



*Research article*

## **Modified Mahalanobis-Taguchi System based on proper orthogonal decomposition for high-dimensional-small-sample-size data classification**

**Ting Mao<sup>1</sup>, Lanting Yu<sup>1</sup>, Yueyi Zhang<sup>1,\*</sup> and Li Zhou<sup>2</sup>**

<sup>1</sup> School of Economics and Management, China Jiliang University, Hangzhou 310000, China

<sup>2</sup> China National Institute of Standardization, Beijing 100000, China

\* **Correspondence:** Email: 02a0702026@cjlu.edu.cn.

**Abstract:** Mahalanobis-Taguchi System (MTS) is an effective algorithm for dimensionality reduction, feature extraction and classification of data in a multidimensional system. However, when applied to the field of high-dimensional small sample data, MTS has challenges in calculating the Mahalanobis distance due to the singularity of the covariance matrix. To this end, we construct a modified Mahalanobis-Taguchi System (MMTS) by introducing the idea of proper orthogonal decomposition (POD). The constructed MMTS expands the application scope of MTS, taking into account correlations between variables and the influence of dimensionality. It can not only retain most of the original sample information features, but also achieve a substantial reduction in dimensionality, showing excellent classification performance. The results show that, compared with expert classification, individual classifiers such as NB, RF, k-NN, SVM and superimposed classifiers such as Wrapper + RF, MRMR + SVM, Chi-square + BP, SMOTE + Wrapper + RF and SMOTE + MRMR + SVM, MMTS has a better classification performance when extracting orthogonal decomposition vectors with eigenvalues greater than 0.001.

**Keywords:** Mahalanobis-Taguchi System; proper orthogonal decomposition; high-dimensional-small-sample-size data; classification

---

### **1. Introduction**

Mahalanobis-Taguchi System (MTS) is a new pattern recognition method proposed by Japanese

quality engineer Genichi Taguchi [1,2]. This method utilizes Mahalanobis distance and Taguchi robust design to perform dimensionality reduction, feature extraction, classification and diagnosis in a multi-dimensional system [2]. MTS uses Mahalanobis distance as the covariance distance. It can effectively measure the importance of each variable [4] and has unique robustness. Therefore, MTS is widely used in dimensionality reduction and classification of large-sample-size data and unbalanced data, such as dimensionality reduction and classification of agricultural crop [3], pre-indicators of delirium [4], and health monitoring [5].

In recent years, more and more high-dimensional-small-sample-size data require dimensionality reduction and classification, such as pilot randomized controlled designs in clinical research [6], energy prediction and optimization in the petrochemical industry [7], and rapid-cycle quality improvement projects [8]. The sample numbers of datasets from these fields are often much smaller than their dimension numbers [9]. Therefore, it is difficult to apply statistical methods for parameter estimation and classification, and the constructed predictive models are prone to problems such as poor learning performance, large uncertainty, and low accuracy, which will interfere with the final decision [10]. Some scholars proposed to perform principal component analysis (PCA) based on variance measurement to achieve dimensionality reduction and classification of high-dimensional-small-sample-size data [11]. However, PCA only reduces the dimensionality of a single dataset and does not consider correlations between datasets [12]. To address this issue, Ma et al. improved PCA and proposed sparse principal component analysis (SPCA), which sparsifies load factors so that most of them are 0, thereby rendering the principal component a stronger interpretation ability [13]. However, SPCA requires the calculation of a large number of eigenvalues, which is computationally costive [14]. In addition, some scholars used penalized techniques to reduce the dimensionality of high-dimensional small sample data, such as least absolute shrinkage and selection operator (LASSO) regression or smoothly clipped absolute deviation (SCAD) and minimum concave penalty functions. However, such techniques suffer from so-called oracle inequalities, multiple local minimums, and computational difficulties [15]. The minimum-redundancy-maximum-relevance (MRMR) is used as one of the effective methods of dimensional reduction because of its high accuracy, but it has high computational cost and is significantly influenced by the number of features [16]. Wrapper is also an effective dimensional reduction method, but often suffers from over-fitting problems [17].

In view of the unique effectiveness of MTS in dimensionality reduction and classification, Xiao et al. [18] optimized MTS to perform classification on high-dimensional-small-sample-size data. MTS uses definite thresholds to identify valid features and build models, thus exhibiting good dimensionality reduction, feature extraction and classification performances. However, when using MTS to process high-dimensional-small-sample-size data, the solution of an inverse problem is not unique, there will be an infinite number of solutions compatible with the data [19]. Therefore, the singularity of covariance matrix must be considered.

To overcome the covariance matrix singularity problem, and to inherit and innovate the traditional variance measurement-based dimensionality reduction and classification methods, in this study, we introduce the idea of proper orthogonal decomposition (POD) from the traditional methods to construct a modified MTS (MMTS) for dimensionality reduction and classification of high-dimensional-small-sample-size data. POD was originally proposed by K. Pearson [20] and H. Hotelling [21]. It can reduce a large number of interdependent variables to a much smaller number of variables [22]. Therefore, it

can effectively overcome the variance matrix singularity problem. Compared with the adjoint matrix method or the Schmidt orthogonalization method used in the traditional MTS, the POD method has several advantages in terms of building fewer matrix elements, greatly reducing computation, and improving computing speed. When applied in high-dimensional-small-sample-size data, POD-based MMTS can not only solve the covariance matrix singularity problem in MTS, improve the architecture and algorithm of the traditional MTS, expand the application scope of MTS, but also inherit and innovate the traditional dimensionality reduction methods by overcoming the poor robustness and weak interpretation ability of PCA and reducing the computational cost of SPCA. Therefore, MMTS is a method that can quickly and efficiently extract valid features and achieve the goal of dimensionality reduction and classification.

The structure of this article is as follows. In Section 2, we introduce the theory and method of the traditional MTS. In Section 3, we first introduce POD and then systematically explain the MMTS constructed based on the idea of POD. In Section 4, we introduce a dataset and the evaluation metrics, and demonstrate through experiments that the MMTS exhibits good classification accuracy. Section 5 summarizes this research.

## 2. Mahalanobis-Taguchi System

MTS is one of the most cutting-edge methods in the field of quality engineering. It utilizes Mahalanobis distance and Taguchi robust design to perform diagnosis and forecast in a multi-dimensional system [1]. It applies the whole set of ideas of Taguchi signal-to-noise ratio (SNR) experimental design into characteristic variable selection of pattern recognition [23]. This method uses orthogonal tables and SNR gains to test the validity of characteristic variables and select influential characteristic variables, thereby optimizing and simplifying characteristic variables [1] and effectively reducing dimensionality. It realizes minimizing the number of original features in the system without losing the quality of recognition [24].

Since MTS uses Mahalanobis distance as the covariance distance, it can effectively measure the importance of attribute sets [25]. Mahalanobis distance, which constitutes a general indication of the degree of divergence in the means of the sample characteristics in multivariate space [26], is used to eliminate the influence of dimensions [27]. The Mahalanobis distance facilitates the construction of multi-dimensional measure scales for the purpose of classification [28]. Therefore, MTS has great advantages on dimensionality reduction. In addition, MTS does not require any data distribution assumption, requires only small sample sizes, and is easy to operate [2]. In view of these characteristics, MTS is widely used in the medical, engineering, and financial fields [29].

The basic operation steps of MTS are as follows:

Step1: Construction of the Mahalanobis space

Collect a sample dataset with a sample number of  $n$  and a characteristic variable number of  $p$  from the normal group. Let  $x_{ij}$  be the value of the  $j$ th characteristic variable of the  $i$ th sample in the normal group, where  $i=1,2,\dots,n$ ,  $j=1,2,\dots,p$ .

Calculate the mean  $\bar{x}_j$  and the standard deviation  $s_j$ , and standardize the sample dataset:

$$\begin{aligned}
 \bar{x}_j &= \frac{1}{n} \sum_{i=1}^n x_{ij} \\
 s_j &= \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2} \\
 Z_{ij} &= \frac{x_{ij} - \bar{x}_j}{s_j}
 \end{aligned} \tag{1}$$

Step2: Calculation of the Mahalanobis distance and SNR

Calculate the Mahalanobis distances of the normal and abnormal groups. Mahalanobis distances is a square distance, recorded as  $D^2$ :

$$MD_i = D^2 = Z_i R^{-1} (Z_i)^T \tag{2}$$

where  $R$  is the correlation coefficient matrix of data.  $Z_i$  is the all characteristic values of the  $i$  th sample that have been standardized,  $Z_i = [Z_{i1}, Z_{i2}, \dots, Z_{ip}]$ .

If the Mahalanobis distance of the normal group is greater than that of the abnormal group, the Mahalanobis space is valid, and an appropriate orthogonal table is selected for orthogonal experiments and SNR calculation. This study takes larger-the-better characteristics as an example:

$$SNR = -10 \lg \left( \frac{1}{m} \sum_{i=1}^m \frac{1}{MD_i^2} \right) \tag{3}$$

Where  $m$  is the sample number of the abnormal group.

Step 3: Selection of characteristic variables

For a  $j$  th characteristic variable,  $\overline{SNR}_{level1}$  represents the mean SNR value of experiments using this characteristic variable, while  $\overline{SNR}_{level2}$  represents the mean SNR value of experiments not using this characteristic variable. If  $SNR_{gain} = \overline{SNR}_{level1} - \overline{SNR}_{level2} > 0$ , the  $j$  th characteristic variable is selected, otherwise the variable is removed.

### 3. Modified Mahalanobis-Taguchi System

This section introduces the classification of high-dimensional-small-sample-size data using an MMTS. First, we introduce the POD proposed by K. Pearson [20] and H. Hotelling [21]. Then we describe an improved POD proposed by Zhu et al. [30], which is suitable for matrices with a large number of variables and a relatively small number of samples. Finally, we introduce the algorithm and operation steps of the constructed MMTS in details.

#### 3.1. Proper Orthogonal Decomposition

POD can obtain uncorrelated principal components [31] and is therefore widely used in multi-

dimensional and multi-variable scenarios, such as comparison of stock return volatility patterns [32] and environmental research [33].

Set  $X$  as the data matrix,  $t_i$  as the orthogonal decomposition vector of matrix  $XX'$  and  $r_i$  as the orthogonal decomposition vector of matrix  $XX$ . The principle of POD is to perform  $t_i = Xr_i (i = 1, 2, \dots, n)$  orthogonal transformation on data matrix  $X$ , so that the variance of vector  $t_i$  is the largest, that is  $\max(t_i' t_i)$ . When the variance corresponding to  $t_i$  is smaller, the vector is considered to contain more noise components and needs to be removed. The model is:

$$\begin{aligned} & \max(t_i' t_i) \\ & \text{s.t.} \begin{cases} t_i = Xr_i (i = 1, 2, \dots, n) \\ r_i' r_i = 1 (i = 1, 2, \dots, n) \\ t_i' t_j = 0 (i \neq j) \end{cases} \end{aligned} \quad (4)$$

To enhance the calculation efficiency of the iterative algorithm, Zhu et al. [30]. improved the algorithm. Instead of using matrix  $(XX)$ , the improved algorithm uses matrix  $(XX')$ , which is more suitable for matrices with a large number of variables and a relatively small number of samples.

Multiply  $X$  by  $r_i = \frac{1}{\lambda_i} (XX)_i r_i$ , we get:

$$\begin{aligned} r_i &= \frac{1}{\lambda_i} (XX)_i r_i \\ Xr_i &= \frac{1}{\lambda_i} X (XX)_i r_i \\ \therefore Xr_i &= t_i \\ \therefore t_i &= \frac{1}{\lambda_i} X (XX)_i r_i \\ \lambda_i t_i &= (XX')_i t_i \end{aligned} \quad (5)$$

Where  $\lambda_i$  is the characteristic value corresponding to  $r_i$ .

Perform the following transformations on  $(XX)_{i+1} = (XX)_i - \lambda_i r_i r_i'$ :

$$\begin{aligned} (XX)_{i+1} &= (XX)_i - \lambda_i r_i r_i' \\ X(XX)_{i+1} &= X(XX)_i - \lambda_i Xr_i r_i' \\ (XX')_i X_{i+1} &= (XX')_i X_i - \lambda_i t_i t_i' \\ (XX')_i X_{i+1} X' &= (XX')_i X_i X' - \lambda_i t_i t_i' X' \\ (XX')_i (XX')_{i+1} &= (XX')_i (XX')_i - \lambda_i t_i t_i' \\ (XX')_i (XX')_{i+1} &= (XX')_i (XX')_i - (XX')_i t_i t_i' \\ (XX')_{i+1} &= (XX')_i - t_i t_i' \end{aligned} \quad (6)$$

Perform the following transformation on  $\lambda_i r_i$ :

$$\lambda_i r_i = (X'X)r_i = X't_i \quad (7)$$

It is demonstrated that the results obtained by this algorithm (including  $r_i, t_i, \lambda_i$ ) are exactly the same as those obtained by the traditional algorithm.

### 3.2. A modified Mahalanobis-Taguchi System based on the idea of POD

MTS has unique advantages in dimensionality reduction and classification. It utilizes orthogonal tables to design experimental plans, takes the SNR gain of the Mahalanobis distance as the evaluation metric, and uses the data analysis method of orthogonal experiment to optimize the selection of characteristic variables. Therefore, this method is novel, ease to operate and effective [23]. However, when the dimension number of data is greater than the number of samples, the Mahalanobis distance cannot be calculated [34], thereby making MTS unusable. This research introduces the idea of POD into MTS and proposes an MMTS, thereby eliminating the system structural and algorithmic defects of MTS.

The algorithm and operation steps of the POD-based MMTS are as follows:

Step 1: Data standardization

Let the input data matrix  $X$  be an  $n \times p$  matrix, and that  $n < p$ .  $x_{ij}$  is an element in the  $i$  th row and  $j$  th column of  $X$ . Its actual meaning is the value of the  $j$  th characteristic variable of the  $i$  th sample, where  $i=1,2,\dots,n$ ,  $j=1,2,\dots,p$ .

Standardize the data as follows:

$$\begin{aligned} \bar{x}_j &= \frac{1}{n} \sum_{i=1}^n x_{ij} \\ s_j &= \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2} \\ Z_{ij} &= \frac{x_{ij} - \bar{x}_j}{s_j} \end{aligned} \quad (8)$$

Step 2: Iteratively calculate the variance of  $ZZ'$  matrix of the normal samples until  $\frac{\lambda^{(k+1)} - \lambda^{(k)}}{\lambda^{(k)}} < 10^{-8}$

is satisfied.  $\lambda^{(k)}$  is the characteristic value of the  $k$  th iteration. At this point, all useful information has been extracted by default.

Calculate the variance of  $ZZ'$  matrix of the normal samples:

$$(ZZ')_1 = ZZ' \quad (9)$$

where  $Z_i = [Z_{i1}, Z_{i2}, \dots, Z_{ip}]$ ,  $Z = [Z_1, Z_2, \dots, Z_n]$ .

Select  $\lambda_j$  to make vector  $t_j$  a unit vector:

$$t_j = \frac{1}{\lambda_j} (ZZ')_k t_j \quad (10)$$

Where  $\lambda_j$  is the characteristic value and  $t_j$  is the orthogonal decomposition vector of matrix  $(ZZ')_k$ .

Select  $\lambda_j$  to make vector  $r_j$  a unit vector,  $t_j = Zr_j$ , and calculate vector  $r_j$ :

$$r_j = \frac{1}{\lambda_j} Z' t_j \quad (11)$$

Update the variance matrix:

$$(ZZ')_{k+1} = (ZZ')_k - t_j t_j' \quad (12)$$

The abnormal samples are extracted according to the orthogonal decomposition vectors extracted from the normal samples.

Step 3: Calculate the correlation matrix and Mahalanobis distance and verify the validity of the Mahalanobis space.

Let the dimension number obtained by POD be  $q$ . Set  $\lambda_{newj}$  as the characteristic value and  $\xi_j$  as the orthogonal decomposition vector of the data matrix,  $j=1,2,\dots,q$ . Note that  $\xi_j$  needs to be unitized.

$$R = \sum_{j=1}^q \lambda_{newj} \xi_j \xi_j' \quad (13)$$

Calculate the Mahalanobis distance:

$$MD_i = Z_i' \hat{R}^{-1} Z_i = \sum_{j=1}^q \frac{(\hat{\xi}_j' Z_i)^2}{\hat{\lambda}_{newj}} \quad (14)$$

Where  $Z_i = [Z_{i1}, Z_{i2}, \dots, Z_{ip}]$ .

If the Mahalanobis distance of the abnormal samples is greater than that of the normal samples, the Mahalanobis space is valid.

Step 4: Calculate the SNR gain, extract valid features, and complete dimensionality reduction.

Use an appropriate orthogonal table to carry out orthogonal experiments, and select a valid orthogonal decomposition variable by calculating the SNR. The SNR used in this article has larger-the-better characteristics:

$$SNR = -10 \lg \left( \frac{1}{m} \sum_{i=1}^m \frac{1}{MD_i^2} \right) \quad (15)$$

Where  $m$  is the number of samples in the abnormal group.

When designing the orthogonal table, "1" indicates that the characteristic variable is used, while "2" indicates that the characteristic variable is not used. The calculation formula of SNR gain is as follows:

$$gain = \overline{SNR}_{level1} - \overline{SNR}_{level2} \quad (16)$$

If  $gain = \overline{SNR}_{level1} - \overline{SNR}_{level2} > 0$ , then the characteristic variable is considered to be worth keeping; if  $gain = \overline{SNR}_{level1} - \overline{SNR}_{level2} < 0$ , it is considered that the characteristic variable should be

removed.

Step 5: Use the dimensionality-reduced characteristic variable to verify the validity of the Mahalanobis space again. If the Mahalanobis space is valid, the dimensionality-reduced characteristic variable can be used for other analyses (such as prediction and classification) of unknown samples.

Take classification as an example. Before classifying the test set  $X$ , it is necessary to standardize its data matrix. Let the element of the unknown dataset be  $X_{ij}$ , the mean value and standard deviation of the matrix of “normal” samples be  $\bar{x}_j$  and  $s_j$ , respectively, and the processed matrix element be  $XD_{ij}$ , then:

$$XD_{ij} = \frac{X_{ij} - \bar{x}_j}{s_j} \quad (17)$$

Extract  $q$  dimensions of valid variables from the processed data matrix, calculate the Mahalanobis distance  $MD_i$ , and classify the test set according to the threshold.

$$MD_i = XD_i R^{-1} (XD_i)^T \quad (18)$$

Where  $R^{-1}$  is the inverse matrix of the correlation matrix.

The main contribution of this paper is to integrate the idea of POD into MTS, thereby perfecting MTS's own theoretical system, solving the problems of covariance matrix singularity and the eigenvalue estimation, and expanding the application scope of MTS. In addition, our research establishes a reliable and effective method for dimensionality reduction of high-dimensional-small-sample-size data, thereby providing new ideas for academic research on high-dimensional-small-sample-size data and feature extraction and dimensionality reduction in real life scenarios.

## 4. Experiments and results

### 4.1. Dataset

In recent years, radiomics and texture analysis have received extensive attention from clinical and academic circles [35–37]. Therefore, in this section, we select a dataset from the UCI database showing colonoscopy video information of gastrointestinal lesions [38]. The dataset includes 76 cases, 698 characteristic variables, and three types of lesions, namely hyperplastic lesions, adenomas and serrated adenomas. After deleting the missing values, the dimension of dataset is 415.

From the perspective of a binary classification problem, hyperplastic lesions belong to the “benign” category with a sample size of 21, while adenomas and serrated adenomas together belong to the “malignant” category with a sample size of 55. During binary classification of this dataset, the average classification accuracy of seven clinicians (including four experts and three beginners) is 79.4956%. The average classification accuracy of the expert group is 79.6052%, and the average classification accuracy of the beginner group is 79.3860%. The data features used in this study include narrow band imaging (NBI), color, 3D shape, and textural features. The software used in this study includes MATLAB 2018b, SPSS25, and Minitab17.

Our experiments compare the results of MMTS with the results of individual classifiers such as naive bayes (NB), random forests (RF), k-nearest neighbor (k-NN) and support vector



machine (SVM) techniques that have been commonly used in the gastroenterology field to facilitate high-dimensional-small-sample-size data classification. Since Wrapper [39], MRMR [40] and Chi-square features selection [41] are common methods in terms of feature selection and dimensionality reduction, and the synthetic minority oversampling technique (SMOTE) [42,43] is widely used in case of small samples to increase the number of observations and thus improve classification accuracy, our experiments also compare the results of MMTS with the results of superimposed classifiers such as Wrapper + RF, MRMR + SVM, Chi-square + BP, SMOTE + Wrapper + RF and SMOTE + MRMR + SVM.

#### 4.2. Metrics

To verify if MMTS can effectively classify high-dimensional-small-sample-size data, this study uses precision, specificity, recall, accuracy, and F-value as evaluation metrics [44].

Table 1 is a confusion matrix for the binary classification problem, where TP (True Positives) represents the number of positive cases that are correctly predicted, TN (True Negatives) represents the number of negative cases that are predicted correctly, FP (False Positives) represents the number of positive cases that are incorrectly predicted, and FN (False Negatives) represents the number of negative cases that are incorrectly predicted.

**Table 1.** Confusion matrix for the binary classification problem.

		Ture class	
		Negative	Positive
Hypothesis output	Negative	TN	FN
	Positive	FP	TP
	Sum	Nn	Np

T: True; F: False; N: Negative; P: Positive

$$precision = \frac{TP}{TP + FP} \quad (19)$$

$$recall = \frac{TP}{TP + FN} \quad (20)$$

$$specificity = \frac{TN}{TN + FP} \quad (21)$$

$$accuracy = \frac{TP + TN}{Nn + Np} \quad (22)$$

$$F = \frac{(1 + \beta^2) \times Recall \times Precision}{\beta^2 \times Recall + Precision} \quad (23)$$

F-value is the weighted harmonic average of “precision” and “recall”. It is usually assumed that both metrics are equally important, that is  $\beta=1$ , then  $F = \frac{2\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}$ . The larger the F-value (the closer it is to 1), the better the classification effect.

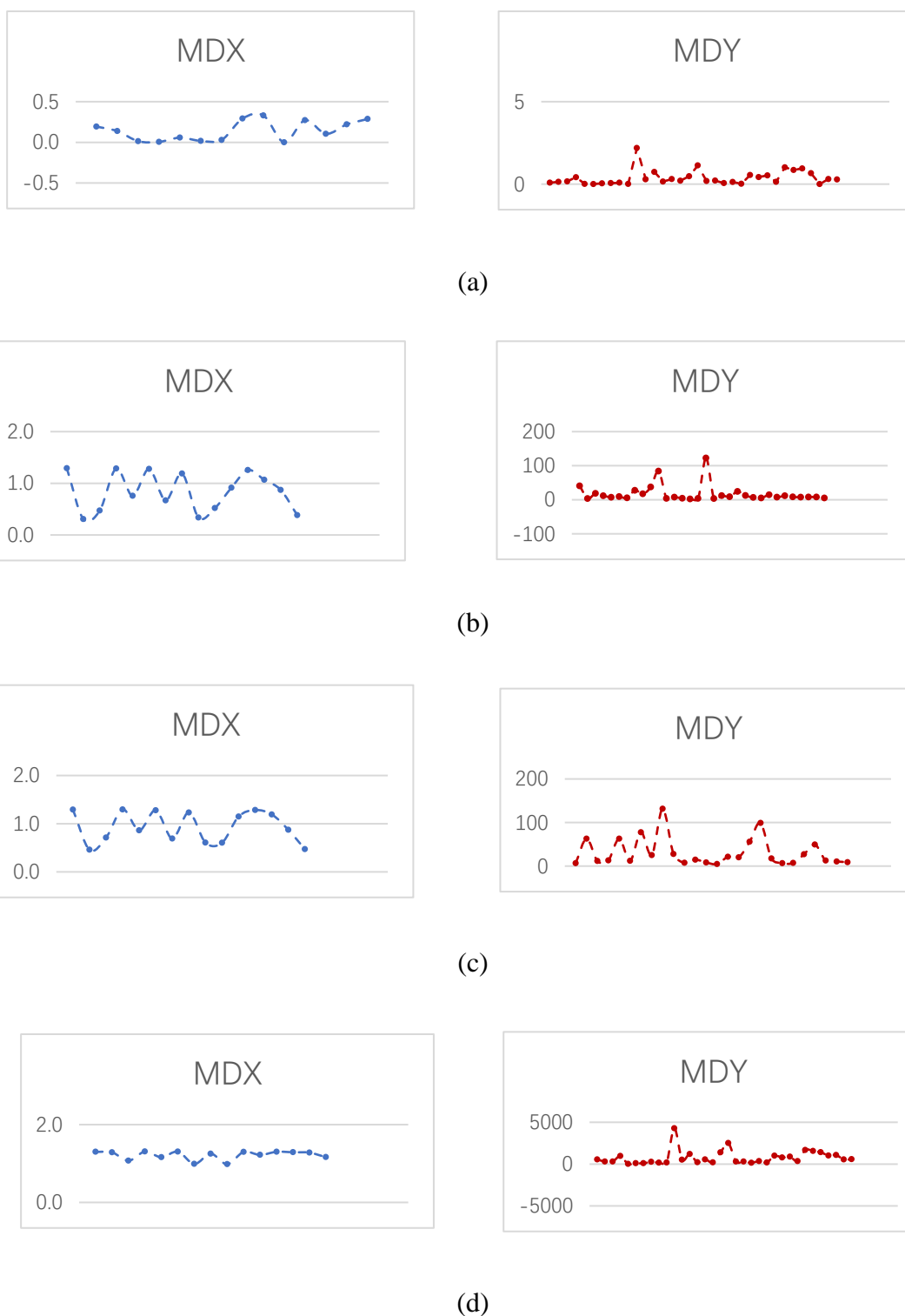
#### 4.3. Results

Iterative calculation is performed on a training set  $ZZ'$  randomly selected from normal samples to obtain the orthogonal decomposition results shown in Table 2.

**Table 2.** Orthogonal decomposition vectors.

Component	Initial eigenvalue			Extract load sum of square		
	Total	Variance percentage (%)	Cumulative variance percentage (%)	Total	Variance percentage (%)	Cumulative variance percentage (%)
1	14.40442	96.02944	96.029	14.40442	96.029	96.029
2	0.40432	2.69545	98.725	0.40432	2.695	98.725
3	0.12969	0.86459	99.589	0.12969	.865	99.589
4	0.03494	0.23296	99.822	0.03494	.233	99.822
5	0.01601	0.10670	99.929	0.01601	.107	99.929
6	0.00350	0.02331	99.952	0.00350	.023	99.952
7	0.00315	0.02098	99.973	0.00315	.021	99.973
8	0.00181	0.01206	99.985	0.00181	.012	99.985
9	0.00110	0.00736	99.993	0.00110	.007	99.993
10	0.00059	0.00396	99.997	0.00059	.004	99.997
11	0.00020	0.00136	99.998	0.00020	.001	99.998
12	0.00013	0.00089	99.999	0.00013	.001	99.999
13	0.00008	0.00054	100.000	0.00008	.001	100.000
14	0.00004	0.00026	100.000			
15	0.00002	0.00014	100.000			

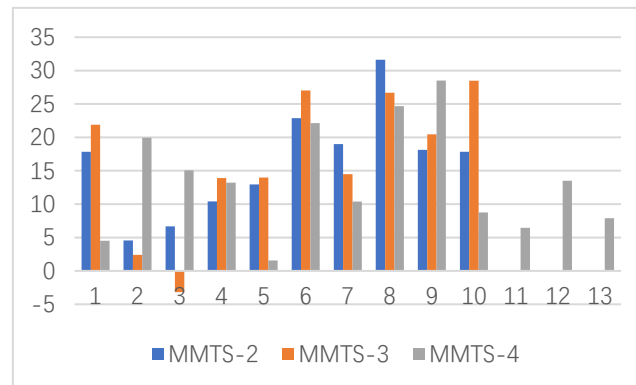
For high-dimensional-small-sample-size data, the optimal extraction number of orthogonal decomposition vector is still an open topic. Therefore, to determine the optimal extraction number of orthogonal decomposition vectors, this study extracts the corresponding number of orthogonal decomposition vectors following four schemes of eigenvalues greater than 1, eigenvalues greater than 0.001, eigenvalues greater than 0.0005, and eigenvalue variance accounting for 100% (hereinafter abbreviated as MMTS-1, MMTS-2, MMTS-3, and MMTS-4, respectively). Figure 1 shows the Mahalanobis distances of normal samples (MDX) and those of abnormal samples (MDY) calculated by the different schemes.



**Figure 1.** MDXs and MDYs under the four schemes of (a) MMTS-1, (b) MMTS-2, (c) MMTS-3, and (d) MMTS-4.

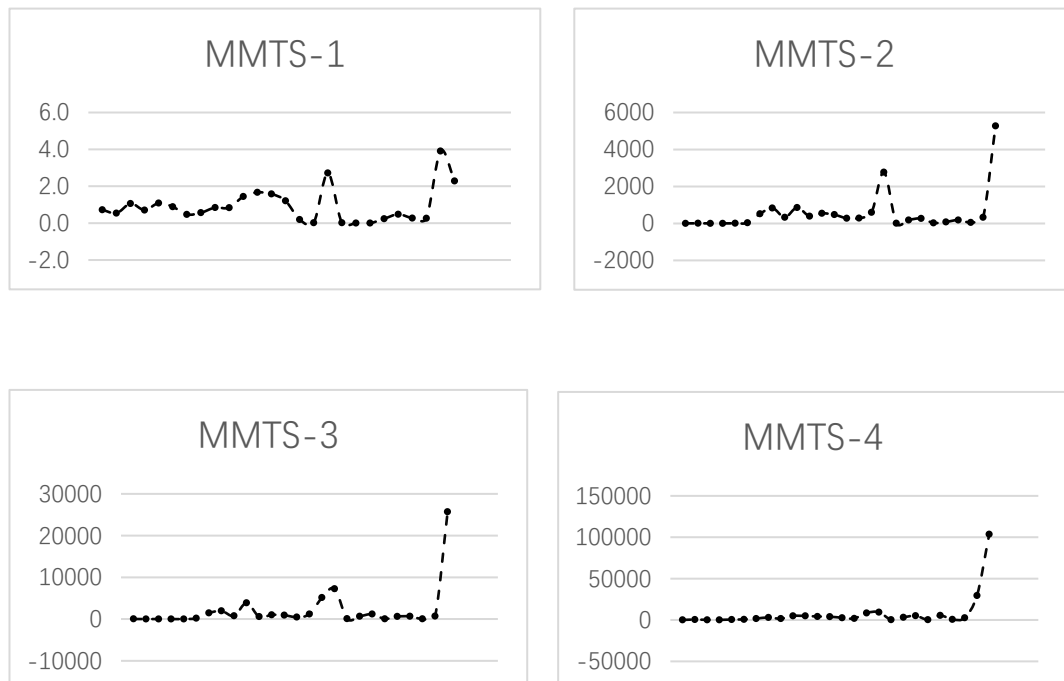
Since MMTS-1 has only one orthogonal decomposition vector and cannot be subjected to dimensionality reduction, we calculate the SNR gains of the other three schemes. Orthogonal tables

$L_{12}(2^9)$ ,  $L_{12}(2^{10})$ , and  $L_{16}(2^{13})$  are chosen according to the corresponding scheme. The calculated SNR gains are shown in Figure 2.



**Figure 2.** SNR gains of MMTS-2, MMTS-3, and MMTS-4.

According to the results, except for the gain of the third eigenvector of MMTS-3, which is less than zero, the SNR gains calculated by other methods are all greater than zero. Therefore, the third eigenvector of MMTS-3 is removed, while all the remaining eigenvectors are reserved. For MMTS-3, after removing the third eigenvector, it is necessary to recalculate the Mahalanobis space, verify the validity of the Mahalanobis space, and recalculate the Mahalanobis distance. Figure 3 shows the test set Mahalanobis distances calculated by the four schemes.



**Figure 3.** Mahalanobis distances of the test set.

Then we classify the samples in the test set according to the Mahalanobis distances, and calculate the precision, specificity, recall, accuracy, and  $F$ -value of the four schemes. The results are summarized in Table 3.

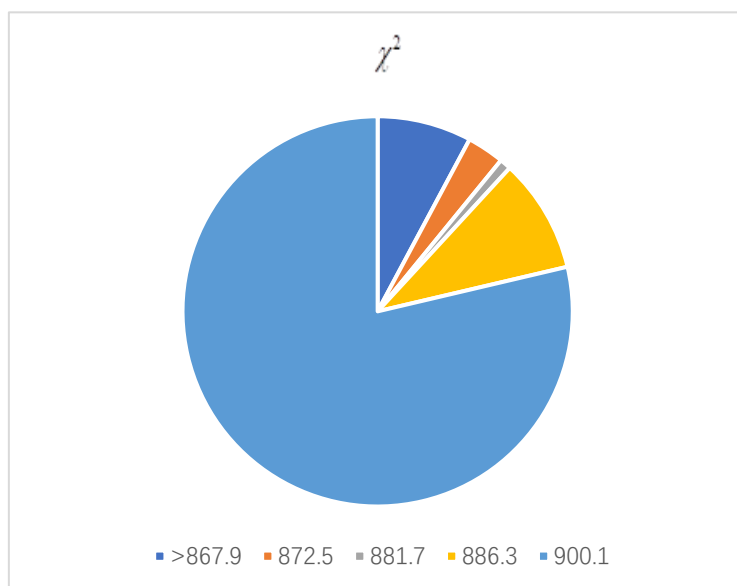
**Table 3.** Summary of classification performance of four schemes.

Number	Method	Number of variables	Precision	Specificity	Recall	Accuracy	F
(a)	MMTS-1	1	28.57%	54.55%	47.62%	52.63%	0.36
(b)	MMTS-2	9	83.33%	95.00%	83.33%	92.31%	0.83
(c)	MMTS-3	9	58.06%	76.36%	85.71%	78.95%	0.69
(d)	MMTS-4	13	44.68%	52.73%	100.00%	65.79%	0.62

Among the four schemes, the optimal orthogonal decomposition vector extraction scheme is MMTS-2, that is, selecting orthogonal decomposition vectors with eigenvalues greater than 0.001. The evaluation metric values of MMTS-2 are  $precision=83.33\%$ ,  $specificity=95.00\%$ ,  $recall=83.33\%$ ,  $accuracy=92.31\%$ , and  $F=0.83$ . In MMTS-1, since only one orthogonal decomposition vector is selected, the component contains relatively less feature information of the original dataset. Therefore, this scheme cannot be used to effectively classify the samples in the test set and the values of its evaluation metrics are relatively low. In MMTS-3 and MMTS-4, orthogonal decomposition vectors with eigenvalues  $> 0.0005$  and with eigenvalue variance accounting for 100% are extracted, respectively, retaining most feature information of the original dataset. However, “overfitting” phenomenon is likely to occur when using these two schemes to process high-dimensional-small-sample-size data, thereby resulting in decreased evaluation metric values and affecting final classification results. MMTS-2 selects orthogonal decomposition vectors with eigenvalues greater than 0.001 and extracts enough feature information from the original dataset without resulting in “overfitting”. Therefore, this scheme exhibits great effectiveness in classification.

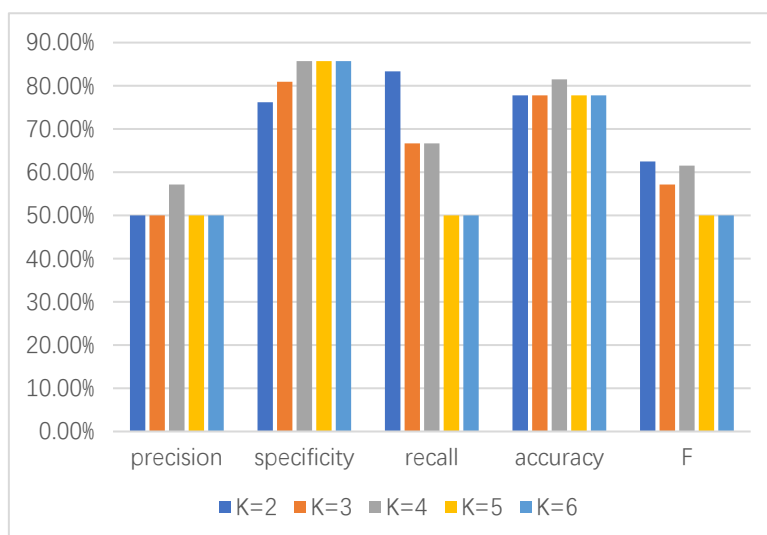
To verify the effectiveness of the proposed method, we compare it with expert classification, individual classifiers such as NB, RF, k-NN, SVM and superimposed classifiers such as Wrapper + RF, MRMR + SVM, Chi-square + BP, SMOTE + Wrapper + RF and SMOTE + MRMR + SVM.

When using Wrapper + RF and MRMR + SVM methods, we select 30 features for classification. When using Chi-square + BP method, we select 332 features based on Chi-square values. Figure 4 shows the Chi-square values distribution.



**Figure 4.** Chi-square values distribution of Chi-square+SVM method.

When using SMOTE + Wrapper + RF and SMOTE + MRMR + SVM methods, in order to explore the best classification performance, we debug the K value in the SMOTE (Select K nearestneighbor when synthesizing samples.),  $K = 2, 3, 4, 5, 6$ . Finally, we set  $K = 2$  to add 20 normal samples. Figure 5 shows classification performance under different K values.



**Figure 5.** Classification performance under different K values.

The classification performance results of NB, RF, k-NN, SVM, Wrapper + RF, MRMR + SVM, Chi-square + BP, SMOTE + Wrapper + RF and SMOTE + MRMR + SVM are summarized in Table 4.

**Table 4.** Summary of classification performance of MMTS-2 and other methods.

Method	Number of variables	Precision	Specificity	Recall	Accuracy	F
MMTS-2	9	83.33%	95.00%	83.33%	92.31%	0.83
NB	415	33.33%	70.00%	50.00%	65.38%	0.40
RF	415	50.00%	75.00%	83.33%	76.92%	0.63
k-NN	415	33.33%	70.00%	50.00%	65.38%	0.40
SVM	415	54.55%	75.00%	100.00%	80.77%	0.71
Wrapper + RF,	30	83.33%	95.00%	83.33%	92.31%	0.83
MRMR + SVM	30	40.00%	70.00%	66.67%	69.23%	0.50
Chi-square + BP	332	60.00%	90.00%	50.00%	80.77%	0.55
SMOTE + Wrapper + RF	30	50.00%	80.00%	66.67%	76.92%	0.57
SMOTE + MRMR + SVM	30	57.14%	85.00%	66.67%	80.77%	0.62

Compared with the methods mentioned above, MMTS-2 shows better dimensionality reduction and classification performance. The accuracy of the MMTS-2 method is higher ( $F = 0.83$ ) as compared to the average accuracy of the other methods. In terms of dimensionality reduction, the number of selected features for the MMTS-2 method is much lower ( $N = 9$ ) as compared to the other methods. Overall, the classification performance of the MMTS-2 method is identical to the classification performance of the Wrapper + RF method, and both are superior to the other methods.

In general, MMTS not only greatly reduces the number of characteristic variables and effectively achieves the goal of dimensionality reduction, but also obtains better results on all the evaluation metrics except for the recall of SVM. Our experiments have proved that MMTS can achieve dimensionality reduction without sacrificing classification accuracy, it still retains a large amount of the most important original feature information.

When used for dimensionality reduction and classification of high-dimensional-small-sample-size data, MMTS can reduce the computational cost while retaining valid variables to effectively achieve the goal of dimensionality reduction. In addition, since this method can enlarge small differences between two types of samples, the differences in Mahalanobis space and Mahalanobis distance between two types of samples will be enlarged. Therefore, the method shows good performance when classifying test set. Taken collectively, our results indicate that the MMTS proposed in this study is suitable for dimensionality reduction and classification of high-dimensional-small-sample-size data.

## 5. Discussions and conclusions

PCA and SPCA are not suitable for processing high-dimensional-small-sample-size data, because they do not consider the impact of data correlation on dimensionality. As one of the promising dimensionality reduction and classification methods in the 21st century, the traditional MTS must consider the covariance

matrix singularity problem when dealing with multi-dimensional-small-sample-size data. In order to solve the problems of traditional dimensionality reduction methods and MTS in dimensionality reduction and classification of high-dimensional-small-sample-size data, this study proposes an MMTS based on the idea of POD.

Our experiments prove that MMTS can achieve the optimal results when extracting decomposition vectors with eigenvalues greater than 0.001. Extracting too much information will lead to the phenomenon of “overfitting”, while extracting too little information can not accurately classify the samples.

In addition, the dimensionality reduction efficiency and classification performance of MMTS is better than those of expert classification, individual classifiers such as NB, RF, k-NN, SVM and superimposed classifiers such as Wrapper + RF, MRMR + SVM, Chi-square + BP, SMOTE + Wrapper + RF and SMOTE + MRMR + SVM.

The classification accuracy of the Wrapper + RF, MRMR + SVM, and Chi-square + BP methods are much higher as compared to the individual classifiers (NB, RF, k-NN, SVM). Our results demonstrate the importance of dimensionality reduction when dealing with high-dimensional-small-sample-size data classification. Dimensionality reduction and feature selection can not only reduce calculations but also improve classification accuracy.

When using SMOTE + Wrapper + RF and SMOTE + MRMR + SVM, in order to obtain better classification performance, we debug the value of  $K$  and set the value to 2 to 6. The classification performance is better when  $k = 2$ , which indicates that when the sample size is small,  $K = 2$  has a better performance in terms of the nearest neighbor selection. Furthermore, we notice that in comparison to the Wrapper + RF method, the classification performance of the SMOTE + Wrapper + RF method is lower. However, in comparison to the MRMR + SVM method, the classification performance of the SMOTE + MRMR + SVM method is higher. Thus, the method of using SMOTE to expand the sample size is not applicable to all classifiers.

In general, compared with NB, RF, k-NN, SVM, Wrapper + RF, MRMR + SVM, Chi-square + BP, SMOTE + Wrapper + RF and SMOTE + MRMR + SVM, MMTS is more suitable for high-dimensional small sample data classification research.

The main contributions of MMTS are: (a) It solves the problems of covariance matrix singularity and eigenvalue estimation difficulty in dimensionality reduction and classification of high-dimensional-small-sample-size data, thereby improving the architecture and algorithm of the traditional MTS. (b) It takes into account the correlation of data, reduces the computational cost, has high robustness, and overcomes the problems of PCA and SPCA in dimensionality reduction. (c) It provides a new idea for processing high-dimensional-small-sample-size data, and achieves quick and effective dimensionality reduction and classification without sacrificing valid sample feature information. In summary, MMTS is of certain reference value, it provides new ideas for academic research on high-dimensional-small-sample-size data dimensionality reduction and classification in real life scenarios.

## Acknowledgments

This work was supported by the China National Social Science Fund Project “Research on the



Implementation Path of Leading Quality Improvement with Standards in the Context of Building a Powerful Manufacturing Country” [grant number 18BJY033]. We would like to thank TopEdit (www.topeditsci.com) for English language editing of this manuscript.

### Conflict of interest

The authors declare there is no conflict of interest.

### References

1. G. Taguchi, S. Chowdhury, Y. Wu, *The Mahalanobis-Taguchi System*, 2001.
2. Z. P. Chang, Y. W. Li, N. Fatima., A theoretical survey on Mahalanobis-Taguchi System, *Measurement*, **136** (2019), 501–510.
3. N. Deepa, K. Ganesan, Mahalanobis Taguchi System based criteria selection tool for agriculture crops, **41** (2016), 1407–1414.
4. B. Buenviaje, J. E. Bischoff, R. A. Roncace, C. J. Willy, Mahalanobis-Taguchi System to identify preindicators of delirium in the ICU, *IEEE J. Biomed. Heal. Informatics*, **20** (2016), 1205–1212.
5. J. Wang, C. Duan, Structural health monitoring using Mahalanobis-Taguchi System, *Proc. 2009 Int. Conf. Inf. Eng. Comput. Sci. ICIECS 2009*, (2009).
6. A. K. Dwivedi, I. Mallawaarachchi, L. A. Alvarado, Analysis of small sample size studies using nonparametric bootstrap test with pooled resampling method, *Stat. Med.*, **36** (2017), 2187–2205.
7. H. F. Gong, Z. S. Chen, Q. X. Zhu, Y. L. He, A monte carlo and PSO based virtual sample generation method for enhancing the energy prediction and energy optimization on small data problem: An empirical study of petrochemical industries, *Appl. Energy*, **197** (2017), 405–415.
8. E. Etchells, M. Ho, K. G. Shojania, Value of small sample sizes in rapid-cycle quality improvement projects, *BMJ Qual. Saf.*, **25** (2016), 202–206.
9. W. Jia, D. Zhao, L. Ding, An optimized RBF neural network algorithm based on partial least squares and genetic algorithm for classification of small sample, *Appl. Soft Comput. J.*, **48** (2016), 373–384.
10. Abdul Lateh, A. K. Muda, Z. I. M. Yusof, N. A. Muda, M. S. Azmi, Handling a small dataset problem in prediction model by employ artificial data generation approach: A review, *J. Phys. Conf. Ser.*, **892** (2017).
11. F. Song, Z. Guo, D. Mei, Feature selection using principal component analysis, *Int. Conf. Syst. Sci.*, **2** (2010).
12. C. Lameiro, P. J. Schreier, A sparse CCA algorithm with application to model-order selection for small sample support, *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. Proc.*, (2017), 4721–4725.
13. Z. Ma, Sparse principal component analysis and iterative thresholding, *Ann. Stat.*, **41** (2013), 772–801.
14. C. M. Feng, Y. L. Gao, J. X. Liu, C. H. Zheng, S. J. Li, D. Wang, A Simple Review of Sparse Principal Components Analysis, *Intell. Comput. Theor. Appl.*, **9772** (2016), 374–383.
15. L. R. Eun, C. Jinwoo, Y. Kyusang, A systematic review on model selection in high-dimensional regression, *J. Korean Stat. Soc.*, **48** (2019), 12.

16. S. Ramírez-Gallego, I. Lastra, D. Martínez-Rego, V. Bolón-Canedo, J. M. Benítez, F. Herrera, et al., Fast-mRMR: Fast minimum redundancy maximum relevance algorithm for high-dimensional big data, *Int. J. Intell. Syst.*, **32** (2017), 134–152.
17. González, J. Ortega, M. Damas, P. Martín-Smith, J.Q. Gan, A new multi-objective wrapper method for feature selection-accuracy and stability analysis for BCI, *Neurocomputing*, **333** (2019), 407–418.
18. X. Xiao, D. Fu, Y. Shi, J. Wen, Optimized Mahalanobis-Taguchi System for high-dimensional small sample data classification, *Comput. Intell. Neurosci.*, **2020** (2020).
19. B. Bayar, N. Bouaynaya, R. Shterenberg, SMURC: High-dimension small-sample multivariate regression with covariance estimation, *IEEE J. Biomed. Heal. Informatics*, **21** (2017), 573–581.
20. K. Pearson, On lines and planes of closest fit to systems of points in space, *Phil. Mag*, **2** (1901).
21. H. Hotelling, Analysis of a complex of statistical variables into principal components, *Educ. Psychol*, **24** (1933), 417–441.
22. G. Kerschen, J. C. Golinval, A. F. Vakakis, L. A. Bergman, The method of proper orthogonal decomposition for dynamical characterization and order reduction of mechanical systems: an overview, *Nonlinear Dyn.*, **41** (2005), 147–169.
23. X. Chen, Research on Several Issues of mahalanobis taguchi System, 2008.
24. L. Cheng, V. Yaghoubi, W. Van Paepegem, M. Kersemans, Mahalanobis classification system (MCS) integrated with binary particle swarm optimization for robust quality classification of complex metallic turbine blades, *Mech. Syst. Signal Process.*, **146** (2021), 107060.
25. Z. Chang, L. Cheng, L. Cui, Interval choquet fuzzy integral multi-attribute decision-making method based on mahalanobis taguchi system, *Control Decis.*, **31** (2016), 180–186.
26. Y. Kikuchi, T. Ishihara, Anomaly detection and prediction of high-tension bolts by using strain of tower shell, *Wind Energy*, (2020), 1–16.
27. Z. Sheng, L. Cheng, Y. Gu, Research on the generation mechanism of mahalanobis space in mahalanobis taguchi system based on control chart[, *Math. Stat. Manag.*, **26** (2017), 1059–1068.
28. Z. Chang, W. Chen, Y. Gu, H. Xu, Mahalanobis-Taguchi System for symbolic interval data based on kernel mahalanobis distance, *IEEE Access*, **8** (2020), 20428–20438.
29. W. Z. A. W. Muhamad, K. R. Jamaludin, S. A. Saad, Z. R. Yahya, S. A. Zakaria, Random binary search algorithm based feature selection in Mahalanobis Taguchi system for breast cancer diagnosis, *AIP Conf. Proc.*, 2018.
30. E. Zhu, X. Wang, A principal component orthogonal decomposition algorithm suitable for processing fingerprint data of traditional Chinese medicine, *J. Xiamen Univ. Natural Sci. Ed.*, **6** (2005), 150–151.
31. K. Lu, Y. Jin, Y. Chen, Y. Yang, L. Hou, Z. Zhang, et al., Review for order reduction based on proper orthogonal decomposition and outlooks of applications in mechanical systems, *Mech. Syst. Signal Process.*, **123** (2019), 264–297.
32. D. Wang, L. He, J. Zhu, Comparison of stock return volatility patterns based on functional adaptive clustering, *Stat. Res.*, **35** (2018), 79–91.
33. V. Penenko, E. Tsvetova, Orthogonal decomposition methods for inclusion of climatic data into environmental studies, *Ecol. Modell.*, **217** (2008), 279–291.

34. M. Ohkubo, Y. Nagata, Anomaly detection in high-dimensional data with the Mahalanobis-Taguchi system, *Total Qual. Manag. Bus. Excell.*, **29** (2018), 1213–1227.
35. R. Valeria, C. Renato, R. Carlo, U. Lorenzo, C. Sirio, V. Francesco, et al., Prediction of tumor grade and nodal status in oropharyngeal and oral cavity squamous-cell carcinoma using a radiomic approach, *Anticancer Res.*, **40** (2020), 271–280.
36. A. Stanzione, C. Ricciardi, R. Cuocolo, V. Romeo, J. Petrone, M. Sarnataro, et al., MRI radiomics for the prediction of fuhrman grade in clear cell renal cell carcinoma: A machine learning exploratory study, *J. Digit. Imaging*, (2020), 1–9.
37. V. Romeo, C. Ricciardi, R. Cuocolo, A. Stanzione, F. Verde, L. Sarno, et al., Machine learning analysis of MRI-derived texture features to predict placenta accreta spectrum in patients with placenta previa Magnetic resonance imaging, **64** (2019), 71–76.
38. P. Mesejo, D. Pizarro, A. Abergel, Computer-aided classification of gastrointestinal lesions in regular colonoscopy, *IEEE Trans. Med. Imaging*, **35** (2016), 2051–2063,.
39. R. L.Babu, S. Vijayan, Wrapper based feature selection in semantic medical information retrieval, *J. Med. Imaging Heal. Informatics*, **6** (2016), 802–805.
40. G. L. Irene, R. V. Esther, Characterization of artifact signals in neck photoplethysmography, *IEEE Trans. Biomed. Eng.*, **67** (2020), 1–1.
41. J. Gardezi, I. Faye, F. Adjed, N. Kamel, M. Hussain, Mammogram classification using chi-square distribution on local binary pattern features, *J. Med. Imaging Heal. Informatics*, **7** (2017) , 30–35.
42. H. A. Khan, W. Jue, M. Mushtaq, M. U. Mushtaq, Brain tumor classification in MRI image using convolutional neural network, *Math. Biosci. Eng.*, **17** (2020), 6203.
43. G. D’Addio, C. Ricciardi, G. Improta, P. Bifulco, M. Cesarelli, Feasibility of Machine Learning in Predicting Features Related to Congenital Nystagmus, In: Henriques J., Neves N., de Carvalho P. (eds XV Mediterranean Conference on Medical and Biological Engineering and Computing MEDICON 2019), IFMBE Proceedings, (2020).
44. M. El-Banna, Modified Mahalanobis Taguchi System for imbalance data classification, *Comput. Intell. Neurosci.*, **2017** (2017).



AIMS Press

©2021 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)