



ARTICLE

A Data-Driven Fault Prediction Method for Bearing Ring CNC Grinding Machines

Yanan Wang, Xiaoying Yang*, Zhijie Pei, Xin Yang and Bo Li

School of Mechatronics Engineering, Henan University of Science and Technology, Luoyang, China

*Corresponding Author: Xiaoying Yang, Email: lyxy111@163.com

Received: 15 April 2026; Accepted: 29 June 2026; Published: 13 August 2026

ABSTRACT: Sudden faults in bearing ring computer numerical control (CNC) grinding machines significantly impact product processing quality and production efficiency, making precise state prediction urgent to avoid downtime risks. However, the numerous operational parameters collected on-site and the focus of existing methods on outputting fault labels without analyzing the evolution trends of the equipment's operational state lead to unclear fault discrimination criteria and weak traceability, making it difficult to provide effective early-warning support during the incipient stages of a fault. To address these issues, this paper constructs a data-driven integrated algorithm adopting a “predict-then-classify” approach. First, the Pearson-ReliefF algorithm is utilized to eliminate redundant features and retain sensitive parameters. Second, the BiGRU-Attention algorithm is employed to capture bidirectional dependencies in time-series data, realizing the prediction of the equipment's operational state trends. Finally, the Sparrow Search Algorithm (SSA) is introduced to optimize core Support Vector Machine (SVM) parameters, achieving precise identification of fault types in bearing ring CNC grinding machines. Experimental results indicate that the proposed algorithm exhibits robust performance in both the prediction and classification stages. The prediction metrics MAE, RMSE, and R^2 are 0.0124, 0.0152, and 0.975, respectively, and the average multi-fault identification accuracy based on 10 repeated experiments with random seeds reaches 98.25%. This verifies the effectiveness of the method, which is of significant importance for ensuring the processing quality of bearing rings, reducing operation and maintenance costs, achieving intelligent online predictive diagnosis, and supporting the preventive maintenance of equipment.

KEYWORDS: Bearing; CNC grinding machine; data-driven; fault prediction

1 Introduction

As core foundational components in the high-end equipment manufacturing industry, the machining precision and quality of bearing rings directly determine the operational performance and production efficiency of downstream equipment. CNC grinding machines serve as the critical core equipment for achieving the precision machining of these bearing rings [1]. In actual production, CNC grinding machines for bearing rings operate for prolonged periods under high-speed and heavy-load conditions, where sudden failures can severely impact product machining quality and production efficiency [2]. With the application of the Industrial Internet and smart sensing technologies in the manufacturing sector, multi-dimensional operational parameters, such as vibration and temperature, can be collected in real time, providing a robust foundation for data-driven equipment fault prediction [3].

The implementation of predictive maintenance technologies for equipment primarily revolves around three core stages: feature dimensionality reduction, condition prediction, and fault diagnosis [4]. However, existing research on bearing ring CNC grinding machines still suffers from a disconnect between prediction and diagnosis processes. Moreover, prediction results often lack a physical evolution basis. Consequently, it remains difficult to achieve proactive anomaly trend prediction and precise fault localization simultaneously [5].

As a crucial prerequisite in equipment fault prediction, feature dimensionality reduction is the core method for resolving issues such as high dimensionality and high redundancy in multi-source monitoring data [6]. Existing dimensionality reduction and feature selection methods mostly focus on single-dimensional considerations, making it difficult to simultaneously balance the elimination of linear redundancy and the retention of non-linear fault-sensitive information. Regarding linear correlation analysis and data redundancy elimination, Zhou et al. [7] employed Pearson correlation analysis to determine input parameters, while Jiang and Hu [8] utilized Principal Component Analysis (PCA) to remove redundant information from high-dimensional time-series data. Although such methods can efficiently quantify and eliminate linear correlations between variables, they struggle to process the complex non-linear features present in industrial equipment operational data [9]. To overcome the application limitations of linear methods, Sun et al. [10] adopted the ReliefF algorithm for feature selection, effectively reducing dimensionality and improving fault diagnosis accuracy; Cui et al. [11] utilized Kernel Principal Component Analysis (KPCA) for the dimensionality reduction of fault features. However, non-linear mapping methods often overlook the inherent linear redundancy within the features themselves [12]. Therefore, this paper proposes a hybrid feature selection algorithm that integrates Pearson correlation analysis and ReliefF, aiming to eliminate linear redundancy while fully retaining fault-sensitive information, thereby providing high-quality data support for subsequent prediction models.

In the field of equipment condition prediction, deep learning has achieved significant progress. Liu et al. [13] constructed a fault prediction model based on CNN-LSTM, achieving accurate prediction by efficiently capturing time-series data; Tan et al. [14] proposed a novel early warning method based on CNN-BiLSTM-AE, which achieves dynamic capture of gearbox operational trends by integrating spatial feature extraction and adaptive reconstruction error analysis; Rama et al. [15] designed an RNN-LSTM model that combines SCADA temporal data to achieve short-term fault prediction for wind turbines. However, models based on LSTM and its variants are constrained by complex gating mechanisms, resulting in high training costs for long sequences. Meanwhile, models incorporating CNNs primarily focus on local feature extraction, making it difficult to effectively characterize the global temporal dependencies of the data. To tackle these algorithmic deficiencies, the gated recurrent unit (GRU), which features fewer parameters and faster convergence, has been widely applied to equipment temporal condition prediction tasks. Ali and Kamal [16] proposed a Transformer-GRU model based on the HHO-WHO optimization framework, significantly enhancing prediction robustness by achieving hyperparameter tuning through global and local optimization; Cheng et al. [17] integrated DenseNet-GRU with an attention mechanism, achieving accurate estimation of tool wear values; Han et al. [18] constructed a multi-model fusion prediction method integrating CNN, LSTM, and GRU, effectively extracting the spatiotemporal features of vibration signals. Nevertheless, conventional unidirectional GRU structures struggle to fully capture the bidirectional evolution patterns of condition features, and lack the capability to dynamically focus on key degradation characteristics when dealing with complex monitoring sequences. To tackle this algorithmic limitation, this paper introduces a BiGRU-Attention condition prediction model. By relying on BiGRU to capture forward and backward contextual information within the data, and integrating an attention mechanism to achieve

adaptive weighting of key features, this approach effectively enhances condition prediction accuracy while ensuring a lightweight model architecture.

Although isolated condition prediction models can characterize the evolution trends in the early stages of faults, they fail to provide explanatory support for the evolutionary mechanisms behind data anomalies, making it difficult to precisely determine fault types. Therefore, it is necessary to integrate fault diagnosis techniques to overcome the application limitations of single prediction models. Lu et al. [19] achieved precise identification of airborne cable arc faults based on the GA-RF method; Li [20] adopted an improved Whale Optimization Algorithm (SAWOA) to optimize key parameters of LSTM, providing a new technical solution for the health management of bearing equipment; He et al. [21] proposed a wind turbine blade fault diagnosis method based on multi-sensor fusion and a CNN-Transformer coupled architecture, which remains effective under noise interference. However, complex deep learning models often rely on massive annotated data, whereas fault samples at industrial sites are scarce, making it difficult to meet their training requirements. In contrast, Support Vector Machine (SVM) has become a commonly used method for industrial equipment fault diagnosis due to its excellent performance in small-sample, non-linear, and high-dimensional pattern recognition tasks. Nevertheless, the classification performance of SVM is highly dependent on the selection of hyperparameters, such as the penalty factor and kernel parameters. To address the issues of high subjectivity and low optimization accuracy associated with manual parameter tuning: Wei et al. [22] employed the Genetic Algorithm (GA) for the adaptive optimization of core SVM hyperparameters; Atal et al. [23] utilized the highly efficient iterative mechanism of the Grey Wolf Optimizer (GWO) to rapidly locate the optimal SVM parameter combination; Zhu et al. [24] leveraged the swarm information-sharing mechanism of Particle Swarm Optimization (PSO) to significantly reduce the time consumed for hyperparameter optimization. Despite these advancements, when processing high-dimensional and complex fault features, conventional optimization algorithms still suffer from slow early-stage convergence and a tendency to fall into local optima in later stages. Therefore, this paper proposes a diagnosis method based on an SVM optimized by the SSA, utilizing the unique population iteration mechanism of SSA to achieve the global dynamic optimization of hyperparameters.

In summary, to address the technical bottlenecks of multi-source data redundancy and the disconnection between prediction and diagnosis processes, this paper proposes an integrated algorithm framework combining Pearson-ReliefF feature dimensionality reduction, BiGRU-Attention state prediction, and SSA-SVM fault diagnosis. By adopting a two-stage integrated structure of “predict-then-classify,” the framework utilizes the feature evolution trends generated by the state prediction model as the input for the fault diagnosis model. Ultimately, this achieves the collaborative implementation of equipment operational state trend prediction and precise fault localization. The main contributions of this paper are as follows:

- (1) Addressing the limitations of existing research where predictive and diagnostic models operate independently, predictive outcomes fail to inform diagnostic decisions, and predictive models cannot provide fault interpretation for anomalous trends, this paper proposes a two-stage integrated fault prediction framework following the principle of “predict first, then diagnose”. By deeply coupling time-series state prediction with fault diagnosis technology, the fault diagnostic model is driven by the predicted future evolution trends of the equipment, achieving a paradigm shift from passive fault identification to proactive anomaly anticipation.
- (2) Addressing the characteristics of multi-source heterogeneous data in bearing ring CNC grinding machines, this paper proposes a Pearson-ReliefF hybrid feature selection algorithm. Pearson correlation analysis is first applied to eliminate linearly redundant features, followed by the ReliefF algorithm to quantify and rank fault sensitivity weights, thereby extracting a fault-sensitive feature subset from the

high-dimensional raw data. This provides a high-quality data foundation for the subsequent prediction and diagnostic models.

- (3) Addressing the practical constraint of scarce fault samples in bearing ring CNC grinding machines, this paper constructs a lightweight prediction-diagnosis model combination adapted to small-sample scenarios. The BiGRU-Attention state prediction model integrates bidirectional temporal feature mining with an attention-based dynamic weighting mechanism to accurately capture the evolution patterns of multi-dimensional parameters. The SSA-SVM fault diagnosis algorithm employs the Sparrow Search Algorithm for global hyperparameter optimization to enhance feature recognition capability under small-sample conditions. The two models work in synergy to effectively improve the accuracy of both fault prediction and diagnosis.

2 Fault Prediction Algorithm

This section constructs a fault prediction framework for grinding machines based on multi-algorithm integration. It systematically elaborates on the methodological logic and implementation processes of Pearson-ReliefF feature selection, BiGRU-Attention time-series prediction, and SSA-SVM fault identification.

2.1 Modeling of the Multivariate Time Series Prediction Problem

A multivariate time series is a sequence that simultaneously records the variations of multiple variables over time. The operating state of a bearing ring CNC grinding machine is collectively represented by multiple interrelated parameters. Through OPC technology and sensors, multi-dimensional parameters, such as lubrication system pressure, hydraulic pump pressure, spindle motor current, and feed axis temperature, are collected synchronously, forming a typical high-dimensional, nonlinear multivariate time series.

To achieve dimensionality reduction for high-dimensional data, this paper employs the Pearson-ReliefF algorithm to screen and obtain a feature vector composed of k key parameters. The equipment operating status parameters at time t can then be expressed as:

$$x_t = [x_{t,1}, x_{t,2}, x_{t,3}, \dots, x_{t,k}]^T \in R^k \quad (1)$$

where $x_{t,i}$ ($i = 1, 2, \dots, m$) represents the monitored value of the i -th key state parameter at time t .

Regarding the multivariate time series prediction problem for bearing ring CNC grinding machines, the mathematical description is as follows: Given the historical sequence data of length L collected by m sensors. At time t , the equipment state data collected by the sensors can be represented as vector $x_t \in R^m$, and the input historical data exists in the form of matrix X , as expressed by Eq. (2).

$$X = \{x_{t-L+1}, x_{t-L+2}, \dots, x_t\} \in R^{L \times m} \quad (2)$$

The objective of the prediction model is to forecast the state evolution trend $\hat{X}$ for the future h time steps based on the historical matrix X , as expressed by Eq. (3).

$$\hat{X} = F_{BiGRU-Attention}(X; \theta) = \{\hat{x}_{t+1}, \hat{x}_{t+2}, \dots, \hat{x}_{t+h}\} \in R^{h \times m} \quad (3)$$

where θ represents the set of model parameters to be optimized, L is the time window size, h is the prediction step, and $\hat{x}_{t+h}$ is the equipment state prediction vector output by the model at time $t + h$.

The multivariate time series output by this prediction model will serve as the core input basis for the subsequent SSA-SVM fault diagnosis model. By extracting trend features from the predicted waveforms, it provides high-quality data support for the precise classification of early faults.

2.2 Feature Selection Algorithm Based on Pearson-ReliefF

Pearson correlation analysis is a statistical method used to measure the degree of linear correlation between two continuous variables. The calculation process is shown in Eq. (4):

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (4)$$

where x_i and y_i represent the i -th sample values of variables x and y , respectively; $\bar{x}$ and $\bar{y}$ represent the sample means of variables x and y , respectively; n is the number of samples; and r is the correlation coefficient, which ranges from $[-1, 1]$. An absolute value closer to 1 indicates a stronger degree of linear correlation, while a value closer to 0 indicates a weaker linear relationship.

The ReliefF algorithm is a feature selection algorithm that screens out key and effective features from a massive number of features by quantifying their contributions to distinguishing different categories. A sample A is randomly selected from the dataset S , and the distances from this sample to its nearest neighbor of a different class and its nearest neighbor of the same class are calculated, respectively. Subsequently, based on the distance differences, the weights are adjusted feature by feature through an iterative updating rule. The iterative updating rule is shown in Eq. (5):

$$W(X) = w(X) - \frac{\sum_{j=1}^k \text{diff}(X, A_i, H_j)}{fk} + \sum_{C \neq \text{Class}(A_i)} \frac{P(C)}{1 - P(\text{Class}(A_i))} \cdot \frac{\sum_{j=1}^k \text{diff}(X, A_i, S_j)}{fk} \quad (5)$$

where $w(X)$ is the weight of feature X , with an initial value of 0; f is the number of samplings; k is the number of nearest neighbors; H_j and S_j represent the nearest neighbor sample of sample A_i within the same class and the nearest neighbor sample in a different class, respectively. $\text{Class}(A_i)$ is the sample class, $P(C)$ is the prior probability of class C ; and $\text{diff}(X, A_i, S_j)$ represents the numerical difference between sample A_i and its same-class neighbor H_j on feature X .

2.3 Time-Series Prediction Algorithm Based on BiGRU-Attention

BiGRU is an extended form of GRU [25]. By constructing two GRU modules in opposite directions, forward and backward, it forms a bidirectional transmission mechanism. While retaining the efficient temporal information processing capability of GRU, it comprehensively captures bidirectional contextual features and fully considers the information from both historical and future moments of the data, thereby improving the learning efficiency and prediction performance of the model. The model structure is shown in Fig. 1.

The calculation processes for the reset gate (r_t), update gate (z_t), candidate hidden state ($\tilde{h}_t$), and final hidden state (h_t) in GRU are shown in Eqs. (6)–(9):

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r) \quad (6)$$

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t] + b_z) \quad (7)$$

$$\tilde{h}_t = \tanh(W_h \cdot [r_t * h_{t-1}, x_t] + b_h) \quad (8)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \quad (9)$$

where x_t represents the input at the current time step; h_{t-1} represents the hidden state at the previous time step; W_r , W_z , W_h represent the corresponding weight matrices; b_r , b_z , b_h represent the bias terms for various

gates; $*$ denotes element-wise multiplication; σ denotes the sigmoid activation function; $\tanh$ denotes the hyperbolic tangent activation function; $(1 - z_t) * h_{t-1}$ represents the retained historical information; and $z_t * \tilde{h}_t$ represents the integrated new information.

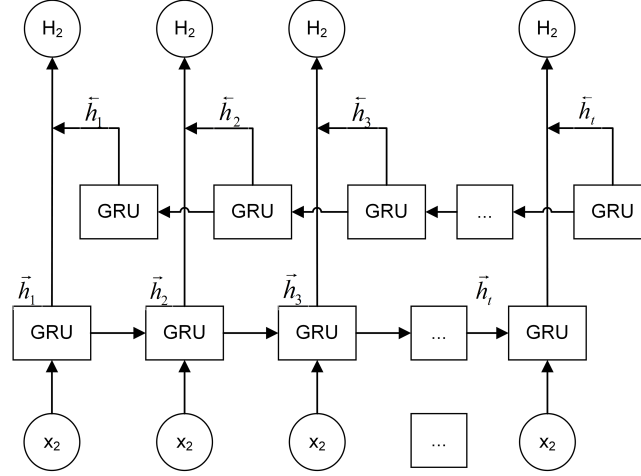


Figure 1: Structure of BiGRU.

Although the BiGRU network can extract time-series features, within its network architecture, each input element is assigned equal weight, resulting in “undifferentiated attention” toward all inputs during data processing. In contrast, the attention mechanism can assign differentiated weights during the prediction process based on the informational value of different parts of the sequence, enabling the model to focus on time steps and features critical for forecasting future trajectories [26]. The model structure is shown in Fig. 2. By performing a weighted sum of the features output by the BiGRU, the attention mechanism allows the model to analyze correlations between different time steps of the input sequence and dynamically allocate weights. This enables the model to focus on important parts of the sequence while ignoring irrelevant or redundant information during prediction or classification, making it more adaptive and precise when dealing with complex temporal problems [27].

First, the attention score e for each variable is calculated:

$$e_t = u \tanh(nh_t + b) \quad (10)$$

The scores are normalized using the softmax function:

$$a_t = \frac{\exp(e_t)}{\sum_{j=1}^T \exp(e_j)} \quad (11)$$

Finally, the weighted sum is performed to obtain the comprehensive information of the time-series state features:

$$A = \sum_{t=1}^T a_t h_t \quad (12)$$

where h_t is the hidden state output of BiGRU at time t ; e_t is the calculated attention score; a_t represents the normalized attention weight distribution; u and w are weight coefficients; b is the bias coefficient; and A is the comprehensive information of the time-series state features after the weighted sum.

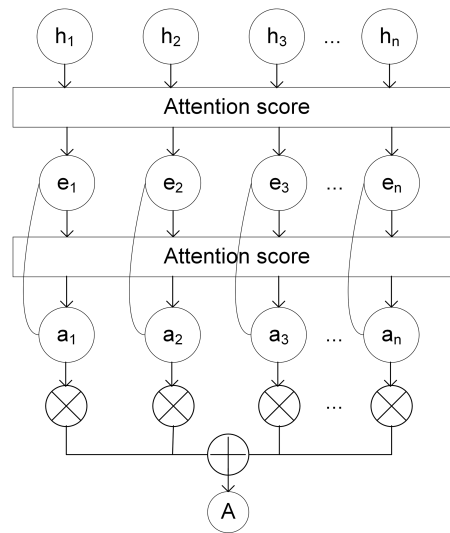


Figure 2: Structure of attention.

2.4 Fault Identification Algorithm Based on SSA-SVM

Proposed by Cortes and Vapnik in 1995, the support vector machine (SVM) is an optimal compromise algorithm based on risk minimization [28]. By searching for the optimal classification hyperplane, it maximizes the sum of the distances from this hyperplane to the two classes of sample sets, thereby maximizing the classification performance. It can effectively classify both linear and nonlinear problems and has demonstrated good application effects in many practical problems. The classification schematic of SVM is shown in Fig. 3.

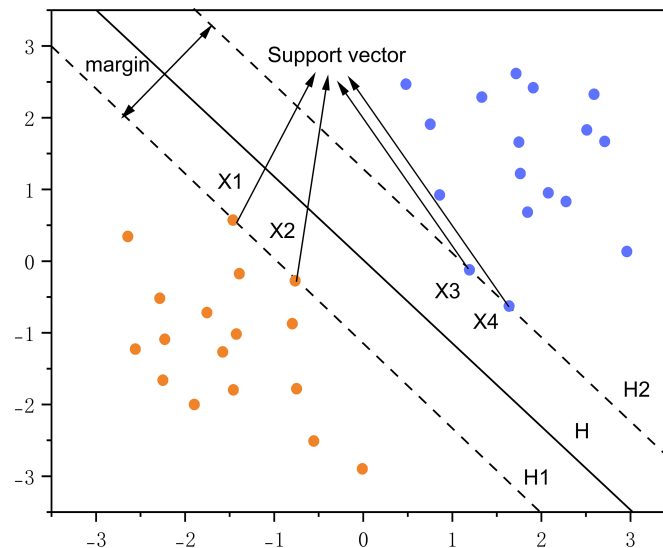


Figure 3: Schematic of the optimal separating hyperplane in SVM.

SVM possesses outstanding modeling capabilities for complex nonlinear decision boundaries and can output high-precision discrimination results in classification tasks. The training of an SVM model requires the setting and tuning of multiple parameters. Its overall performance is highly correlated with the selection

of the kernel function parameter λ and the penalty coefficient C . Improper parameter selection will directly lead to a decline in the model's generalization ability and prediction accuracy [29]. Relying on its excellent global optimization capability and fast convergence characteristics, SSA can effectively achieve the parameter optimization of SVM, providing reliable support for the enhancement of model performance.

Proposed by Xue and Shen [30] in 2020, SSA is a novel swarm intelligence optimization algorithm that simulates the group behaviors of sparrows during natural foraging and predator avoidance. SSA divides the population individuals into three roles: discoverers, joiners, and scouts. Relying on a division of labor and collaboration mechanism, it achieves a dynamic balance between global exploration and local exploitation: The discoverers are responsible for exploring spaces and searching for high-quality foraging areas, providing the core optimization direction for the population; the joiners perform local searches based on the position information of the discoverers, enhancing the convergence efficiency of the optimization; the scouts undertake the responsibility of risk monitoring and will issue early warning signals when the population faces danger, thus preventing the algorithm from falling into local optima.

Simulating the foraging process of the sparrow population, let the sparrow population size be N and the dimension of the search space be d . The position vector of the i -th sparrow at the t -th iteration is denoted as $X_i^t = [x_{i1}^t, x_{i2}^t, \dots, x_{id}^t]$, and its corresponding fitness value is $f(X_i^t)$. The discoverers are responsible for exploring the space and searching for high-quality foraging areas, providing the core optimization direction for the population. The update rule is shown in Eq. (13):

$$X_i^{t+1} = \begin{cases} X_i^t \cdot \exp\left(\frac{-i}{\alpha \cdot T_{\max}}\right), & R_2 < ST \\ X_i^t + Q \cdot L, & R_2 \geq ST \end{cases} \quad (13)$$

where $\alpha \in (0, 1]$ is a random number; $T_{\max}$ is the maximum number of iterations; R_2 is the warning value, and ST is the safety threshold; Q is a random number following a normal distribution; and L is a $1 \times d$ matrix with all elements being 1.

The joiners perform local searches based on the position information of the discoverers to enhance the optimization convergence efficiency. The update rule is shown in Eq. (14):

$$X_i^{t+1} = \begin{cases} Q \cdot \exp\left(\frac{X_{\text{worst}}^t - X_i^t}{i^2}\right), & i > \frac{N}{2} \\ X_{\text{best}}^{t+1} + |X_i^t - X_{\text{best}}^{t+1}| \cdot A \cdot L, & \text{else} \end{cases} \quad (14)$$

where X_{best}^{t+1} is the current optimal position; X_{worst}^{t+1} is the current worst position; and A is a random sign matrix.

The scouts undertake the responsibility of risk monitoring. When the population faces danger, they will issue early warning signals in advance to prevent the algorithm from falling into local optima. The update rule is shown in Eq. (15):

$$X_i^{t+1} = \begin{cases} X_{\text{best}}^t + \beta \cdot |X_i^t - X_{\text{best}}^t|, & f_i > f_g \\ X_i^t + k \cdot \frac{|X_i^t - X_{\text{worst}}^t|}{f_i - f_w + \varepsilon}, & f_i = f_g \end{cases} \quad (15)$$

where β is a step size coefficient following a standard normal distribution; $k \in [-1, 1]$ is a random perturbation factor; f_i is the fitness value of the current individual sparrow; f_g and f_w are the global best and worst

fitness values, respectively; and ε is a minimal constant used to avoid calculation anomalies caused by a zero denominator.

Featuring a fast convergence speed and strong global optimization capability, SSA can effectively avoid premature convergence compared to traditional optimization algorithms. It can precisely accomplish the adaptive optimization of the SVM kernel function parameter γ and penalty coefficient C , thereby enhancing the fault identification capability for bearing ring CNC grinding machines.

2.5 Flow Framework of Fault Prediction for Bearing Ring CNC Grinding Machines

Framework of fault prediction for bearing ring CNC grinding machines proposed in this paper is shown in Fig. 4.

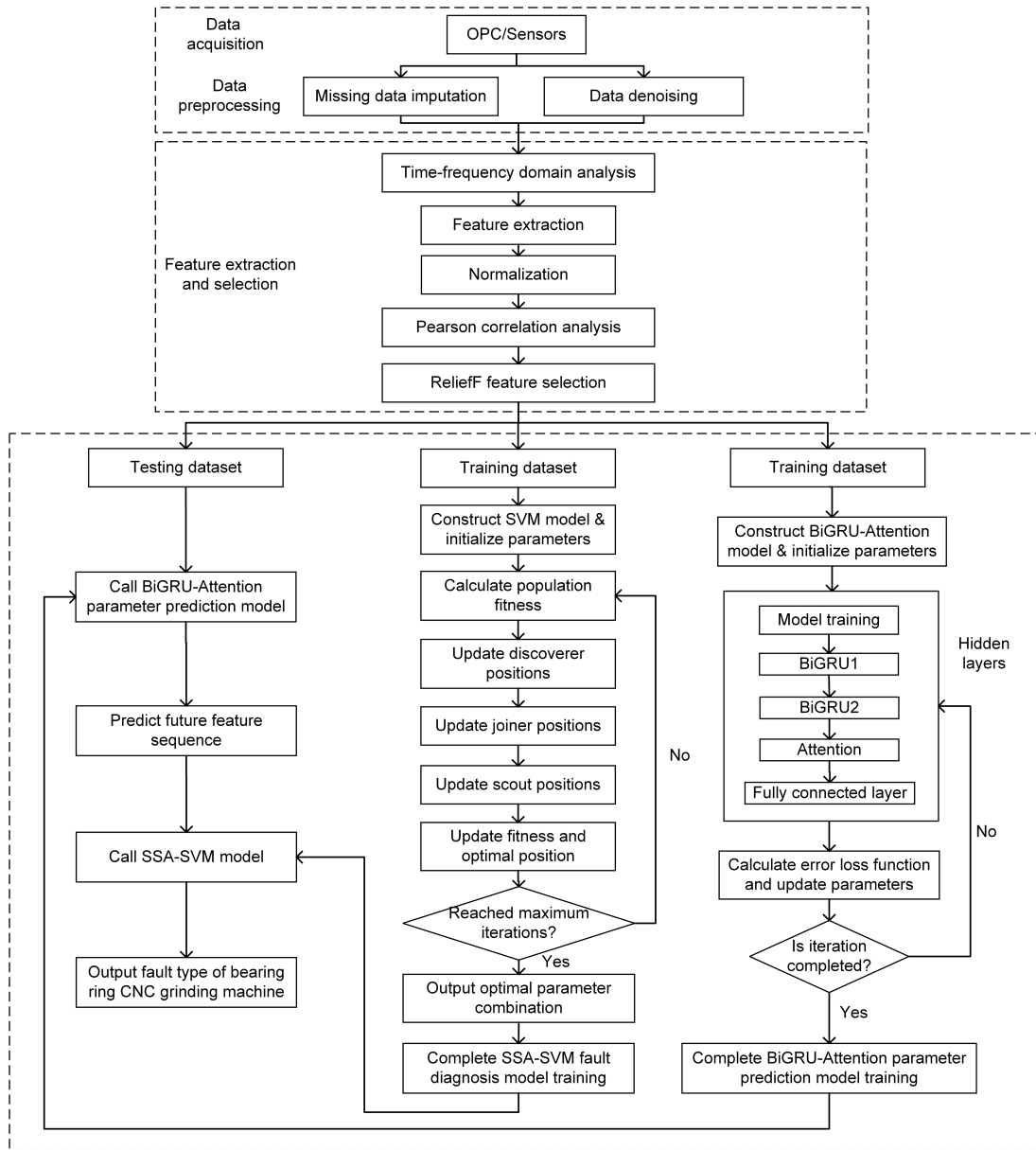


Figure 4: Theoretical framework of the fault prediction algorithm.

Step 1: Data acquisition and preprocessing

Multi-dimensional operational data, such as spindle vibration acceleration, spindle motor temperature, and spindle speed of the bearing ring CNC grinding machine, are collected through OPC technology and sensor interfaces. Missing data are imputed using interpolation methods, and noise is reduced using methods such as filtering and wavelet transform, providing high-quality data for subsequent processes.

Step 2: Feature extraction and selection

Time-frequency domain analysis is performed on the preprocessed data to extract time-domain and frequency-domain features, followed by normalization. Highly redundant features are eliminated through Pearson correlation analysis, and the ReliefF algorithm is then utilized to screen out key feature parameters sensitive to the equipment state, ultimately constructing an efficient feature dataset.

Step 3: Construction of the BiGRU-Attention feature parameter time-series prediction model

Based on the chronological order of data collection under each working condition, the dimensionally reduced data is sequentially divided into a training set and a testing set at a ratio of 7:3, with a validation set partitioned from within the training set at a ratio of 9:1. The parameters of the BiGRU-Attention model are initialized, a two-layer BiGRU network structure is built, and an Attention layer is introduced to dynamically allocate feature weights. This is then connected to a fully connected layer to complete the model architecture. During training, the parameters of the error loss function are updated and iteratively optimized to complete the training of the prediction model for the feature parameter time series.

Step 4: Construction of the multi-class SSA-SVM fault diagnosis model

An SVM model is constructed, the sparrow population is initialized, and individual fitness is cyclically calculated. The individual with the global optimal fitness is selected through iterative evaluation, and the positions of discoverers, joiners, and scouts are updated in stages until the maximum number of iterations is reached. Consequently, the optimal parameter combination is outputted, completing the training of the SSA-SVM model.

Step 5: Fault prediction of the bearing ring CNC grinding machine based on predicted features

The testing set is input into the trained BiGRU-Attention model to output the feature sequence prediction results for the future 30 time steps. This predicted sequence is then used as input for pattern recognition through the SSA-SVM diagnosis model, ultimately outputting the fault type of the bearing ring CNC grinding machine.

3 Case Study

3.1 Experimental Setup

The experimental environment in this study is divided into two parts: the model training environment and the online deployment environment.

The model training was conducted on a computing platform equipped with an Intel Core i5-10400 processor (base frequency 2.90 GHz) and 8 GB of RAM, running the Windows 10 operating system. The integrated development environment (IDE) was PyCharm 2023.3.2, utilizing Python interpreter version 3.12.1. During the experiment, libraries such as pandas, numpy, and matplotlib were primarily utilized to perform tasks including data preprocessing, feature analysis, and result visualization. All model training was executed entirely in a CPU environment.

The online deployment adopts a three-tier data link architecture following the pipeline of “PLC → Industrial Gateway → MES System”. Operational data from the CNC grinding machines is collected in real

time by PLCs, aggregated and transmitted through an industrial-grade IoT gateway (ARM Cortex-A53 quad-core 64-bit processor, 1.6 GHz clock speed, 1 GB LPDDR4 memory, running Ubuntu 20.04), and uploaded to the MES system server. The trained prediction model is deployed on the MES server, which receives data uploaded by the gateway via OPC technology, invokes the model for online inference and equipment health assessment, and feeds the prediction results back to the monitoring interface in real time. This architecture requires no computational upgrades to the on-site PLCs or gateways, demonstrating strong adaptability to industrial field environments.

3.2 Data Preparation

The experimental data in this study were collected from the actual production environment of a railway bearing intelligent factory, with the 3MZ2225TCNC inner ring rib grinding machine selected as the research subject. The bearing production line in the factory is uniformly managed by an MES system, and CNC grinding machines at each workstation perform real-time acquisition of multi-dimensional operational data via OPC communication technology and on-site sensors. The physical appearance of this bearing ring CNC grinding machine is shown in [Fig. 5](#).

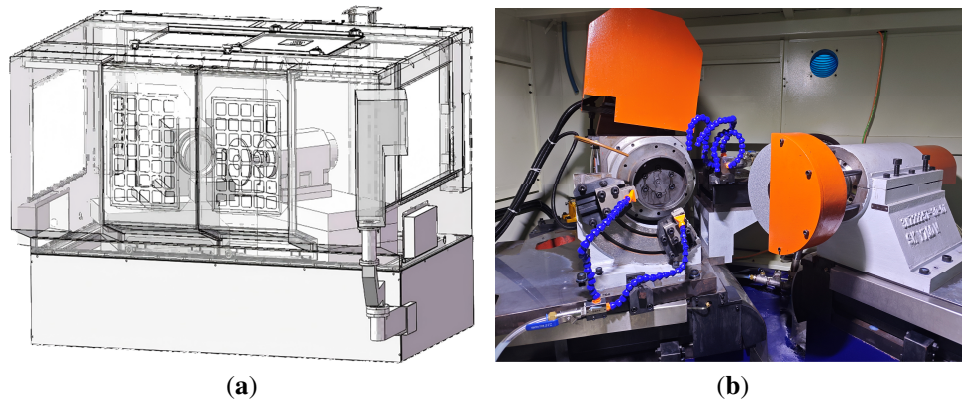


Figure 5: (a) Inner ring rib grinding machine; (b) Internal structure.

In the factory production environment, the equipment operating states are complex, and the monitoring parameters are not independent variables but change dynamically under the mutual influence of multiple factors. Parameters including spindle vibration acceleration, spindle motor temperature, spindle speed, spindle motor current, feed axis vibration acceleration, feed axis lead screw temperature, feed axis motor current, feed position deviation, feed speed, grinding feed speed, rough grinding amount, intermediate dressing compensation amount of the grinding wheel, lubrication system pressure, and hydraulic pump pressure are selected as the feature parameters for predicting faults of the bearing ring CNC grinding machine.

For the spindle vibration acceleration and feed axis vibration acceleration, high-frequency continuous acquisition was performed using piezoelectric accelerometers at a sampling frequency of 25.6 kHz, while the remaining 12 steady-state operational parameters were synchronously acquired via OPC technology at a polling interval of 100 ms. Each operating condition was continuously collected during an independent operating period, with vibration signals and other operational parameters extracted synchronously at 1-s intervals, forming a multivariate feature time series with a temporal resolution of 1 s/step. A total of 600 samples were extracted for each operating condition, yielding 3600 samples in total, with the operating conditions listed in [Table 1](#). The valid data segments for each operating condition were determined by comprehensively referencing MES system operation logs, equipment inspection records, and maintenance

work orders, and were further manually annotated and verified by on-site engineers to ensure the accuracy and reliability of sample labels.

Table 1: Operating conditions and labels.

Condition Label	Operating Condition	Abbreviation	Number of Samples
F0	Normal (Fault-free)	N	600
F1	Grinding wheel dulling	GWD	600
F2	Spindle bearing wear	SBW	600
F3	Ball screw wear	BSW	600
F4	Insufficient lubrication system pressure	ILSP	600
F5	Insufficient hydraulic pump pressure	IHPP	600

3.2.1 Data Preprocessing

In the data monitoring system of the bearing ring CNC grinding machine, data loss may occur at certain time points during the sensor data acquisition process due to hardware failures, transmission anomalies, or improper human operations. This will severely affect the prediction accuracy of the fault prediction model. This paper adopts the adjacent mean method to impute the missing data values, thereby obtaining a complete dataset.

During the operation of the bearing ring CNC grinding machine, multi-dimensional monitoring data are susceptible to interference from complex operating conditions, such as electromagnetic radiation in the workshop, equipment vibration, and the start-stop of the cooling system. This leads to noise and instantaneous abnormal fluctuations. Such noise can mask the true variation patterns of the data and degrade the accuracy of the subsequent fault prediction models.

Therefore, to enhance the continuity of the sample data for the operating parameters of the bearing ring CNC grinding machine, this paper employs a wavelet denoising method to denoise the sample data. A combination of a fixed threshold and a soft threshold function is selected to eliminate the noise from the samples. Taking the spindle speed, a key operating parameter of the bearing ring CNC grinding machine, as an example, the comparison results before and after denoising are shown in Fig. 6.

It can be clearly observed from Fig. 6 that compared with the original data curve, the denoised data curve is significantly optimized. The irregular fluctuations at the peaks and inflection points of the curve are notably weakened, the overall shape is smoother, and the data variation range is substantially reduced. At the same time, the denoised data maintains a high degree of similarity with the original data, which can more accurately characterize the actual variation trends of the operating parameters of the bearing ring CNC grinding machine.

For multi-source heterogeneous monitoring data such as vibration, pressure, and temperature, since each parameter has different physical dimensions and numerical ranges, standardization is required to eliminate dimensional differences. This study adopts the Min-Max normalization method, which maps the original data into a unified dimensional space through linear transformation. The mathematical expression of this method is shown in Eq. (16):

$$y = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (16)$$

where x is the feature value before normalization; y is the feature value after normalization; and $x_{\max}$ and $x_{\min}$ are the maximum and minimum values of the feature value x , respectively.

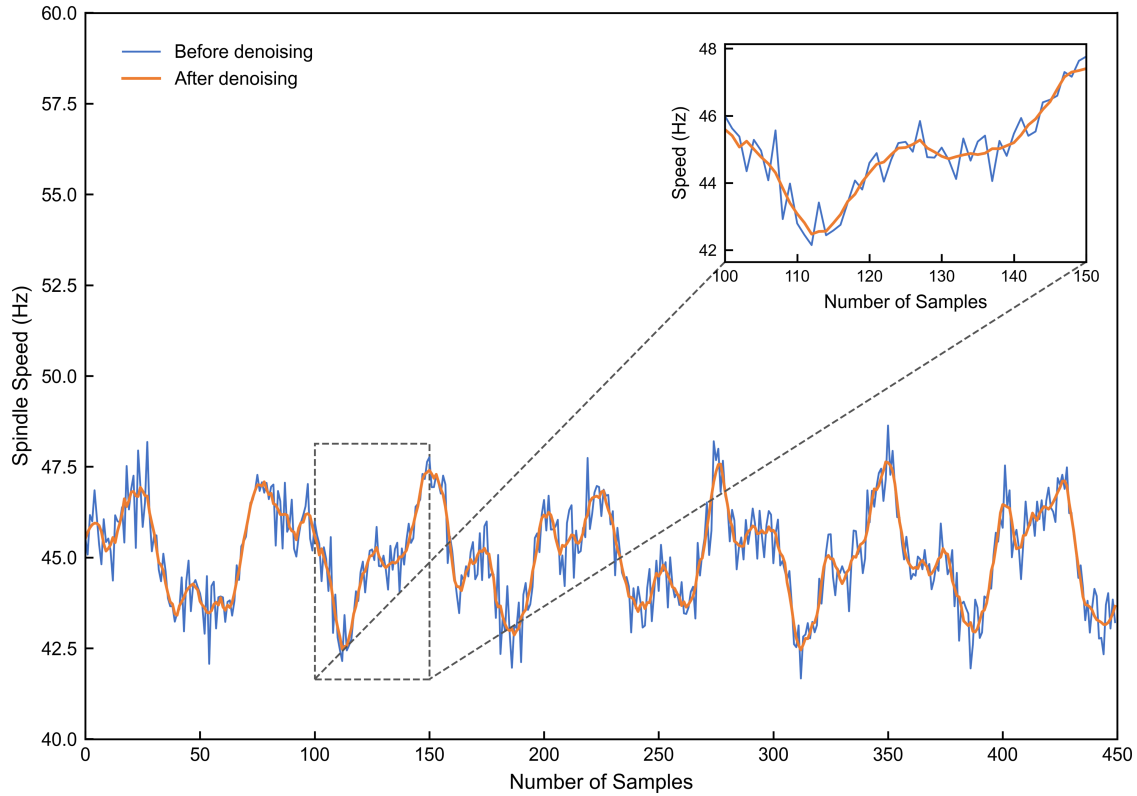


Figure 6: Data denoising results.

3.2.2 Feature Extraction

The original vibration signals of the bearing ring CNC grinding machine exhibit strong non-stationary characteristics due to the influence of operating conditions such as rotational speed fluctuations and sudden load changes, making it difficult to capture the essential characteristics of faults through direct modeling. Therefore, time-domain, frequency-domain, and time-frequency-domain feature extraction is performed on the non-stationary vibration signals of the CNC grinding machine.

For each monitoring signal, 6 time-domain features are extracted, including root mean square, peak-to-peak value, kurtosis, crest factor, impulse factor, and skewness. Additionally, 3 frequency-domain features are extracted, comprising the spectral skewness coefficient, spectral kurtosis coefficient, and spectral power, along with 8 sub-band energy features based on wavelet packet energy decomposition. Thus, a total of 34 features are extracted from the two types of vibration signals. For the other 12 types of stationary signals, data preprocessing is performed first, and then the root mean square feature value is extracted for each type of signal. Ultimately, a 46-dimensional feature set is obtained in total.

3.2.3 Feature Selection and Sensitivity Analysis Based on Pearson-ReliefF

Industrial monitoring datasets cover multi-dimensional feature parameters. Some features have no significant correlation with the fault prediction model. Redundant and irrelevant features can easily cause model underfitting, reduce identification accuracy, and weaken algorithm efficiency. The original feature

set in this study is 46-dimensional, and dimensionality reduction is required to retain key diagnostic information and eliminate redundant features.

First, Pearson correlation analysis is applied to the training set to reduce redundant components among the feature set. Following established standards in the field of feature selection [7], when the absolute value of the correlation coefficient between two features exceeds 0.8, significant linear redundancy is considered to exist. Accordingly, features identified as linearly redundant based on the Pearson coefficient are filtered out during the experiment. Ultimately, 28 mutually non-redundant features are retained, as listed in [Table 2](#).

Table 2: Optimal features.

Feature Category	Feature Content
Spindle vibration acceleration	Root mean square; Kurtosis; Skewness; Spectral skewness coefficient; Spectral kurtosis coefficient; WPE(3, 1); WPE(3, 3); WPE(3, 5); WPE(3, 7)
Feed axis vibration acceleration	Root mean square; Impulse factor; Skewness; Spectral skewness coefficient; Spectral kurtosis coefficient; WPE(3, 2); WPE(3, 4); WPE(3, 6); WPE(3, 8)
Stationary signal features	Root mean square of spindle motor current; Root mean square of feed axis motor current; Mean value of spindle motor temperature; Mean value of feed axis lead screw temperature; Mean value of feed position deviation; Mean value of feed speed; Mean value of lubrication system pressure; Mean value of hydraulic pump pressure; Mean value of grinding wheel spindle speed; Intermediate dressing compensation amount of the grinding wheel

Since Pearson correlation analysis can only measure the degree of linear correlation between features, it cannot effectively identify the nonlinear relationships between features and fault categories. Furthermore, it is difficult to quantify the contribution weight of features to fault diagnosis, which easily leads to the mistaken filtering of some fault-sensitive nonlinear features. Therefore, this study adopts the ReliefF algorithm to precisely quantify the contribution of each feature to fault diagnosis by calculating the correlation weight between the features and the fault categories on the training set. Based on a weight threshold, features with low contributions are eliminated to further compress the feature dimensions. This process screens out the core feature subset that is most sensitive to the grinding machine's fault trends, providing high-quality input parameters for the construction of subsequent fault prediction models.

After calculation, the importance weights of the respective feature variables are shown in [Table 3](#):

Table 3: Importance weights of feature variables.

Rank	Feature Name	Weight	Rank	Feature Name	Weight
1	Mean value of lubrication system pressure	0.185	15	Root mean square of feed axis motor current	0.052
2	Mean value of hydraulic pump pressure	0.178	16	Spindle vibration acceleration-WPE(3, 3)	0.045
3	Root mean square of spindle motor current	0.165	17	Feed axis vibration acceleration-Kurtosis	0.04
4	Mean value of feed position deviation	0.155	18	Spindle vibration acceleration-Spectral kurtosis coefficient	0.035

(Continued)

Table 3 (continued)

Rank	Feature Name	Weight	Rank	Feature Name	Weight
5	Spindle vibration acceleration-Kurtosis	0.138	19	Feed axis vibration acceleration-WPE(3, 2)	0.03
6	Feed axis vibration acceleration-Impulse factor	0.125	20	Feed axis vibration acceleration-Spectral skewness coefficient	0.025
7	Mean value of feed axis lead screw temperature	0.11	21	Spindle vibration acceleration-Skewness	0.02
8	Spindle vibration acceleration-WPE(3, 7)	0.105	22	Feed axis vibration acceleration-Skewness	0.015
9	Spindle vibration acceleration-WPE(3, 5)	0.098	23	Feed axis vibration acceleration-WPE(3, 4)	0.012
10	Intermediate dressing compensation amount of the grinding wheel	0.092	24	Spindle vibration acceleration-WPE(3, 1)	0.01
11	Mean value of spindle motor temperature	0.088	25	Feed axis vibration acceleration-WPE(3, 6)	0.008
12	Feed axis vibration acceleration-WPE(3, 8)	0.075	26	Spindle vibration acceleration-Spectral skewness coefficient	0.005
13	Spindle vibration acceleration-Root mean square	0.065	27	Mean value of grinding wheel spindle speed	0.002
14	Feed axis vibration acceleration-Root mean square	0.058	28	Mean value of feed speed	0.001

As shown in Table 3, the feature weight distribution is highly consistent with the fault degradation mechanisms of the bearing ring CNC grinding machine. The mean lubrication system pressure and mean hydraulic pump pressure have the highest weights of 0.185 and 0.178, respectively, indicating their strong sensitivity to lubrication and hydraulic circuit faults. The root mean square of spindle motor current also shows a high weight of 0.165, as grinding wheel dulling and spindle bearing wear increase spindle load and current fluctuations. The mean feed position deviation and mean feed-axis ball screw temperature reflect the positioning error and frictional temperature rise caused by ball screw wear, jointly characterizing its degradation process. Vibration-related features, such as spindle vibration kurtosis, feed-axis vibration impulse indicator, and wavelet packet energy, can capture impact signals and energy variations induced by bearing damage and grinding wheel anomalies. However, due to their delayed response at early fault stages, their weights remain moderate. Other parameters, including feed vibration indicators, spindle speed, and feed speed, mainly reflect routine operating conditions and show weaker correlations with fault evolution, resulting in lower weights. Overall, the ReliefF weight distribution agrees well with the physical degradation mechanisms of different fault types, confirming the rationality and reliability of the feature selection results.

To further visually verify the discriminative capability of each feature across different fault states, the sample distributions of feature variables under different fault conditions of the bearing ring CNC grinding machine are presented, as shown in Fig. 7.

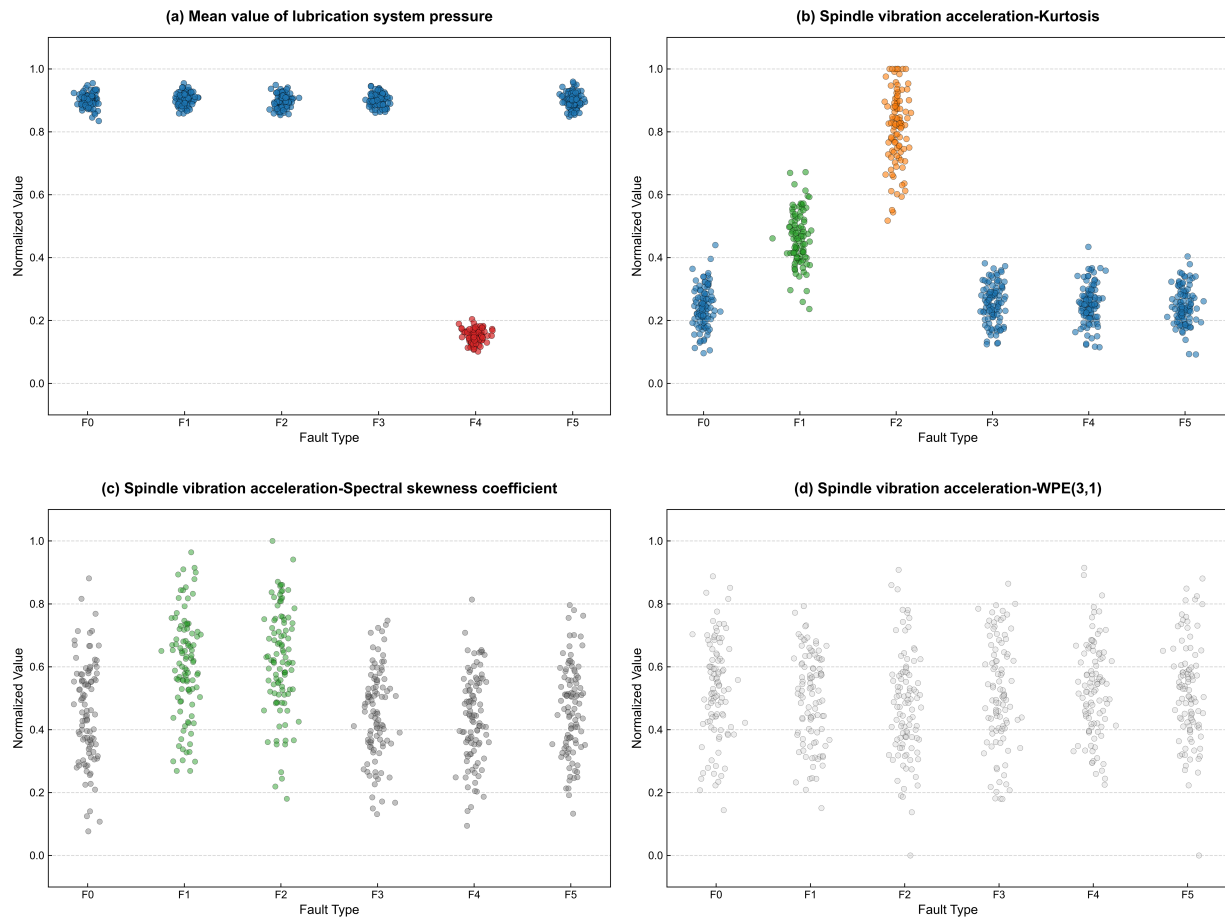


Figure 7: Sample distributions of selected parameters under various fault types.

From the sample distribution diagram of the mean value of the lubrication system pressure, it can be seen that the sample distributions under different fault states have significant distinctiveness. The sample points corresponding to various faults show a clear clustering and separation trend in space, and the sample interval differences between the normal state and fault states are clear. The spindle vibration acceleration-kurtosis can be distinguished, but the dispersion of the data points is slightly larger, leading to a weakened effect. Although the sample distribution of the spindle vibration acceleration-spectral skewness coefficient shows trend differences, there are obvious overlapping areas, and the classification boundaries are blurred. Regarding the sample distribution of the spindle vibration acceleration-WPE (3, 1), the sample points under different fault states are highly aggregated, the intervals overlap severely, and there are no obvious clustering boundaries. It is difficult to distinguish the samples between the fault states and the normal state; thus, its fault separability is poor.

It can be concluded that the ReliefF separability metric can effectively quantify the fault separation capability of features. In subsequent algorithm diagnostics, prioritizing features with high separability metrics and good fault separation degrees will help improve the comprehensive performance of the overall prediction and diagnosis model for bearing ring CNC grinding machines.

To identify the optimal feature subset, features ranked by ReliefF weights from the top 8 to 18 dimensions were sequentially used as inputs, and the classification accuracy was evaluated on the validation set. The results are shown in [Fig. 8](#).

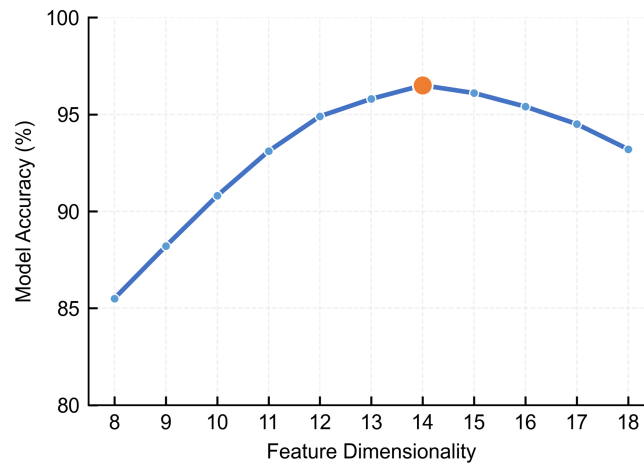


Figure 8: Relationship between feature dimensionality and model accuracy.

As can be seen from Fig. 8, the test accuracy of the model generally shows an upward trend as the feature dimensionality increases from 8 to 14. Therefore, this paper selects the top 14 features ranked by ReliefF weight to construct the final feature vector, denoted as the ReliefF optimal feature set, providing highly identifiable input parameters for the subsequent diagnosis model.

3.3 Simulation and Result Analysis of the BiGRU-Attention Model

In constructing the BiGRU model, a two-layer hidden layer architecture is adopted. The model takes a historical feature sequence of length $L = 30$ steps as input with a prediction horizon of $h = 30$ steps, and training samples are constructed by sequentially sliding with a step size of $S = 1$ step over the feature sequences of each fault condition. The model parameter configuration is shown in Table 4.

Table 4: Parameter configuration of the BiGRU-Attention model.

Network Structure	Parameter	Value
BiGRU-L1	Units	32
	return_sequences	TRUE
	dropout	0.3
BiGRU-L2	Units	16
	return_sequences	TRUE
	dropout	0.3
Attention	Type	Bahdanau
	Output_dim	16
Fully connected output layer	Dense	(14, 'linear')
Batch Size	–	64
Epochs	–	100
Optimizer	–	Adam
Learning rate	–	0.001
Loss function	–	MSE
Early stopping	patience	15

After completing the parameter configuration and training of the BiGRU-Attention model, an independent test set was used to evaluate its predictive performance. The changes in the training and validation losses during model training are shown in Fig. 9. Fig. 10 presents a comparison between the predicted and actual values of the pressure-related feature parameters of the lubrication system in the CNC bearing grinding machine.

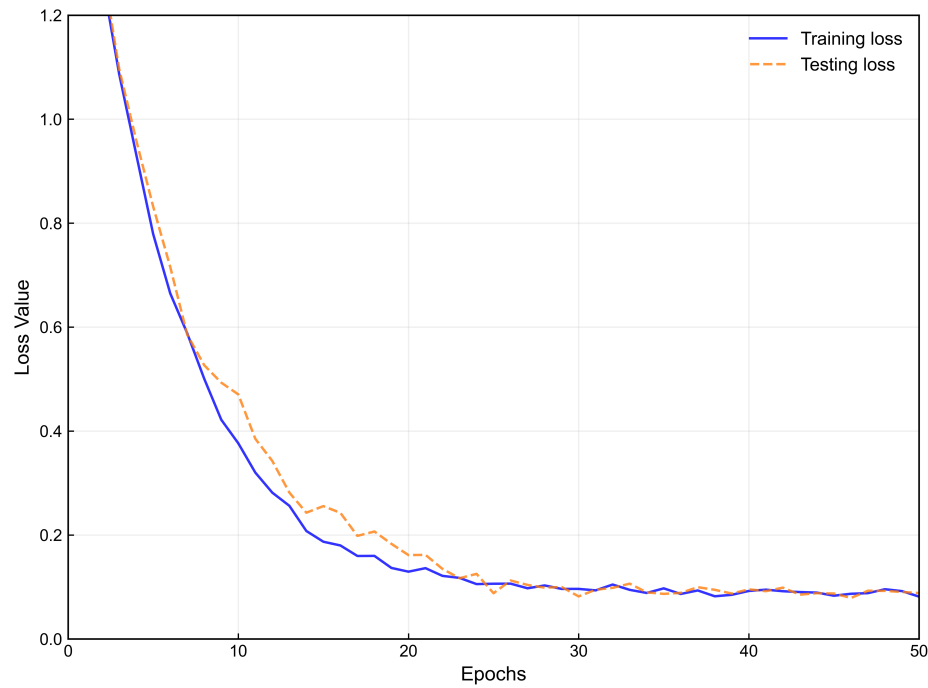


Figure 9: Model loss function.

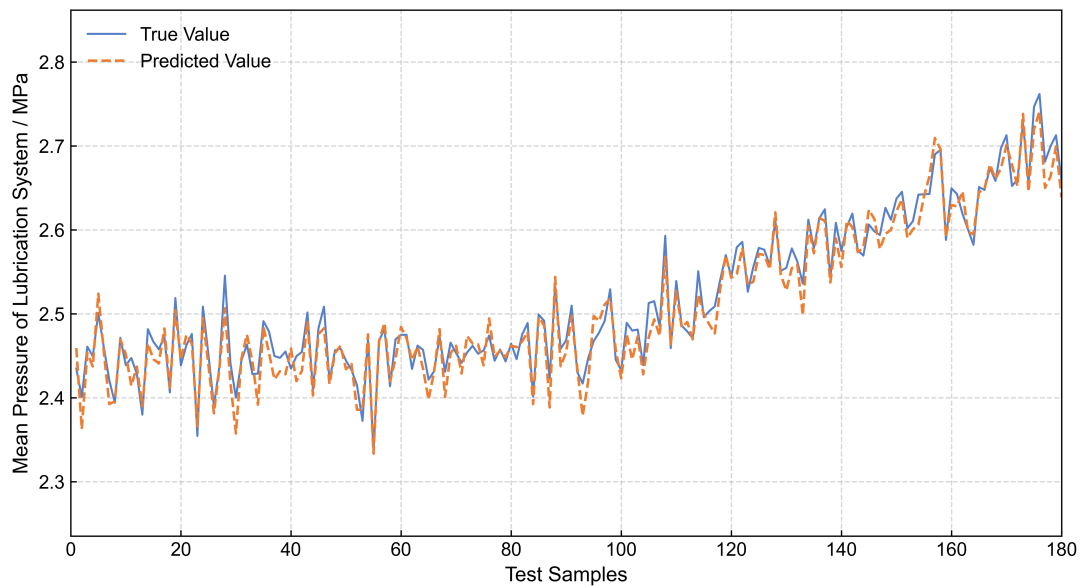


Figure 10: Prediction results of the BiGRU-Attention model.

Based on Fig. 9, both the training loss and validation loss of the BiGRU-Attention model decrease rapidly during the initial training stage and gradually stabilize after approximately 30 epochs. The training and validation loss curves exhibit highly consistent trends, with only a small difference between them, indicating that the model training process converges stably without evident overfitting. As shown in Fig. 10, the predicted values are highly consistent with the actual values, and the two curves show similar dynamic variation trends. This indicates that the model can accurately capture the temporal fluctuation characteristics of industrial parameters, thereby validating its effectiveness. The final predictive performance of the model on the independent test set is evaluated using MAE, RMSE, and R^2 , as reported in Table 5.

Table 5: SSA algorithm parameter settings.

Parameter	Symbol	Value	Parameter	Symbol	Value
Population size	N	30	Maximum iterations	$T_{\max}$	100
Discoverer ratio	PD	0.2	Scout ratio	SD	0.1
Safety threshold	ST	0.8	Fitness function	$f(C, \gamma)$	$1 - \text{Acc}_{\text{val}}$
Penalty factor range	C	[0.1, 100]	Kernel parameter range	γ	[0.001, 10]

To validate the predictive performance of the BiGRU-Attention model, XGBoost, BiLSTM, BiGRU, TCN, BiLSTM-Attention, and PatchTST are selected as baseline comparison models. To ensure fairness in comparison, all deep learning models adopt a unified data splitting strategy, identical input sequence lengths and prediction horizons, and are trained with the Adam optimizer, MSE loss function, and 100 training epochs, with each model's network architecture parameters independently determined through grid search on the validation set. XGBoost is configured with 100 trees, a maximum depth of 6, and a learning rate of 0.1, with a sliding window mechanism applied to ensure consistency in historical information with the sequential models. The predictive performance of all models is comprehensively evaluated using three metrics: MAE, RMSE, and R^2 , with the comparison results presented in Table 6.

Table 6: Comparison of prediction performance among different prediction models.

Prediction Model	MAE	RMSE	R^2
XGBoost	0.0342	0.0415	0.885
BiLSTM	0.0215	0.0268	0.938
BiGRU	0.0198	0.0245	0.945
TCN	0.0181	0.0224	0.952
BiLSTM-Attention	0.0156	0.0189	0.966
PatchTST	0.0148	0.0183	0.971
BiGRU-Attention	0.0124	0.0152	0.975

As shown in Table 6, the BiGRU-Attention model achieves the best results across all three metrics of MAE, RMSE, and R^2 . The advantage is most pronounced compared to XGBoost, with MAE and RMSE reduced by 0.0218 and 0.0263, respectively. Compared to recurrent networks such as BiLSTM and BiGRU, the incorporation of the attention mechanism further improves prediction accuracy. Compared to advanced architectures such as TCN and PatchTST, the proposed method still maintains a performance advantage in small-sample industrial scenarios. These results demonstrate that through the synergistic effect of bidirectional temporal modeling and attention-based dynamic weighting, BiGRU-Attention outperforms

all baseline models in both prediction accuracy and goodness of fit under small-sample industrial conditions, exhibiting stronger adaptability to industrial deployment scenarios.

3.4 Simulation and Result Analysis of the SSA-SVM Model

This paper introduces the SSA algorithm to perform adaptive optimization of the core parameters of the SVM model. The key algorithm parameter configurations are listed in Table 6. The population size is set to $N = 30$, the maximum number of iterations to $T_{\max} = 100$, the proportions of discoverers and scouts to 0.2 and 0.1, respectively, and the safety threshold to $ST = 0.8$. During the optimization process, the fitness function is defined as $f(C, \gamma) = 1 - \text{Acc}_{\text{val}}$, where Acc_{val} denotes the classification accuracy of the SVM on the validation set.

Within the search space of $C \in [0.1, 100]$ and $\gamma \in [0.001, 10]$, the SSA algorithm converges at the 22nd iteration, ultimately determining the globally optimal parameter combination as $C = 38.5$ and $\gamma = 0.56$. The optimal parameters are then fed into the SVM model, yielding the final model configuration as shown in Table 7.

Table 7: Optimized parameters of the SVM model.

Parameter	Value	Parameter	Value
Kernel	'rbf'	Probability	TRUE
Penalty coefficient (C)	38.5	Kernel function parameter (γ)	0.56

After SSA parameter optimization, the obtained optimal parameter combination is input into the SVM model to construct the SSA-SVM fault diagnosis model. Once the model is determined, the previously obtained predicted parameter values are used as inputs for the SVM multi-classification model for fault diagnosis, identifying potential future faults of the bearing ring CNC grinding machine. The fault classification results are displayed through a fault identification diagram, as shown in Fig. 11.

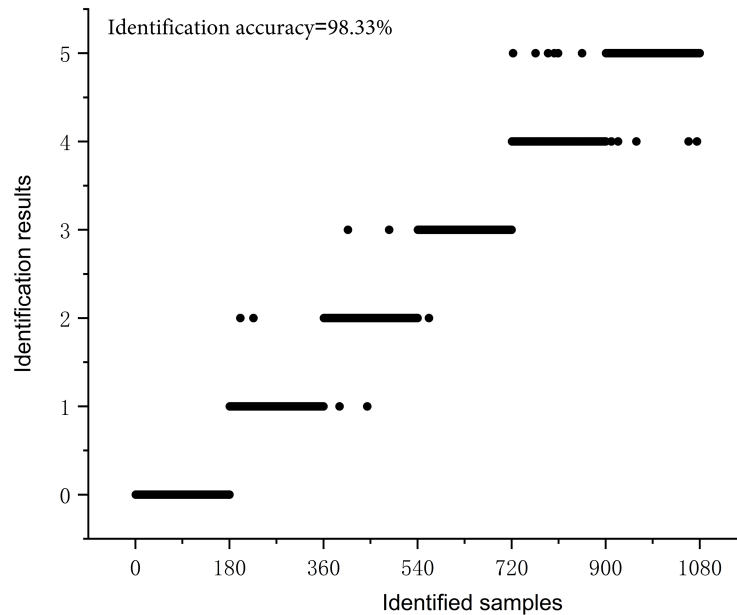


Figure 11: Fault identification diagram.

To ensure the robustness of the algorithmic model, repeated experiments are conducted using 10 different random seeds, while keeping the 7:3 temporal split unchanged. The classification performance of the SSA-SVM model on the same test set is recorded, with results presented in [Table 8](#).

Table 8: Results of 10 repeated experiments with different random seeds for SSA-SVM model.

No.	Random Seed	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
1	seed = 0	98.33	98.12	98.06	98.09
2	seed = 1	98.04	97.90	97.72	97.81
3	seed = 2	98.50	98.29	98.26	98.28
4	seed = 3	97.98	97.75	97.69	97.72
5	seed = 4	98.41	98.21	98.14	98.18
6	seed = 5	98.27	98.03	97.96	98.00
7	seed = 6	98.64	98.54	98.58	98.63
8	seed = 7	97.87	97.72	97.54	97.63
9	seed = 8	98.30	98.03	98.15	98.08
10	seed = 9	98.16	97.94	97.90	97.92
Mean	—	98.25	98.05	98.00	98.03
Std	—	± 0.24	± 0.25	± 0.31	± 0.29

As shown in [Table 8](#), across 10 repeated experiments with different random seeds, the SSA-SVM model achieves a mean accuracy of 98.25% with a standard deviation of only $\pm 0.24\%$, with individual results consistently distributed within the range of 97.87% to 98.64%. The negligible fluctuation range demonstrates that the proposed method exhibits strong stability and robustness, yielding reliable conclusions.

To further investigate the performance of the SSA-SVM model and validate the effectiveness and superiority of the SSA algorithm in SVM hyperparameter optimization, BPNN, RF, PSO-SVM, XGBoost, and SVM are constructed as baseline comparison models, each subjected to the same 10 repeated experiments with different random seeds. To ensure fairness in comparison, SVM is modeled with default parameters; BPNN, RF, and XGBoost are each tuned via grid search combined with validation set evaluation under the same temporal data splitting strategy as SSA-SVM; PSO-SVM and SSA-SVM adopt the same rbf kernel function and consistent fitness evaluation mechanism, with optimization conducted within a unified parameter search space of $C \in [0.1, 100]$ and $\gamma \in [0.001, 10]$. The confusion matrices and performance metric comparisons for all models are presented in [Fig. 12](#) and [Table 9](#), respectively.

As shown in [Table 9](#), SSA-SVM achieves the best performance across all four metrics, with a mean accuracy of 98.25% and a standard deviation of only $\pm 0.24\%$. It outperforms the unoptimized SVM by 12.04 percentage points, surpasses the heuristic-optimized PSO-SVM by 3.50 percentage points with superior stability, and exceeds the ensemble models RF and XGBoost by 5.29 and 1.90 percentage points, respectively, demonstrating the optimization advantage of the SSA algorithm and the robustness of the proposed model. Furthermore, paired-sample t -tests conducted on 10 repeated experimental results indicate that the accuracy differences between SSA-SVM and all baseline models reach a highly significant level ($p < 0.001$), confirming that the performance improvements achieved by the proposed method are statistically significant.

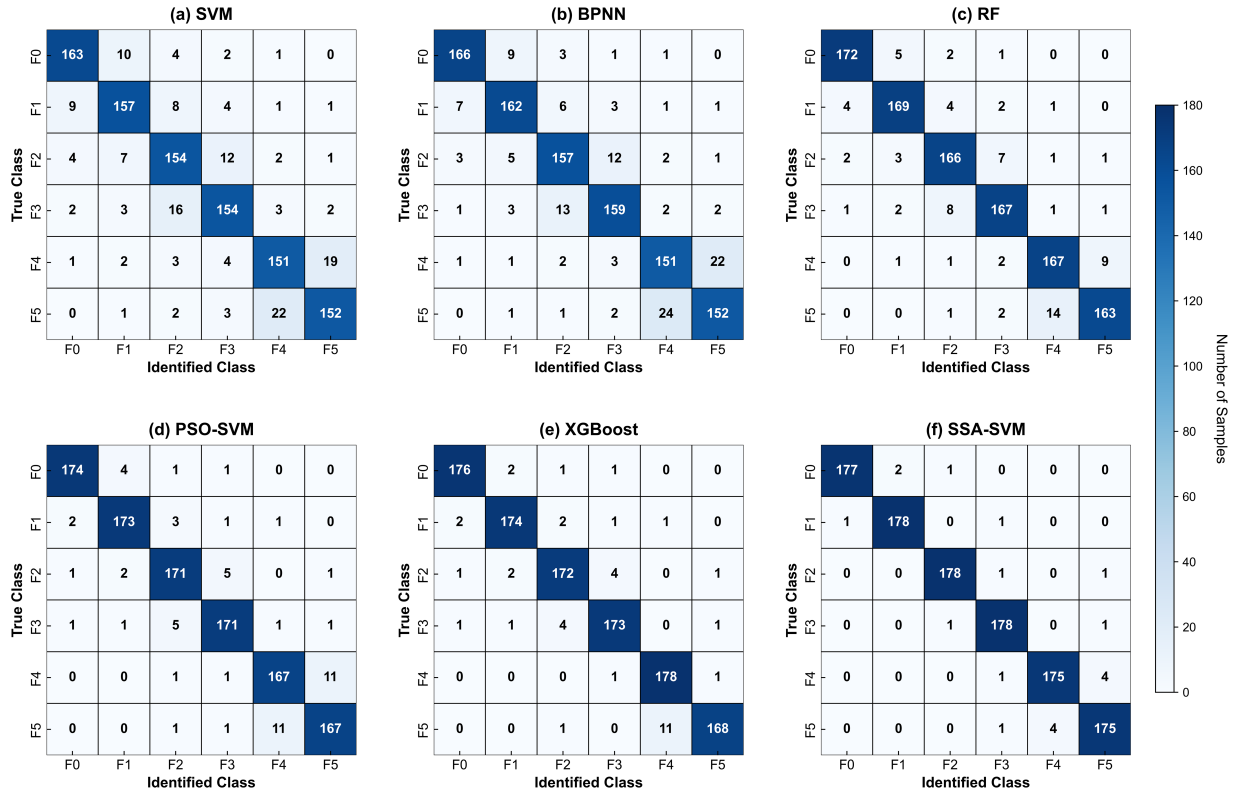


Figure 12: Confusion matrices of different models.

Table 9: Performance evaluation metrics of different models (%).

Classification Model	Accuracy (Mean $\pm$ Std)	Precision (Mean $\pm$ Std)	Recall (Mean $\pm$ Std)	F1 (Mean $\pm$ Std)	t	p	Significance
SVM	86.21 $\pm$ 0.61	85.75 $\pm$ 0.57	86.03 $\pm$ 0.64	85.88 $\pm$ 0.59	59.1	<0.001	***
BPNN	87.71 $\pm$ 0.64	88.12 $\pm$ 0.62	88.10 $\pm$ 0.66	87.83 $\pm$ 0.63	41.5	<0.001	***
RF	92.96 $\pm$ 0.47	93.15 $\pm$ 0.55	92.74 $\pm$ 0.49	93.03 $\pm$ 0.61	32.6	<0.001	***
PSO-SVM	94.75 $\pm$ 0.48	94.63 $\pm$ 0.61	94.96 $\pm$ 0.52	94.79 $\pm$ 0.44	20.4	<0.001	***
XGBoost	96.35 $\pm$ 0.36	96.56 $\pm$ 0.42	95.98 $\pm$ 0.29	96.27 $\pm$ 0.31	18.1	<0.001	***
SSA-SVM	98.25 $\pm$ 0.24	98.05 $\pm$ 0.25	98.00 $\pm$ 0.31	98.03 $\pm$ 0.29	—	—	—

3.5 Ablation Study

To verify the contribution of each module to the overall model performance, this section takes the full model as the baseline and adopts a stepwise removal strategy, sequentially eliminating the Pearson-ReliefF feature selection, the attention mechanism, and the SSA optimization algorithm to construct three ablation models. All other conditions remain unchanged across models, and the experimental results are presented in Table 10.

Table 10: Results of the ablation study.

Model	Dim	MAE	RMSE	R ²	Accuracy	F1
Model I	14	0.0124	0.0152	0.975	98.25%	98.03%
Model II	46	0.0241	0.0293	0.929	92.37%	92.15%
Model III	14	0.0198	0.0245	0.945	94.82%	94.65%
Model IV	14	0.0124	0.0152	0.975	85.42%	85.32%

Among them, Model I represents the complete framework proposed in this paper (Pearson-ReliefF + BiGRU-Attention + SSA-SVM); Model II removes the Pearson-ReliefF feature selection from Model I; Model III removes the attention mechanism from Model I; and Model IV removes the SSA optimization algorithm from Model I.

As shown in Table 10, the complete framework Model I achieves the best overall performance, with a classification accuracy and F1 score of 98.25% and 98.03%, respectively. The ablation results for each module indicate that: removing the Pearson-ReliefF feature selection causes interference from redundant features to significantly increase prediction error, resulting in the largest performance drop of 5.88 percentage points in fault identification accuracy; removing the attention mechanism weakens the model's ability to focus on key temporal features, leading to a decline in prediction precision and a reduction of 3.43 percentage points in fault identification accuracy; removing the SSA optimization algorithm leaves prediction-stage performance unaffected, but causes a substantial drop of 12.83 percentage points in fault identification accuracy.

In summary, the three modules contribute at the levels of feature dimensionality reduction, key information enhancement, and parameter optimization, respectively, working in a mutually supportive and synergistic manner. The Pearson-ReliefF method ensures the quality of model inputs; the attention mechanism strengthens the model's ability to perceive critical degradation-related temporal patterns; and the SSA algorithm significantly improves classifier accuracy through optimization. All three components are indispensable, jointly ensuring high accuracy and strong robustness of the fault diagnosis model.

3.6 Engineering Applicability Analysis

This section systematically validates the applicability of the proposed method in practical industrial deployment scenarios from two dimensions: prediction error propagation and step size selection, and noise robustness.

3.6.1 Prediction Error Propagation and Step Size Selection

In the fault diagnosis stage, the output of the BiGRU-Attention model is used as the input to the SSA-SVM classifier; therefore, prediction errors may propagate downstream and affect diagnostic performance. To quantitatively evaluate this uncertainty propagation effect, this study conducts a joint analysis from two aspects. First, diagnostic metrics under different prediction horizons are evaluated on the test set to measure the impact of error accumulation on fault identification. Second, Monte Carlo Dropout (MC Dropout) is used to perform approximate uncertainty quantification for the prediction results of the BiGRU-Attention model, and prediction interval evaluation metrics are combined to analyze the effect of predictive uncertainty on the downstream SSA-SVM diagnostic performance.

Specifically, MC Dropout keeps the Dropout layers active during inference with a dropout rate of 0.3, and performs ($T = 100$) stochastic forward passes for each input sequence. The predictive mean is used as the final predicted value, while the standard deviation is used to quantify epistemic uncertainty and construct

the 95% prediction interval. The predictive mean is then fed into the trained SSA-SVM classifier to obtain the diagnostic results, and accuracy, precision, recall, and F1 score are recorded under different prediction horizons. Meanwhile, Prediction Interval Coverage Probability (PICP) and Mean Prediction Interval Width (MPIW) are used to further evaluate the quality of the prediction intervals and analyze the effect of prediction uncertainty on downstream diagnostic performance. The corresponding formulations are given in Eqs. (17) and (18):

$$PICP = \frac{1}{N} \sum_{i=1}^N c_i, \quad c_i = \begin{cases} 1, & y_i \in [L_i, U_i] \\ 0, & otherwise \end{cases} \quad (17)$$

$$MPIW = \frac{1}{N} \sum_{i=1}^N (U_i - L_i) \quad (18)$$

where N is the total number of samples, y_i denotes the true value, and L_i and U_i represent the lower and upper bounds of the prediction interval, respectively. PICP reflects the proportion of samples whose true values fall within the prediction interval; a value closer to the 95% target coverage probability indicates more reliable interval calibration. MPIW denotes the mean width of the prediction interval; under the premise of guaranteed coverage, a smaller value indicates a more precise uncertainty estimation. Under prediction step sizes of $h \in \{10, 20, 30, 40, 50\}$, all performance metrics are comprehensively recorded, with results presented in Table 11.

Table 11: Diagnostic performance analysis at different prediction step sizes.

Prediction Step	RMSE	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)	PICP	MPIW
10	0.0108	98.71	98.53	98.47	98.50	0.938	0.043
20	0.0133	98.46	98.28	98.21	98.24	0.946	0.061
30	0.0152	98.22	98.04	97.96	98.00	0.953	0.087
40	0.0203	96.85	96.61	96.54	96.57	0.961	0.121
50	0.0267	94.31	94.05	93.87	93.96	0.967	0.149

As shown in Table 11, as the prediction horizon increases, RMSE rises from 0.0108 to 0.0267, while accuracy and F1 score decrease from 98.71% and 98.50% to 94.31% and 93.96%, respectively, indicating progressive error accumulation over longer horizons. Meanwhile, MPIW increases from 0.043 to 0.149, and PICP rises from 0.938 to 0.967, remaining generally close to the 95% target coverage probability. This suggests that the MC Dropout-based prediction intervals become wider with longer horizons but maintain good calibration.

Specifically, when $h = 10\text{--}30$, fault diagnosis accuracy decreases by only 0.49 percentage points, and MPIW increases moderately, indicating that both diagnostic performance and predictive uncertainty remain within acceptable ranges. When $h = 30\text{--}50$, the accuracy decline expands to 3.91 percentage points, and MPIW increases more rapidly, showing a consistent stepwise degradation in diagnostic performance and uncertainty control.

Considering early-warning lead time, diagnostic performance, and predictive uncertainty, this study selects ($h = 30$) as the prediction horizon. Under this setting, the model achieves an accuracy of 98.22%, an F1 score of 98.00%, a PICP of 0.953, and an MPIW of 0.087, providing a 30-s advance warning while maintaining accurate fault diagnosis and reliable prediction interval estimation. In addition, the single-inference latency

is approximately 98 ms, which is lower than the sampling interval, enabling real-time deployment on the MES server without upgrading the existing industrial computing hardware.

3.6.2 Noise Robustness Validation

In practical industrial deployments, sensor signals are susceptible to noise fluctuations caused by factors such as electromagnetic interference in workshops, equipment vibration, and the start-stop cycles of cooling systems. To evaluate the robustness of the proposed model under noisy conditions, Gaussian white noise at varying signal-to-noise ratios (SNRs) was superimposed onto the test set feature sequences, with SNR values set to 20, 15, 10, and 5 dB in sequence, simulating interference environments ranging from mild to severe. A systematic comparison of fault classification performance was conducted across the conventional SVM, BPNN, RF, PSO-SVM, XGBoost, and the proposed SSA-SVM model, with results presented in [Table 12](#).

Table 12: Fault identification accuracy of different models under various noise levels (%).

Noise Condition	SVM	BPNN	RF	PSO-SVM	XGBoost	SSA-SVM
No noise	85.42	87.84	92.60	94.80	95.45	98.25
SNR = 20 dB	82.84	85.72	91.14	93.83	94.24	97.65
SNR = 15 dB	76.99	80.65	87.95	91.42	91.62	96.10
SNR = 10 dB	71.09	75.75	84.70	89.10	89.00	94.55
SNR = 5 dB	62.98	68.88	80.20	85.88	85.37	92.42

As shown in [Table 12](#), as the level of noise interference increases, the classification accuracy of all models declines to varying degrees; however, the SSA-SVM proposed in this study consistently maintains the best classification performance and the smallest degradation margin across all noise levels. Under light and moderate interference conditions at SNR = 20 and 15 dB, the SSA-SVM achieves accuracies of 97.65% and 96.10%, respectively, with a maximum decline of only 2.15 percentage points relative to the noise-free baseline. By contrast, the conventional SVM's accuracy drops from 85.42% to 76.99% at SNR = 15 dB, a degradation of 8.43 percentage points, at which point the performance gap between the two models begins to emerge. When conditions deteriorate to SNR = 5 dB, the accuracies of SVM and BPNN fall sharply to 62.98% and 68.88%, respectively, and PSO-SVM, a similarly optimized model of the same category, also declines to 85.88%. In contrast, the SSA-SVM still maintains a high recognition rate of 92.42%, with an overall maximum degradation of only 5.83%, demonstrating exceptionally strong robustness against interference.

These results indicate that the Pearson-ReliefF feature selection stage effectively filters out low-contribution redundant features, improving the overall signal-to-noise ratio of the model inputs. The global dynamic optimization mechanism of the SSA algorithm enables the SVM hyperparameters to maintain an optimal classification boundary even under noisy conditions, preventing the model from converging to local optima on corrupted data. As a result, the model sustains stable robustness under typical industrial noise interference scenarios and can accommodate signal quality fluctuations encountered in real-world deployments.

3.6.3 Cross-Device Applicability Validation

To further examine the generalization capability of the proposed fusion-based prediction framework under different equipment and operating conditions, this study selects the inner-ring raceway grinding machine and the outer-ring rabbit grinding machine as extended validation objects, in addition to the

inner ring rib grinding machine. The inner-ring raceway grinding machine belongs to the same inner-ring machining line as the original equipment but corresponds to a different process, whereas the outer-ring rabbet grinding machine belongs to the outer-ring machining line, with the machining object changing from the inner ring to the outer ring. These devices are used to evaluate the model's applicability in cross-process, cross-production-line, and cross-machining-object scenarios.

All three devices are CNC grinding machines for bearing rings, with consistent key components, sensor configurations, data acquisition methods, and fault labeling rules. The fault types include normal operation and typical failures of the spindle, feed screw, grinding wheel, lubrication circuit, and hydraulic circuit, with labels denoted as F0–F5. During the experiments, the framework structure, input window length, prediction horizon, and main hyperparameters remain unchanged.

The cross-device applicability is evaluated through direct transfer and small-sample temporal fine-tuning. In the direct transfer stage, the target-device test data are directly input into the trained model without introducing target-device training samples. In the fine-tuning stage, the first 15% of the target-device data in chronological order is used for fine-tuning, and the remaining 85% is used for testing. The original model parameters are used for initialization, and both the prediction and diagnostic models are optimized. To reduce the influence of randomness, fine-tuning experiments are repeated using 10 different random seeds under the fixed chronological split. The equipment generalization validation results are shown in Table 13.

Table 13: Cross-device generalization validation results.

Equipment	Validation Method	R ²	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
Inner ring rib grinding machine	Original validation	0.975	98.25 ± 0.24	98.05 ± 0.25	98.00 ± 0.31	98.03 ± 0.29
Inner-ring raceway grinding machine	Direct transfer	0.928	92.36	92.14	91.87	92.00
	Small-sample temporal fine-tuning	0.958	96.18 ± 0.38	95.96 ± 0.42	95.88 ± 0.45	95.92 ± 0.41
Outer-ring rabbet grinding machine	Direct transfer	0.889	88.42	88.16	87.91	88.03
	Small-sample temporal fine-tuning	0.936	94.27 ± 0.46	94.05 ± 0.50	93.86 ± 0.54	93.95 ± 0.49

As shown in Table 13, under direct transfer conditions, the diagnostic accuracies of the inner-ring raceway grinding machine and the outer-ring rabbet grinding machine are 92.36% and 88.42%, respectively, representing decreases of 5.89 and 9.83 percentage points compared with the 98.25% accuracy of the original equipment. Nevertheless, both still maintain good recognition capability, indicating that the prediction and diagnostic models developed in this study possess a certain degree of cross-device transferability. Specifically, the inner-ring raceway grinding machine belongs to the same inner-ring production line as the original equipment, resulting in a smaller decline; whereas the outer-ring rabbet grinding machine involves cross-production-line and cross-machining-object conditions, leading to a larger decline. This further verifies the influence of operating-condition similarity on transfer performance.

After small-sample temporal fine-tuning, the mean diagnostic accuracies of the inner-ring raceway grinding machine and the outer-ring rabbet grinding machine increase to 96.18% and 94.27%, respectively, representing improvements of 3.82 and 5.85 percentage points compared with direct transfer. R^2 also increase from 0.928 and 0.889 to 0.958 and 0.936, respectively. These results indicate that a small number of continuous time-series samples from the target equipment can effectively alleviate the distribution shift caused by differences in equipment structure and operating conditions, thereby validating the cross-device generalization capability of the proposed fusion-based prediction framework in cross-process and cross-production-line scenarios.

4 Limitations and Future Work

Although the data-driven fault prediction method for bearing ring CNC grinding machines proposed in this paper achieves promising experimental results, several limitations remain that warrant further improvement in future work.

First, the data scale and scenario diversity are relatively limited. The experimental data were collected from specific production periods of bearing CNC grinders. Although the samples cover six types of fault operating conditions, each type contains only 600 samples, resulting in a relatively small overall dataset.

Second, the validation of cross-temporal generalization remains insufficient. Although cross-process and cross-production-line equipment validation has been conducted in [Section 3.6.3](#), and the results indicate that the model possesses a certain degree of cross-device transferability and fine-tuning adaptability, the existing data are still mainly derived from relatively short production periods. The stability of the model under cross-seasonal conditions, long-term operation, and gradual operating-condition drift remains to be further examined.

Third, external validation and long-term demonstration of large-scale deployment are lacking. Due to practical constraints in industrial data collection, no publicly available industrial datasets were incorporated for independent external validation, leaving the long-term stability of the proposed method under large-scale industrial deployment and varying workshop environments in need of further verification.

In response to the above limitations, future work will focus on three directions. First, the data scale and scenario coverage will be expanded by collecting data from multiple grinding machines of the same model across different production periods and factory environments. Data augmentation methods such as CGAN and SMOTE will also be introduced to alleviate fault-sample scarcity and improve generalizability. Second, transfer learning and domain adaptation will be explored to enable rapid model transfer using richly annotated source-equipment data and limited target-equipment samples, thereby improving robustness across different equipment, time periods, and operating conditions. Third, semi-supervised learning will be investigated to better exploit large amounts of unlabeled industrial data, reduce dependence on manual annotation, and enhance the long-term applicability of the method in large-scale industrial deployment.

5 Conclusion

This paper addresses the problems of high-dimensional monitoring parameters, serious information redundancy, production line downtime and reduced production efficiency caused by sudden faults of bearing ring CNC grinding machines under high-speed and heavy-load working conditions. A “prediction-first, diagnosis-later” method is proposed. By using the efficient feature selection ability of the Pearson-ReliefF hybrid algorithm, the model can accurately eliminate redundant parameters and retain fault-sensitive features, laying a high-quality data foundation for subsequent experiments. The BiGRU-Attention model is adopted to capture the time-series evolution laws of parameters and realize trend prediction of weak

early fault symptoms. Combined with the powerful classification performance of SSA-SVM, the accurate identification and classification of potential fault types are achieved.

Experimental results show that the prediction algorithm proposed in this paper achieves significant improvements in both prediction accuracy and fitting performance compared with mainstream benchmark models such as XGBoost and BiLSTM. It can realize the advance prediction of the equipment feature evolution trend within the future 30 s, and the average fault recognition accuracy reaches 98.25%. By deeply coupling the time-series state prediction model with the fault diagnosis model, this study effectively overcomes the shortcomings of single models and realizes the collaborative implementation of early anomaly warning and accurate fault location for bearing ring CNC grinding machines, providing reliable technical support for ensuring the stable operation of machine tools and reducing economic losses caused by sudden failures.

Acknowledgement: Not applicable.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: Yanan Wang; data collection: Zhijie Pei; analysis and interpretation of results: Xin Yang; draft manuscript preparation: Bo Li; refinement of study design, methodological guidance, and project supervision: Xiaoying Yang. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: Data not available due to commercial restrictions.

Ethics Approval: Not applicable.

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

References

1. Kumar S, Park HS, Nedelcu D. Development of real-time grinding process monitoring and analysis system. *Int J Precis Eng Manuf.* 2021;22(8):1345–55. doi:10.1007/s12541-021-00539-5.
2. Shi C, Chen B, Shi Y, Zha J. Surface roughness prediction of bearing ring precision grinding based on feature extraction. *Appl Sci.* 2025;15(11):6027. doi:10.3390/app15116027.
3. Guo D. Motor bearing fault diagnosis based on industrial Internet of Things and transfer learning. *Front Mech Eng.* 2025;11:1647310. doi:10.3389/fmech.2025.1647310.
4. Luo W, Hu T, Ye Y, Zhang C, Wei Y. A hybrid predictive maintenance approach for CNC machine tool driven by Digital Twin. *Robot Comput Integr Manuf.* 2020;65(1):101974. doi:10.1016/j.rcim.2020.101974.
5. Bai X, Liang H, Jia H, Fu Y, Pan Z. Multi-fault prediction of electromechanical actuators based on SCI-BiGRU. *IEEE Access.* 2025;13(3):168374–92. doi:10.1109/ACCESS.2025.3612910.
6. Bai Y, Zhao H, Tang H, Jia Z, Liu Y. Shield machine anomaly detection using hierarchical dimensionality reduction and multi-scale spatio-temporal attention residual network. *Measurement.* 2026;261(6):119971. doi:10.1016/j.measurement.2025.119971.
7. Zhou L, Zhao Q, Zeng J, Zhu A, Long X, He L. Fault diagnosis and data reconstruction of temperature sensors for wind turbine stator winding. *Shock Vib.* 2025;2025(1):4713545. doi:10.1155/vib/4713545.
8. Jiang X, Hu D. PGRS-FaultNet: a fusion model for fault detection and prediction in intelligent manufacturing. *J Circuits Syst Comput.* 2025;34(11):2550243. doi:10.1142/s0218126625502433.
9. Hu H, Feng F, Han J, Zhu J, Song C. Fault warning technology based on multivariate statistical analysis. *J Ind Eng Manag.* 2024;2(1):64–70. doi:10.62517/jiem.202403110.
10. Sun Y, Cao Y, Li P, Su S. Fault diagnosis for railway point machines using VMD multi-scale permutation entropy and ReliefF based on vibration signals. *Chin J Electron.* 2025;34(1):204–11. doi:10.23919/cje.2023.00.258.

11. Cui P, Li G, Zhang Q, Qi Z. Multiple domain identification of fault arc based on KPCA-LSTM method. *Comput Electr Eng*. 2025;123(1):110171. doi:10.1016/j.compeleceng.2025.110171.
12. Yang Y, Zhai J, Wang H, Xu X, Hu Y, Wen J. An improved fault diagnosis method for rolling bearing based on relief-F and optimized random forests algorithm. *Machines*. 2025;13(3):183. doi:10.3390/machines13030183.
13. Liu B, Wang X, Zhang Z, Zhao Z, Wang X, Liu T. Fault prediction of hydropower station based on CNN-LSTM-GAN with biased data. *Energies*. 2025;18(14):3772. doi:10.3390/en18143772.
14. Tan R, Wang Z, Chen Y, Zhou X, Zhu P, AL-Bukhaiti K, et al. CNN-BiLSTM-Autoencoder hybrid for prognostics of gearbox Over-Temperature faults in offshore wind turbines. *Sustain Energy Technol Assess*. 2026;85(4):104806. doi:10.1016/j.seta.2025.104806.
15. Rama VSB, Hur SH, Yang JM. Short-term fault prediction of wind turbines based on integrated RNN-LSTM. *IEEE Access*. 2024;12:22465–78. doi:10.1109/ACCESS.2024.3364395.
16. Ali AR, Kamal H. Hybrid HHO-WHO optimized transformer-GRU model for advanced failure prediction in industrial machinery and engines. *Sensors*. 2026;26(2):534. doi:10.3390/s26020534.
17. Cheng Y, Xue J, Lu M, Zhou S, Gai X, Guan R. MS-DenseNet-GRU tool wear prediction method based on attention mechanism. *Struct Health Monit*. 2025;24(1):517–33. doi:10.1177/14759217241240663.
18. Han K, Wang W, Guo J. Research on a bearing fault diagnosis method based on a CNN-LSTM-GRU model. *Machines*. 2024;12(12):927. doi:10.3390/machines12120927.
19. Lu Y, Chen F, Deng M, Jiang Z, Wu X, Wang B. Research on arc fault classification and identification of airborne ACIPDS based on GA-RF. *IEEE Access*. 2024;12(3):184381–92. doi:10.1109/ACCESS.2024.3510938.
20. Li Y. Research on bearing equipment fault diagnoses via SAWOA-LSTM. *J Meas Eng*. 2025;13(3):655–68. doi:10.21595/jme.2025.24805.
21. He L, Huang Z, Shao H, Hu Z, Wang Y, Mei J, et al. Fault diagnosis of wind turbine blades based on multi-sensor weighted alignment fusion in noisy environments. *Comput Mater Contin*. 2026;86(3):59. doi:10.32604/cmc.2025.073227.
22. Wei Y, Zhao J, Yang Z, Wang P, Zeng Z, Wang X. Fault detection method for flexible DC grid based on CEEMDAN multiscale entropy and GA-SVM. *Electr Eng*. 2024;106(5):6171–83. doi:10.1007/s00202-024-02349-0.
23. Atal JK, Rana SS, Dash RN, Sen A, Mishra SS. Robust fault classification of three phase induction motor using GWO-SVM with simulation and hardware validation. *IEEE Access*. 2025;13:203519–35. doi:10.1109/ACCESS.2025.3639068.
24. Zhu S, Chen P, Li X, Deng Q, Yan F. Mechanical fault diagnosis of high-voltage disconnectors via multi-domain energy features of vibration signals in power systems. *Processes*. 2025;13(10):3254. doi:10.3390/pr13103254.
25. Jiao J, Xie Z, Ding J, Sun Z, Huang S. A hybrid harmonic analysis and IVMD-BiGRU model for tide level forecasting. *Ocean Eng*. 2026;343(2):123569. doi:10.1016/j.oceaneng.2025.123569.
26. Zhang C, Li J, Xu L, Gao F. Rolling bearing fault diagnosis method based on TCN-BiGRU-Attention. In: *Proceedings of the 2024 3rd International Conference on Electronics and Information Technology (EIT)*; 2024 Sep 20–22; Chengdu, China. p. 327–30. doi:10.1109/EIT63098.2024.10762286.
27. Xu Z, Li YF, Huang HZ, Deng Z, Huang Z. A novel method based on CNN-BiGRU and AM model for bearing fault diagnosis. *J Mech Sci Technol*. 2024;38(7):3361–9. doi:10.1007/s12206-024-0610-2.
28. Cortes C, Vapnik V. Support-vector networks. *Mach Learn*. 1995;20(3):273–97. doi:10.1007/BF00994018.
29. Lei Z, Lin H, Tang X, Xiong Y, Wen H. Wind turbine blade fault detection method based on TROA-SVM. *Sensors*. 2025;25(3):720. doi:10.3390/s25030720.
30. Xue J, Shen B. A novel swarm intelligence optimization approach: sparrow search algorithm. *Syst Sci Control Eng*. 2020;8(1):22–34. doi:10.1080/21642583.2019.1708830.