A Modified Mahalanobis Distance Metric Used to Optimize Mahalanobis Space and Improve Classification Performance of MTS

Zhirong Sheng 1,2 and Longsheng Cheng 1

1 School of Economics and Management, Nanjing University of Science and Technology, Nanjing 210094, China
2 Taizhou College, Nanjing Normal University, Taizhou 225300, China

Correspondence should be addressed to Zhirong Sheng; szr7711033@126.com

Received 24 December 2021; Accepted 22 June 2022; Published 7 July 2022

Copyright © 2022 Zhirong Sheng and Longsheng Cheng. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The Mahalanobis–Taguchi system (MTS) is a diagnostic and forecasting technique in a multidimensional system that integrates the Mahalanobis distance (MD) and robust engineering of Taguchi. To implement MTS, a set of observations from a normal group is selected to construct the Mahalanobis space (MS). With this MS as a reference, new observations from an unknown group can be judged to be normal or abnormal, and the degree of abnormality can be determined. MD is very sensitive to data changes, so the data quality of normal samples used to construct the MS directly affects the accuracy of classification. In practical applications, the selection of normal samples depends on the experience and subjective judgment of experts and lacks an objective selection mechanism. In this paper, a modified MD metric is proposed, which is combined with the individual control chart to obtain a robust MS. First, the initial MS is constructed according to the normal samples selected by experts, and the MD of each normal sample is calculated by the initial MS. Then, the MD of each normal sample in the corresponding reduced MS is computed, and the incremental MD is used as the new distance metric to establish the individual control charts. The stability rules of the control chart are employed to eliminate abnormal points, and the MS of the stable state is obtained. To evaluate the effectiveness of the modified MD metric, a numerical simulation experiment is implemented, and the results show that the proposed method is effective and improves the classification performance of MTS. Finally, the improved MTS method is applied to a real medical diagnosis case.

1. Introduction

The Mahalanobis–Taguchi system (MTS), developed by Dr. Genichi Taguchi, is a pattern recognition technology that is mainly used for diagnosis and prediction in multidimensional systems [1, 2]. The first core idea of this method is to use MD as the distance metric, which not only considers the correlation between variables but also eliminates the influence of the dimensions and transforms a multivariate system problem into a unitary problem with MD as the object of analysis and decision. Another core idea is to use two-level orthogonal arrays (OAs) and signal-to-noise ratios (SNRs) for variable selection to achieve the aim of system optimization. The MTS methodology does not have any specific assumptions about the probability distribution of the input variables; it is completely data analytic and can be easy to understand and implement. In many fields, such as fire warning, disease diagnosis, software fault diagnosis, company financial crisis prediction, enterprise lean production evaluation, and personal credit assessment, MTS has been successfully applied [3–8]. Compared with other well-known methods, such as neural networks, support vector machines, and decision tree, MTS shows good performance not only in classification accuracy but also in variable selection efficiency [9–11].

Some shortcomings in MTS have also attracted the attention of scholars. Woodall et al. reviewed the methods and identified some conceptual, operational, and technical issues with MTS [12]. Subsequently, many scholars carried out several improvement studies from the aspects of variable selection and threshold determination.
Rajesh et al. advised that, if other variable selection methods could improve variable selection results better than OAs and SNRs, they could be applied to MTS [13]. The OA and SNR methods used for variable selection are mainly based on economic considerations. Through a few representative variables, good classification results can be obtained under the premise of a constant or slight decline in classification performance. To replace the variable selection method of OAs and SNRs in MTS, various mathematical optimization models have been established. In most of these models, the weighted sum of the error classification and the minimum number of variables to achieve the error classification are regarded as objective functions. Various intelligent optimization algorithms, such as particle swarm optimization (PSO), genetic algorithm (GA), binary ant colony optimization (BACO), binary particle swarm optimization (BPSO), and Gompertz binary particle swarm optimization (GBPPO), are used to solve programming problems [14–19]. Using the optimization algorithm to select variables not only considers the economy of the simplified model but also the classification accuracy, which becomes an important optimization objective.

The determination of a suitable threshold for distinguishing between normal and abnormal observations is another challenge to implementing MTS in practice. Taguchi introduced the threshold method based on a quadratic loss function, which is subjective and often not easy to implement because it is difficult to estimate the relative cost or loss. Su et al. proposed a probabilistic threshold method that uses the Chebyshev theorem to determine the threshold [20]. Das et al. established statistical control limits as thresholds by combining a statistical control chart with MTS based on the chi-squared distribution and beta distribution, and applied them to the fault diagnosis of rolled steel plates [21]. Jin and Chow performed Box-Cox transformation on MD and used three times the standard deviation of the transformed normal distribution to set the threshold and applied it to the motor fault diagnosis of cooling fans [22]. Ramilie et al. compared the performance of the four most common thresholding methods, namely, the Type I-Type II error method, the probabilistic thresholding method, the receiver operating characteristics (ROC) curve, and the Box-Cox transformation method, in 20 different datasets and recommended the use of the Type I-Type II error method due to its lower computational complexity [23].

Obtaining a robust MS is very crucial in MTS. The first step in MTS is to define the “normal” group or “healthy” group to construct the MS according to the selected multivariate variables. Taguchi pointed out that this step requires experts to obtain normal data based on their professional knowledge and historical experience. Due to the sensitivity of MD to changes in the reference group, the quality of normal data used to construct the MS has an important impact on the accuracy of subsequent classification results. When the normal group is chosen to construct the MS and an abnormal group is selected to verify the validity of the MS, the ideal situation where the two types of data can be completely distinguished rarely occurs in practical application, usually overlapping inevitably occurs, which leads to inaccurate application results of MTS. In Taguchi’s method, the MS was constructed by calculating MDs of all samples and then deleting individuals with large MD values and using the remaining data to reconstruct the MS. The process was repeated until a robust MS was obtained. Yang and Cheng used statistical control charts to choose normal samples. They selected normal samples to construct the MS by using the upper and lower specification limits of product attributes and univariate mean control charts with two standard deviations and three standard deviations, respectively, and compared the classification performance of these three types of MS [24]. Wang et al. verified Yang’s method by using rolling bearing failure data and concluded that a better classification result would obtain better classification results [25]. Wang et al.’s method ignored the principle of a control chart; that is, the control of each single variable cannot guarantee overall control in the case of multiple variables. In addition, they regarded the products within the control limit as the normal group and the products outside the control limit as the abnormal group. According to the classified data, they verified the classification performance of MTS and naturally obtained good conclusions, lacking a general conclusion that could be generalized for other datasets. Das et al. proposed an unsupervised clustering algorithm based on MD for small-sample and large-sample datasets to construct the MS [26]. Liparas et al. used two-step cluster analysis to establish MS. Their method can eliminate outliers in the MS, thus improving the accuracy of classification. Liparas et al. used the partitioning around medoid (PAM) clustering method to select normal data from the normal group in the training set and selected the class leading to the minimum classification overlap to construct the MS [27]. In recent reviews on MTS, the selection of normal samples for constructing MS has become a hot topic that needs more attention and research [28].

In this paper, a modified MD metric is introduced. Combined with the stability rules of individual control charts, the robust MS is obtained by eliminating abnormal data. Compared with the existing methods, the advantages of this proposed method in optimizing the MS and improving the classification performance of MTS are studied when normal data and abnormal data overlap.

This paper is organized into five sections. Section 2 summarizes the implementation steps of MTS, the principle of the individual control chart, and gives the MS generation mechanism of MTS, which combines the modified MD metric and the rule for judging stability of individual control chart. A numerical simulation experiment is conducted in Section 3. Section 4 applies the improved MTS method to a practical case of medical diagnosis, and the last section gives the conclusion of the paper.

2. Methods

2.1. Mahalanobis-Taguchi System. The MTS combines MD and Taguchi’s design of experiment. The MD corresponding to the data is measured, and a measurement scale is constructed. Then, OAs and SNRs are used as system optimization tools. MTS has generally developed into the following four stages:
Stage 1: Construct MS as a reference. This stage includes four steps:

(1) Define $p$-dimensional variables in the multivariate system and identify the normal group according to these $p$ variables.

(2) Collect a sample in the normal group.

(3) Standardize the normal sample:

$$z_{ij} = \frac{v_{ij}}{s_j}, \quad i = 1, 2, \ldots, n, \quad j = 1, 2, \ldots, p, \quad (1)$$

where

$$v_j = \frac{1}{n} \sum_{i=1}^{n} v_{ij},$$

$$s_j = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (v_{ij} - \bar{v_j})^2}. \quad (2)$$

It represents the mean and standard deviation of the $jth$ variable, respectively. $v_{ij}$ is the $i$th observation of the $jth$ variable.

(4) Calculate the MD for the normal sample as shown in the following equation:

$$MD_i = \frac{1}{p} Z_i^T S^{-1} Z_i, \quad i = 1, 2, \ldots, n, \quad (4)$$

where $Z_i = [z_{i1}, z_{i2}, \ldots, z_{ip}]^T$ is the $i$th standardized observation and $Z_i^T$ is the transpose of the vector $Z_i$. $S$ is the correlation coefficient matrix that is calculated as follows, and $S^{-1}$ is the inverse matrix of $S$:

$$S = \frac{1}{n-1} \sum_{i=1}^{n} Z_i Z_i^T. \quad (5)$$

The MS is composed of the mean value, standard deviation, correlation coefficient, and the MDs calculated by the normal sample, which is used as a reference frame for the decision-making of MTS.

Stage 2: Verify the validity of the MS, which involves four steps:

(1) Define the abnormal conditions and collect the abnormal sample. In accordance with the principle of MTS, it is assumed that the normal sample and abnormal sample can be clearly distinguished.

(2) The mean and standard deviation of the normal sample calculated in stage 1 are used to standardize the abnormal data.

(3) Calculate the MDs of all abnormal data using (4), where $S$ is still the correlation coefficient matrix of the normal sample.

(4) If the MD of each individual in the abnormal sample is significantly greater than that in the MS, the MS constructed in stage 1 is considered effective. Otherwise, normal samples should be collected again to construct an effective MS.

Stage 3: Optimization of the MS (also known as feature screening or variable selection).

Dr. Taguchi believed that not all original variables contributed to data classification, so OAs and SNRs were used to screen the useful variables.

We design a two-level OA and arrange $p$ variables in the forefront column of the OA; every row represents a run of the experiment. Each variable contains two levels, indicating whether the variable is used in the construction of the MS. Level 1 means the variable should be included to construct the MS, and Level 2 means the variable should not be included to construct the MS. The MS is reconstructed according to the selected variables in each row, and the MDs of abnormal data are calculated and denoted by $MD_k$, $k = 1, 2, \ldots, m$. We calculate the larger-the-better SNRs:

$$SNR_j = -10 \log_{10} \left[ \frac{1}{m} \sum_{k=1}^{m} \frac{1}{MD_k} \right]. \quad (6)$$

After running all experiments, the mean SNRs are calculated under two conditions. For variable $v_j$, $SNR_j$ is used to represent the average SNRs when the variable participates in the construction of the MS, and $SNR_j$ is used to represent average SNRs when the variable does not participate in the construction of the MS. The difference between these two kinds of average SNRs is known as the “gain” and is calculated as shown in the following equation:

$$Gain = SNR_j - \bar{SNR_j}. \quad (7)$$

If the gain is positive, the variable is considered useful and will be kept to construct the MS; otherwise, it will be excluded. The remaining useful variables are used in Stage 4 for diagnosis or prediction.

Stage 4: Diagnosis or prediction future observations.

Classification rules are constructed according to the optimized MS to diagnose and predict future observations. In this process, it is necessary to determine a suitable threshold to balance or reduce the losses caused by the two types of classification errors.

2.2. Variable Control Chart for Individuals. A statistical process control chart is a simple process control system that is used to monitor the abnormal fluctuation of the process to maintain the stability of quality. Assuming that abnormal fluctuations have been eliminated in the process, the quality characteristics of the process should be subject to normal distribution. The control chart constructs the corresponding control boundaries according to the 3-sigma principle of normal distribution. When the observations are placed on the control chart in sequence, they can be used for process control. If abnormal fluctuation occurs, the dot will fall out of the boundaries. Therefore, the essence of the control chart is to distinguish accidental factors and abnormal factors, and the control limits of the control chart are to distinguish the
boundary between normal fluctuation and abnormal fluctuations.

When determining the control limits, the controlled sample is used to estimate the unknown parameters. However, in practice, it may be uncertain whether the sample is completely controlled or not, so the control limits obtained are only the control bounds of the initial attempt. These control limits can be used to check this set of samples; if the sample points all fall in the region within the control limits, this can determine the sample is controlled. If one or more points fall outside the control limits, these sample points should be removed, and the control limits are reestimated with the remaining sample until all sample points fall within the control limits, which is called the statistical controlled state or stable state.

There are many kinds of control charts, and a variable control chart for individuals is used in this paper. Let the sample observations be \( x_1, x_2, \ldots, x_n \), and \( n \geq 25 \) is usually required. The moving range is calculated as follows:

\[
R_i = |x_i - x_{i-1}|, \quad i = 2, 3, \ldots, n.
\]  

Then, we calculate the mean of this sample and the mean of the moving range:

\[
\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i, \quad (9)
\]

\[
\bar{R} = \frac{1}{n-1} \sum_{i=2}^{n} R_i.
\]  

The upper and lower control limits of the variable control chart for individuals are defined as equations (10) and (11), respectively:

\[
UCL = \bar{x} + 2.66\bar{R}, \quad (10)
\]

\[
LCL = \bar{x} - 2.66\bar{R}. \quad (11)
\]

2.3. MS Generation Mechanism Combining a Modified MD Metric with Individual Control Charts. We assume that \( X_i (i = 1, 2, \ldots, n) \) is the initial normal sample selected from a multidimensional system according to the experts’ subjective experience. Due to the ambiguity of judgment, there may be a small number of abnormal items mixed in this sample, especially in the case of overlap between normal data and abnormal data.

The “initial MS” is constructed with the initial normal sample, and the MD in the initial MS is calculated for each individual, denoted as \( MD_{1i} (i = 1, 2, \ldots, n) \). Then, for each individual \( X_i \), its MD is calculated in the corresponding “reduced MS” and denoted as \( MD_{2i} \), where the reduced MS refers to the remaining \( n - 1 \) samples in the initial MS after removing \( X_i \) itself.

If \( X_i \) belongs to the normal group, then the reduced MS after removing itself differs by only one normal sample from the initial MS. The MD calculated in these two MS, \( MD_{1i} \) and \( MD_{2i} \), should not differ much, which can be regarded as normal fluctuations caused by sampling errors. If the sample \( X_i \) in the initial MS does not belong to the normal group, then the reduced MS after removing itself will differ from the initial MS by an abnormal sample. The MD calculated in these two MS, \( MD_{1i} \) and \( MD_{2i} \), should differ greatly, which can be regarded as abnormal fluctuations.

The incremental MD (IMD) for each sample in the initial MS and in the corresponding reduced MS is calculated as follows:

\[
IMD_i = MD_{2i} - MD_{1i}, \quad i = 1, 2, \ldots, n. \quad (12)
\]

If \( IMD_i \) is small, the sample \( X_i \) shall be retained as normal data; if \( IMD_i \) is relatively large, the sample \( X_i \) shall be removed as abnormal data.

In this paper, the variable control chart for individuals is adopted, and the incremental MD, namely, IMD, is taken as the monitoring object. With these IMDs, the control limits of the variable control chart for individuals are calculated, and these IMDs are dotted on the control chart in turn. If one or more points fall outside the control limits, these points will be regarded as abnormal samples and removed. With the retained sample, a new MS is constructed, the IMD of each sample in the new MS is calculated, and new control limits are computed, and these points are dotted on the control chart to determine whether there are any outliers. The above steps are repeated until all sample points fall within the control limits. At this point, these sample points under statistical control are regarded as normal samples and are used to construct the robust MS. The generation mechanism of MS based on the variable control chart for individuals and the modified MD metric is shown in Figure 1.

3. Simulation Study

3.1. Datasets and Research Method. In this section, we perform a simulation study to evaluate the performance of the proposed new MD measurement scale on constructing robust MS. It is also compared with the conventional approach to verify the effectiveness of the method in this paper. Using the R package MixSim, we simulate a 4-dimensional Gaussian mixture model with two components. The two components have a total of 100 observations with an average 0.03 overlap between them, and two outliers in the dataset (see Figure 2).

The first step in applying MTS for classification is to select normal samples to construct the MS. An initial normal class is selected, which includes 45 sets of data from the first component of the Gaussian mixture model, assuming that due to the ambiguity of the judgment, it also includes 5 sets of data located in the overlapping region of the two components, and 2 sets of outlier data (see Figure 3). The purpose of the experiment is to construct a robust MS from this initial normal class to improve the diagnosis of the anomalous data.

3.2. Result Analysis. For each individual in the initial normal class, we first calculate its MD on the initial MS, calculate its MD on their corresponding reduced MS, and then calculate...
the IMD. With the IMD as the monitoring object, the control limits of the individual control chart are calculated, and these 52 incremental values are continuously dotted on the control chart (see Figure 4(a)). Three points fall above the upper control limit, 2 of which are data points originally located in the overlap region and 1 is an anomaly. We remove these 3 points from the initial MS, reconstruct the updated MS with the remaining 49 points, calculate their respective MDs on the two kinds of MS, and re-establish the individual control chart to monitor the IMDs. The results are shown in Figure 4(b), with 2 points falling above the upper control limit, one of which is a point originally located in the overlap region and one is a point originally belonging to the normal class. Repeating the above steps, after 4 iterations, 10 points are finally identified as anomalies in the initial normal class, including 2 original anomalies, 5 points in the overlap region, and 3 original normal points (see Figures 4(c) and 4(d)). Finally, the remaining 42 points are used to recalculate the IMD and construct the individual control chart with all points in a statistically controlled state (see Figure 4(e)). Using these 42 points, a robust MS can be obtained.

For the MS in the steady state, the MD of each normal individual is calculated and the individual control chart is established. The upper control limit of the control chart can be used as the classification threshold of the normal and abnormal classes. In this case, the threshold is 3.03. Therefore, when the MD of a sample to be judged is greater than 3.03, it is classified as abnormal class; when the MD of a sample to be judged is less than 3.03, it is classified as normal class.

3.3. Discussions. The above simulation results show that using the incremental MD in the initial and corresponding reduced MS as the new measurement scale to
Figure 4: The individual control charts for monitoring the IMDs. (a) Three points are judged to be abnormal. (b) Two points are judged to be abnormal. (c) Three points are judged to be abnormal. (d) Two points are judged to be abnormal. (e) The remaining 42 points are in controlled status.
establish the individual control charts, the mechanism of generating MS based on the steady-state rules is effective to eliminate the abnormal data in the initial MS and obtain the robust MS.

In order to compare the effectiveness of the modified measurement scale for outlier checking, the following individual control chart is created directly using the MD as the measurement scale. It has been in a stable state after two iterations, and only 3 points are judged as anomalies, including the original one abnormal point and two points in the overlapping area (see Figure 5). Another abnormal point and the remaining data points in the overlapping area have not been detected.

The reason is that for normal samples in the initial MS, the difference between their MDs on the initial MS and the corresponding reduced MS is small, while for abnormal samples in the initial MS, the difference between their MDs on the initial MS and the corresponding reduced MS is relatively large. This property coincides with the essence of the control chart to distinguish between normal and abnormal fluctuations, a property not possessed by the MD on a single MS. It is also noted that in Figure 4, three samples that originally belonged to the normal class are eliminated as anomalies, because the control chart-based MS generation mechanism starts entirely from the data itself and does not consider the original normal or abnormal property, and in practical applications, it is generally not fully determined in advance whether it is normal or not.

In order to test the robustness of the MS obtained by the method in this paper and the effect on improving the anomaly detection ability of the MTS, 50 sets of data from the second component of the Gaussian mixture model in the simulation dataset are used as test set, and three Mahalanobis spaces are considered:

\[
\text{UCL} = 3.365
\]

\[
\text{UCL} = 3.060
\]

\[
\text{UCL} = 3.030
\]
MS1, the initial MS constructed by 52 sets of data in the initial normal class;

MS2, using the MD as the measurement scale to construct the individual control chart, remove the three runaway points in the control chart, and obtain the MS constructed from 49 sets of steady state data;

MS3, the MS obtained using the method of this paper, that is, the MS consisting of 42 sets of data obtained after removing 10 runaway points.

The thresholds are all determined using the control chart method, and the upper control limits are used as the thresholds to distinguish the normal and abnormal classes. The classification results using different MS are shown in Figure 6.

To compare the robustness of the MS, the minimum covariance determinant (MCD) is used as an evaluation metric, and a robust MS should have a smaller covariance determinant. The initial MS, that is, MS1, is mixed with a certain proportion of class overlap data and anomalous data and therefore has the worst diagnostic effect on anomalous classes. The robustness and anomaly diagnosis effect of the MS2 generated by the control chart using the MD as the measurement scale are not significantly improved. The robustness and anomaly diagnosis of the MS3 obtained by the improved IMD measurement scale has been significantly improved (see Table 1).

4. A Case Study on Vertebral Column Data

Vertebral column dataset consists of measurements of orthopedic patients: normal, disk hernia, and spondylolisthesis. Each patient is represented by six biomechanical attributes derived from the shape and orientation of the pelvis and lumbar spine: pelvic incidence, pelvic tilt, lumbar lordosis angle, sacral slope, pelvic radius, and grade of spondylolisthesis. In this paper, 100 normal data and 150 abnormal (spondylolisthesis) data are selected for a two-class classification study.

Now, we randomly select 60% of the data from the normal category and the abnormal category, respectively, as the training set, where the normal category data in the training set are used to construct the initial MS and the abnormal category data are used to verify the validity of the MS. The remaining 40% of the data in the normal and abnormal categories are used as the test set to verify the classification performance of the MTS.
Table 1: Comparison of the robustness of different Mahalanobis spaces and their effects on the anomaly detection of MTS.

<table>
<thead>
<tr>
<th></th>
<th>Covariance determinant</th>
<th>Misclassification rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>MS1</td>
<td>$2.80 \times 10^{-5}$</td>
<td>0.84</td>
</tr>
<tr>
<td>MS2</td>
<td>$1.06 \times 10^{-5}$</td>
<td>0.38</td>
</tr>
<tr>
<td>MS3</td>
<td>$2.25 \times 10^{-6}$</td>
<td>0.06</td>
</tr>
</tbody>
</table>

Table 2: Confusion matrix for the classification of the training set on the initial MS.

<table>
<thead>
<tr>
<th>True class</th>
<th>Normal</th>
<th>Abnormal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>58</td>
<td>2</td>
</tr>
<tr>
<td>Abnormal</td>
<td>9</td>
<td>81</td>
</tr>
</tbody>
</table>

Figure 7: Continued.
The confusion matrix of the training set classified on the initial MS is shown in Table 2. The results show that the classification of normal classes is good, but the detection of abnormal classes is low, and the MS needs to be optimized. In the following, the initial MS is optimized using the modified MD measurement scale. Using the IMD as the monitoring object and the control chart judging stability rule, after 6 iterations, the 9 anomalies in the initial MS are removed and the updated MS is constructed with the remaining samples (see Figure 7).

Recalculating the confusion matrix of the training set on the optimized robust MS (see Table 3), there was an increase in the classification error rate for the normal class, but a significant improvement in the classification error rate for the abnormal class, after all, improving the diagnostic ability of the abnormal class is in a more concerned position in medical diagnosis, and then, the overall misclassification rate also decreased. The classification results on the training set show that the established MS is valid and can be used for prediction. The confusion matrix for the test set classification is shown in Table 4, and the results show that the detection of anomaly classes is very good.

5. Conclusions

MD is used as a measurement scale in MTS. As MD is very sensitive to the changes in sample data that constitute the MS, the reasonable selection of normal samples to construct the MS is a prerequisite to ensure the classification accuracy of MTS. Aiming at this problem, this paper proposes a generating mechanism of robust MS combining a modified MD metric with the steady rules of the control chart. The modified MD metric is the incremental MD calculated in the initial MS and in the corresponding reduced MS. Then, the incremental MD is used as the new distance metric to establish the individual control charts. The simulation experiment and actual case show the effectiveness of this method in optimizing MS. By comparing the classification performance in different MS, it is shown that the proposed method can improve the classification accuracy and the implementation efficiency of MTS in practical applications.

Data Availability

The simulation data used to support the findings of this study are available from the corresponding author upon request. The vertebral column dataset is derived from UCL public dataset.

Conflicts of Interest

The authors declare that they have no conflicts of interest.
Acknowledgments

This work received support from the Project of the National Natural Science Foundation of China (11601234).

References