

Research Article

A Novel Efficient Feature Dimensionality Reduction Method and Its Application in Engineering

Zhun Cheng and Zhixiong Lu 

College of Engineering, Nanjing Agricultural University, Nanjing 210031, China

Correspondence should be addressed to Zhixiong Lu; luzx@njau.edu.cn

Received 27 February 2018; Revised 27 July 2018; Accepted 27 August 2018; Published 8 October 2018

Academic Editor: Bernhard C. Geiger

Copyright © 2018 Zhun Cheng and Zhixiong Lu. 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.

In the engineering field, excessive data dimensions affect the efficiency of machine learning and analysis of the relationships between data or features. To render feature dimensionality reduction more effective and faster, this paper proposes a new feature dimensionality reduction approach combining a sampling survey method with a heuristic intelligent optimization algorithm. Drawing on feature selection, this method builds a feature-scoring system and a reduced-dimension length-scoring system based on the sampling survey method. According to feature scores and reduced-dimension lengths, the method selects a number of features and reduced-dimension lengths that are ranked in the front with high scores. This feature dimensionality reduction method allows for in-depth optimal selection of features and reduced-dimension lengths with high scores using an improved heuristic intelligent optimization algorithm. To verify the effectiveness of the dimensionality reduction method, this paper applies it to road roughness time-domain estimation based on vehicle dynamic response and gene-selection research in bioengineering. Results in the first case show that the proposed method can improve the accuracy of road roughness time-domain estimation to above 0.99 and reduce measured data of the vehicle dynamic response, reducing the experimental workload significantly. Results in the second case show that the method can select a set of genes quickly and effectively with high disease recognition accuracy.

1. Introduction

The curse of dimensionality is a common problem in engineering. Ineffective and unreasonable feature dimensionality reduction will compromise machine-learning efficiency, pattern recognition accuracy, and data-mining efficiency while increasing the workload of measured data experiments to some extent. A set of high-dimension features possesses the following problems: too many features but few samples, too many features with few or no relations to the mining task, and excessive redundancy among features [1–4].

Feature dimensionality reduction methods can be classified into two types: feature selection and feature projection. Feature projection is also called *feature reparameterization* and transforms the original feature space into one with a smaller dimensionality and more independent dimensions [5–7]. Feature projection generally uses principal component analysis (PCA), kernel principal component analysis, or

other improved principal component analysis methods for linear or nonlinear transformations. Although these methods can realize feature dimensionality reduction to some extent, they require extensive calculations and high computational complexity. These dimensionality reduction methods also cannot reduce necessary experimental data and are inconvenient when identifying features heavily correlated to mining tasks. By contrast, feature selection obtains low-dimension features more effectively compared to machine learning or data mining and carries low computational complexity [8–10]. However, existing feature selection methods are still time-intensive with respect to calculations, and some data dimensionality reduction problems in engineering remain difficult to solve using these methods.

The road roughness time-domain estimation of vehicle dynamic response is taken as an example in this paper, capturing feature dimensionality reduction in multivariate time series. Acquiring road roughness information based

on vehicle dynamic response is an economical and practical method, but most researchers have only improved the precision of road roughness estimation by using different neural networks or enhanced estimation methods [11–14]; they have thus far failed to evaluate the types and quantities of dynamic responses to be assessed and the positions of points to be measured in the vehicle. Measurement points can change positions freely in the vehicle, and each position can offer three dynamic responses (vertical vibration acceleration, speed, and displacement); as such, the example has theoretically infinite feature dimensions. In this case, to improve the accuracy and speed of road roughness estimation based on neural networks, key features must be selected. Additionally, the subsequent experiment needs to test and measure the selected features, which guides the experiment and reduces the workload substantially.

For instance, in gene-selection research on cancer, the mining of gene expression profile data can identify cancer types. The gene expression profile has many data dimensions but small samples. For instance, leukemia has 7129 gene dimensions but only 72 samples for analysis, which increases the challenges of feature dimensionality reduction. Various methods exist for gene selection, but they are computationally costly and complex [15–18]. Many researchers have introduced information to facilitate gene selection, but these methods cannot properly reflect the length of a set of gene features, namely, the number of dimensions after dimensionality reduction. Overall, engineering projects in road roughness time-domain estimation based on vehicle dynamic response and biological gene selection suffer from a curse of dimensionality with an extensive number of features. In addition to their computational cost and complexity, common dimensionality reduction methods (e.g., PCA and cluster analysis) fail to effectively reduce dimensionality (i.e., experimental data) and thus cannot improve estimation accuracy. If researchers use feature selection, uncertainty persists regarding the number of features selected such that the length of the final feature set is unclear. Several variables must be measured to determine which are uncertain.

To solve the problem of excessive dimensions, this paper proposes a new feature selection method based on a sampling survey method and heuristic intelligent optimization algorithm. The general process is as follows: first, we build a feature-scoring system and a reduced-dimension length-scoring system based on the sampling survey method; second, we sort features and reduced-dimension lengths according to their scores and select the features and reduced-dimension lengths that are ranked in the front; third, to apply selection optimization to features and reduced-dimension lengths simultaneously, we redefine the meaning of the information carried by individuals in the heuristic intelligent optimization algorithm population and improve the algorithm accordingly. To better explain and verify the effect of the dimensionality reduction method, this paper conducts feature dimensionality reduction using road roughness time-domain estimation of vehicle dynamic response and gene-selection research in bioengineering. Results show that the new feature dimensionality reduction method proposed can select features quickly and effectively:

the accuracy of road roughness time-domain estimation was generally higher than 0.99 and effectively reduced vehicle dynamic response measurement data, and a reasonable number of genes was selected.

2. New Feature Dimensionality Reduction Method Based on Sampling Survey Method and Improved Heuristic Intelligent Optimization Algorithm

2.1. Scoring System Based on Sampling Survey Method. The sampling survey method takes part of the population as samples for analysis [19]. In a typical dimensionality problem, given a vast number of features, the final number remains unknown after dimensionality reduction; therefore, using enumeration to analyze all samples requires an intensive workload and may fail. By employing the sampling survey method and a certain sample scale, this paper analyzes the features and number of them after dimensionality reduction. All samples must have the same chance of being selected. To obtain useful information for feature dimensionality reduction, we propose a feature-scoring system and a scoring system for reduced-dimension length.

The feature-scoring system operates as follows:

- Step 1 Draw n_1 sets of features with any number of dimensions taken randomly from features with equal probability. Calculate the fitness of each sample based on classification recognition accuracy, data-mining efficiency, or neural network estimation accuracy.
- Step 2 Set a threshold value p and select all samples with fitness larger than p ; these samples comprise the set of valid feature scores.
- Step 3 Score each feature based on the set of valid feature scores. The score of each feature is composed of the following two parts:

$$F_{\text{feature_score}_1} = \frac{\sum_{i=1}^{N_1} (N_1 - R_i) - F_{\text{feature_score}_1_min}}{F_{\text{feature_score}_1_max} - F_{\text{feature_score}_1_min}}, \quad (1)$$

where $F_{\text{feature_score}_1}$ is the first part of the score (Score 1) of a feature, N_1 is the number of sets of valid feature scores, R_i is the fitness ranking by the i^{th} set of valid feature scores, $F_{\text{feature_score}_1_min}$ is the minimum of all Score 1s, and $F_{\text{feature_score}_1_max}$ is the maximum of all Score 1s.

$$F_{\text{feature_score}_2} = \frac{\sum_{i=1}^{N_1} (f_{\text{score}_2}(i)) - F_{\text{feature_score}_2_min}}{F_{\text{feature_score}_2_max} - F_{\text{feature_score}_2_min}}, \quad (2)$$

where $F_{\text{feature_score}_2}$ is the second part of the score (Score 2) of a feature; $f_{\text{score}_2}(i)$ is a piecewise function such that if the i^{th} set of valid feature scores contains the feature, then $f_{\text{score}_2}(i) = 1$ and $f_{\text{score}_2}(i) = 0$ otherwise; $F_{\text{feature_score}_2_min}$ is the minimum of all Score 2s; and $F_{\text{feature_score}_2_max}$ is the maximum of all Score 2s.

The total score of a feature is represented by the following formula:

$$F_{\text{feature_score}} = F_{\text{feature_score}_1} + F_{\text{feature_score}_2}, \quad (3)$$

where $F_{\text{feature_score}}$ is the total feature score.

According to Formulas (1) and (2), $F_{\text{feature_score}_1}$ and $F_{\text{feature_score}_2}$, strictly speaking, each belong to the range of $[0, 1]$. $F_{\text{feature_score}_1}$ reflects the degree of contribution of a feature to fitness. $F_{\text{feature_score}_2}$ reflects the frequency of a feature in a set of valid feature scores. A feature with a higher $F_{\text{feature_score}_1}$ or $F_{\text{feature_score}_2}$ can realize a better fitness value.

Step 4 Rank features according to their total scores and choose the first n_2 features for the set of valid features.

The scoring system for reduced-dimension length is as follows:

Steps 1 and 2 These steps are the same as in the feature-scoring system

Step 3 Score the length of each reduced dimension according to the set of valid feature scores. The score of each dimension length is as follows:

$$F_{\text{dimension_score}} = \sum_{i=1}^{N_2} f_{\text{fitness}}(i), \quad (4)$$

where $F_{\text{dimension_score}}$ is the total reduced-dimension length score, N_2 is the number of sets with the same dimension length in a set of valid feature scores, and f_{fitness} is the fitness of each sample.

Step 4 Rank each dimension length according to $F_{\text{dimension_score}}$, select those with a proportion of cumulative gross scores greater than 80% out of the total score of all dimension lengths, and identify reasonable dimension lengths after dimensionality reduction.

To verify that a scoring system based on the sampling survey method can be used to effectively select features with the greatest influence on results, this paper uses a calculation example for simple mathematical analysis: the classical problem of Hald cement. Hald cement contains four chemical components: $3\text{CaO}\cdot\text{Al}_2\text{O}_3$ (x_1), $3\text{CaO}\cdot\text{SiO}_2$ (x_2), $4\text{CaO}\cdot\text{Al}_2\text{O}_3\cdot\text{Fe}_2\text{O}_3$ (x_3), and $2\text{CaO}\cdot\text{SiO}_2$ (x_4). Table 1 lists data on relations between heat y (calorie) released by each gram of cement and the content of these four components.

In statistics, variance reflects the degree to which two or more data points are different. Then, using the idea of calculating the degree of contribution in PCA, this paper takes the

TABLE 1: Data on relations between content x and heat y of four components of Hald cement.

Sample no.	x_1	x_2	x_3	x_4	y
1	7	26	6	60	78.5
2	1	29	15	52	74.3
3	11	56	8	20	104.3
4	11	31	8	47	87.6
5	7	52	6	33	95.9
6	11	55	9	22	109.2
7	3	71	17	6	102.7
8	1	31	22	44	72.5
9	2	54	18	22	93.1
10	21	47	4	26	115.9
11	1	40	23	34	83.8
12	11	66	9	12	113.3
13	10	68	8	12	109.4

proportion of variance of one variable from the total variance of all variables to represent the degree of influence of that variable:

$$d_i = \frac{\text{Var}(x_i)}{\sum \text{Var}(x_i)}, \quad (5)$$

where d_i is the degree of influence of the i^{th} feature, $\text{Var}(x_i)$ is the variance of the i^{th} feature, and $\sum \text{Var}(x_i)$ is the total variance of all features.

Formula (5) can be used to calculate the degrees of influence of four components of Hald cement (x_1 , x_2 , x_3 , and x_4), which are 0.0579, 0.4050, 0.0686, and 0.4686, respectively. x_2 and x_4 have high degrees of influence, x_3 has less influence, and x_1 has the least influence. To further reflect the degree of influence of these variables on heat y , we employed the radial basis function neural network (RBFNN) and considered several parts of the four components of Hald cement as input features to estimate y , the heat released by each gram of each component. If a feature does not have typical representativeness, then the estimation accuracy of RBFNN will be low. Therefore, we can rank the four features in terms of their influence degree according to the estimation accuracy after RBFNN machine learning. Considering a single feature as input into the neural network and the coefficient of determination as the evaluation index for the degree of closeness of heat y 's predicted value to the actual value, we obtained coefficients of determination for x_1 , x_2 , x_3 , and x_4 as 0.7765, 0.9580, 0.8454, and 0.9495, respectively. The results are similar to the variance analysis; therefore, we can conclude that x_2 and x_4 exert great influences on heat y , x_3 has less influence, and x_1 has the least influence.

According to Step 1 of the scoring system based on the sampling survey method proposed in this paper, we must also calculate the fitness of each sample by considering the accuracy of classification recognition, data-mining efficiency, or neural network estimation accuracy as the fitness to analyze the four features. To better compare their degrees of

TABLE 2: Scores of four components of Hald cement.

x_1	x_2	x_3	x_4
0.0920	1.8962	0.6642	1.1559

influence, we considered the output approximation degree of RBFNN when using the scoring system; Table 2 lists the results.

When ranking the four features by degree of influence in descending order, we obtained x_2 , x_4 , x_3 , and x_1 . The results of dimensionality length show that when the dimensionality is 2, we can estimate heat y using the RBFNN with zero error. In this case, two features input into the neural network must contain either x_2 or x_4 . The above analysis confirms that the proposed scoring system based on the sampling survey method can be used to evaluate the degree of feature importance to select an ideal feature-set length.

2.2. Improving the Heuristic Intelligent Optimization Algorithm to Enhance Algorithm Execution Efficiency. This paper uses an artificial fish swarm algorithm (AFSA) and particle swarm optimization (PSO) to explain and verify the new feature dimensionality reduction method [20–24]. To improve the rate of convergence and precision of the algorithm, we refine AFSA and PSO.

AFSA is improved as follows:

- (1) Add the memory function to all AFSA behaviors and store historical optimal positions and corresponding optimal values of individuals and the entire fish swarm
- (2) Introduce the speed and position calculation formulas of PSO into swarm foraging behavior and update and calculate the movement speed and position of the next-generation fish swarm
- (3) According to the idea of variation in a genetic algorithm, if the global optimum remains the same in multiple records, we shall disturb multiple fish in the swarm with a certain probability

The PSO is improved as follows:

- (1) To avoid the prematurity phenomenon in PSO, introduce the probability judgment Metropolis rule of a simulated annealing (SA) algorithm

$$P = \exp\left(\frac{-\Delta E}{T}\right), \quad (6)$$

where P is the probability of accepting a poor solution, ΔE is the difference in fitness between two generations, and T is the temperature variable in SA, which declines as the number of iterations increases.

- (2) Change the update criterion of temperature variable T to

$$T_i = (T_0 - T_{\text{end}}) \left(\frac{N-i}{N}\right) + T_{\text{end}}, \quad (7)$$

where T_i is the value of T of the i^{th} iteration, T_0 is the initial value of T , T_{end} is the final value of T , and N is the total number of iterations.

- (3) Taking the nonuniform mutation idea of a genetic algorithm as reference, disturb learning parameter C of flight speed in PSO to intensify the local search ability. The following formula calculates learning parameter C :

$$C_{i+1} = C_i \cdot \text{rand} \cdot a_1 \cdot \left(\frac{N-i}{N}\right), \quad (8)$$

where C_{i+1} is the value of C of the $(i+1)^{\text{th}}$ iteration, C_i is the value of C of the i^{th} iteration, and a_1 is a constant, such that $a_1 = 1$ in this paper.

To theoretically confirm the excellent optimum searching performance of the improved heuristic intelligent optimization algorithm, this paper analyzes convergence of the improved PSO and improved AFSA. Solis and Wets [25] studied the stochastic optimization algorithm and provided convergence criteria for stochastic optimization. For a type of stochastic optimization algorithm to converge to the global optimal solution, the following two conditions must be satisfied:

- (1) Condition H1: $f(D(x_k, \xi)) \leq f(x_k)$, and if $\xi \in A$, then $f(D(x_k, \xi)) \leq f(\xi)$. In this condition, A is the feasible solution space, D is the stochastic optimization algorithm, x_k is the result of the k^{th} iteration, $x_{k+1} = D(x_k, \xi)$, and ξ is the solution searched by the algorithm in this iteration.

Condition H1 ensures that the fitness $f(x_k)$ in each iteration of the stochastic optimization algorithm is non-incremental. The stochastic optimization algorithm with convergence ability can get sequence $\{f(x_k)\}_{k=1}^{\infty}$ with a probability of 1, and the sequence converges to a certain value. A stochastic optimization algorithm satisfying condition H1 has convergence ability; although the final convergence value is not always a globally optimal solution, it may be locally optimal.

- (2) Condition H2: For A 's arbitrary Borel subset B , if its measure > 0 , then $\prod_{k=0}^{\infty} (1 - \mu_k(B)) = 0$. $\mu_k(B)$ is the probability measure of the algorithm's k^{th} iteration result in set B .

Condition H2 indicates that if a stochastic optimization algorithm can search and find the globally optimal solution, the algorithm goes through the globally optimal solution at

least once when the number of iterations approaches infinity; that is, the algorithm has ergodicity. The convergence of the improved PSO is confirmed below.

In the PSO, we define each particle by updating its speed and position based on the following formulas:

$$v_{k+1} = c_0 v_k + c_1 (p_{\text{best}} - x_k) + c_2 (g_{\text{best}} - x_k), \quad (9)$$

$$x_{k+1} = x_k + v_{k+1}, \quad (10)$$

where v_{k+1} is the speed of a particle in the $(k+1)^{\text{th}}$ iteration, c_0 is the inertia weight factor, c_1 and c_2 are learning factors, p_{best} is the historical best position of an individual, g_{best} is the historical best position of the swarm, and x_{k+1} is the position of a particle in the $(k+1)^{\text{th}}$ iteration.

The PSO stores the historical optimal position of a swarm, g_{best} , thus ensuring that the optimization result obtained in each step of the algorithm is nonincremental; hence, the PSO satisfies condition H1.

Then, we judge the convergence of particle speed and position according to convergence in probability. With an inertia weight c_0 and the proposed learning factor calculation formula, when the number of iterations approaches infinity, we have $\lim_{k \rightarrow \infty} P\{|v_k - 0| < \varepsilon\} = 1$, where ε is an arbitrarily small positive value. Suppose x^* is a point of convergence in the PSO; then, we have $\lim_{k \rightarrow \infty} P\{|x_k - x^*| < \varepsilon\} = 1$.

If the number of iterations is sufficiently large, then the particle's position and update speed will converge to x^* and 0, respectively, with a probability of 1.

Additionally, we analyze and confirm the convergence of PSO using the difference equation [26]. Although v_k and x_k are multidimensional variables, the dimensions are mutually independent; therefore, we simplify algorithm analysis to be a 1D case. To simplify the calculation, we suppose that p_{best} , the position of optimal solution of a particle, and g_{best} , the position of the optimal solution of the whole swarm, are constant.

Then, according to Formulas (9) and (10), we get

$$v_{k+2} = c_0 v_{k+1} + c_1 (p_{\text{best}} - x_{k+1}) + c_2 (g_{\text{best}} - x_{k+1}), \quad (11)$$

$$x_{k+2} = x_{k+1} + v_{k+2}. \quad (12)$$

Substituting Formulas (9)–(11) into Formula (12), we get

$$x_{k+2} + (-c_0 + c_1 + c_2 - 1)x_{k+1} + c_0 x_k = c_1 p_{\text{best}} + c_2 g_{\text{best}}. \quad (13)$$

It is a second-order constant-coefficient nonhomogeneous difference equation, and this paper uses the characteristic equation method to solve it.

Equation (13)'s characteristic equation is $\lambda^2 + (-c_0 + c_1 + c_2 - 1)\lambda + c_0 = 0$. A quadratic equation with one unknown quantity can be divided into the following three cases: $\Delta = 0$, $\Delta > 0$, and $\Delta < 0$ ($\Delta = (-c_0 + c_1 + c_2 - 1)^2 - 4c_0$).

If $\Delta = 0$, then we have $\lambda_1 = \lambda_2 = -(-c_0 + c_1 + c_2 - 1)/2$. In this case, we obtain the following equation:

$$x_k = \left(x_0 + \frac{(1 - c_1 - c_2)x_0 + c_0 v_0 + c_1 p_{\text{best}} + c_2 g_{\text{best}}}{\lambda_1} k \right) \lambda_1^k, \quad (14)$$

where x_0 is the initial value of the algorithm iteration and v_0 is the initial speed of the iteration.

If $\Delta > 0$, then we have $\lambda_1 = (c_0 - c_1 - c_2 + 1 + \sqrt{\Delta})/2$ and $\lambda_2 = (c_0 - c_1 - c_2 + 1 - \sqrt{\Delta})/2$. In this case, we get the following equations:

$$\begin{aligned} x_k &= \frac{c_1 p_{\text{best}} + c_2 g_{\text{best}}}{c_1 + c_2} + \frac{\lambda_2 b_1 - b_2}{\lambda_2 - \lambda_1} \lambda_1^k + \frac{b_2 - \lambda_1 b_1}{\lambda_2 - \lambda_1} \lambda_2^k, \\ b_1 &= x_0 - \frac{c_1 p_{\text{best}} + c_2 g_{\text{best}}}{c_1 + c_2}, \\ b_2 &= (1 - c_1 - c_2)x_0 + c_0 v_0 + c_1 p_{\text{best}} + c_2 g_{\text{best}} \\ &\quad - \frac{c_1 p_{\text{best}} + c_2 g_{\text{best}}}{c_1 + c_2}. \end{aligned} \quad (15)$$

If $\Delta < 0$, then we have $\lambda_1 = (c_0 - c_1 - c_2 + 1 + i\sqrt{-\Delta})/2$ and $\lambda_2 = (c_0 - c_1 - c_2 + 1 - i\sqrt{-\Delta})/2$. x_k 's calculation formula is the same as the solution in the case of $\Delta > 0$.

If x_k has a limit when k approaches infinity, then x_k requires convergence in all three cases. We get $|\lambda_1| < 1$ and $|\lambda_2| < 1$. Essentially, the PSO's convergence conditions are $c_0 < 1$, $c_1 + c_2 > 0$, and $2c_0 - c_1 - c_2 + 2 > 0$.

We propose an improved PSO to improve learning factors c_1 and c_2 . With an increase in the number of iterations, c_1 and c_2 approach 0. Therefore, the improved PSO surely satisfies the above convergence conditions only if the inertia factor c_0 has a proper value.

We prove the convergence of the improved PSO proposed using the three methods above. However, if the algorithm is required to converge to the global optimal solution, it must satisfy condition H2. In this case, the algorithm goes through the global optimal solution at least once when its number of iterations approaches infinity; that is, the particle in the PSO should demonstrate ergodicity. Apparently, standard PSO cannot guarantee identification of the global optimal value, resulting in the aforementioned prematurity phenomenon. We introduce SA into the evolutionary criteria of PSO because SA has been theoretically proven to converge to the global optimal solution with a probability of 1 in certain conditions [27].

Next, this paper proposes modifying the learning factor to provide the algorithm with a fast flying speed in early iterations and then obtaining a large search scope in the search space to search and find the global optimal solution when the search scope declines as the search proceeds. According to the derivation of reference [28], the standard AFSA is convergent. Proposed improvements introduce the speed and position calculation formulas of PSO into the foraging behaviors of the fish swarm. This way, in addition to clustering and

tailgating behaviors, the entire fish swarm can share information about the historical optimal position of the swarm and the optimal positions of individuals during foraging behavior. Moreover, with the first improvement to AFSA (i.e., storing the historical optimal positions of individuals and the whole fish swarm and corresponding optimal values) and the convergence analysis of PSO, we can ensure that the improved AFSA satisfies the convergence judging condition H1 in the stochastic optimization algorithm.

Because we use the speed and position update formulas of PSO in the fish swarm's foraging behaviors where the fish swarm may also exhibit clustering and tailgating behaviors, all of which move in the direction with higher fitness, we suppose that x^* is the convergence point of AFSA. Then, we have $\lim_{k \rightarrow \infty} P\{|x_k - x^*| < \varepsilon\} = 1$ where ε is an arbitrarily small positive value; therefore, the AFSA possesses convergence.

Similarly, to satisfy condition H2, the position of the fish swarm should have ergodicity in AFSA. Like PSO, the standard AFSA also suffers from the prematurity phenomenon. If a super individual appears in the early iterations of the algorithm, then the whole fish swarm will converge quickly to the value. Therefore, we propose a third AFSA improvement: if the historical optimal solution does not change in multiple iterations, then the algorithm randomly changes some individuals in the fish swarm. When the iteration approaches infinity, the fish swarm position will take all possible values in the definition domain with a theoretic probability of 1. Essentially, the improved AFSA is globally convergent with a certain probability.

To verify the computational complexity of the improved PSO and improved AFSA, we choose the following test function [29]:

$$\max f(x) = x \sin(10\pi x) + 2, \quad (16)$$

where x is an independent variable within the domain of $[-1, 2]$.

For analysis, we used a Lenovo computer with an Intel(R) Core(TM) i5-4590 processor, CPU at 3.30 GHz, and 8 GB memory. The PSO and AFSA used 20 particles over 30 iterations. Figure 1 presents the algorithm test results.

The standard AFSA and standard PSO were trapped in the area of the local optimal value in early iterations and slowly identified the global optimal value as the number of iterations increased, whereas the improved AFSA and improved PSO quickly identified the area of global optimal value in early iterations and finally converged to the global optimal value. The improved PSO and improved AFSA spent only 0.0904 s and 0.0719 s, respectively, converging to the global optimal value, whereas the standard PSO and standard AFSA took 0.0990 s and 0.5321 s, respectively.

2.3. Improving the Heuristic Intelligent Optimization Algorithm Based on the Need for Feature Dimensionality Reduction. To apply the heuristic intelligent optimization algorithm to optimal feature selection, we improved the algorithm based on classification recognition accuracy, data-mining efficiency, or neural network estimation precision as

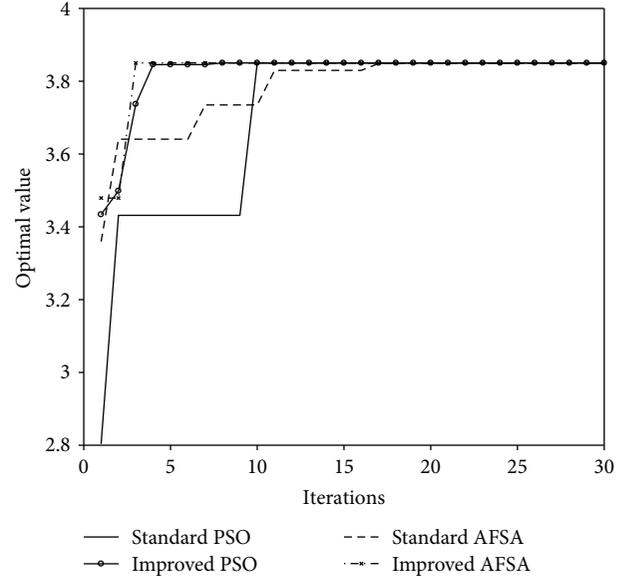


FIGURE 1: Iterative evolution curves of test function.

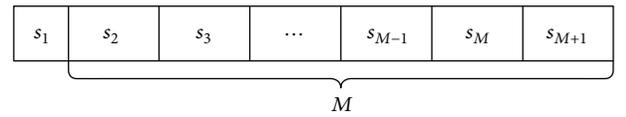


FIGURE 2: Information carried by individuals in the algorithm.

TABLE 3: Selected intervals of reasonable dimension length after dimensionality reduction.

Lower limit of interval	Upper limit of interval	Reasonable dimension length
$b_1 = 0$	c_1	L_1
b_2	c_2	L_2
\dots	\dots	\dots
b_n	$c_n = 1$	L_n

the target value of algorithm optimization. We redefined the individuals of each population in the intelligent optimization algorithm (i.e., redefined information carried by either the fish swarm in AFSA or each particle in PSO). By imitating the transcription and translation process of genetic information in biology, information carried by individuals is defined as shown in Figure 2.

Here, S_1 is the first information of an individual within a value range of $[0, 1]$. M is the total number of elements in the valid feature set. Table 3 lists reasonable dimension lengths after dimensionality reduction.

In this case, L_1 , L_2 , and L_n represent reasonable dimension lengths after the first, second, and n^{th} dimensionality reduction (i.e., the number of reasonable features after dimensionality reduction), respectively. Parameters are $b_1 = 0$, $c_1 = b_2$, $c_2 = b_3$, \dots , $c_{n-1} = b_n$, and $c_n = 1$. $c_1 - b_1$ is

equal to the proportion of the total score of L_i out of the total score of all reasonable dimension lengths. Similarly, $c_n - b_n$ is equal to the proportion of the total score of L_n out of the total score of all reasonable dimension lengths. The following is the calculation formula of $c_i - b_i$:

$$c_i - b_i = \frac{F_{\text{dimension_score_}L_i}}{\sum_{j=1}^n F_{\text{dimension_score_}L_j}}, \quad (17)$$

where c_i and b_i are the upper and lower limits L_i , respectively, of the i^{th} reasonable dimension length; $F_{\text{dimension_score_}L_i}$ is the total score of L_i , and n is the total number of reasonable dimension lengths.

If the value of s_i is within the interval of the i^{th} line in Table 1, then the corresponding reasonable dimension length L_i is selected. Moreover, the heuristic intelligent algorithm will extract information from s_2 to s_{L_i+1} according to the value of L_i , and the information will be used to calculate the target value of the algorithm.

According to continuous iterative calculation of the heuristic intelligent optimization algorithm, the information carried by each individual will evolve in the direction of the optimal target value to obtain the most reasonable dimension length after dimensionality reduction and feature selection.

3. Road Roughness Time-Domain Estimation Based on Proposed Dimensionality Reduction Method

Researchers generally measure contacting or noncontacting road roughness to determine roughness; however, this method returns measurement results slowly and requires expensive equipment. A method combining a dynamic response and neural network can quickly estimate road roughness. Yet little research exists about the types and quantity of dynamic response and the points to be measured in the vehicle. Because measurement points can change positions freely in the vehicle, and each position can offer three dynamic responses (i.e., vertical vibration acceleration, speed, and displacement), this example theoretically has infinite feature dimensions.

3.1. Building a Vehicle Vibration Model with Freely Changing Measurement Points in the Vehicle. Assume the vehicle is symmetric along the longitudinal axis, and the left and right wheels are under the same road conditions while driving. Given these assumptions, we simplify the vehicle model into a half-vehicle model and the vehicle body into a rigid rod. The vibration model in this paper only considers degrees of freedom: the vehicle body's vertical vibration, pitch-angle vibration, vertical vibration of the seat system, vertical vibration of front wheel, and vibration of the rear wheel (Figure 3). We consider the points in the vehicle, of which the distance to the center of mass is e , as the dynamic response measurement points and build the vibration equation based on Lagrange's second equation.

In Figure 3, Z_1, Z_2, Z_3, Z_c , and α denote the vertical vibration displacements (m) of the front wheel, rear wheel, seat

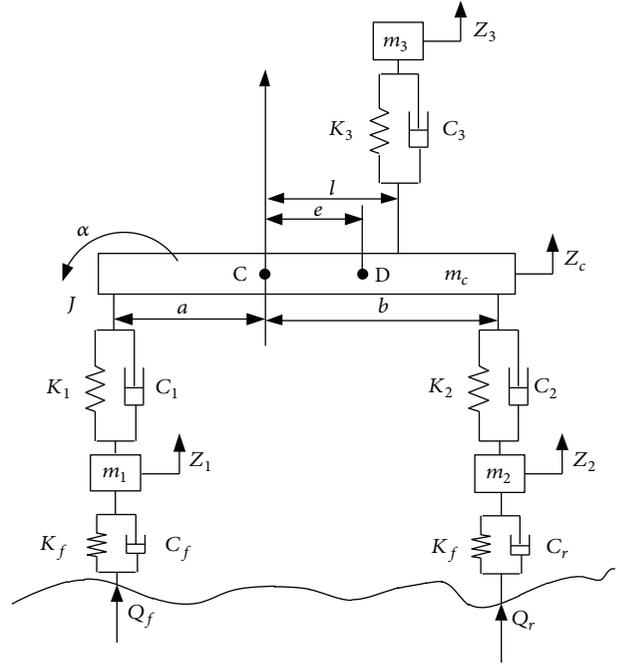


FIGURE 3: Vehicle vibration model with five degrees of freedom.

system, center of mass of the vehicle body, and pitch-angle displacement of the vehicle body (rad), respectively; m_1, m_2, m_3, m_c , and J are the mass of the front wheel, rear wheel, seat system, vehicle body (kg), and rotational inertia of the vehicle body's mass center around the horizontal axis ($\text{kg}\cdot\text{m}^2$), respectively; a, b, l , and e represent the distance from the front axle, back axle, seat system, and measurement point to the center of mass (m), respectively; C_1, C_2, C_3, C_f , and C_r are the damping of the front suspension, rear suspension, seat system, front wheel, and rear wheel ($\text{N}\cdot\text{s}/\text{m}$), respectively; K_1, K_2, K_3, K_f , and K_r are the rigidity of the front suspension, rear suspension, seat system, front wheel, and rear wheel (kN/m), respectively; and Q_f and Q_r are the vertical displacement of road roughness to the front wheel and rear wheel (m), respectively.

The following is Lagrange's second equation:

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_i} \right) - \frac{\partial L}{\partial q_i} + \frac{\partial U}{\partial \dot{q}_i} = Q_i, \quad (18)$$

where q denotes the generalized coordinates; $L = T - V$ represents the difference between the expression of kinetic energy T and potential energy V ; U is the dissipated energy, referring to energy consumed by the damping element; and Q_i is the generalized force.

Because the pitch angle has a smaller vibration, the paper takes $\sin \alpha \approx \alpha$. We assume the vertical vibration displacement of the measurement point to be Z_D and the distances from the measurement point to the front axle and the rear axle to be l_1 and l_2 , respectively.

The kinetic energy of the system is

$$T = \frac{1}{2} m_c (\dot{Z}_D + e\dot{\theta}_D)^2 + \frac{1}{2} J \alpha^2 + \frac{1}{2} m_1 \dot{Z}_1^2 + \frac{1}{2} m_2 \dot{Z}_2^2 + \frac{1}{2} m_3 \dot{Z}_3^2. \quad (19)$$

The potential energy of the system is

$$V = \frac{1}{2} K_1 [(Z_D - l_1 \alpha) - Z_1]^2 + \frac{1}{2} K_2 [(Z_D + l_2 \alpha) - Z_2]^2 + \frac{1}{2} K_3 [(Z_D + l \alpha) - Z_3]^2 + \frac{1}{2} K_f (Z_1 - Q_f)^2 + \frac{1}{2} K_r (Z_2 - Q_r)^2. \quad (20)$$

The dissipated energy of the system is

$$U = \frac{1}{2} C_1 [(\dot{Z}_D - l_1 \dot{\alpha}) - \dot{Z}_1]^2 + \frac{1}{2} C_2 [(\dot{Z}_D + l_2 \dot{\alpha}) - \dot{Z}_2]^2 + \frac{1}{2} C_3 [(\dot{Z}_D + l \dot{\alpha}) - \dot{Z}_3]^2 + \frac{1}{2} C_f (\dot{Z}_1 - \dot{Q}_f) + \frac{1}{2} C_r (\dot{Z}_2 - \dot{Q}_r). \quad (21)$$

Substituting Equations (19)–(21) into Equation (18), we obtain the vehicle vibration differential equation in the example. Figure 4 illustrates the simulation model built in Matlab/Simulink.

According to the vibration simulation model, we can obtain dynamic response data conveniently. The vibration model uses the filtered white noise method to produce the road roughness signal, meeting experiment requirements, and considers a level-B road as the road to be estimated.

3.2. Research on Feature Dimensionality Reduction in Road Roughness Estimation Using the Proposed Method. This example uses the improved AFSA for optimal feature selection. We number the vertical accelerated speed, vertical speed, and vertical displacement of the positions of measurement point, center of the front wheel, center of the rear wheel, and center of the driver's seat as shown in Table 4.

Numbers 2, 7, and 12 are the pitch-angle accelerated speed, pitch-angle speed, and pitch angle of the measurement point, respectively.

The example uses the improved AFSA and assumes information carried by the i^{th} fish to be X_i .

$$X_i = [x_{i1}, x_{i2}, x_{i3}, \dots, x_{i17}], \quad (22)$$

where x_{i1} is the value of the first information, within the range of $[0, 1]$; x_{i2}, \dots, x_{i16} are positive integers within $[1, 15]$, the values of which represent the corresponding dynamic response numbers; and x_{i17} is the distance from the measurement point to the center of mass, e , and $e \in [-1.25, 1.51]$ in the example.

In the iteration process, the improved AFSA finds the corresponding reasonable dimension length L_i from Table 3 according to the value of x_{i1} and selects the first L_i elements from x_{i2}, \dots, x_{i16} accordingly.

The example uses the RBFNN for time-domain estimation of road roughness. We used the improved AFSA and conducted 12 total optimization experiments; results appear in Table 5.

In Table 5, R^2 is the coefficient of determination. The feature dimensionality reduction method proposed can choose features with great influence on the precision of road roughness estimation. The sets of features selected in Experiments 3 and 4 were applied in Experiments 7 and 9; thus, we used the features selected in Experiments 3 and 4 for time-domain estimation of road roughness. Figures 5 and 6 display the estimation results.

The actual value was close to the estimated value. The coefficient of determination exceeded 0.99 in both tests, indicating that the proposed feature dimensionality reduction method can select a feature set with high contributions to road roughness using only four sets of features; thus, we can accurately estimate road roughness using only four dynamic response values. To further confirm this result, we used the random forest method to estimate road roughness based on the results of Experiments 3 and 4. Estimation results are presented in Figures 7 and 8.

The road roughness estimation results obtained with feature sets from Experiments 3 and 4 based on the random forest method were 0.9489 and 0.9708, respectively. Although the estimation accuracy of the random forest method was slightly lower than that of RBFNN, it was still quite high, suggesting that the set of dynamic response features selected with the proposed method can match other estimation data methods perfectly. When the proposed feature selection method was combined with a prediction method, only a few dynamic response parameters were needed to estimate road roughness accurately.

4. Gene-Selection Research in Bioengineering Using the Proposed Dimensionality Reduction Method

The experimental subjects in the example consisted of four open microarray datasets on leukemia, the colon, SRBCT, and brain cancer. Leukemia [30] contained 72 samples, each with 7129 genes. Acute myeloid leukemia (AML) contained 25 samples, and acute lymphoblastic leukemia contained 47. The colon [31] contained 62 samples (40 of tumor tissue and 22 of normal tissue), each with 2000 genes. SRBCT [32] included 83 samples, each with 2308 genes: 29 samples of the Ewing family of tumors, 18 samples of neuroblastoma, 25 samples of rhabdomyosarcoma, and 11 samples of Burkitt lymphomas.

Brain cancer [33] contained 60 samples, each with 7219 genes: 46 samples of patients with classic brain cancer and 14 samples with desmoplastic brain cancer. Although brain cancer only had two classes, the dataset

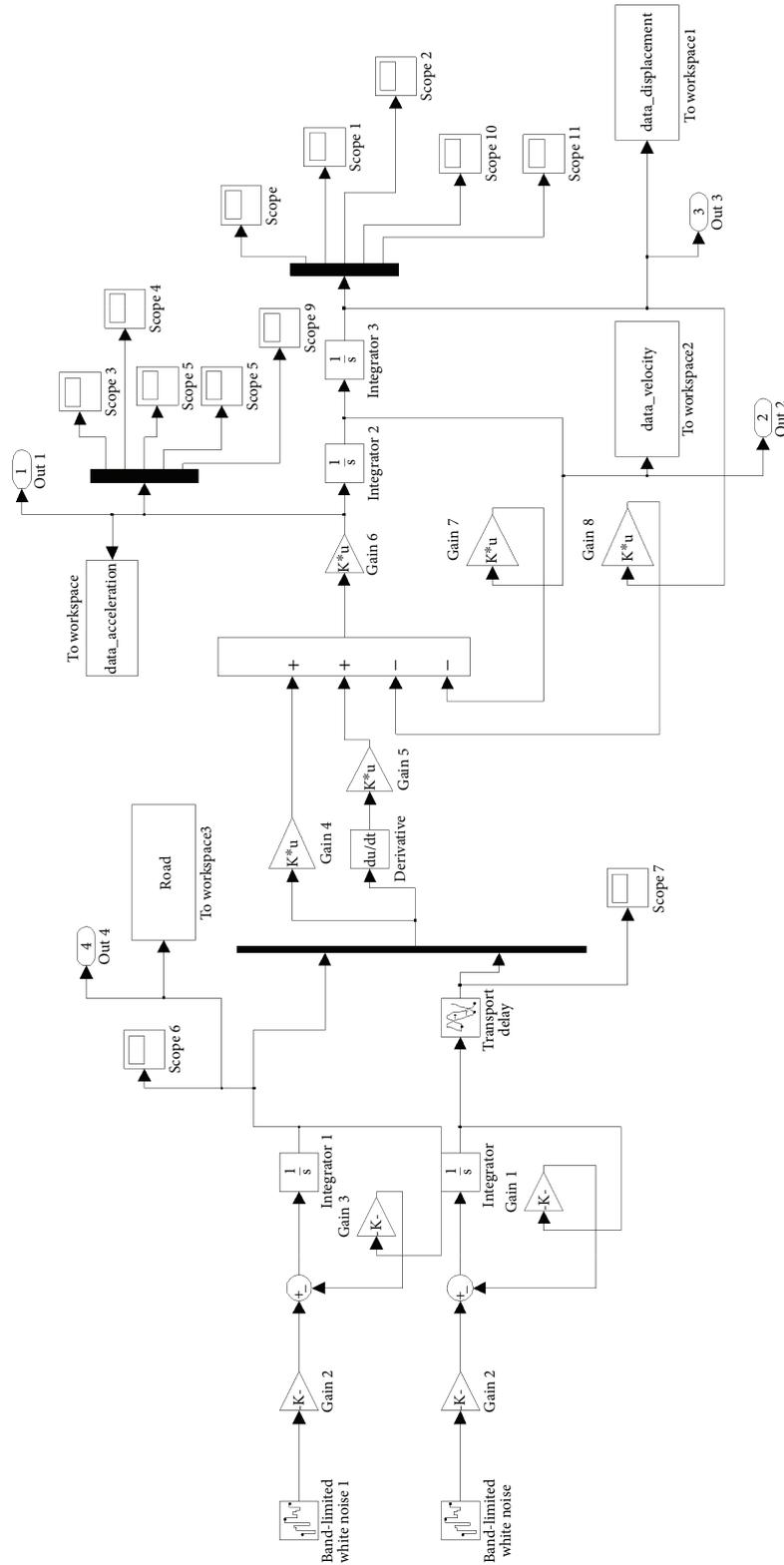


FIGURE 4: Vibration simulation model.

has proven difficult to classify in current research with classification accuracy lower than 85% in much of the literature. Lung cancer [34] included 203 samples, each with

3312 genes. Samples were classified into five subtypes: 139 adenocarcinoma, 20 pulmonary carcinoids, 21 squamous-cell lung carcinomas, six small-cell lung carcinomas, and 17

TABLE 4: Numbers of dynamic responses.

	Measurement point	Front wheel	Rear wheel	Driver's seat
Accelerated speed	1	3	4	5
Speed	6	8	9	10
Displacement	11	13	14	15

TABLE 5: Optimization selection results of dynamic response features.

Test number	Features selected	R^2	Distance e
1	[1, 3, 4, 7, 8, 12, 13]	0.9959	-0.5574
2	[1, 3, 4, 12, 13, 15]	0.9934	0.3557
3	[2, 3, 4, 13]	0.9939	0.9049
4	[1, 4, 8, 9]	0.9948	-1.4350
5	[2, 3, 4, 13]	0.9944	-0.0074
6	[1, 3, 4, 8, 9, 14]	0.9890	-1.2500
7	[2, 3, 4, 13]	0.9944	-0.5321
8	[2, 3, 4, 6, 10, 12]	0.9904	-1.0486
9	[1, 4, 8, 9]	0.9932	-1.4712
10	[1, 3, 4, 9, 11, 13]	0.9823	-0.3755
11	[2, 4, 7, 13, 14]	0.9892	-0.9510
12	[3, 4, 8, 13, 14, 15]	0.9649	0.4839

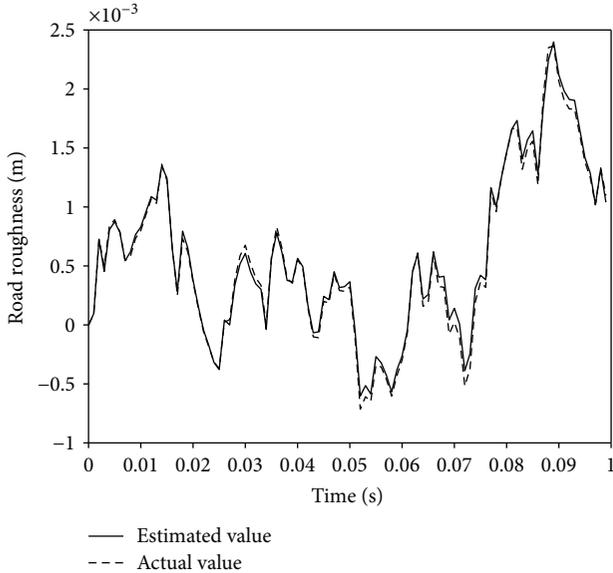


FIGURE 5: Road roughness estimation results with features in Experiment 3.

normal lungs. Lymphoma had 58 samples, each with 7129 genes. The samples included two subtypes: 32 cured patients and 26 not cured.

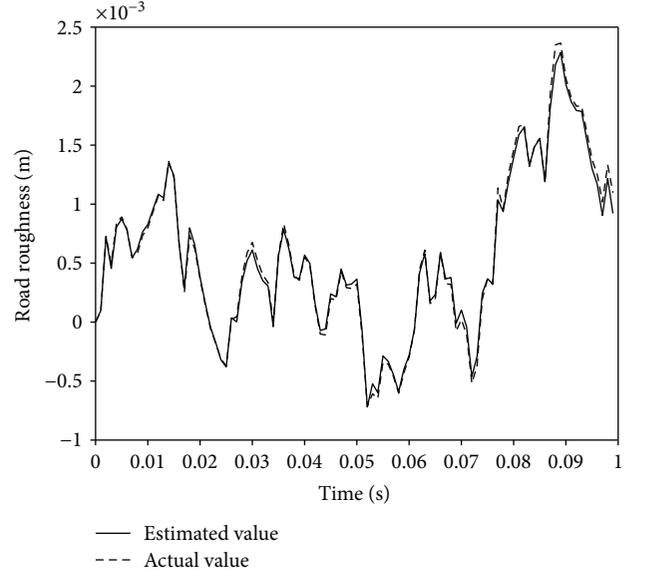


FIGURE 6: Road roughness estimation results with features in Experiment 4.

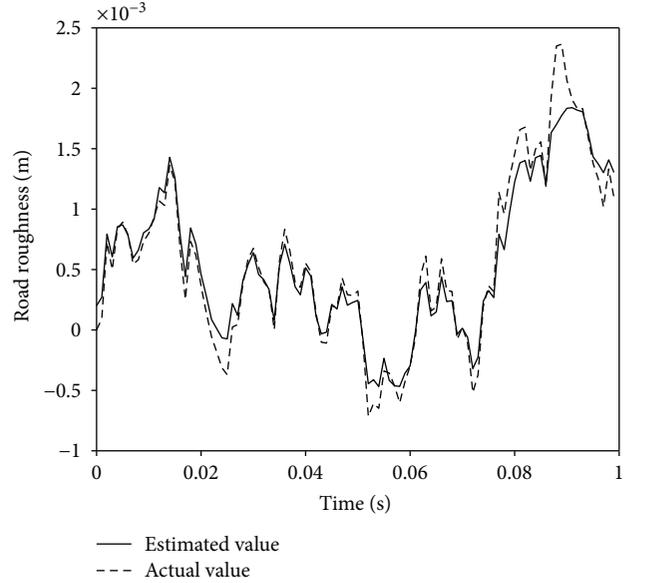


FIGURE 7: Road roughness estimation results based on the random forest method and features in Experiment 3.

Because each type of tumor disease is related to many genes, we roughly extracted genes using the information index to classification (IIC) [34]. This example uses the improved PSO and assumes information carried by the i^{th} particle to be Y_i .

$$Y_i = [y_{i1}, y_{i2}, y_{i3}, \dots, y_{in_1}], \quad (23)$$

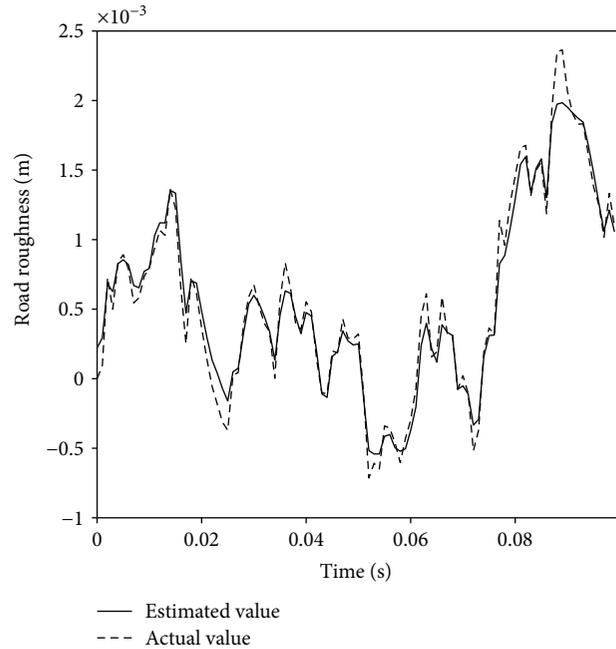


FIGURE 8: Road roughness estimation results based on the random forest method and features in Experiment 4.

TABLE 6: Comparisons between proposed feature dimensionality reduction method and six other methods.

	Mean accuracy (%) + std (%) (number of genes)			
	Colon	Brain cancer	Lung cancer	Lymphoma
BPSO-ELM	93.34 ± 1.99 (9)	85.45 ± 2.33 (7)	94.80 ± 0.72 (11)	83.50 ± 2.72 (8)
K-Means-PSO-ELM	93.94 ± 1.17 (5)	86.55 ± 2.35 (5)	95.67 ± 0.72 (11)	83.72 ± 2.33 (6)
K-Means-BPSO-ELM	93.50 ± 2.02 (9)	87.23 ± 2.34 (8)	95.64 ± 0.56 (12)	85.14 ± 2.87 (6)
BPSO-GCSI-ELM	97.83 ± 1.33 (9)	89.89 ± 2.23 (7)	96.13 ± 0.42 (12)	—
K-Means-GCSI-MBPSO-ELM	97.61 ± 1.37 (6)	88.63 ± 2.16 (3)	97.10 ± 0.63 (11)	86.97 ± 2.44 (8)
Proposed method	99.08 ± 1.23 (13)	93.73 ± 2.64 (4)	98.67 ± 0.43 (11)	93.60 ± 2.33 (7)

where y_{i1} is the value of the first information, within the range of $[0, 1]$; y_{i2}, \dots, y_{in_1} are positive integers within the range of $[1, n_1]$, and their values represent corresponding gene numbers; and n_1 is the number of roughly extracted genes based on IIC.

The example uses the extreme learning machine (ELM) to recognize tumor subtypes and takes the mean and standard deviations of the 5-fold CV accuracy of 100 experiments to verify the accuracy of extracted genes.

To demonstrate the advantage of the proposed method in sample classification accuracy, we chose five methods with high gene-selection accuracy, including binary particle swarm optimization BPSO-ELM, K-means-PSO-ELM, K-means-BPSO-ELM, BPSO-GCSI-ELM, and K-means-GCSI-MBPSO-ELM, to compare with the proposed method. Results of 100 independent repeated experiments were taken from [16]; Table 6 shows the comparison results.

BPSO is a PSO algorithm that solves discrete optimization problems using particles formed by the binary system [35]. BPSO-ELM optimally selects disease-related genes

using BPSO and evaluates the classification results of genes selected each time using ELM [36]. K-Means-PSO-ELM and K-means-BPSO-ELM each complete clustering analysis on all disease-related genes using K-means first followed by optimal selection with standard PSO and BPSO combined with the ELM classifier [37]. GCSI refers to gene-to-class sensitivity information; the BPSO-GCSI-ELM method generates an initial particle swarm using the GCSI of each gene and influences the updating and evolutionary process of particles [38]. Combining the above methods, K-means-GCSI-MBPSO-ELM first completes K-means clustering analysis on all disease-related genes and then introduces the GCSI into the modified BPSO to optimally select genes in combination with the ELM classifier [39].

Because other gene extraction methods have demonstrated good effects on gene extraction of leukemia and SRBCT, this paper does not compare the results of these two types of diseases. Tables 6 and 7 indicate that we can select genes with greater influences on tumor subtype

TABLE 7: Gene sets selected using the proposed feature dimensionality reduction method with classification accuracy.

Dataset	Selected gene sets	Accuracy of test set (%) \pm std	5-fold CV accuracy (%) \pm std
Leukemia	4050, 2642, 2121	100 \pm 0.00	100 \pm 0.00
	4050, 2642, 1882	100 \pm 0.00	100 \pm 0.00
	4050, 2642, 3258	100 \pm 0.00	100 \pm 0.00
	2335, 2642, 1843, 4050	100 \pm 0.00	100 \pm 0.00
Brain cancer	3041, 3052, 3692	94.36 \pm 0.043	92.47 \pm 0.021
	3052, 973, 3041, 3692, 4796	92.00 \pm 0.023	91.78 \pm 0.046
	3041, 3052, 1014, 3692	92.42 \pm 0.042	92.73 \pm 0.026
	7129, 2881, 3052, 865, 1970, 2935, 4871	92.78 \pm 0.012	91.88 \pm 0.019
Colon	493, 792, 739, 513, 280, 1346, 1334, 1954, 1110, 1563, 223, 175, 1668	97.13 \pm 0.028	99.08 \pm 0.012
	493, 792, 377, 1976, 280, 1039, 1334, 698, 1110, 1563, 175, 1668, 1617	98.18 \pm 0.024	98.34 \pm 0.009
	792, 1423, 14, 1976, 1909, 1110, 1589, 102, 107, 1916, 175, 1151, 251	93.73 \pm 0.031	98.71 \pm 0.013
	792, 14, 1976, 765, 1909, 1524, 1110, 175, 43, 53, 1293, 1740, 251	96.86 \pm 0.033	99.05 \pm 0.011
SRBCT	742, 1003, 1954, 430, 2050, 123	100 \pm 0.00	100 \pm 0.00
	545, 1955, 1434, 509, 971, 255	100 \pm 0.00	100 \pm 0.00
	1003, 545, 1911, 153, 123, 1489, 2161	100 \pm 0.00	100 \pm 0.00
	1955, 2050, 545, 2144, 2045, 123, 1489	100 \pm 0.00	100 \pm 0.00
Lung cancer	1765, 2779, 2841, 1474, 2045, 3191, 2763, 2817, 525, 1552, 1630	100 \pm 0.00	98.27 \pm 0.004
	525, 1493, 607, 2763, 792, 580, 867, 368, 3279, 2158, 1225	100 \pm 0.00	98.39 \pm 0.003
	1765, 883, 2763, 792, 580, 867, 985, 3279, 2988, 2045, 814	100 \pm 0.00	98.67 \pm 0.004
	1765, 525, 2763, 2841, 1474, 2583, 867, 985, 2045, 814, 918	100 \pm 0.00	98.67 \pm 0.00
Lymphoma	5279, 4862, 6965, 2374, 1855, 2060	93.45 \pm 0.025	93.39 \pm 0.018
	2828, 2437, 152, 1855, 4998, 2416	97.55 \pm 0.028	92.62 \pm 0.019
	1855, 2828, 152, 2437, 80, 530, 1102	92.30 \pm 0.041	92.36 \pm 0.029
	5279, 4687, 4940, 5449, 1133, 1855, 4519	90.50 \pm 0.031	93.60 \pm 0.023

classification using the proposed feature dimensionality reduction method. Gene sets selected with the proposed method exhibited high classification accuracy, substantially higher than other methods. The standard deviation of classification accuracy of the proposed method was smaller than that of the other methods except brain cancer. For the four diseases, the average accuracy of the proposed method was higher than the highest average accuracy of the other methods by 1.25%, 3.84%, 1.57%, and 6.63%, respectively (from left to right in Table 6). For brain cancer and lymphoma, the average classification accuracy improved to approximately 93%; for colon and lung cancer, the average classification accuracy reached approximately 99%. The classification accuracy was fairly high, greatly improving the practicability of gene-based disease detection.

To compare the ELM classifier with other classifiers based on disease classification results, we selected the random forest classifier and support vector machine (SVM) classifier to identify tumor diseases based on the gene sets in Table 7. We used the average and standard deviation of 5-fold CV accuracy from 100 tests to evaluate the classification results of different classifiers. Table 8 shows the results of the random forest classifier and SVM classifier.

In terms of tumor disease recognition accuracy based on the selected gene sets using the proposed method, the

random forest classifier and SVM classifier were generally poorer than the ELM classifier. For lung cancer, however, the SVM classifier's average accuracy was slightly higher than that of the ELM classifier and its standard deviation was smaller.

5. Conclusions

- (1) This paper proposes a new feature dimensionality reduction method based on a sampling survey method and heuristic intelligent optimization algorithm. First, we build a feature-scoring system and a reduced-dimension length-scoring system based on the sampling survey method and select features ranking in the front with reasonable dimension lengths. Then, we select features according to the improved heuristic intelligent optimization algorithm. Our method has only two steps with simple operation and low computational complexity.
- (2) We successfully apply the proposed feature dimensionality reduction method to road roughness time-domain estimation research. Results show that we only need as few as four dynamic responses to estimate road roughness accurately. The experimental results of 12 feature dimensionality reductions reveal

TABLE 8: Tumor disease recognition results of random forest classifier and SVM classifier.

Dataset	Gene sets selected	Random forest classifier (%) \pm std	SVM classifier (%) \pm std
Leukemia	4050, 2642, 2121	98.06 \pm 0.72	100 \pm 0.00
	4050, 2642, 1882	95.56 \pm 1.10	99.44 \pm 0.72
	4050, 2642, 3258	96.39 \pm 1.34	98.75 \pm 0.79
	2335, 2642, 1843, 4050	96.53 \pm 1.18	97.50 \pm 0.88
Brain cancer	3041, 3052, 3692	82.65 \pm 3.15	65.48 \pm 5.18
	3052, 973, 3041, 3692, 4796	81.75 \pm 3.52	72.75 \pm 3.88
	3041, 3052, 1014, 3692	82.68 \pm 3.28	75.83 \pm 3.76
	7129, 2881, 3052, 865, 1970, 2935, 4871	82.20 \pm 3.42	84.95 \pm 3.28
Colon	493, 792, 739, 513, 280, 1346, 1334, 1954, 1110, 1563, 223, 175, 1668	85.32 \pm 2.99	94.52 \pm 2.18
	493, 792, 377, 1976, 280, 1039, 1334, 698, 1110, 1563, 175, 1668, 1617	79.52 \pm 0.038	90.16 \pm 2.46
	792, 1423, 14, 1976, 1909, 1110, 1589, 102, 107, 1916, 175, 1151, 251	83.55 \pm 5.03	93.55 \pm 2.74
	792, 14, 1976, 765, 1909, 1524, 1110, 175, 43, 53, 1293, 1740, 251	85.65 \pm 3.52	94.35 \pm 2.77
SRBCT	742, 1003, 1954, 430, 2050, 123	94.12 \pm 2.17	99.26 \pm 0.52
	545, 1955, 1434, 509, 971, 255	95.41 \pm 1.78	98.94 \pm 1.04
	1003, 545, 1911, 153, 123, 1489, 2161	94.72 \pm 1.72	97.63 \pm 1.15
	1955, 2050, 545, 2144, 2045, 123, 1489	94.78 \pm 1.64	98.01 \pm 1.10
Lung cancer	1765, 2779, 2841, 1474, 2045, 3191, 2763, 2817, 525, 1552, 1630	94.91 \pm 0.91	98.74 \pm 0.33
	525, 1493, 607, 2763, 792, 580, 867, 368, 3279, 2158, 1225	94.26 \pm 1.12	98.55 \pm 0.48
	1765, 883, 2763, 792, 580, 867, 985, 3279, 2988, 2045, 814	94.83 \pm 1.03	99.16 \pm 0.31
	1765, 525, 2763, 2841, 1474, 2583, 867, 985, 2045, 814, 918	94.52 \pm 1.00	99.18 \pm 0.43
Lymphoma	5279, 4862, 6965, 2374, 1855, 2060	79.22 \pm 3.61	83.71 \pm 2.67
	2828, 2437, 152, 1855, 4998, 2416	73.19 \pm 4.06	78.98 \pm 3.03
	1855, 2828, 152, 2437, 80, 530, 1102	72.76 \pm 4.65	71.57 \pm 3.63
	5279, 4687, 4940, 5449, 1133, 1855, 4519	73.24 \pm 4.50	75.47 \pm 3.56

the mean accuracy of road roughness time-domain estimation to be 0.9897.

- (3) We successfully apply the proposed feature dimensionality reduction method to gene-selection research in bioengineering. Results show that the gene sets selected with the proposed method have high classification accuracy with mean classification accuracy substantially higher than that of other gene-selection methods.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This research was supported by the Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX17_0647) and National Key R&D Program of China (2016YFD0701103).

References

- [1] A. Jalilvand and N. Salim, "Feature unionization: a novel approach for dimension reduction," *Applied Soft Computing*, vol. 52, pp. 1253–1261, 2017.
- [2] H. Itoh, A. Imiya, and T. Sakai, "Dimension reduction and construction of feature space for image pattern recognition," *Journal of Mathematical Imaging and Vision*, vol. 56, no. 1, pp. 1–31, 2016.
- [3] W. Zhao and S. Du, "Spectral-spatial feature extraction for hyperspectral image classification: a dimension reduction and deep learning approach," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 8, pp. 4544–4554, 2016.
- [4] C. Liu, W. Wang, M. Konan et al., "A new validity index of feature subset for evaluating the dimensionality reduction algorithms," *Knowledge-Based Systems*, vol. 121, pp. 83–98, 2017.
- [5] J. Liu, "Feature dimensionality reduction for myoelectric pattern recognition: a comparison study of feature selection and feature projection methods," *Medical Engineering & Physics*, vol. 36, no. 12, pp. 1716–1720, 2014.
- [6] M. Yu, L. Shao, X. Zhen, and X. He, "Local feature discriminant projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 9, pp. 1908–1914, 2016.
- [7] H. X. Vo and L. J. Durllofsky, "Regularized kernel PCA for the efficient parameterization of complex geological models," *Journal of Computational Physics*, vol. 322, pp. 859–881, 2016.

- [8] S. Guo, D. Guo, L. Chen, and Q. Jiang, "A L1-regularized feature selection method for local dimension reduction on microarray data," *Computational Biology and Chemistry*, vol. 67, pp. 92–101, 2017.
- [9] J. Liu, Y. Lin, M. Lin, S. Wu, and J. Zhang, "Feature selection based on quality of information," *Neurocomputing*, vol. 225, pp. 11–22, 2017.
- [10] S. Goswami, A. Chakrabarti, and B. Chakraborty, "An efficient feature selection technique for clustering based on a new measure of feature importance," *Journal of Intelligent & Fuzzy Systems*, vol. 32, no. 6, pp. 3847–3858, 2017.
- [11] H. Imine, Y. Delanne, and N. K. M'Sirdi, "Road profile input estimation in vehicle dynamics simulation," *Vehicle System Dynamics*, vol. 44, no. 4, pp. 285–303, 2006.
- [12] A. González, E. J. O'Brien, Y. Y. Li, and K. Cashell, "The use of vehicle acceleration measurements to estimate road roughness," *Vehicle System Dynamics*, vol. 46, no. 6, pp. 483–499, 2008.
- [13] Z. Q. Gu, Y. F. Zhu, S. Zhang, and X. K. Ma, "Identification of mining road roughness based on GA-BP neural network," *China Mechanical Engineering*, vol. 25, no. 23, pp. 3232–3238, 2014.
- [14] Y. C. Qin, J. F. Guan, L. Gu, Y. Li, and R. Liu, "Road profile input estimation using adaptive-neuro fuzzy inference system," *Transactions of Beijing Institute of Technology*, vol. 35, no. 5, pp. 481–489, 2015.
- [15] W. Sun and F. Han, "Gene selection method based on gene-to-class sensitivity information and binary particle swarm optimization," *Application Research of Computers*, vol. 31, no. 9, pp. 2648–2664, 2014.
- [16] F. Han, W. Sun, and Q. H. Ling, "A novel strategy for gene selection of microarray data based on gene-to-class sensitivity information," *PLoS One*, vol. 9, no. 5, article e97530, 2014.
- [17] P. W. Novianti, V. L. Jong, K. C. B. Roes, and M. J. C. Eijkemans, "Meta-analysis approach as a gene selection method in class prediction: does it improve model performance? A case study in acute myeloid leukemia," *BMC Bioinformatics*, vol. 18, no. 1, p. 210, 2017.
- [18] Y. X. Wang, J. X. Liu, Y. L. Gao, C. H. Zheng, and J. L. Shang, "Differentially expressed genes selection via Laplacian regularized low-rank representation method," *Computational Biology and Chemistry*, vol. 65, pp. 185–192, 2016.
- [19] S. Liu and X. Liu, "A sample survey based linguistic MADM method with prospect theory for online shopping problems," *Group Decision and Negotiation*, vol. 25, no. 4, pp. 749–774, 2016.
- [20] W. Hu and F. Y. Xu, "Self-tuning of PID parameters based on improved particle swarm optimization," *Application Research of Computers*, vol. 29, no. 5, pp. 1791–1794, 2012.
- [21] J. S. Wang, J. C. Wang, and W. Wang, "Self-tuning of PID parameters based on particle swarm optimization," *Control and Decision*, vol. 20, no. 1, pp. 73–81, 2005.
- [22] A. H. Halassi, "An attractor-based multiobjective particle swarm optimization," *International Journal of Applied and Computational Mathematics*, vol. 3, no. 2, pp. 1019–1036, 2017.
- [23] C. Y. Wu, "A new improved artificial fish swarm optimization algorithm," *CAAI Transactions on Intelligent Systems*, vol. 10, no. 3, pp. 1–6, 2015.
- [24] Q. Duan, M. Mao, P. Duan, and B. Hu, "An improved artificial fish swarm algorithm optimized by particle swarm optimization algorithm with extended memory," *Kybernetes*, vol. 45, no. 2, pp. 210–222, 2016.
- [25] F. J. Solis and R. J. B. Wets, "Minimization by random search techniques," *Mathematics of Operations Research*, vol. 6, no. 1, pp. 19–30, 1981.
- [26] S. Gao, K. Tang, X. Jiang, and J. Yang, "Convergence analysis of particle swarm optimization algorithm," *Science Technology and Engineering*, vol. 6, no. 12, pp. 1625–1631, 2006.
- [27] Y. Feng, *Research and Application of Simulated Annealing Algorithm*, Kunming University of Science and Technology, Kunming, China, 2004.
- [28] G. Huang, J. Liu, and Y. Yao, "Global convergence proof of artificial fish swarm algorithm," *Computer Engineering*, vol. 38, no. 2, pp. 204–206, 2012.
- [29] Z. Cheng and Z. Lu, "Research on the PID control of the ESP system of tractor based on improved AFSA and improved SA," *Computers and Electronics in Agriculture*, vol. 148, pp. 142–147, 2018.
- [30] Y. Wang, F. S. Makedon, J. C. Ford, and J. Pearlman, "Hyk-Gene: a hybrid approach for selecting marker genes for phenotype classification using microarray gene expression data," *Bioinformatics*, vol. 21, no. 8, pp. 1530–1537, 2005.
- [31] U. Alon, N. Barkai, D. A. Notterman et al., "Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, no. 12, pp. 6745–6750, 1999.
- [32] J. Khan, J. S. Wei, M. Ringnér et al., "Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks," *Nature Medicine*, vol. 7, no. 6, pp. 673–679, 2001.
- [33] Y. Y. Leung, C. Q. Chang, Y. S. Hung, and P. C. Fung, "Gene selection for brain cancer classification," in *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 5846–5849, New York, NY, USA, September 2006.
- [34] Y. Li and R. Xiaogang, "Feature selection for cancer classification based on support vector machine," *Journal of Computer Research and Development*, vol. 42, no. 10, p. 1796, 2005.
- [35] J. Kennedy and R. C. Eberhart, "A discrete binary version of the particle swarm algorithm," in *1997 IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation*, pp. 4104–4108, Orlando, FL, USA, October 1997.
- [36] D. Tang, F. Han, and Z. Cheng, "Gene selection method based on gene scoring strategy and improved particle swarm optimization," *Computer Engineering and Design*, vol. 39, no. 3, pp. 710–715, 2018.
- [37] S. Yang, F. Han, and J. Guan, "A hybrid gene selection and classification approach for microarray data based on clustering and PSO," *Communications in Computer and Information Science*, vol. 375, pp. 88–93, 2013.
- [38] C. Yang and F. Han, "A gene selection method based on BPSO with prior information," *Software Guide*, vol. 14, no. 7, pp. 36–40, 2015.
- [39] F. Han, C. Yang, Y. Q. Wu et al., "A gene selection method for microarray data based on binary PSO encoding gene-to-class sensitivity information," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 14, no. 1, pp. 85–96, 2017.

