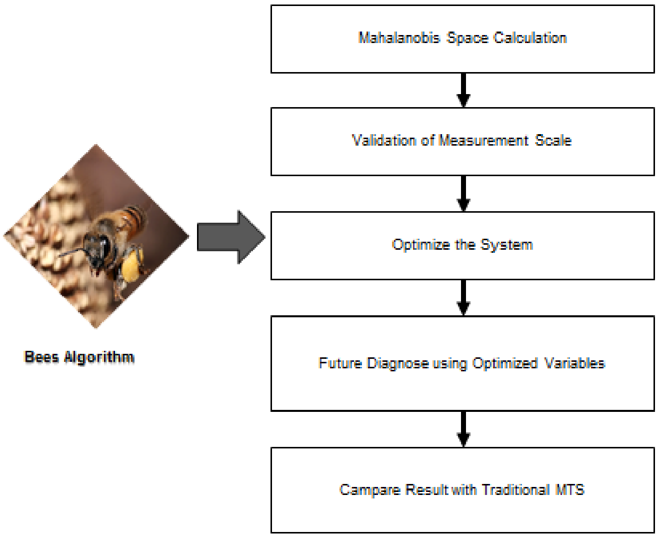


Figure 3: Conceptual flow chart of the study

Table 6: BA parameters setting for this study

Parameter	Setting
n	6
m	3
e	1
nep	4
nsp	2
nop	3

6.1. BA parameter setting

The BA parameters as mentioned in Sub-Section 5.1, are set according to Table 6. The parameters are set based on a trial and error approach, which provides the best algorithm performance (i.e. faster running time) for the same set of problem.

6.2. Neighbourhood search structure

A neighborhood structure can be represented as a function defined as $N : S \rightarrow 2^k - 1$, which assigns to every $s \in S$ set of neighbours $N(s) \subseteq S$ where S is the global search landscape while s is the intended neighbourhood search space. $N(s)$ is termed as the neighbourhood of s. The introduction of a neighbourhood structure brings forward the definition of the concept of locally maximum solutions. A locally maximum solution (or local maximum) with respect to a neighborhood structure N is a solution s such that

■ Enumerate all possible combination set of solutions and evaluate the SNR for the full set of variable:

*[Solution enumeration for the search space, **S** = 2^n - 1, where **n** is the number of original variable set]*

[SNR $_{original\ variable\ set}$ for benchmarking]

■ Initialize population of ***ns*** scout bees for random search

■ Evaluate the ***ns*** scout bees' fitness functions (for maximization)

$$SNR_{LTB} = f(S)_{max} = -10log_{10}\left[\frac{1}{t}\Sigma_{i=1}^{t}\left(\frac{1}{D_i^2}\right)\right]$$

■ While Stopping Criteria (***stlim*** not met)

  *[f (**S** ') > f(**S**)]*

  [minimum half of the population reaching a common solution]

// forming new population (update population)

  ■ Select sites (***ne***, ***nb***) for neighbourhood search *N* (***s***)

    *[N (**s**) = 2^s -1 : where **s** is the number of variable combination found so far]*

  ■ Recruit bees for selected neighbourhood sites (***nre***, ***nrb***)

  ■ Select the fittest (the most maximum SNR value) bee from each neighborhood site

    s *' ← Choose Best of (N(**s**))*

    if *f(**s** ') > f(**S**)*

    then *(if better solution is found in N(**s**))*

      S *←* ***s'***

    else *(**s'** is a local maximum)*

      S *← remains*

    end if

  ■ Assign remaining bees to search globally for other possible solution and evaluate fitness

    *[Globally random search in **S** and compute f(**S** ')]*

■ End While

Figure 4: The pseudo code of the proposed algorithm

$\forall s \in N(s') : f(s') \geq f(s)$, where s' is called a strict locally maximum solution if $f(s') > f(s), \forall s \in N(s')$ [30].

6.3. BA stopping criterion

The stopping criterion is set when the optimum solution is reached with the same value by at least half of the scout bees population. In other words, when the saturated values have been reached, the scout bees can no longer find other better solutions than what they have found so far; therefore, it is denoted as the final solution before stopping the algorithm. Figure 4 illustrates the pseudo code of the proposed algorithm.

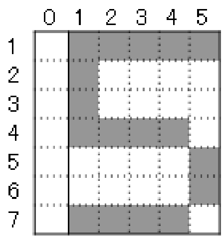
7. Case Study

In this research, optimization is performed on feature selection of character recognition as proposed by [18] using both OA and BA techniques.

7.1. Taguchi's character recognition

The feature selection technique in character recognition concept proposed by [18] is based on the instances of variation and abundance items in a character [22]. Figure 5 illustrates an example of variation and abundance instances of character '5'.

Variation is defined as the number of switches between white-to-grey or grey-to-white as represented by the small circle; while abundance is the number of square grey



		Notation for analysis	
Variation	Abundance	Variation	Abundance
1	5	A	H
2	1	B	I
2	1	C	J
2	4	D	K
1	1	E	L
1	1	F	M
2	4	G	N

Figure 5: Example of variation and abundance instances of character '5'

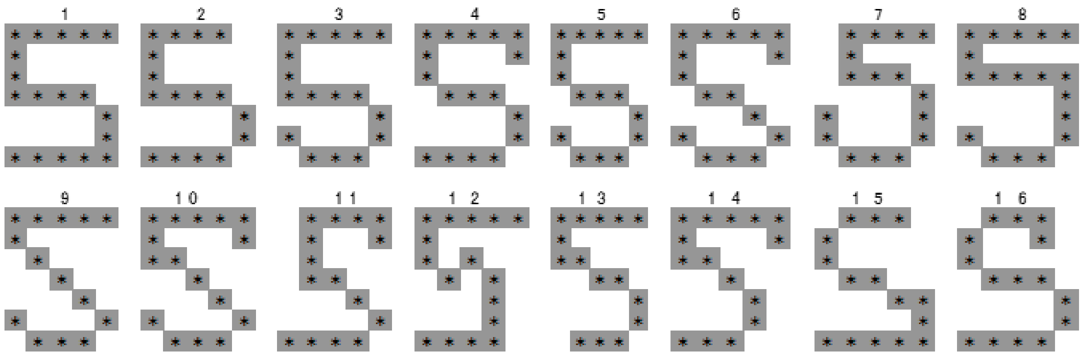


Figure 6: Unit space for pattern '5'

boxes as the arrow passes through each row in the index (see Figure 5). These variation and abundance items act as the variables of interest in MTS for the purposed of feature selection. [22] provide a detailed explanation of these concepts and examples of how they are deployed in the MTS methodology. In this paper, pattern recognition for character '5'is selected for analysis and comparison studies between BA and OA in optimization.

7.2. Unit and Signal Samples

The unit and signal samples are selected based on the work by [22]. There are 16 different samples of character '5'known as normal units and four different samples known as the signal.

The samples with their corresponding item of variations and abundances based on [22] are provided in Figure 6, Table 7, Figure 7, and Table 8.

7.3. Threshold value

A threshold value of 4 (an MD value) is selected based on suggestions by [22], who claimed that the possibility of a sample falling into the unit group with an MD value greater than 4 is small and even smaller as the MD value continues increasing and as it is becomes further from this threshold value. [1] and [22] provide further discussion on this threshold value determination.

Table 7: Data of reference space for the features of the 16 unit samples readable as ‘5’

	Variation							Abundance							MD
Items Label	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	
1	1	2	2	2	1	1	1	5	1	1	4	1	1	5	1.07
2	2	2	2	2	1	1	2	4	1	1	4	1	1	4	1.07
3	1	2	2	2	1	3	2	5	1	1	4	1	2	3	0.88
4	1	3	2	2	1	1	2	5	2	1	3	1	1	4	1.07
5	1	2	2	2	1	3	2	5	1	1	3	1	2	3	0.97
6	1	3	2	2	2	3	2	5	2	1	2	1	2	3	0.59
7	1	2	2	1	3	3	2	4	1	3	1	2	2	3	1.07
8	1	2	1	1	1	3	2	5	1	5	1	1	2	3	1.06
9	1	2	2	2	2	3	2	5	1	1	1	1	2	3	1.06
10	1	3	2	2	2	3	2	5	2	2	1	1	2	3	0.75
11	1	3	2	2	2	1	2	4	2	1	2	1	1	4	1.07
12	1	2	4	4	2	2	2	5	1	2	2	1	1	4	1.07
13	1	2	2	2	1	1	2	5	1	2	2	1	1	3	1.07
14	1	3	2	2	2	2	2	5	2	2	1	1	1	3	1.07
15	2	2	2	2	1	1	1	3	1	1	2	2	1	3	1.07
16	2	4	2	2	1	1	2	3	2	1	3	1	1	4	1.07
Average															1.00

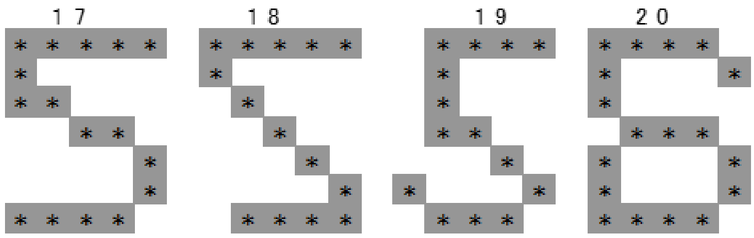


Figure 7: Patterns of the signal ‘5’

Table 8: Feature values of the signal samples

	Variation							Abundance							MD
Items Label	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	
17	1	2	2	2	1	1	2	5	1	2	2	1	1	4	1.82
18	1	2	2	2	2	1	1	5	1	1	1	1	1	4	4.06
19	1	2	2	2	2	3	2	4	1	1	2	1	2	3	3.21
20	2	3	2	2	3	3	2	4	2	1	3	2	2	4	1.07

Table 9: Example of SNR Gain calculation with a positive gain

SNR Optimal System	13.98 dB
SNR Original System	11.52 dB
Gain	2.46 dB

8. Results and Discussion

The optimization algorithms for both OA and BA were constructed using Visual Basic language platform. The programming algorithms were then compiled on a 64-bit Hewlett Packet EliteBook 2540p notebook, which was run on Intel i7 (2.13 GHz) chip with 4 Gigabytes RAM.

In this study, two performance criteria are used to evaluate the performance of the proposed algorithm, which are SNR gain and MD improvement.

8.1. SNR Gain

In MTS, the SNR is not only used to identify useful variables, but it can also be used to measure improvement in the functionality of the system [1] based on the gain in the SNR value. An SNR gain is defined as a value having a positive real number when the SNR value of the original system is subtracted by SNR value of the optimized set. A positive gain value denotes an improvement in the optimized system. The higher the positive value, the better the optimized system is in terms of its functionality and performance compared to the original (not optimized) system. A positive SNR gain relates to variability (noise) reduction in the optimized system [1]. Hence, an accurate recognition and prediction decision is obtained with less computing costs. Table 9 illustrates an example of an SNR gain computation.

8.2. MD Improvement

Similar to the SNR gain concept, MD improvement is calculated based on the MD value of the optimized variable obtained from each tested sample. Again, this value is subtracted by the value of MDs on samples from the original variable set where a positive difference denotes an improvement to the MDs of the sample. An optimized system with MD improvement signifies better recognition accuracy compared to the original set (before optimization) which enhances its discriminant ability.

8.3. Exhaustive Search

An exhaustive search algorithm was conducted as a comparative study (an analogy to a full factorial in design of experiment context). The exhaustive search technique is guaranteed to find the true optima of any given optimization problem [30] since it will search and evaluate each and every single possible solution. However, the drawback of this strategy is that it requires large computational efforts (i.e. long computing time).

Table 10: Optimized variables obtained

Sample No.	Quantity	The Variables	SNR (dB)	Gain (dB)	Time (sec)
Original	14	A B C D E F G H I J K L M N	5.50	-	-
Exhaustive	12	A B C D E F G H I K L N	5.861	0.321	7757.237
BA	12	A B C D E F G H I K L N	5.861	0.321	956.591
OA	10	A D C G H I K L M	5.510	-0.030	10.849

Table 11: MD values obtained based on optimized variables

Sample No.	Original Set of Variable	Variable Set Optimized via BA	Variable Set Optimized via OA
17	1.820	2.069	0.850
18	4.060	4.026	3.258
19	3.210	3.402	2.153
20	110.180	71.498	7.130

8.4. Optimization results

Table 10 depicts the optimized variables obtained using the exhaustive search, OA, and BA techniques. The original variables (before optimization) are also included in the table for reference.

From Table 10, BA found a greater number of optimized variables compared to OA with 12 and 10 variables, respectively. The types of variables optimized by both optimizers were also different with only the variable ‘J’being considered insignificant by both optimizers. The other insignificant variable/s detected by BA was ‘M’, while the variables ‘B’, ‘C’and ‘N’were considered insignificant by OA.

8.5. Validation with optimized variables

To validate the performance of the optimizers with the new sets of optimized variables, the discriminant ability based on MD measurement was determined. The same signal samples illustrated in Figure 7 were used to calculate the MD; however the variables used in the calculation will be based on the optimized sets. The MD computational results were tabulated in Table 11.

It was evident that with the gain in SNR, a slightly greater MD value was expected with the system optimized via BA, which showed an improvement in its discriminant power. However, an exception was given to sample number 20 (i.e. character ‘6’, see Figure 7) since the MD value was lower than the MD value calculated using the original variable set. Nevertheless, the MD value for this particular sample was still considerably large enough and reliable to discriminate the sample from number ‘5’, and it was comparable to the MD value of the original set.

On the other hand, no discriminant power improvement was shown for the system optimized via OA. The MD values obtained using the variable set optimized via OA were lower than that of the original set. This is expected as a result of the negative SNR gain value obtained (see Table 10).

9. Conclusion

The ability to classify objects with a fewer number of significant variables and better accuracy is of major concern for MTS. This work has demonstrated a hybrid MTS approach using BA as the feature selection strategy in the optimization solution of MTS with the use of SNR as the basis of its selection measurement. This paper provides a comparative study using the traditional MTS optimization method of OA to identify and optimize character recognition based on a study demonstrated by [22]. It is shown that optimization using OA in this type of study is inappropriate compared to the heuristic search optimization scheme via BA. Therefore, the results suggest that a Swarm Intelligence technique such as BA offers a promising optimization solution and has the potential to replace OA for optimizing the system. Moreover, to the authors' knowledge, this is the first attempt made to deploy BA into the MTS optimization scheme with a sound Taguchi approach. As a way forward, despite the promising result, future research should focus on further improvements to the methodology particularly in terms of its computing speed. Future research should also be conducted on additional sample tests for further test and refine the proposed integrated methodology.

References

[1] G. Taguchi and R. Jugulum, *The Mahalanobis Taguchi strategy: A pattern technology system*. John Wiley & Sons, 2002.

[2] P. C. Mahalanobis, "On the generalized distance in statistics," *Proceedings of the National Institute of Sciences (Calcutta)*, vol. 2, pp. 49–55, 1936.

[3] E. A. Cudney, K. Paryani, and K. M. Ragsdell, "Applying the mahalanobis–taguchi system to vehicle handling," *Concurrent Engineering*, vol. 14, no. 4, pp. 343–354, 2006.

[4] R. S. Kestle and E. A. Cudney, "An evaluation of mahalanobis-taguchi system experiment metrics," in *IIE Annual Conference. Proceedings*, p. 1, Institute of Industrial Engineers-Publisher, 2011.

[5] G. Taguchi, R. Jugulum, and S. Taguchi, *Computer-based robust engineering: essentials for DFSS*. ASQ Quality Press, 2004.

[6] D. M. Hawkins, "Discussion," *Technometrics*, vol. 45, no. 1, pp. 25–29, 2003.

[7] W. H. Woodall, R. Koudelik, K.-L. Tsui, S. B. Kim, Z. G. Stoumbos, and C. P. Carvounis, "A review and analysis of the mahalanobis–taguchi system," *Technometrics*, vol. 45, no. 1, pp. 1–15, 2003.

[8] B. Abraham and A. M. Variyath, "Discussion," *Technometrics*, vol. 45, no. 1, pp. 22–24, 2003.

[9] A. Pal and J. Maiti, “Development of a hybrid methodology for dimensionality reduction in mahalanobis–taguchi system using mahalanobis distance and binary particle swarm optimization,” *Expert Systems with Applications*, vol. 37, no. 2, pp. 1286–1293, 2010.

[10] R. Jugulum, G. Taguchi, S. Taguchi, and J. O. Wilkins, “Discussion,” *Technometrics*, vol. 45, no. 1, pp. 16–21, 2003.

[11] J. Häcker, F. Engelhardt, and D. D. Freyw, “Robust manufacturing inspection and classification with machine vision,” *International journal of production research*, vol. 40, no. 6, pp. 1319–1334, 2002.

[12] P. Saraiva, N. Faísca, R. Costa, and A. Gonçalves, “Fault identification in chemical processes through a modified mahalanobis-taguchi strategy,” *Computer Aided Chemical Engineering*, vol. 18, pp. 799–804, 2004.

[13] P. Das and S. Datta, “Developing an unsupervised classification algorithm for characterization of steel properties,” *International Journal of Quality & Reliability Management*, vol. 29, no. 4, pp. 368–383, 2012.

[14] J.-l. NIU and L.-s. CHENG, “Classification using improved mahalanobis-taguchi system based on omni-optimizer [j],” *Systems Engineering-Theory & Practice*, vol. 6, p. 020, 2012.

[15] E. Reséndiz, L. A. Moncayo-Martínez, and G. Solís, “Binary ant colony optimization applied to variable screening in the mahalanobis–taguchi system,” *Expert Systems with Applications*, vol. 40, no. 2, pp. 634–637, 2013.

[16] E. Reséndiz and C. A. Rull-Flores, “Mahalanobis–taguchi system applied to variable selection in automotive pedals components using gompertz binary particle swarm optimization,” *Expert Systems with Applications*, vol. 40, no. 7, pp. 2361–2365, 2013.

[17] A. S. Iquebal and A. Pal, “Artificial bee colony optimisation-based enhanced mahalanobis taguchi system for classification,” *International Journal of Intelligent Engineering Informatics*, vol. 2, no. 2-3, pp. 181–194, 2014.

[18] G. Taguchi, “Method for pattern recognition,” Nov. 4 1997. US Patent 5,684,892.

[19] I.-J. Hwang and G.-J. Park, “A multi-objective optimization using distribution characteristics of reference data for reverse engineering,” *International Journal for Numerical Methods in Engineering*, vol. 85, no. 10, pp. 1323–1340, 2011.

[20] S. Feng, O. Hiroyuki, T. Hidennori, K. Yuichi, and S. Hu, “Qualitative and quantitative analysis of gmaw welding fault based on mahalanobis distance,” *International Journal of Precision Engineering and Manufacturing*, vol. 12, no. 6, pp. 949–955, 2011.

[21] W. Guo, R.-z. Yin, G. Li, and N. Zhao, “Research on selection of enterprise management-control model based on mahalanobis distance,” in *The 19th International Conference on Industrial Engineering and Engineering Management*, pp. 555–564, Springer, 2013.

[22] S. Teshima, Y. Hasegawa, and K. Tatebayashi, *Quality Recognition & Prediction: Smarter Pattern Technology with the Mahalanobis-Taguchi System*. Momentum Press, 2012.

[23] A. S. Hedayat, N. J. A. Sloane, and J. Stufken, *Orthogonal arrays: theory and applications*. Springer Science & Business Media, 2012.

[24] S. Hedayat, NJA and J. Stufken, "Orthogonal arrays: Theory and applications," 1999.

[25] S. Park, *Robust design and analysis for quality engineering*. Boom Koninklijke Uitgevers, 1996.

[26] M. S. Phadke, *Quality Engineering Using Robust Design*. PTR Prentice-Hall, Inc, 1989.

[27] R. Dolah and Z. Miyagi, "Effect of peel side on optimum condition for measuring flexible film peel strength in t-peel adhesion test," *Journal of Testing and Evaluation*, vol. 42, no. 1, pp. 1–13, 2013.

[28] D. Pham, A. Ghanbarzadeh, E. Koc, S. Otri, S. Rahim, and M. Zaidi, "The bees algorithm–a novel tool for complex optimisation," in *Intelligent Production Machines and Systems-2nd I* PROMS Virtual International Conference 3-14 July 2006*, p. 454, Elsevier, 2011.

[29] D. T. Pham and M. Castellani, "The bees algorithm: modelling foraging behaviour to solve continuous optimization problems," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 223, no. 12, pp. 2919–2938, 2009.

[30] C. Blum and A. Roli, "Metaheuristics in combinatorial optimization: Overview and conceptual comparison," *ACM Computing Surveys (CSUR)*, vol. 35, no. 3, pp. 268–308, 2003.