processes

Article

Research on Valve Life Prediction Based on PCA-PSO-LSSVM

Mingjiang Shi [1,*], Peipei Tan [1], Liansheng Qin [2] and Zhiqiang Huang [1]

[1] School of Mechatronic Engineering, Southwest Petroleum University, Chengdu 610500, China
[2] School of Electrical and Information Engineering, Southwest Petroleum University, Chengdu 610500, China
* Correspondence: swpushi@126.com

Abstract: The valve is a key control component in the oil and gas transportation system, which, due to the environment, transmission medium, and other factors, is susceptible to internal leakage, resulting in valve failure. Conventional testing methods cannot judge the service life of valves. Therefore, it is important to carry out valve life prediction research for oil and gas transmission safety. In this work, a valve service life prediction method based on the PCA-PSO-LSSVM algorithm is proposed. The main factors affecting valve service life are obtained by principal component analysis (PCA), the least squares support vector machine (LSSVM) is used to predict the valve service life, the parameters are optimized by using particle swarm optimization (PSO), and the valve service life prediction model is established. The results show that the predicted valve service life based on the PCA-PSO-LSSVM algorithm is closer to the actual value, with an average relative error (MRE) of 16.57% and a root mean square error (RMSE) of 1.2636. Valve life prediction accuracy is improved, which provides scientific and technical support for the maintenance and replacement of valves.

Keywords: ball valve; life prediction; principal component analysis; particle swarm optimization; least squares support vector machine

1. Introduction

As an important part of the oil and gas transportation pipeline [1], valves are widely used in pipeline networks. However, due to improper operation and production processes, defects and other factors can lead to valve spool pitting, surface damage, or overall fracture, resulting in valve damage and failure. Many factors can affect valve life. In this paper, the causes of valve failures in certain pipelines in China were obtained through statistical analysis and are shown in Table 1.

Table 1. Reasons Affecting Valve Life.

Reason		Result
(a)	Problems in the design and manufacturing process	Poor sealing leads to valve leakage or continuous discharge of small flow rate
(b)	Gate plate and sealing surface deformation	
(c)	Damage during manufacture, transportation, inspection, installation and use	
(d)	Solid impurity in medium	
(e)	Containing corrosive substances such as H_2S, Cl^- ions in the medium	

Through the disassembly of more than 28 large-diameter ball valves that failed on multiple gas transmission lines, the morphological characteristics of the internal leakage failure sites of the ball valves were recorded in detail. By analyzing the morphological characteristics of the internal leakage sites, we found that the main manifestation of valve failure is related to ball damage and seat seal damage, with specific damage forms shown in

check for updates

Citation: Shi, M.; Tan, P.; Qin, L.; Huang, Z. Research on Valve Life Prediction Based on PCA-PSO-LSSVM. *Processes* **2023**, *11*, 1396. https://doi.org/10.3390/pr11051396

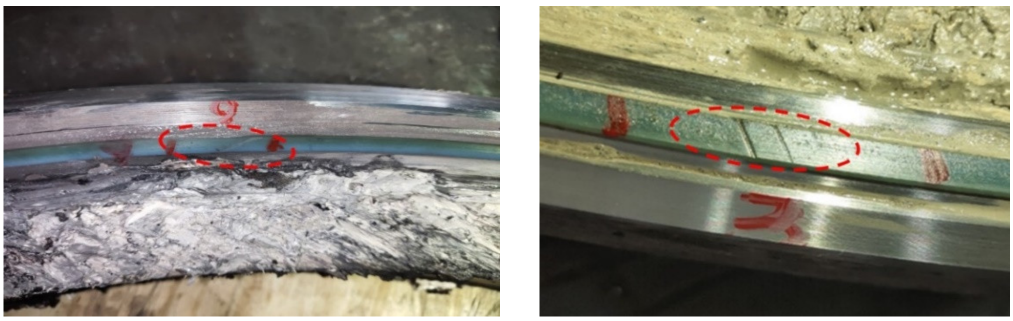
Academic Editor: Qingbang Meng

Received: 29 March 2023
Revised: 20 April 2023
Accepted: 24 April 2023
Published: 5 May 2023

Table 2. Comparing the damage to oil and gas pipeline ball valves at each station, we found that the proportion of valve internal leakage at oil transmission stations is high, accounting for 46.4%, while internal leakage of valve at gas transmission stations is relatively low, accounting for 10.7%. The main form of failure is damage to the seal of the valve seat, with the vast majority of gas pipeline ball valve seal damage resulting from varying degrees of scratching and wear. Seal shear fracture and seal tear-off are the most prevalent types of damage. Examples of different degrees of scoring and wear, shear fracture, and tearing of the seal are shown in Figures 1–3, respectively. Valve failures lead to oil and gas leakage [2], resulting in accidents [3] and causing serious economic losses and even casualties. Valve failure will cause the valve to be ineffective and ultimately unusable. At this point, the service life of the valve is the time to valve failure. Therefore, in order to prevent accidents caused by valve leakage, it is important to carry out valve life prediction research to achieve timely maintenance and replacement of valves and avoid the impacts caused by valve leakage.

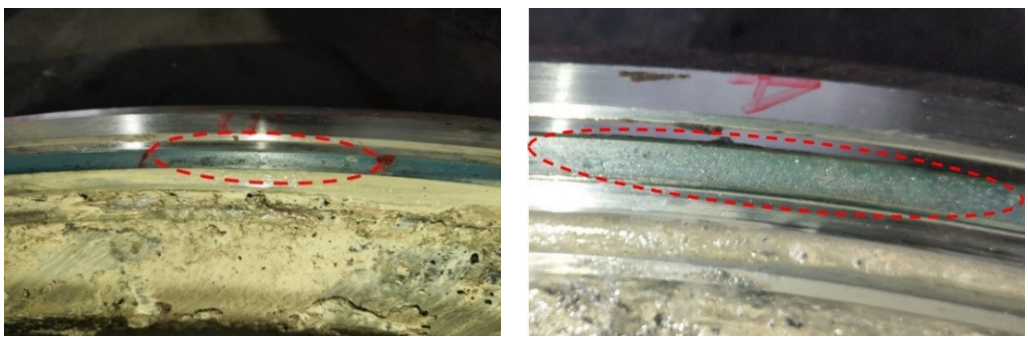
Table 2. Failure forms of internal leakage of ball valves.

Main Failure Mode	The Specific Embodiment of Failure Form
Sphere damage	(a) The ball switch is not in place (b) Severely scratched spherical surface (c) Spherical coating abrasion (d) Foaming of spherical coating
Damaged valve seat seal	(a) Shear fracture of seal ring (b) Scratch of seal ring (c) Ring tearing (d) Sealing ring wear



(**a**) (**b**)

Figure 1. Scratched seal. (**a**) Minor scratches. (**b**) Severe scratches.



(**a**) (**b**)

Figure 2. Worn seals. (**a**) Slight wear. (**b**) Heavy wear.

(a) (b)

Figure 3. Shear fracture and tear off-of sealing ring. (**a**) Shear fracture of the seal. (**b**) Tear off of the sealing ring.

The pipeline environment is complex and variable, and there are many different reasons for valve failure, resulting in different service life timeframes. Thus, the relationship between the service life of the valve and the cause is nonlinear, and it is impossible to directly calculate the service life of the valve with a single mathematical model. Therefore, it is necessary to predict the service life based on a large amount of data and information. Artificial intelligence (AI) is an activity dedicated to making machines intelligent, and intelligence is the quality of a system that is necessary to have foresight and provide appropriate functions in the environment in which it is placed. At present, big-data-driven artificial intelligence technology can obtain highly accurate results by training big data, learning processes, and learning functions. AI can effectively retain complex nonlinear relationships between parameters and efficiently and accurately build predictive models of nonlinear parameters [4,5]. As a classic machine learning method, SVM (Support Vector Machine) can solve nonlinear problems. It is suitable for learning small sample data and is widely used in fault diagnosis and life prediction. The research on Remaining Useful Life (RUL) was used by Nuhic et al. [6] through SVM for prediction. Research on SVM's prediction of battery life cycle was also put forward by Venugopal, Deepak et al. [7]. The results of Nicholas Kwong Howe Su et al. [8] showed that SVM can provide higher RUL accuracy than LSTM and artificial neural networks. De Cooman et al. [9] proposed a transformer fault diagnosis method based on the improved gray wolf optimization algorithm and support vector machine (SVM) to achieve optimization of penalty factors and kernel parameters to reduce the error detection rate. Zhang et al. [10] proposed a fault diagnosis method for high-voltage circuit breakers based on multi-class correlation vector machines. Least squares support vector machine (LSSVM), as one of the SVM methods, has the simplest equation, the fastest solution speed and the highest accuracy, and is widely used to deal with regression analysis [11]. Zhang et al. [12] developed a sparse learning machine based on Least Squares Support Vector Machine (LSSVM). However, the prediction performance of LSSVM is very sensitive to the choice of characteristic parameters. The traditional parameter choice is carried out through an iterative experiment, which depends on the user's experience, and prediction accuracy is greatly influenced by human factors; thus, results are far below the target accuracy. In addition, a large number of measurement data must be recorded to obtain the optimal feature parameters of LSSVM, which further increases the prediction time. Fan et al. [13] proposed a novel deep structure using continuous restricted Boltzmann machine and support vector machine (SVM) and optimized the SVM using particle swarm optimization (PSO) algorithm model parameters. Wang et al. [14] presented the research results of a hybrid fault diagnosis technique, which utilizes and improves the particle swarm optimization (PSO) algorithm to perform further based on qualitative reasoning through knowledge-based initial diagnosis and sample data provided by an online simulation model diagnosis.

Valves are affected by various factors during use that will affect its service time, such as the transport media, sealing materials, and valve connection methods. Therefore, valve failure data collection takes a long time and information is limited. In view of the characteristics of the small samples of available valve data and the multiple factors affecting valve service life, the least squares support vector machine (LSSVM), which is appropriate for small sample data analysis, is used as the basis of the valve life prediction model. Principal component analysis (PCA) is used to reduce the dimension of feature variables, and then particle swarm optimization (PSO) is used to optimize the parameters of the LSSVM algorithm to improve prediction accuracy and establish the PCA-PSO-LSSVM valve life prediction model.

2. Method

2.1. Principle of Principal Component Analysis

There are seven factors that influence valve life prediction: transportation medium, pipeline, functional position, valve type, sealing materials, connection methods, and leakage classification. There is some information overlap between these factors, which affects the accuracy of the prediction model. Therefore, the characteristic variables affecting valve life are first processed using principal component analysis to reduce the dimensionality of the input variables and serve to improve the calculation [15].

Principal component analysis (PCA) is a multivariate statistical method that is used to study multiple correlated variables. By making full use of the original observation information when studying complex system problems and simplifying multiple related variables into unrelated principal component variables through data dimensionality reduction operations, the information redundancy problem can be effectively solved and algorithm efficiency improved. The main processes include data standardization, determination of principal components, calculation of variance contribution of principal components and cumulative variance contribution, and selection of principal components.

(1) Data standardization: In a specific complex system, n sample data affected by m variables $\begin{bmatrix} X_1 & X_2 & \cdots & X_m \end{bmatrix}$ form the original observation information matrix:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} = \begin{bmatrix} X_1 & X_2 & \cdots & X_m \end{bmatrix} \quad (1)$$

where m is the input sample dimension of valve life and n is the valve life sample dimension.

Considering the differences in the criteria used among variables, those with large variance will cause greater interference in the principal component analysis; thus, the original observations need to be standardized. The calculation formula is as follows:

$$x_{ij}^* = \frac{(x_{ij} - \overline{x_j})}{s_j} \quad (2)$$

where x_{ij} is the value of the valve service life input variable x_j in the ith sample, $\overline{x_j} = \sum\limits_{i=1}^{n} x_{ij}/n$ is the sample average value of the valve service life input variable x_j, $s_j = \sqrt{\frac{1}{n-1}\sum\limits_{i=1}^{n} (x_{ij} - \overline{x}_j)^2}$ is the valve service life input sample standard deviation, and x_{ij}^* is the standardized value of x_{ij} to form a standardized data matrix $X_{n\times m}^*$.

(2) Principal component determination: the standardized observation information matrix X^* contains all the original observation values for valve service life, and the linear combination of the input variables can obtain m uncorrelated new variables, that is, m principal components, using the formula:

$$\left.\begin{array}{l} y_1 = c_{11}X_1 + c_{12}X_2 + \cdots + c_{1m}X_m \\ y_2 = c_{21}X_1 + c_{22}X_2 + \cdots + c_{2m}X_m \\ \vdots \\ y_m = c_{m1}X_1 + c_{m2}X_2 + \cdots + c_{mm}X_m \end{array}\right\} \quad (3)$$

where $y_1, y_2, y_3, \cdots, y_m$ are the $1st, 2nd, \cdots, mth$ principal components of the original variables $X_1, X_2, \cdots, X_m$ respectively, $c_{ij}(i = 1, 2, \cdots m; j = 1, 2, \cdots m)$ are the original valve service life input variables, $X_i(i = 1, 2, \cdots m)$ is located in the principal component $y_i(i = 1, 2, \cdots m)$ and is also the eigenvector corresponding to each eigenvalue of the correlation coefficient matrix, which satisfies $c_{i1}^2 + c_{i2}^2 + \cdots + c_{im}^2 = 1$.

(3) Principal component variance contribution and cumulative variance contribution are calculated: the characteristic root of the correlation coefficient matrix is equal to the variance of the corresponding principal component, which reflects the proportion of the information contained in the corresponding principal component in the original observed information. In this paper, the principal component variance contribution and cumulative variance contribution are calculated based on the magnitude of the characteristic roots, as shown in Equations (4) and (5):

$$q_j = \lambda_j / \sum_{j-1}^{m} \lambda_j \quad (4)$$

$$Q_p = \sum_{j=1}^{p} q_j \quad (5)$$

where $\lambda_i(i = 1, 2, \cdots, m)$ is the characteristic root of the correlation coefficient matrix.

From Equation (3), it can be seen that the expressiveness of the combined information of the valve lifetime input variables is proportional to the variance contribution rate and, combined with Equation (4), if the cumulative variance contribution rate of the first i principal components is larger, the more information the first i principal components contain about the original observation.

(4) In this paper, we make full use of the original observation information by principal component analysis and solve the network parameter redundancy problem by reducing the data dimensionality. Therefore, it is necessary to extract the principal components while ensuring the cumulative variance contribution. Through principal component analysis, the initial data set can be filtered out to cover most of the information, which ensures no data loss while also reducing the complexity of model training and achieving the purpose of reducing computational resources. Generally, no more than five or six principal components are selected, and the cumulative variance contribution is not less than 80%.

2.2. Least Squares Support Vector Machine

Least Squares Support Vector Machine (LSSVM) is a deformation algorithm based on SVM proposed by Suykens et al. [16]. LSSVM deforms the optimization formulation of the objective function using two paradigms and transforms the inequality constraint in SVM into an equation constraint. Therefore, LSSVM turns the original quadratic programming problem into a linear system of equations, which not only simplifies the computational complexity but also reduces computational time. The basic principle is as follows [17–21]. To perform nonlinear regression on the training sample $\{x_i, y_i\}$, $i = 1, 2, \cdots, n, x_i \in R^n$, a nonlinear mapping function $\varphi(x)$ is introduced to map the training sample to a high-dimensional feature space for linear regression. The LSSVM model in the feature space can be expressed as:

$$y(x) = w^T \varphi(x) + b \quad (6)$$

where w is the weight vector and b is the bias. Its objective function is:

$$\min J(w, \xi) = \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{i=1}^{n} \xi_i^2 \quad (7)$$

where ξ is the training set prediction error variable and $\lambda > 0$ is the normalization parameter. The constraints to be satisfied are:

$$y_i = w^T \varphi(x_i) + b + \xi_i,\ i = 1, 2, \cdots, n \tag{8}$$

Then, the Lagrange multiplier is introduced to convert the data into the Lagrange function:

$$L(w, \xi, \alpha, b) = J(w, \xi) - \sum_{i=1}^{n} \alpha_i [w^T \varphi(x_i) + b + \xi_i - y_i] \tag{9}$$

According to optimization conditions:

$$\frac{\partial L}{\partial \mathrm{w}} = 0, \frac{\partial L}{\partial \xi_i} = 0, \frac{\partial L}{\partial \alpha_i} = 0, \frac{\partial L}{\partial b} = 0$$

the following system of linear equations can be obtained:

$$\begin{bmatrix} 0 & 1 & 1 & \cdots & 1 \\ 1 & K(x_1, x_1) + \gamma^{-1} & K(x_1, x_2) & \cdots & K(x_1, x_n) \\ 1 & K(x_2, x_1) & K(x_2, x_2) + \gamma^{-1} & \cdots & K(x_2, x_n) \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & K(x_n, x_1) & K(x_n, x_2) & \cdots & K(x_n, x_n) + \gamma^{-1} \end{bmatrix} \begin{bmatrix} b \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} 0 \\ y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \tag{10}$$

where $K(x, x_i) = \varphi(x)^T \varphi(x_i)$ is the kernel function, which is the inner product of the high-dimensional feature space. The regression function of LSSVM can be obtained as:

$$y(x) = \sum_{i=1}^{n} \alpha_i K(x, x_1) + b \tag{11}$$

The radial basis function $K(x, x_i) = \exp(-\|x - x_i\|^2 / (2\sigma^2))$ is selected as the kernel function of LSSVM. For the LSSVM model of the RBF kernel function, the only parameters to be determined are the kernel parameter σ and the normalization parameter γ.

2.3. Particle Swarm Optimization Algorithm

The particle swarm optimization (PSO) algorithm simulates birds in a flock by designing a massless particle [22], so the particle only has two properties: speed v and position u, where velocity represents the speed of movement and position represents the direction of movement. The particle swarm updates the above two properties by tracking two extreme values in the search process: the first extreme value is called the individual extreme value p_{best}, which is the optimal solution searched by the individual particle, and the other extreme value is called the global extreme value g_{best}, which is the optimal solution found by the entire population so far. The formulas for updating speed and position are:

$$v_i = \omega v_i + c_1 r_1 (p_{best} - x_i) + c_2 r_2 (g_{best} - x_i) \tag{12}$$

$$u_i = u_i + v_i \tag{13}$$

where v_i is the velocity of the ith particle, u_i is the current position of the ith particle, c_1, c_2 are the acceleration constants, where c_1 represents the weight of the particle and its own historical optimal value, so that its local search ability is expanded, c_2 represents the weight of the optimal value of the particle tracking group, which represents the process of assistance and information sharing among individuals of the group, reflecting the global search capability of the particle swarm. r_1 and r_2 are random numbers between [0, 1]. ω is the inertia weight, which is used to maintain the original velocity and plays a great role in the convergence of the PSO algorithm. The larger the value of ω, the larger the particle leap, and the easier it is to miss the local search ability, whereas the stronger the global search ability. Conversely, the stronger the local search ability, the weaker the global search ability. Therefore, the inertia factor is set larger at the beginning of the iteration and then gradually decreases during the iteration. The decay formula is:

$$\omega = \omega_{\max} - \frac{n(\omega_{\max} - \omega_{\min})}{n_{\max}} \tag{14}$$

where n is the current number of iterations and $n_{\max}$ is the total number of iterations. The fitness function f of the PSO algorithm is expressed as the mean square error of the particle swarm, as:

$$f(u) = \frac{1}{n}\sum_{i=1}^{n}(y_u(x_i) - y_i)^2 \tag{15}$$

2.4. PCA-PSO-LSSVM Model Construction

Although the PSO algorithm can optimize the relevant parameters of LSSVM to find the global optimal solution, when there are many network input parameters and there is a strong correlation between the parameters and if the PSO-LSSVM model is directly used for prediction, it will lead to network input that is too complex and over-fitting. This will result in the possible introduction of model interference and lead to low model accuracy. The PCA algorithm can effectively deal with the correlation between variables. On the premise of retaining enough original observation information, the main factors can be extracted and the data dimension can be reduced, thereby reducing network input and achieving the purpose of simplifying the network. The LSSVM model in this paper adopts the radial basis kernel function. The parameters to be determined are the kernel parameter σ and the normalization parameter γ, so it needs to be optimized. The PCA algorithm is used to reduce the dimensionality of the original sample data, and the processed data is used as the input of the LSSVM model. The PSO algorithm is then used to optimize the parameters of the LSSVM model, which improves the analysis of the LSSVM model. The specific steps are shown in Figure 4.

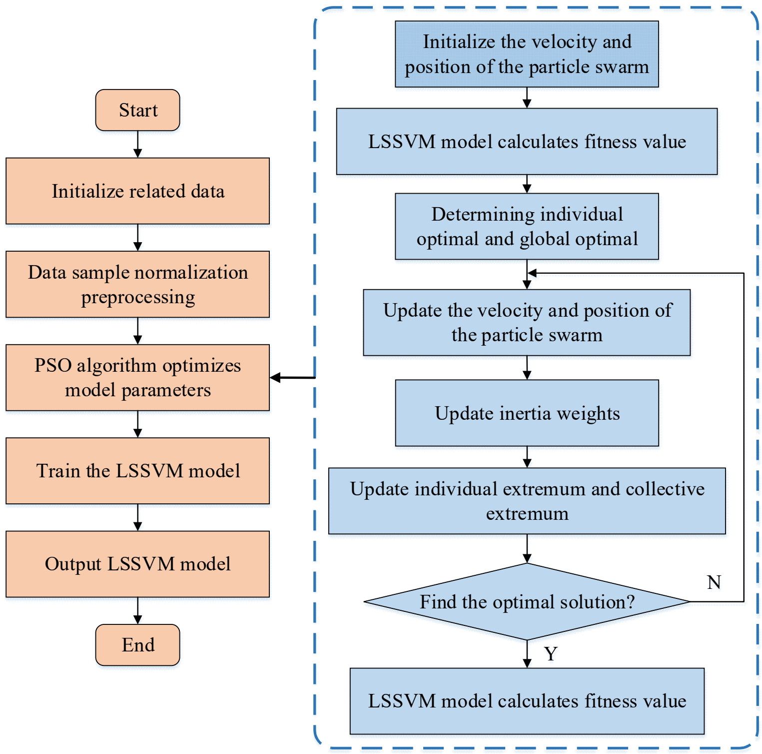


Figure 4. PCA-PSO-LSSVM flow diagram.

Step 1: Set up the training sample set and test a sample set of the model. After pre-processing the factors affecting valve service life, principal component analysis is performed, and the extracted main factors are used as inputs to the valve service life prediction model.

Step 2: The initial parameters of the PSO and LSSVM algorithms are set, and the particle swarm $u_i = [\sigma_i, \gamma_i], i = 1, 2, \cdots, M$ with initialized M LSSVM model parameters is obtained randomly. The initial velocity $v_i = [v_{i1}, v_{i2}], i = 1, 2, \cdots, M$ of the particle swarm is obtained.

Step 3: The training samples are trained using LSSVM, and the fitness function is the mean squared error of prediction for each particle. The initial fitness value is used as the current optimal fitness value for each particle, and the current position is recorded as the individual optimal position. The best initial fitness value is taken as the current global optimal fitness value, and the position of the current best initial fitness value is recorded as the global optimal position.

Step 4: The velocity and position of the particle swarm are updated within a limited range of velocity and position, and the adaptation value is calculated based on the currently updated position.

Step 5: Update the individual optimum. Compare the current fitness value with the particle optimal fitness value, and, if the current fitness value is better, the current position of the particle is taken as its individual optimal position and the global optimum is updated.

Step 6: If the predefined maximum number of iterations is reached or the predefined precision is reached, and if it is satisfied, then the search is finished; if it is not satisfied, then repeat steps 4 and 5 to continue the search.

Step 7: The PCA-PSO-LSSVM model is built using the parameters corresponding to the optimal position.

3. Data Forecasting and Discussion

3.1. Test Data

The data set used in this paper includes many pipelines with valve diameters of 4~40 inches, and the usage statistics of pipeline valves under different conditions (such as transporting natural gas, refined oil and crude oil on trunk lines or bypass lines). The statistical results are shown in Figure 5. The factors related to valve life, such as transmission medium and functional pipeline, are taken as the input data set of the prediction model, and the actual service life is taken as the output. The data set is divided into training samples and test samples of the valve life prediction model to verify the feasibility of the method proposed in this paper.

Two types of valves were investigated, of which ball valves accounted for 94% and gate valves accounted for 6%. According to the statistics of valve usage, seven variables such as conveying medium, pipeline, and valve functional position, are regarded as the influencing factors of valve service life. These variables include 3 types of conveying media, 13 pipelines, 2 types of functional positions, 4 types of sealing materials, 2 types of valves, 2 types of valve connections, and 2 levels of leakage classification. With the factors influencing valve service life as the input of the model, the use time of the valve as the output of the model, and month as the measurement unit of the use time, the valve service life prediction model is established.

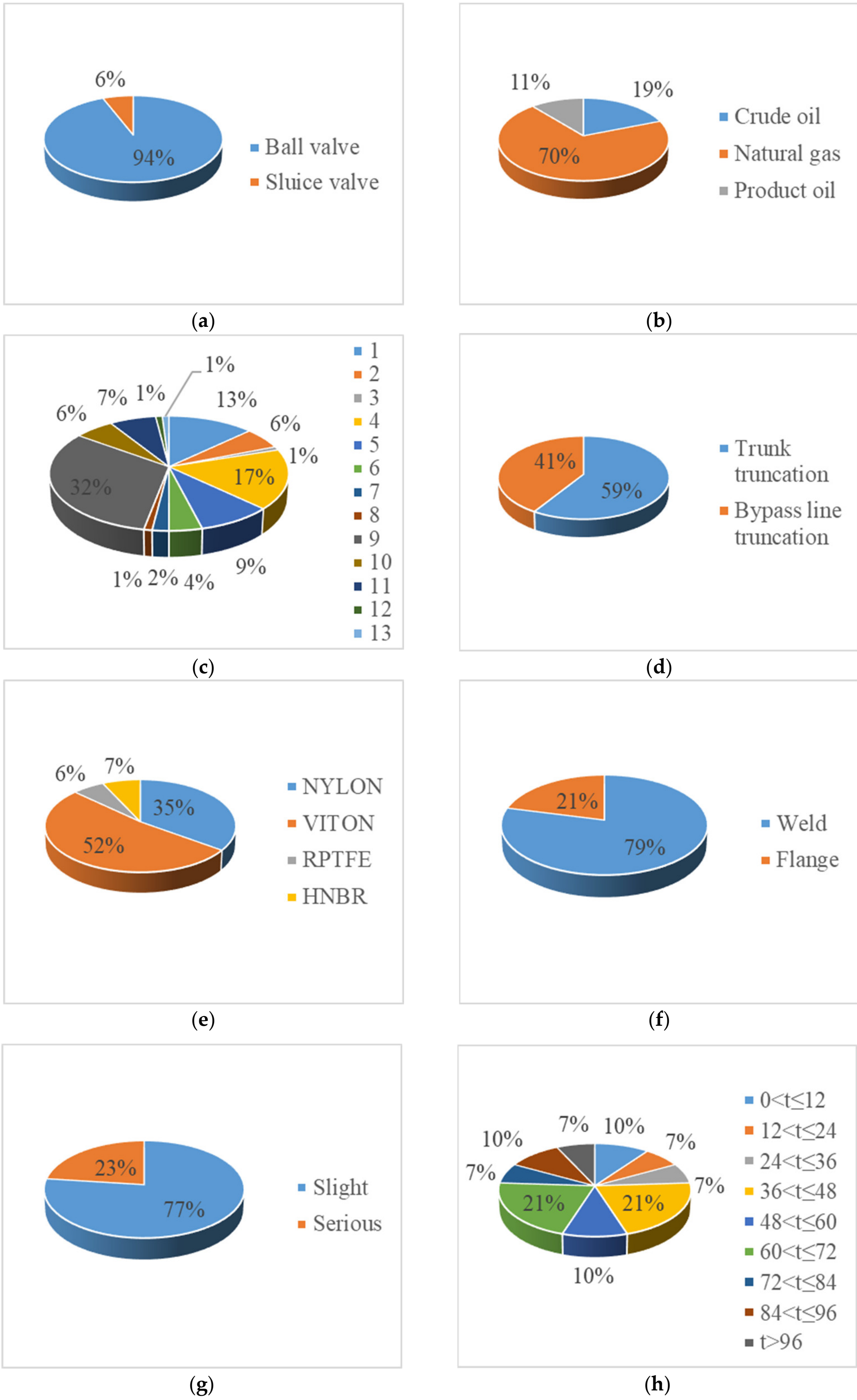


Figure 5. Statistical chart of the actual situation of the valve. (**a**) Valve type; (**b**) Conveying medium; (**c**) Affiliated pipelines; (**d**) Function and Location; (**e**) Sealing materials; (**f**) Connection method; (**g**) Leakage classification; (**h**) Usage time.

3.2. Principal Component Analysis

To achieve data dimensionality reduction, it is essential to conduct principal component analysis (PCA) to determine the main factors affecting the valve service life. The eigenvalues of the covariance matrix of the seven indicator variables are shown in Table 3, and Figure 6 shows the histogram of the principal component contribution. From Table 3 and Figure 6, it can be seen that the cumulative variance contribution rate is 96.80% when 5 principal components are selected. At this point, 5 principal components are sufficient to reflect the characteristic information of the original 7 variables. Compared with the 7 input features of the original sample data, the extraction of principal components by PCA can effectively reduce the dimensionality of the feature variable space, and at the same time, retain most of the information of the original variables and avoid information redundancy caused by cross-correlation of the original variables.

Table 3. Descending eigenvalues and contribution rates.

Main Ingredient	Eigenvalues	Contribution Rate/%	Cumulative Contribution Rate/%
1	2.37657	33.95104	33.95104
2	2.00350	28.62149	62.57253
3	1.26736	18.10506	80.67759
4	0.62783	8.96897	89.64656
5	0.50074	7.15337	96.79993
6	0.22401	3.20007	99.99995
7	7.18×10^{-17}	1.03×10^{-15}	100

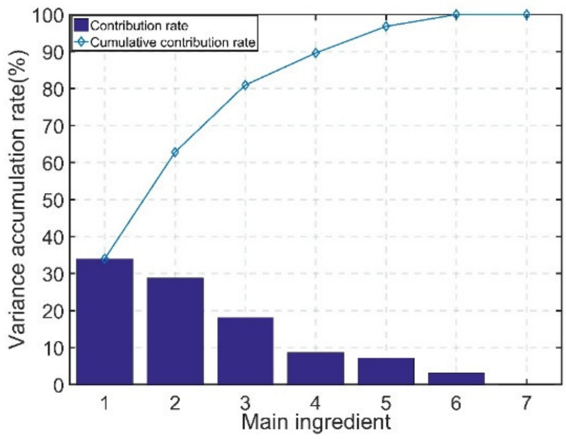


Figure 6. Principal component contribution rate histogram.

The linear expression of the principal components can be obtained from Equation (3) as Equation (16). The original feature input is reduced from 7 feature variables to 5, thereby reducing the input of sample data.

$$\left.\begin{array}{l} y_1 = 0.3414X_1 + 0.3800X_2 + 0.5744X_3 + 0.1358X_4 \\ -0.0344X_5 + 0.2293X_6 + 0.5744X_7 \\ y_2 = -0.4188X_1 - 0.3415X_2 + 0.3133X_3 - 0.3926X_4 \\ +0.5967X_5 - 0.0398X_6 + 0.3133X_7 \\ y_3 = -0.2904X_1 - 0.2879X_2 - 0.0251X_3 + 0.5195X_4 \\ +0.0495X_5 + 0.7478X_6 - 0.0251X_7 \\ y_4 = 0.2287X_1 - 0.0624X_2 - 0.0243X_3 + 0.6643X_4 \\ +0.5584X_5 - 0.4353X_6 - 0.0243X_7 \\ y_5 = -0.5670X_1 + 0.7828X_2 - 0.0899X_3 + 0.0903X_4 \\ +0.2037X_5 - 0.0011X_6 - 0.0899X_7 \end{array}\right\} \quad (16)$$

On this basis, the original data vector with high-dimensional redundancy is dimensionally reduced. The orthogonal transformation in linear programming is used to reduce the original correlated variables to a few uncorrelated composite variables, which lays the foundation for the subsequent model prediction of valve life.

3.3. Model Analysis

Data processed by PCA is used as the input of the valve prediction model, and the valve service life is used as the model output to establish the valve service life prediction model. The valve service life prediction results based on PCA-PSO-LSSVM are shown in Figure 7.

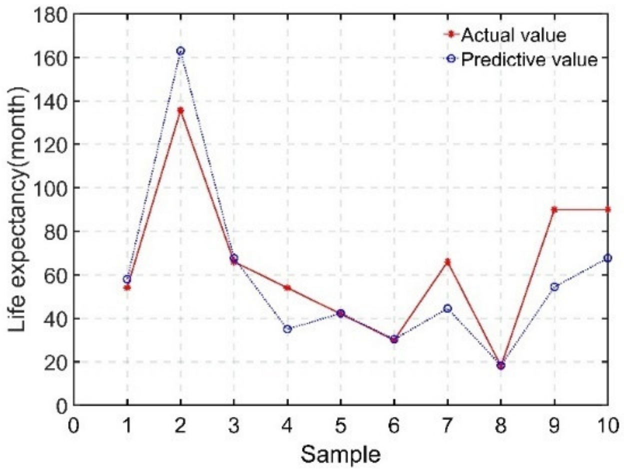


Figure 7. PCA-PSO-LSSVM prediction result graph.

In order to verify the prediction effect of PCA-PSO-LSSVM, the PCA-PSO-LSSVM model is compared with the PSO-LSSVM model, and the prediction results of the PSO-LSSVM model are shown in Figure 8. The relative error obtained is shown in Figure 9. As can be seen from Figure 9, the PCA-PSO-LSSVM model has a smaller relative error associated with the prediction results compared with the PSO-LSSVM model, indicating that the PCA algorithm for data dimensionality reduction can effectively improve the prediction accuracy of the model.

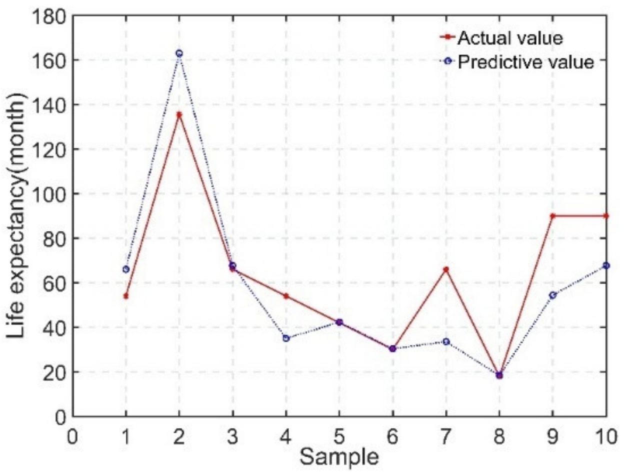


Figure 8. PSO-LSSVM prediction result graph.

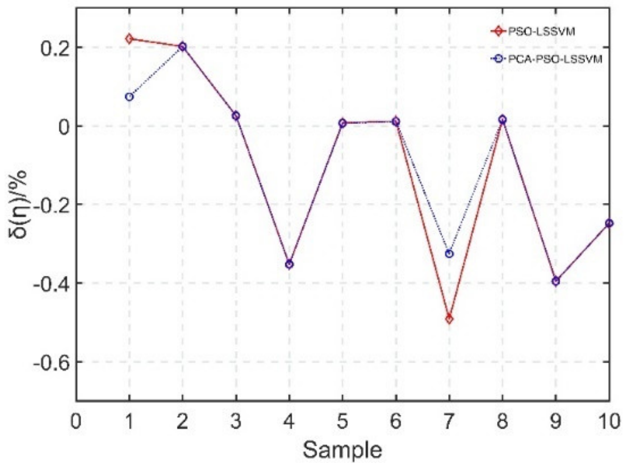


Figure 9. Comparison of relative error results of PCA-PSO-LSSVM and PSO-LSSVM.

At the same time, in order to verify the optimization effect of the particle swarm optimization (PSO) algorithm, the prediction results of the PSO-LSSVM and LSSVM models are compared. The prediction results of the LSSVM model are shown in Figure 10, and the relative error is shown in Figure 11. As can be seen from Figure 11, the PSO-LSSVM model has a smaller relative error associated with the prediction results than the LSSVM model, indicating that the PSO algorithm can effectively optimize the model parameters and establish a more reliable and effective model.

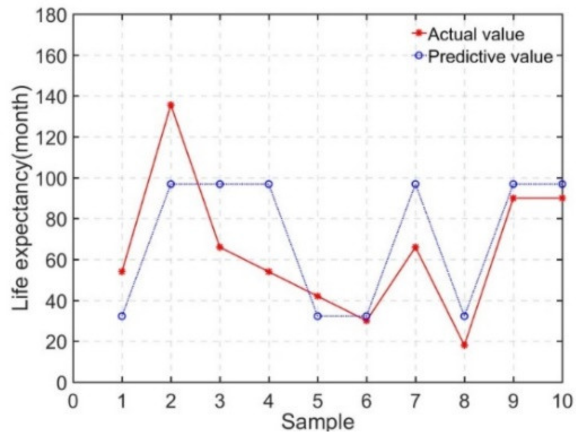


Figure 10. Prediction result graph of LSSVM.

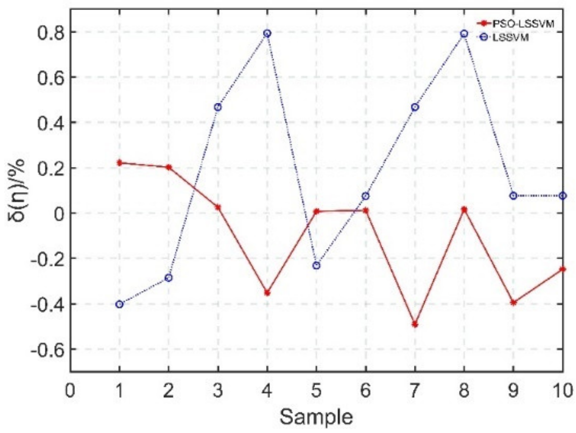


Figure 11. Comparison of relative error results of PSO-LSSVM and LSSVM.

The mean relative error (MRE) and root mean square error (RMSE) of PCA-PSO-LSSVM, PSO-LSSVM, and LSSVM are obtained, and the prediction results of PCA-PSO-LSSVM, PSO-LSSVM, and LSSVM are compared, as shown in Table 4. From Table 4, it can be seen that the MRE and RMSE of PSO-LSSVM are improved by 16.97% and 3.0943, respectively, compared with LSSVM. The MRE of PCA-PSO-LSSVM is 16.57%, which is improved by 3.15% and 20.12%, respectively, compared with PSO-LSSVM and LSSVM. The RMSE of PCA-PSO-LSSVM is 1.2636, which is improved by 2.5201 and 5.6144 compared to PSO-LSSVM and LSSVM, respectively. It can be seen that PCA-PSO-LSSVM can more accurately predict the service life of valves.

Table 4. Comparison of evaluation indicators of various prediction methods.

Method	MRE	RMSE
PCA-PSO-LSSVM	16.57%	1.2636
PSO-LSSVM	19.72%	3.7837
LSSVM	36.69%	6.8780

4. Conclusions

This work investigated the method of valve service life prediction. Firstly, through analysis of the reasons for valve failure, the prediction model of influencing factors and valve life is established, and the service life of the valve is predicted based on the PCA-PSO-LSSVM algorithm. A PCA algorithm was used to perform principal component analysis on the original sample data to obtain five features containing the main information to realize data dimensionality reduction; the five-dimensional data after dimensionality reduction contained more than 95% of the information of the original sample data. Then, the parameters of the LSSVM model were optimized using a PSO algorithm to improve the prediction accuracy of the model, and a valve service life prediction model based on PCA-PSO-LSSVM was established. The PCA-PSO-LSSVM, PSO-LSSVM, and LSSVM models are compared. The results show that the valve life prediction values based on PCA-PSO-LSSVM are closer to the actual values, and the MRE of the prediction results are improved by 3.15% and 20.12% compared with PSO-LSSVM and LSSVM, respectively; the RMSE is improved by 2.5201 and 5.6144 compared with PSO-LSSVM and LSSVM, respectively. Results show that the proposed method has higher accuracy than PSO-LSSVM and LSSVM, which can improve the reliability of valve life prediction, and provide guidance and suggestions for the maintenance and replacement of valves. Thus, the utilization rate of valves can be effectively improved to avoid accidents.

Author Contributions: Conceptualization, M.S. and Z.H.; Methodology, M.S. and Z.H.; Data Curation, L.Q. and P.T.; Writing—Original Draft Preparation, L.Q. and P.T.; Writing—Review and Editing, M.S.; Project Administration, M.S. and L.Q.; Funding Acquisition, M.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Sichuan Province Science and Technology Support Program (2022YFQ0062).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in the study are included in the article, and further inquiries can be directed to the corresponding author.

Acknowledgments: The authors acknowledge Sichuan Province Science and Technology Support Program (2022YFQ0062).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kaewwaewnoi, W.; Prateepasen, A.; Kaewtrakulpong, P. Investigation of the relationship between internal fluid leakage through a valve and the acoustic emission generated from the leakage. *Measurement* **2010**, *43*, 274–282. [CrossRef]
2. Li, Z.L.; Zhang, H.F.; Tan, D.J.; Chen, X.; Lei, H.X. A novel acoustic emission detection module for leakage recognition in a gas pipeline valve. *Process Saf. Environ. Prot.* **2017**, *105*, 32–40. [CrossRef]
3. Zhou, M.F.; Zhang, Q.; Liu, Y.W.; Sun, X.F.; Cai, Y.J. Haitian Pan. An Integration Method Using Kernel Principal Component Analysis and Cascade Support Vector Data Description for Pipeline Leak Detection with Multiple Operating Modes. *Processes* **2019**, *7*, 648. [CrossRef]
4. Kamali, M.R.; Mirshady, A.A. Total organic carbon content determined from well-logs using ΔLogR and Neuro Fuzzy techniques. *J. Pet. Sci. Eng.* **2004**, *45*, 141–148. [CrossRef]
5. Tan, M.J.; Song, X.D.; Yang, X.; Wu, Q.Z. Support-vector-regression machine technology for total organic carbon content prediction from wireline logs in organic shale: A comparative study. *J. Nat. Gas Sci. Eng.* **2015**, *26*, 792–802. [CrossRef]
6. Nuhic, A.; Terzimehic, T.; Soczka-Guth, T.; Buchholz, M.; Dietmayer, K. Health diagnosis and remaining useful life prognostics of lithium-ion batteries using data-driven methods. *J. Power Sources* **2013**, *239*, 680–688. [CrossRef]
7. Deepak, V.; Pranav, V.S.; Lakshmi, B.S.; Vishnu, K. Lifetime Prediction of Lithium-Ion Battery Using Machine Learning for E-Vehicles. *J. Phys. Conf. Ser.* **2021**, *1916*, 012200.
8. Su, N.K.H.; Juwono, F.H.; Wong, W.K.; Chew, I.M. Review on Machine Learning Methods for Remaining Useful Lifetime Prediction of Lithium-ion Batteries. In Proceedings of the International Conference on Green Energy, Computing and Sustainable Technology (GECOST), Miri Sarawak, Malaysia, 26–28 October 2022.
9. De Cooman, T.; Vandecasteele, K.; Varon, C.; Hunyadi, B.; Cleeren, E.; van Paesschen, W.; van Huffel, S. Personalizing heart rate-based seizure detection using supervised SVM transfer learning. *Front. Neurol.* **2020**, *11*, 145. [CrossRef]
10. Zhang, Y.; Jiang, Y.; Chen, Y.; Zhang, Y. Fault diagnosis of high voltage circuit breaker based on multi-classification relevance vector machine. *J. Electr. Eng. Technol.* **2020**, *15*, 413–420. [CrossRef]
11. Muhammad, A.R.; Liang, Z.M.; Salim, H.; Mohammad, Z.K.; Ozgur, K.; Li, B.Q. Least square support vector machine and multivariate adaptive regression splines for streamflow prediction in mountainous basin using hydro-meteorological data as inputs. *J. Hydrology* **2020**, *586*, 124371.
12. Zhang, L.; Li, K.; Du, D.; Guo, Y.; Fei, M.; Yang, Z. A sparse learning machine for real-time SOC estimation of li-ion batteries. *IEEE Access* **2020**, *8*, 156165–156176. [CrossRef]
13. Fan, W.; Si, F.; Ren, S.; Yu, C.; Cui, Y.; Wang, P. Integration of continuous restricted Boltzmann machine and SVR in NOx emissions prediction of a tangential firing boiler. *Chemom. Intell. Lab. Syst.* **2019**, *195*, 103870. [CrossRef]
14. Wang, H.; Peng, M.-J.; Hines, J.W.; Zheng, G.-Y.; Liu, Y.-K.; Upadhyaya, B.R. A hybrid fault diagnosis methodology with support vector machine and improved particle swarm optimization for nuclear power plants. *ISA Trans.* **2019**, *95*, 358–371. [CrossRef] [PubMed]
15. Yuan, C.S.; Li, X.T.; Wu, Q.M.J.; Li, J.; Sun, X.M. Fingerprint liveness detection from different fingerprint materials using convolutional neural network and principal component analysis. *Comput. Mater. Contin.* **2017**, *53*, 357–372.
16. Al-Anazi, A.F.; Gates, I.D. Support vector regression to predict porosity and permeability: Effect of sample size. *Comput. Geosci.* **2012**, *39*, 64–76. [CrossRef]
17. Al-Anazi, A.; Gates, I. A support vector machine algorithm to classify lithofacies and model permeability in heterogeneous reservoirs. *Eng. Geol.* **2010**, *114*, 267–277. [CrossRef]
18. Al-Anazi, A.; Gates, I. Support vector regression for porosity prediction in a heterogeneous reservoir: A comparative study. *Comput. Geosci.* **2010**, *36*, 1494–1503. [CrossRef]
19. El-Sebakhy, E.A. Forecasting PVT properties of crude oil systems based on support vector machines modeling scheme. *J. Petrol. Sci. Eng.* **2009**, *64*, 25–34. [CrossRef]
20. Chapelle, O.; Vapnik, V.; Bengio, Y. Model selection for small sample regression. *Mach. Learn.* **2002**, *48*, 9–23. [CrossRef]
21. Ghosh, A.; Chatterjee, P. Prediction of cotton yarn properties using support vector machine. *Fibers Polym.* **2010**, *11*, 84–88. [CrossRef]
22. Li, B.; Tian, X.T. An effective PSO-LSSVM-based approach for surface roughness prediction in high-speed precision milling. *IEEE Access* **2021**, *9*, 80006–80014. [CrossRef]