Journal of Computer Science and Electrical Engineering
Print ISSN: 2663-1938
Online ISSN: 2663-1946
DOI: https://doi.org/10.61784/jcsee3164

FAULT PREDICTION OF FASTENER PRODUCTION EQUIPMENT VIA LIGHTWEIGHT REGRESSION MODELING WITH SCIKIT-LEARN: EVIDENCE FROM THE HAIYAN FASTENER INDUSTRY

Hang Shu*, TongYue Zhai, JiaYan Wang, Kun Hu
Haiyan College, Jiaxing Vocational and Technical College, Jiaxing 314036, Zhejiang, China.
**Corresponding Author: Hang Shu*

Abstract: In fastener manufacturing, an unexpected breakdown of a cold heading machine stops the entire production line, yet most small and medium-sized enterprises (SMEs) still maintain their equipment by rule of thumb. Targeting the SMEs in the Haiyan fastener industry cluster, this paper takes cold heading machines as the object and builds a low-cost equipment fault prediction method based on regression modeling and Scikit-learn. The remaining useful life (RUL) of key components is taken as the prediction target; four models, namely linear regression, random forest, gradient boosting regression tree (GBRT), and support vector regression (SVR), are compared, and grid search is used for parameter tuning. Since such enterprises generally lack usable historical failure data, the method is validated on real degradation data from the public XJTU-SY accelerated life test dataset of rolling bearings, which share the same data form as the key rotating components of cold heading machines. The results show that GBRT performs best (RMSE = 6.25 min, R^2 = 0.511), that degradation information is concentrated in a few amplitude-type statistical features such as mean amplitude, standard deviation and root mean square, and that in the early-warning replay on two test bearings, alarms are all raised ahead of failure with an average lead time of 8.0 min and no false alarm. The study indicates that lightweight AI tools are sufficient to support equipment fault prediction in county-level manufacturing SMEs, and the whole procedure can be replicated directly.
Keywords: Fault prediction; Regression modeling; Scikit-learn; Fastener manufacturing; Remaining useful life

1 INTRODUCTION

Unplanned downtime caused by sudden equipment failure is one of the most troublesome sources of loss in the production management of small and medium-sized manufacturing enterprises. The idea of predictive maintenance is to model equipment operating data continuously and to issue warnings before a failure occurs, and existing studies have shown that it can significantly reduce downtime losses and maintenance costs [1-2]. However, this approach is not easy to implement in county-level manufacturing clusters. Haiyan is one of the three largest fastener production bases in China; public information indicates that the county hosts more than 400 enterprises along the fastener industry chain, most of which are SMEs. Their cold heading machines operate continuously under heavily fluctuating loads, while maintenance still relies on repair-after-failure or scheduled overhaul. Statistical links between cold heading process parameters and equipment degradation have been documented [3-4], yet these data are rarely exploited. Meanwhile, the intelligent maintenance solutions adopted by large enterprises depend on costly sensing retrofits and professional algorithm teams, which SMEs can hardly afford [5].

A review of existing studies further reveals a clear gap. Reported achievements in equipment fault prediction concentrate on large-scale industrial scenarios such as hydropower, nuclear power and mining [6]. Machine learning has become the mainstream technical route of machine fault diagnosis [7], and data-driven prognostics, covering the full chain from data acquisition to remaining useful life (RUL) prediction, has been systematically reviewed [8]. Among these methods, deep learning models such as long short-term memory networks and multi-task networks have achieved high accuracy [9-10], but the data scale, computing resources and professional operation they demand are exactly what SMEs lack [11]. Comparative experiments also show that when the sample size is limited, classical machine learning models and deep models each have their own applicable boundaries, and lightweight models are by no means simply a low-configuration substitute [6]. Besides accuracy, SMEs care whether a model is understandable and easy to deploy: tree models combined with SHAP-based interpretability analysis have been used for machine tool fault diagnosis [12], and decision trees have realized online monitoring of motor vibration states [13], both showing the practical advantages of lightweight models in industrial sites. At the tool level, Scikit-learn covers the complete workflow of data preprocessing, feature engineering, model training and hyperparameter optimization through a unified API, runs on an ordinary office computer without dedicated hardware, and therefore matches the low-cost and easy-to-replicate constraints of SMEs [14]; a comparative study conducted in a manufacturing SME likewise confirms the suitability of classical machine learning models for predictive modeling in such settings [15]. Nevertheless, whether these lightweight tools work well on fastener production equipment data has not been empirically examined, and replicable

© By the Author(s) 2026, under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0).

cases in which vocational colleges apply mathematical modeling and lightweight AI tools to real data from local industries remain scarce.

Facing the maintenance demand of cold heading machines in Haiyan fastener SMEs, this paper follows a lightweight route of operating-parameter collection, regression modeling and fault warning. The problem is formulated as supervised regression: the operating state of the equipment at minute t is described by a feature vector x(t) composed of time-domain statistics extracted from vibration monitoring, the RUL at that moment, measured in minutes until failure, is taken as the label y(t), and a mapping from x(t) to y(t) is learned on training equipment with known life histories; once the predicted RUL first falls below a preset threshold T, an early warning is triggered. Four representative regression models, namely linear regression, random forest, gradient boosting regression tree (GBRT) and support vector regression (SVR), are implemented and compared in Scikit-learn, with grid search for parameter tuning and a warning-threshold replay to examine practical effectiveness. Considering that such enterprises generally have no usable historical failure data, method validation is carried out on the XJTU-SY rolling bearing accelerated life dataset [16], a widely recognized benchmark in equipment prognostics. Bearings are key rotating parts of the spindle and transmission systems of cold heading machines, and their vibration degradation data are consistent in form with what enterprise vibration monitoring can provide, so the research procedure remains standardized and reproducible. The experimental results show that GBRT achieves an RMSE of 6.25 min and an R^2 of 0.511 on the test set; both test bearings are successfully warned, with an average lead time of 8.0 min and no false alarm, indicating that lightweight AI tools are feasible for fault prediction of fastener production equipment in SMEs.

2 MODEL CONSTRUCTION AND EVALUATION METRICS

From the perspective of mathematical modeling, equipment degradation is essentially a continuous process, and taking the RUL as the dependent variable for regression modeling can directly characterize this process. This section describes the construction of the four candidate models, following the same logic in each case: the model assumption and mathematical form first, then the training workflow and operating mechanism, and finally the hyperparameters to be tuned. Linear regression serves as the baseline to test how far the data are linearly explainable; random forest and GBRT represent ensemble tree models that capture nonlinear coupling among monitoring features; SVR represents kernel methods, which are usually robust on small samples. All four models are built with the unified API of Scikit-learn, so their differences come only from the modeling mechanisms rather than the implementation pipeline.

2.1 Linear Regression

Linear regression assumes that the RUL is a linear combination of the input features:

$$\hat{y} = \beta_0 + \sum_{j=1}^{p} \beta_j x_j \tag{1}$$

where x_j is the j-th input feature, β_j its coefficient, and β_0 the intercept. Training amounts to estimating the coefficients by the ordinary least squares criterion, which minimizes the sum of squared residuals between predicted and observed values:

$$min_\beta \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2 \tag{2}$$

The model has two merits: no hyperparameter requires tuning, and each coefficient can be read directly as the marginal contribution of the corresponding feature, which makes the result easy to interpret for engineers. Its limitation is equally clear: if the degradation process is strongly nonlinear, a hyperplane cannot fit the curved mapping from features to RUL, and the prediction error will grow systematically. In this paper linear regression serves as the baseline, and how far it lags behind the other models measures how nonlinear the bearing degradation data are.

2.2 Random Forest Regression

Random forest is a bagging-type ensemble of decision trees, and its modeling process consists of three steps. The first step is bootstrap sampling: B training subsets are drawn randomly with replacement from the original training set, each of the same size, so that every tree sees a slightly different view of the data. The second step is tree growth: each tree is grown on its own bootstrap sample, and at each node only a random subset of the features is considered as split candidates; the split that yields the largest reduction of variance in the child nodes is chosen, and the tree grows until a stopping rule is met. This two-layer randomness, over samples and over features, decorrelates the individual trees. The third step is aggregation: for a new observation x, every tree outputs a leaf prediction, and the ensemble prediction is their arithmetic mean:

$$\hat{y} = \frac{1}{B} \sum_{b=1}^{B} T_b(x) \tag{3}$$

where $T_b(x)$ denotes the output of the b-th tree and B is the number of trees. The operating mechanism explains why random forest suits industrial monitoring data: averaging over decorrelated trees reduces variance markedly without noticeably increasing bias, so the model is robust to noise and outliers; no assumption is made about the functional form of the mapping, and no feature standardization is needed. In addition, the total impurity decrease brought by each feature across all trees gives a natural feature importance measure, which helps engineers identify the monitoring indicators that carry the degradation information. The main hyperparameters are the number of trees and the maximum

depth: too few trees make the ensemble unstable, while overly deep trees may overfit the training bearings; both are fixed by grid search in this paper.

2.3 Gradient Boosting Regression Tree

The gradient boosting regression tree (GBRT) is also an ensemble of decision trees, but it follows the boosting strategy in which trees are built sequentially rather than independently. The modeling process starts from an initial model that outputs a constant, typically the mean of the training labels. Then, at each round m, the algorithm computes the negative gradient of the loss function with respect to the current predictions at every training sample; under the squared-error loss, this negative gradient reduces to the residual:

$$r_{im} = y_i - F_{m-1}(x_i) \tag{4}$$

where $F_{m-1}(x_i)$ is the prediction of the ensemble built in the previous rounds. A new regression tree h_m is then fitted to these residuals, and the ensemble is updated by adding the new tree, scaled by a shrinkage coefficient ν called the learning rate:

$$F_M(x) = \sum_{m=1}^{M} \nu h_m(x) \tag{5}$$

The operating mechanism is therefore sequential error correction: each tree concentrates on the samples and feature combinations that the current ensemble still mispredicts, so later trees can repair the systematic errors left by earlier ones along specific feature interactions. This makes GBRT particularly good at capturing coupling relationships among statistical features. The learning rate controls the contribution of each tree: a large ν converges fast but tends to overfit, while a small ν needs more trees to reach the same fit and usually generalizes better. Compared with random forest, GBRT typically achieves higher accuracy on tabular industrial data at the cost of more sensitive hyperparameters. In this paper the learning rate and the number of trees are tuned jointly by grid search.

2.4 Support Vector Regression

Support vector regression (SVR) is a kernel method rooted in statistical learning theory. Its model assumption is a function $f(x) = \langle w, \varphi(x) \rangle + b$, where $\varphi(x)$ maps the input features into a high-dimensional space. Instead of penalizing every deviation as in least squares, SVR introduces an ε-insensitive tube: prediction errors within ε of the true value incur no loss, and only samples outside the tube are penalized through the slack variables ξ_i and ξ_i^*. Training amounts to solving the following optimization problem:

$$min_{w,b,\xi} \frac{1}{2}||w||^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \tag{6}$$

Subject to the constraints that each observation lies within the tube up to its slack variables. The first term keeps the function as flat as possible and controls the model complexity, while the second term penalizes the samples outside the tube; the penalty coefficient C controls the trade-off between flatness and tolerance of large errors. In practice the problem is solved in its dual form, where the mapping $\varphi(x)$ never appears explicitly and only inner products are needed; these inner products are replaced by a kernel function, which implicitly corresponds to the high-dimensional space. This paper adopts the radial basis function kernel:

$$K(x_i, x_j) = exp(-\gamma||x_i - x_j||^2) \tag{7}$$

The operating mechanism gives SVR two properties relevant to this study. On the one hand, the final prediction function depends only on the samples that lie on or outside the tube, the so-called support vectors, so the model remains robust when the sample size is small. On the other hand, the kernel trick allows a linear fit in the high-dimensional space to correspond to a flexible nonlinear fit in the original feature space. A practical caveat is that SVR is sensitive to feature scales, so its input is standardized before training, and the penalty coefficient C is tuned by grid search.

2.5 Hyperparameter Optimization and Evaluation Metrics

All four models are implemented and tuned with Scikit-learn. Hyperparameter optimization uses grid search under 3-fold cross-validation (GridSearchCV), a module that has been used for standardized parameter tuning of industrial diagnostic models [10]: the number of trees and the maximum depth are searched for random forest, the learning rate and the number of trees for GBRT, and the penalty coefficient C for SVR. The split is made by individual bearing, so that adjacent sampling points of the same bearing never appear in training and testing at the same time; this avoids the inflated accuracy caused by information leakage between neighboring records.

Regression performance is evaluated by three metrics: the root mean square error RMSE, the mean absolute error MAE, and the coefficient of determination R^2:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{8}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{9}$$

$$R^2 = 1 - \frac{\Sigma(y_i - \hat{y}_i)^2}{\Sigma(y_i - \bar{y}_i)^2} \quad (10)$$

The three metrics characterize the error from different angles. RMSE squares the deviations and therefore amplifies outliers, so a markedly large RMSE indicates that the prediction is far off at some time points; MAE answers the most intuitive question of how many minutes the prediction is off on average; R^2 measures how much of the life variation is explained by the model, and the closer it is to 1 the better. On top of the regression model, a warning threshold T = 10 min is set: a warning is triggered when the predicted RUL first falls below T, and the practical warning capability of a model is evaluated by the lead time from the alarm point to the actual failure, together with the false alarm rate.

3 RESULTS AND ANALYSIS

3.1 Data Source and Preprocessing

The modeling object of this study is the vibration monitoring record of key components of continuously operating equipment such as cold heading machines. Since SMEs rarely have usable historical failure data at hand, method validation adopts the XJTU-SY rolling bearing accelerated life dataset jointly released by Xi'an Jiaotong University and Changxing Shengyang Technology [16]. This dataset is chosen for two reasons. First, bearings are among the most common failing parts in the spindle system of cold heading machines, and their run-to-failure vibration data are consistent in form with what an enterprise can obtain by monitoring key rotating components. Second, the dataset comes from real accelerated life tests rather than simulation; it contains complete vibration records of 15 bearings under three operating conditions from normal operation to failure, and it is widely used as a benchmark in remaining useful life prediction. This paper uses the data of five LDK UER204 bearings under the 35Hz12kN condition (rotational speed 2100 r/min, radial load 12 kN): the sampling frequency is 25.6 kHz, sampling is performed once per minute for 1.28 s each time (32768 points), and the horizontal vibration signal, which carries richer degradation information, is used, giving 616 sampling files in total. An overview of the experiment is given in Table 1.

Table 1 Overview of the XJTU-SY Experiment under the 35Hz12kN Condition

Bearing	Files	Service life (min)	Failed component	Role
Bearing1_1	123	123	Outer race	Test
Bearing1_2	161	161	Outer race	Training
Bearing1_3	158	158	Outer race	Training
Bearing1_4	122	122	Cage	Training
Bearing1_5	52	52	Inner and outer races	Test

Note: The service life and failed component information are provided by the dataset publisher [16]; each sampling file corresponds to a 1.28 s vibration snapshot at one minute of operation.

Preprocessing consists of three steps. The first step is time-domain statistical feature extraction: for each vibration snapshot, ten statistics are computed, namely mean amplitude, standard deviation, root mean square, peak, peak-to-peak value, kurtosis, skewness, crest factor, impulse factor and clearance factor, so that the high-frequency waveform is compressed into one row of features per minute; this is consistent with the enterprise practice of sampling on schedule and storing statistics. The second step is scale unification. The magnitudes of the ten statistics differ considerably (kurtosis is only a few units, while the peak-to-peak value can exceed ten), and feeding them directly would let the large-magnitude features dominate training, so each feature is linearly compressed into the interval [0,1] by

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (11)$$

where the two boundary values x_{min} and x_{max} are computed only from the three training bearings, and the test bearings are transformed with the same boundaries, so that no information from the test data leaks into the training process. The third step adds sliding-window features: for each feature, a rolling mean with window length $w = 5$ min is taken to smooth measurement noise and extract the local trend:

$$\bar{x}_t = \frac{1}{w}\sum_{i=0}^{w-1} x_{t-i} \quad (12)$$

The original features and the window-mean features are concatenated into a 20-dimensional input. The label is processed piecewise. The healthy period of a bearing lasts tens of minutes, during which the features show almost no trend, and forcing the model to fit the true RUL in this period would only add noise. The RUL is therefore capped at 30 min, that is, $y = min(RUL, 30)$; samples whose true RUL exceeds 30 min are all labeled 30 min, so the model concentrates on learning the final degradation stage.

3.2 Model Comparison

Modeling and tuning are carried out entirely with Scikit-learn. The training set contains all 441 sampling records of the three training bearings, and the test set contains the 175 records of the two test bearings. The test results of the four regression models under the unified data pipeline are listed in Table 2.

Table 2 Prediction Performance of the Four Regression Models

Model	RMSE (min)	MAE (min)	R^2
Linear regression	10.50	6.29	-0.380
Random forest	6.43	3.76	0.482
GBRT	**6.25**	**3.71**	**0.511**
SVR	7.12	4.02	0.364

Note: Bold values are the best in each column; the optimal hyperparameters found by grid search are: random forest (150 trees, unlimited maximum depth), GBRT (learning rate 0.05, 150 trees), and SVR (C = 10).

All three machine learning models clearly outperform the linear baseline, and the R^2 of linear regression is even negative, which indicates that the bearing degradation process is strongly nonlinear and linear modeling alone is insufficient. GBRT leads on all three metrics (RMSE = 6.25 min, MAE = 3.71 min, R^2 = 0.511); the reason is roughly that the sequential fitting of residuals by successive regression trees can effectively capture the coupling among the statistical features. SVR, after standardization and tuning, performs in the middle and still beats linear regression. It should be noted that the large difference in degradation paths among individual bearings is a recognized difficulty in remaining useful life prediction [16]; an average absolute error of about 3.7 min is acceptable relative to the life span of 52-161 min, and the error is concentrated in the middle of the life, while the final stage that warning cares about is fitted tightly, as shown below. The whole experiment finishes within a few minutes on an ordinary office computer: for SMEs, sufficient accuracy at a low threshold matters more than the highest accuracy, and this is exactly where the value of the lightweight route lies.

3.3 Prediction Behavior of the Best Model

The predicted values of GBRT at all 175 sampling points of the two test bearings are compared with the true RUL in Figure 1. The scatter points distribute along the ideal prediction line as a whole, and the fit is especially tight in the final life stage where the RUL is below 15 min, which is exactly the region that warning decisions care about most. At the high-RUL end, the points cluster near the truncation cap of 30 min, which is an expected phenomenon under the piecewise label and has no substantial impact on the warning application.

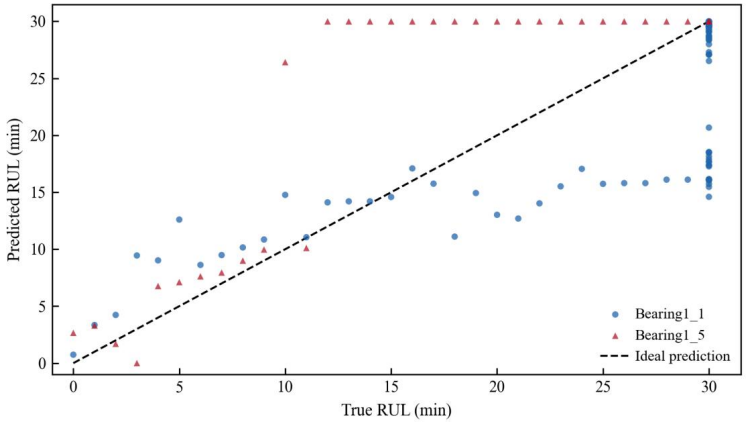


Figure 1 Predicted versus True RUL of the GBRT Model

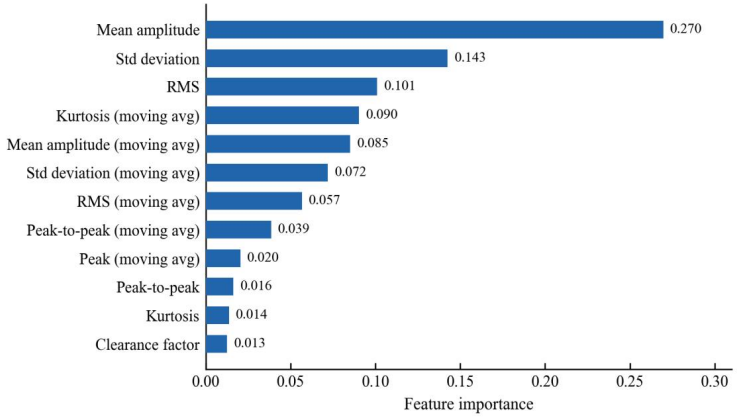


Figure 2 Feature Importance Ranking of the Random Forest (Top 12)

The feature importance analysis (Figure 2) shows that amplitude-type statistical features, including their window means, contribute about 81% of the total importance, among which the mean amplitude alone contributes 27%; in other words, the degradation information of the equipment is concentrated in these few amplitude-type statistics. Vibration energy keeps rising as bearing degradation intensifies, a conclusion consistent with the degradation mechanism of bearings and with related studies [16], and also in line with the engineering experience in cold heading production that load fluctuation is the first signal of equipment condition change [3-4]; in enterprise implementation, it is sufficient to guarantee the data quality of vibration amplitude indicators. The behavior of impulse-type indicators is also interesting: after window smoothing, their importance rises markedly (the window mean of kurtosis contributes 9.0%, far above the 1.4% of raw kurtosis), which shows that the sliding-window construction in Section 3.1 does stabilize noisy impulse features.

3.4 Early-Warning Validation and Discussion

To examine how useful the model is for warning in practice, the complete degradation process is replayed on the two test bearings: the threshold is set to T = 10 min, and a warning is triggered when the predicted RUL first falls below the threshold. The results show that warnings are successfully triggered for both bearings, 7 min and 9 min ahead of the actual failure respectively (8.0 min on average), with no false alarm. Figure 3 shows the warning process of Bearing1_1 (outer race failure, service life 123 min): the predicted value stays near the truncation cap in the early life, decreases smoothly after about 80 min of operation as degradation intensifies and approaches the true RUL, and crosses the threshold 7 min before the actual failure to issue the warning. For enterprises, a minute-level lead time in an accelerated test corresponds to a much more generous response window under actual working conditions, which is sufficient for arranging spare parts procurement and maintenance scheduling.

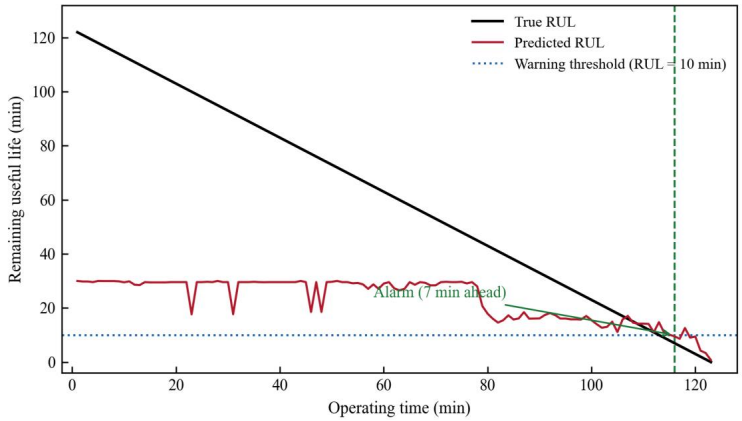


Figure 3 Early-Warning Replay for a Typical Bearing

This study certainly has limitations. The validation sample consists of only five bearings under a single operating condition, so the sample size for warning statistics is limited, and whether the model remains robust under the multiple working conditions and strong noise of an enterprise site awaits further examination. Moreover, the accelerated life test compresses months of degradation into tens of minutes under heavy load, so converting the warning lead time into calendar time requires calibration against the actual load spectrum and production rhythm of the enterprise equipment. Both points await further verification after field data are collected from enterprises.

4 CONCLUSIONS AND FUTURE WORK

This paper verifies the feasibility of the lightweight technical route of regression modeling with Scikit-learn for equipment fault prediction in fastener SMEs. The GBRT model achieves a prediction accuracy of RMSE = 6.25 min and R^2 = 0.511 on real bearing degradation data, and the whole workflow finishes within minutes in an ordinary office environment, a balance between accuracy and cost that matches the reality of SMEs. Feature importance analysis further shows that degradation information is concentrated in a few amplitude-type statistics such as mean amplitude, standard deviation and root mean square, and that the sliding window stabilizes impulse-type indicators, so enterprises can focus data collection on a small number of key monitoring indicators. The warning replay shows that both test bearings are warned successfully with an average lead time of 8.0 min and no false alarm, indicating that the complete procedure of data collection, preprocessing, regression modeling and threshold warning is replicable. Enterprises are advised to start by exporting operating parameters from existing equipment controllers and installing low-cost vibration monitoring, guarantee the data quality of key indicators, build a regression warning model following this workflow, and embed threshold warning into the daily inspection routine.

The limitations of this study should not be overlooked. The validation is based on five bearings under a single accelerated-life condition, the sample size for warning statistics is small, and the lead time measured in minutes still needs calibration against real load spectra before field deployment. Future work will proceed in two directions:

cooperating with fastener enterprises in Haiyan to test and calibrate the method on real production data, and extending the research object from numerical operating parameters to appearance images of parts, so as to explore the application of lightweight image recognition tools in defect detection.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

FUNDING

This work was supported by the 2025 Open Competition Mechanism Cultivation Project of Haiyan College, Jiaxing Vocational and Technical College.

REFERENCES

[1] Carvalho T P, Soares F A A M N, Vita R, et al. A systematic literature review of machine learning methods applied to predictive maintenance. Computers & Industrial Engineering, 2019, 137: 106024.
[2] Lan J F, Zhou Y H, Gao Z L, et al. Machine learning based prediction model for degradation trend of hydropower units. Journal of Hydroelectric Engineering, 2022, 41(12): 135-144.
[3] Tong J F, Feng Z G, Jiang Y L, et al. Simulation and experimental study on cold heading of TB9 titanium alloy mandrel. Forging & Stamping Technology, 2023, 48(8): 32-40.
[4] Tang F, Xing B Y, He X C, et al. Sensitivity analysis and optimization of clinching process parameters. Journal of Plasticity Engineering, 2022, 29(6): 119-124.
[5] Cinar Z M, Nuhu A A, Zeeshan Q, et al. Machine learning in predictive maintenance towards sustainable smart manufacturing in Industry 4.0. Sustainability, 2020, 12(19): 8211.
[6] Zhang G H, Luo Z P. Data-driven fault prediction of the main feedwater pump in nuclear power plants. Atomic Energy Science and Technology, 2024, 58(9): 1982-1988.
[7] Lei Y, Yang B, Jiang X, et al. Applications of machine learning to machine fault diagnosis: A review and roadmap. Mechanical Systems and Signal Processing, 2020, 138: 106587.
[8] Lei Y, Li N, Guo L, et al. Machinery health prognostics: A systematic review from data acquisition to RUL prediction. Mechanical Systems and Signal Processing, 2018, 104: 799-834.
[9] Zhao R, Yan R, Chen Z, et al. Deep learning and its applications to machine health monitoring. Mechanical Systems and Signal Processing, 2019, 115: 213-237.
[10] Huang R Y, Li J P, Wang Z, et al. Intelligent diagnosis and life prediction of equipment based on multi-task learning. Scientia Sinica (Technologica), 2022, 52(1): 123-137.
[11] Jin L J, Xu X, Wu S W, et al. Current status and prospects of deep learning driven predictive maintenance of equipment. Manufacturing Automation, 2025, 47(2): 123-131.
[12] Li K P, Zhang F L, Huang Z G, et al. Interpretable fault diagnosis method based on XGBoost and SHAP analysis. Manufacturing Technology & Machine Tool, 2025(6): 199-208.
[13] Jiang H G, Tang Y, Tang L F, et al. Research and application of motor vibration state monitoring based on decision tree. Computer Applications and Software, 2026, 43(4): 90-95, 109.
[14] Fathi K, van de Venn H W. Data, models, and performance: A comprehensive guide to predictive maintenance in industrial settings. Recent Topics in Maintenance Management. London: IntechOpen, 2024.
[15] Igbokwe N C, Nwamekwe C O, Aguh P S. Predictive modeling of manufacturing defects using machine learning: A comparative performance study in a manufacturing SME. African Journal of Advances in Engineering and Technology, 2025, 1(2): 93-115.
[16] Wang B, Lei Y, Li N, et al. A hybrid prognostics approach for estimating remaining useful life of rolling element bearings. IEEE Transactions on Reliability, 2020, 69(1): 401-412.