Journal of Network Intelligence
Taiwan Ubiquitous Information
 ISSN 2414-8105
Volume **10**, Number **2**, May **2025**

Prediction of Rolling Bearing Aging Trend of Cargo Ropeway Based On A Hybrid Deep Learning Model

Jun Zhang*

Qinghai Power Transmission and Transformation Engineering Company, Xining, Qinghai, 810001, P. R. China
jiaxx1982@163.com

Yong Zhou

Qinghai Huanghua Power Supply Company, Xining, Qinghai, 810001, P. R. China
zy7800114yqww@163.com

Feng Xue

State Grid Qinghai Electric Power Company, Xining, Qinghai, 810001, P. R. China
qhxuef@126.com

Hao Li

Krirk University, Bangkhen, Bangkok, 10220, Thailand
007.lihao@163.com

*Corresponding author: Jun Zhang

Received March 14, 2024; revised July 21, 2024; accepted December 4, 2024.

Abstract. *An increasing number of current transmission line projects are adopting cargo ropeway transportation technology. However, there are common issues with the rolling bearings, which are critical components, limiting the practical application and widespread adoption of cargo ropeway transportation systems. Traditional feature selection methods for predicting the degradation trend of rolling bearings heavily rely on manual expertise and single evaluation algorithms, often leading to the under-selection or misselection of features. Moreover, single deep learning networks are insufficient for fully mining the performance degradation information embedded in the data, resulting in lower predictive accuracy of the models. To address these issues, we propose a rolling bearing degradation trend prediction method based on Adaptive Mahalanobis Space (AMS) and a fusion of deep learning networks. First, we develop a multi-objective feature selection algorithm based on AMS to automatically optimize features, reducing human dependence, enhancing adaptability, and generalization. We combine the Mahalanobis distance (MD) with the Exponential Weighted Moving Average (EWMA) method to effectively characterize the bearing performance degradation trend. Finally, by improving the structure of the Variational Autoencoder (VAE) and integrating Long Short-Term Memory (LSTM), we use the system reconstruction error to represent bearing degradation trends, achieving nonlinear prediction. Experimental results demonstrate that our approach effectively selects optimal features, and the VAE-LSTM model exhibits a strong capability to automatically learn bearing degradation features, along with good generalization and overfitting resistance. It yields satisfactory predictive results with higher accuracy compared to other comparative methods.*

Keywords: Transmission lines; Cargo ropeway transportation; Bearing aging trend; Adaptive Mahalanobis space; Variational autoencoder; Long short-term memory

1. **Introduction.** With the rapid development of China's power grid construction projects, transmission line projects are gradually extending to remote and mountainous areas. Construction of transmission lines in mountainous regions presents challenges such as high altitudes, rugged terrain, and difficult transportation access [1]. In particular, in western regions, these transmission line construction projects often pass through uninhabited mountainous areas where conventional transportation methods, including manual labor and road-based transport, are impractical due to the substantial weight of materials such as tower components, conductors, and infrastructure required for transmission line construction. As a result, cable transportation systems, known as cargo ropeways, have been increasingly adopted to transport materials between construction sites and remote or difficult-to-reach areas [2, 3].

Cargo ropeways exhibit strong adaptability to natural terrain, high climbing capabilities, and the ability to navigate steep slopes, cross gorges, rivers, and other natural obstacles [4]. Compared to other transportation methods, they offer significant advantages, including minimal environmental impact, reduced land usage, lower pollution, landscape preservation, lower investment costs, and lower energy consumption. However, in practical use, there are still some issues, particularly related to the rolling bearings, which are critical components, limiting the widespread practical application of cargo ropeways [5].

Bearings serve as critical components in a wide array of industrial equipment, ranging from machinery in manufacturing plants to the turbines in power generation stations. Their significance cannot be overstated, as they are the linchpin that facilitates the smooth rotation of axles, shafts, and other pivotal components [6]. The operational state of bearings is intrinsically linked to the overall reliability and functionality of rotating machinery, making them not just components, but rather guardians of seamless operations. The slightest deviation from their ideal condition can set in motion a cascade of mechanical issues, making bearings the primary instigator of equipment failures in many industrial settings [7]. Throughout their operational lifecycle, bearings traverse a spectrum of performance states, from optimal and healthy operation to an eventual state of failure. This journey can be characterized by the gradual deterioration of bearing health, marked by subtle shifts in their vibration patterns, temperature profiles, and lubrication conditions. The ability to detect these subtle changes and interpret their significance in real-time is paramount for proactive maintenance and the prevention of catastrophic bearing failures [8].

The real-time monitoring of bearing operation status has evolved significantly over the years, with the advent of advanced sensor technologies and data analytics techniques [9]. Continuous data streams, encompassing parameters like vibration levels, temperature, and lubricant quality, are collected and analyzed to provide insights into bearing health. These insights empower maintenance teams to make data-driven decisions and initiate maintenance activities precisely when required, rather than relying on fixed schedules or waiting for conspicuous signs of impending failure [10].

By leveraging historical operational data, maintenance professionals can identify trends, anomalies, and early warning signs of potential bearing issues. This data-driven approach to maintenance allows for predictive and preventive measures to be taken, ultimately averting equipment breakdowns [11]. The economic ramifications of such proactive maintenance are substantial, as it not only mitigates the direct costs associated with bearing replacements and machinery downtime but also prevents collateral damage to adjacent components that often occurs during catastrophic failures. Moreover, the cost savings extend beyond just avoiding expensive repairs. The reduction in unplanned downtime translates into increased production efficiency and product quality, positively impacting a company's bottom line. Labor costs are optimized as well, as maintenance activities

are strategically planned and executed, reducing the need for emergency callouts and overtime work [12].

In conclusion, the role of bearings in industrial machinery is one of immense responsibility and consequence. Their vigilance in ensuring the uninterrupted operation of rotating machinery makes them indispensable. Through real-time monitoring and data-driven maintenance strategies, the bearing's capacity to anticipate and prevent failures is harnessed, leading to significant economic benefits and operational efficiency across industries. The era of predictive maintenance, facilitated by advanced technologies and informed decision-making, has ushered in a new era of reliability and cost-effectiveness in the realm of industrial machinery.

The accurate prediction of bearing degradation trends relies on selecting suitable degradation features and constructing effective predictive models [13]. With the emergence of new methods and technologies, the variety of features has become increasingly diverse. While multi-domain features can comprehensively represent the operational state of bearings, high-dimensional variables can increase model complexity, leading to information redundancy and reduced computational efficiency, ultimately resulting in lower predictive accuracy. Therefore, feature selection is a key research area in degradation trend prediction. [14] employed correlation-based evaluation criteria combined with filter evaluation models to achieve favorable feature selection results. Yang et al. [15] used the Fisher ratio method to select features and found that a higher Fisher ratio indicates stronger feature discriminative power. Zhang et al. [16] conducted comprehensive feature evaluations using correlation, monotonicity, and robustness algorithms, and the selected features proved effective for degradation trend prediction. The above methods all evaluate features using a single algorithm, which can only analyze the excellence of features in a specific aspect. Moreover, they are susceptible to subjective factors and lack adaptability and generalization. Therefore, they have limitations in removing redundant features, and the performance degradation metrics they construct cannot accurately reflect the operational state of bearings.

In recent years, methods for predicting bearing degradation trends can be broadly categorized into two main approaches: model-driven methods and data-driven methods. Model-driven methods necessitate the development of mathematical models based on degradation mechanisms. However, in practical industrial applications, especially concerning complex mechanical equipment, establishing precise degradation models proves challenging [17]. With the rapid advancement of artificial intelligence, data-driven methods have emerged as a thriving research area. These methods leverage condition monitoring data and machine learning techniques for predictive purposes, overcoming the constraints associated with model-driven approaches. Notable data-driven methodologies include various neural networks, Hidden Markov Models, Support Vector Machines, among others. Among these, autoencoders (AEs) represent an important class of unsupervised learning models in deep learning. They excel at capturing meaningful features from extensive unlabeled data. AEs offer advantages such as a simple model structure, strong learning capabilities, unsupervised learning, and low application costs, rendering them well-suited for anomaly diagnosis in industrial equipment, particularly in industrial scenarios. For example, Li et al. [18] employed local features of fault characteristics in both the time and frequency domains. They used backpropagation networks to learn meaningful features from signals of varying scales, ultimately improving fault diagnosis accuracy. In separate studies [19], AE and Denoising Autoencoders (DAE) were applied to rolling bearing fault diagnosis, leveraging their deep structures to extract abstracted features from vibration signals. This overcame the limitations of traditional manual feature extraction and enhanced fault diagnosis accuracy. Yu [20] introduced a selective stacked

Denoising Autoencoder (SDAE) ensemble model based on negative correlation learning. This model effectively extracted fault features from noisy vibration signals and achieved a stable and well-generalized performance through negative correlation learning and particle swarm optimization. Zhao et al. [21] incorporated Variational Autoencoders (VAE) into the fault diagnosis framework. They expanded the minority class of vibration signal samples to balance training data and employed a CNN for classification, thereby promoting diagnostic accuracy. However, due to the nonlinear and non-stationary characteristics of vibration signals, a single deep learning network may exhibit limited nonlinear learning capabilities. This limitation results in the inadequate exploration of deep degradation features in bearings, lower model generalization performance, and the presence of redundant input data, ultimately affecting network efficiency and predictive accuracy. In a recent study [22], data features were initially used to construct a health index (HI). Subsequently, a nonlinear prediction framework was designed for bearing health state detection based on the HI index. Han et al. [23] proposed a bearing aging automatic detection framework, combining data features using stacked autoencoders (SAE) and utilizing Long Short Term Memory (LSTM) for lifetime prediction.

In response to the aforementioned issues, this paper proposes a novel approach for predicting rolling bearing degradation trends based on the Adaptive Mahalanobis Space (AMS) and a fusion of deep learning networks. Firstly, by combining the advantages of Laplacian Score (LS), correlation, monotonicity, and robustness (CMR) algorithms, we construct a multi-objective feature selection algorithm based on AMS. This algorithm automatically selects the optimal degradation features, effectively addressing issues of strong human dependence, poor adaptability, and low generalization often associated with traditional feature selection methods. Additionally, to mitigate the adverse impact of outliers on degradation trend prediction, we introduce the Exponential Weighted Moving Average (EWMA) method to smooth the Mahalanobis Distance (MD) under the AMS framework, with the smoothed MD serving as a performance degradation metric. Secondly, building upon the VAE model, we propose an end-to-end hybrid VAE-LSTM model that combines the strengths of both models in feature learning. This approach delves deeper into learning degradation information from feature data, ultimately enhancing the predictive accuracy of long time-series data.

2. **Optimal feature selection.**

We propose a novel multi-objective feature selection algorithm based on the AMS. This algorithm combines the strengths of various feature selection methods, including LS and 3σ algorithms. During the construction of AMS, we introduce an evaluation algorithm based on the MS, composed of three criteria and a confusion matrix. This MS evaluation algorithm adaptively selects the optimal features. The adaptive capability of AMS helps overcome the shortcomings of feature selection errors caused by subjective experiences, thereby enabling the identification of rational input features for predictive models and effectively improving predictive accuracy.

2.1. **Basic Principles of MS.**

MS is composed of MD calculated from normal samples, and its establishment process can be divided into three steps:

Step 1: Construction of Initial MS. (1) Identify m important features in the multivariate system and divide them into normal and abnormal samples. (2) Compute the mean and standard deviation of the normal sample features and standardize the normal sample data X_i, $i = 1, 2, \ldots, n$. (3) Calculate the MD_i of normal sample data X_i in the initial MS:

$$MD_i = \frac{1}{M} Z_i^T R^{-1} Z_i \tag{1}$$

where Z_i represents the standardized vector of the i-th sample. R represents the correlation matrix of normal samples.

Step 2: Validation of MS Effectiveness. (1) Standardize the abnormal sample using the mean and standard deviation of normal sample features. (2) Calculate the MD of the abnormal sample in the MS. If the MD of the abnormal sample is significantly greater than that of the normal sample, then the MS is considered effective. Otherwise, a new MS needs to be determined.

Step 3: Calculation of MD for Test Samples Based on the constructed MS, using the mean and standard deviation of the features of normal samples, calculate the MD for the test sample.

2.2. **Evaluation Metrics for MS.** The effectiveness of MS is crucial for the accuracy of predicting rolling bearing degradation trends. Therefore, an AMS evaluation algorithm is constructed using three criteria and a confusion matrix to adaptively select degradation features that effectively represent the bearing's performance, thus establishing the optimal MS.

The 3σ criteria [24] are common methods for determining thresholds in normal distribution data, and the confusion matrix is a method for evaluating model accuracy, describing the relationship between the true attributes of samples and the recognition results. To determine if the MS constructed with the selected features is effective, one can classify values smaller than the threshold as Positive and those greater than the threshold as Negative. Then, the numbers of true positive, false positive, true negative, and false negative instances are counted to create the confusion matrix. Based on the statistical results from the confusion matrix, four performance metrics are used to evaluate the effectiveness of MS: Accuracy, Precision, Recall, and F1-Score. These metrics transform the statistical results from the confusion matrix into values between 0 and 1, with 1 indicating that the MS constructed with the selected features performs best and 0 indicating the poorest performance.

2.3. **Multi-Objective Feature Selection.** To enhance the accuracy and effectiveness of predictive models, it is essential to employ appropriate methods for removing redundant variables, retaining features that effectively characterize the degradation of bearings, and integrating the selected features to construct performance degradation indicators for rolling bearings.

LS is an algorithm used to evaluate and select features based on their importance [25]. It offers advantages such as not requiring supervised information and high computational efficiency. Features are rearranged in ascending order of scores, where a smaller LS score indicates stronger discriminative and local-preserving capabilities of the feature, signifying greater importance. The calculation formula for LS is as follows [25]:

$$L_r = \frac{\sum_{ij}(f_{ri} - f_{rj})^2 S_{ij}}{Var(f_r)} \tag{2}$$

where f_{ri} represents the r-th feature of the r-th sample. S_{ij} represents the similarity between the i-th and j-th samples. $Var(f_r)$ is the estimated variance of the r-th feature.

However, reasonable degradation features typically exhibit good correlation, monotonicity, and robustness. Correlation measures the linearity between features and time, monotonicity assesses the consistency of feature change trends, and robustness reflects the tolerance of features to outliers. LS only considers the correlation between features, thus requiring the incorporation of correlation, monotonicity, and robustness for further screening. Using EWMA [26], the feature $f(t)$ at time t is decomposed into a stationary

trend $f_T(t)$ and a random residual $f_R(t)$. Subsequently, the three indicators of correlation $Corr$, monotonicity Mon, and robustness Rob are defined as follows:

$$Corr = \frac{\left|\sum_{t=1}^{L}(f(t)-\bar{f})(t-\bar{t})\right|}{\sqrt{\sum_{t=1}^{L}(f(t)-\bar{f})^2\sum_{t=1}^{L}(t-\bar{t})^2}} \tag{3}$$

$$Mon = \left|\frac{\sum_{t=1}^{L} diff(f) - \sum_{t=1}^{L} -diff(f)}{L-1}\right| \tag{4}$$

$$Rob = \frac{1}{L}\sum_{t=1}^{L}\exp\left(-\left|\frac{f_R(t)}{f(t)}\right|\right) \tag{5}$$

where L represents the time duration. f and t are the means of the features and time, respectively. $diff(f)$ denotes the difference between two consecutive points. To comprehensively consider the three indicators, a weighted linear combination is proposed as the final criterion for feature selection. The formula for calculation is as follows:

$$Cri = \omega_1 Corr + \omega_2 Mon + \omega_3 Rob \tag{6}$$

where Cri represents the composite index. ω_i denotes the weights assigned to each indicator. In this study, $\omega_1 = 0.3$, $\omega_2 = 0.4$, and $\omega_3 = 0.3$ are used.

2.4. **Feature Selection Algorithm Workflow.** The specific workflow of the multi-objective feature selection algorithm based on AMS is illustrated in Figure 1. Initially, the original feature space is established. And feature importance is ranked independently using both LS and 3σ algorithms. Next, the common features that rank in the top 50% in both algorithms are selected as the base set, while the remaining features form the candidate set. Candidate features are iteratively added to the base set to construct AMS, and the results of MS evaluation metrics are recorded. The effectiveness of adding a new feature to the base set, based on improved MS performance, is assessed. If it enhances the performance, the feature is retained in the base set, which is subsequently updated. Conversely, if the performance deteriorates, the last-added feature is removed. Finally, it is determined whether all the candidate features have been traversed. If so, the updated base set, representing the optimal subset of features, is output. Using this optimal feature subset, MS is constructed, and the MD of the test sample is calculated. The EWMA method is applied to smooth the MD, correcting minor deviations and mitigating the adverse impact of outliers. The smoothed MD serves as the performance degradation indicator for rolling bearings.

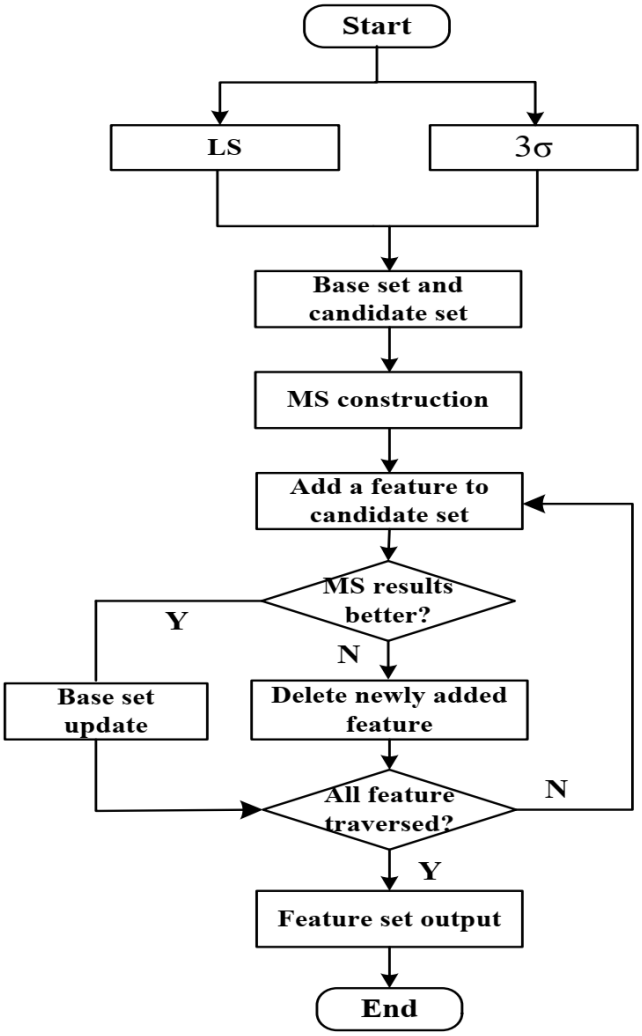

Figure 1. Workflow of the optimal feature selection.

3. **Bearing Degradation Trend Prediction Model based on VAE and LSTM.**

3.1. **Variational Autoencoder (VAE).** The VAE is an important variant of the AE. VAE was originally introduced as a generative model with the aim of learning the sample distribution, approximating the true data distribution through an estimated distribution, and subsequently generating samples similar to the original ones [27]. The architecture of the VAE model is illustrated in Figure 2.

In Figure 2, z represents latent variables. μ and σ denote the mean and standard deviation of the latent variable z respectively, which are calculated through a network layer. Typically, z is assumed to follow a Gaussian distribution, i.e., $P(z) \sim \mathcal{N}(0, I)$. The loss function of VAE is given as follows [27]:

$$L_{VAE} = \mathbb{E}_{z\sim Q(z|x)}[\log(P(x,z))] + D_{KL}(Q(z|x)\|P(z)) \tag{7}$$

where $Q(z|x)$ represents the approximate posterior distribution. $P(x|z)$ is the conditional distribution that needs to be learned during the decoding process. $P(z)$ is the prior distribution. D_{KL} denotes the Kullback-Leibler (KL) divergence. The first term in the loss function can be chosen as either binary cross-entropy or mean squared error, depending on the specific requirements. The second term, the KL divergence, is calculated as:

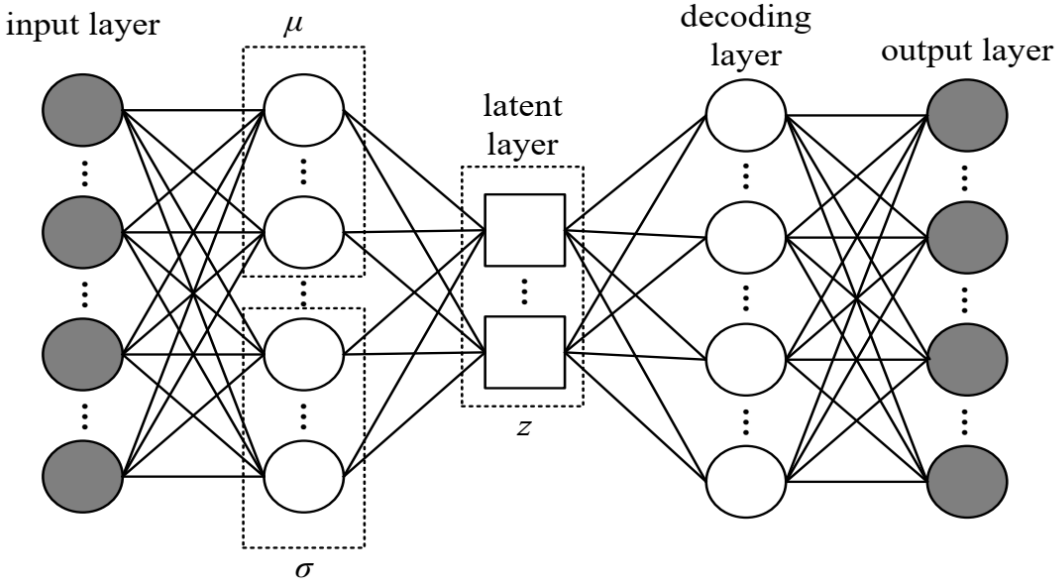

Figure 2. Structure of the VAE model.

$$D_{KL}(Q(z|x)\|P(z)) = -\frac{1}{2}\sum(1 + \log\sigma^2 - \mu^2 - \sigma^2) \tag{8}$$

To address the issue of the model's inability to use backpropagation due to random sampling, VAE employs the reparameterization technique. Firstly, a parameter $\varepsilon \sim \mathcal{N}(0, I)$ is introduced and sampled. Then, z is obtained through the following linear transformation:

$$z = \mu + \varepsilon\sigma \tag{9}$$

3.2. **LSTM.** The LSTM is a variant of Recurrent Neural Networks (RNN) and a type of neural network equipped with memory functionality, capable of learning temporal dependencies within time series data [28]. LSTM addresses the issues of gradient explosion and vanishing gradients that are encountered by RNN models during training with long sequences. As a result, LSTM has found wide application in fields related to time series data. Its inherent properties make it an ideal choice for tasks involving anomaly detection in time series and nonlinear data flows. LSTM introduces the concept of a cell state and utilizes gate structures to control and propagate information states. The architecture of an LSTM unit is depicted in Figure 3. In the figure, σ represents the sigmoid activation function with values in the range $[0, 1]$, the 'tanh' function keeps the state values within the range of [-1, 1]. $\otimes$ denotes the Hadamard product, and $\oplus$ represents matrix addition.

3.3. **Proposed VAE-LSTM Model.** The proposed VAE-LSTM model is an improved version of VAE incorporating LSTM, and it involves several key modifications: (a) LSTM units replace the ordinary neurons in the hidden layer to enhance the model's capacity for learning from time series data. (b) Two additional encoding layers are introduced before the layers that compute the mean and standard deviation to improve the model's feature extraction capabilities. (c) An extra hidden layer is added during the decoding phase to enhance the model's decoding ability, effectively transforming VAE into a deep VAE. The model structure is illustrated in Figure 4, where L represents the LSTM unit.

Before inputting the optimized features into the VAE-LSTM network, data normalization is performed because the scales of various indicators may differ. In this study, the

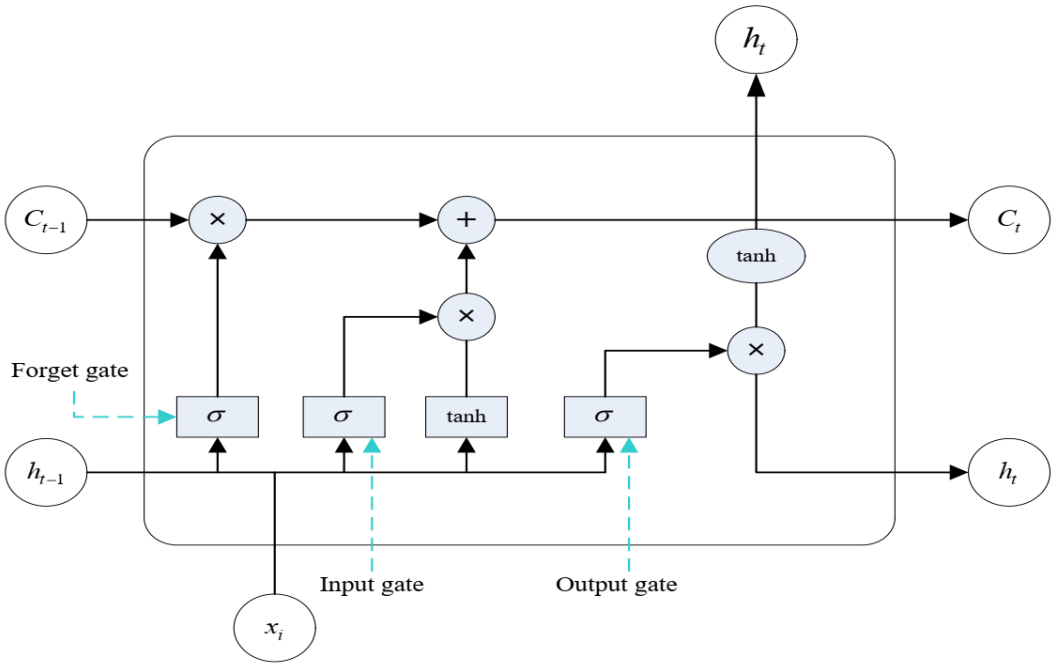

Figure 3. Structure of the LSTM cell.

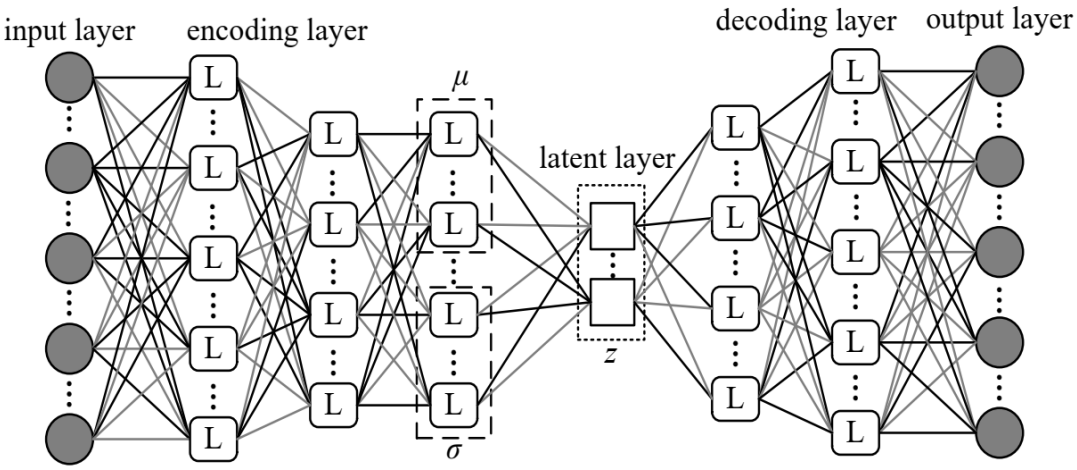

Figure 4. Architecture of the proposed hybrid model VAE-LSTM.

maximum value method is used for data standardization, with the following formula:

$$\tilde{x} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{10}$$

where x is the normalized data. $x_{\min}$ and $x_{\max}$ represent the minimum and maximum values in the sample data.

Through experimental investigation, this study improved the loss function employed in the VAE by increasing the proportion of the KL divergence term in the loss function, from -1/2 to -1/20. In cases involving standard deviation σ, the term $1 + \log \sigma^2 - \mu^2 - \sigma^2$ was modified to $1 + \sigma^2 - \mu^2 - e^{\sigma}$, enhancing the σ component, as shown in Equation (11). Here, N represents the dimensionality of the input data:

$$L_{VAE-LSTM}(\mathbf{W}, \mathbf{b}) = N \times \frac{1}{2}\left(\|\hat{\mathbf{x}} - \mathbf{x}\|_2^2\right) - \frac{1}{20}\sum\left(1 + \sigma^2 - \mu^2 - e^{\sigma}\right) \tag{11}$$

where $\mathbf{W}$ and $\mathbf{b}$ are weight and bias parameters, respectively. $\hat{\mathbf{x}}$ and $\mathbf{x}$ represent the output signal and input signal, respectively.

Furthermore, the method for sampling latent variables was enhanced, transitioning from linear sampling to nonlinear sampling, by replacing σ with $\frac{\sigma}{e^2}$. This change increases the sampled values of the latent variable z:

$$z = \mu + \varepsilon \cdot e^{\frac{\sigma}{2}} \tag{12}$$

Lastly, the standard deviation of parameter ε during sampling was reduced. Originally, ε was sampled from a standard Gaussian distribution with mean $\mu_\varepsilon = 0$ and standard deviation $\sigma_\varepsilon = 1$. After reduction, the standard deviation σ_ε became 0.05, while the mean μ_ε remained unchanged at 0, resulting in denser ε sampling.

4. **Experiments and Analysis.**

4.1. **Dataset description.** The experimental data for this study was obtained from the Case Western Reserve University (CWRU) Bearing Data Center [29]. These data were sampled at a frequency of 12 kHz from accelerometers within the motor-driven mechanical systems. Four different types of bearing defects were diagnosed, including rolling element damage, outer race damage, inner race damage, and composite damage, with respective defect diameters of 0.007, 0.014, and 0.021 inches. In total, there were ten fault categories, including normal bearings. These classification is different from image classification [30] and transfer learning [31]. Figure 5 presents examples of bearings with various types of damage. Data was collected under different load conditions corresponding to different horsepower (HP) ratings. Horizontal vibration signals were chosen for analysis as it has been observed that they typically contain richer degradation information compared to vertical vibration signals.

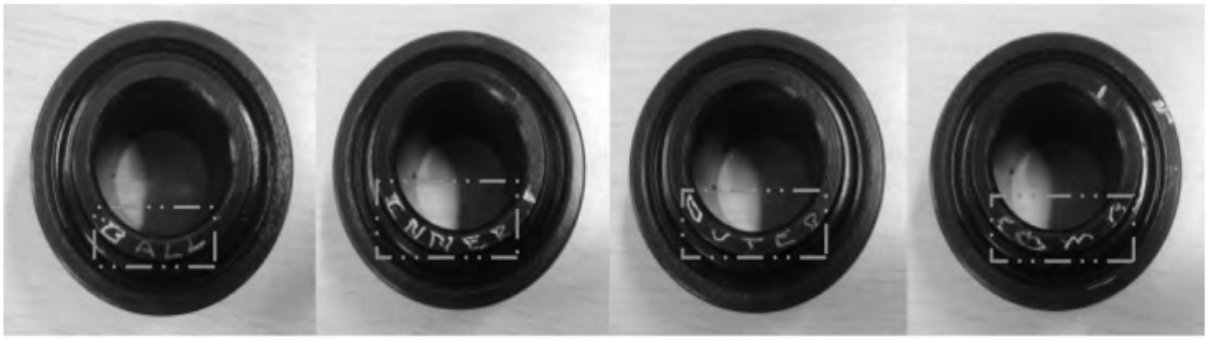

Figure 5. Different types of bearing fault conditions are shown from left to right, including rolling element damage, outer race damage, inner race damage, and composite damage.

To create the training set for the bearing fault diagnosis model, a random selection of N_{train} data points was made from the first 70% of the ordered signal data $S^l = [s_0, s_1, ..., s_m]$ for each fault class l, where m is the length of the original signal. For signal sampling, a signal sample length of 250 was set, which was extracted from each data point s_i to form a single signal sample $x_i = [s_i : s_{i+250}]$. Finally, N_{train} signal samples of length 250 constituted the training dataset, with the last 20% of the training dataset set aside as the validation set. Similarly, from the last 30% of the data, a testing dataset of size N_{test} was obtained using the same method. In this experiment, 1,000 signal samples of length 250 were sampled for each fault category. Consequently, for the ten fault types, a total of 10,000 signal samples were used for training, validation, and testing.

4.2. **Experimental Environment Setup.** In this study, the experimental software environment consisted of TensorFlow 2.5 and Keras 2.4, both implemented using the Python 3.8 programming language. As for hardware specifications, the processor used was an Intel Core i5-9400 operating at 3.2 GHz with four cores, and the system had 16 GB of Samsung memory. In the VAE-LSTM model, rectified linear unit (ReLU) was employed as the activation function, and the Adam optimizer was selected. The model's initialization method was based on glorot_uniform from the Keras deep learning framework. This method uses values sampled from a uniform distribution within the range of [-limit, limit], where "units_in" represents the number of input neurons, and "units_out" represents the number of output neurons. The learning rate was set to 0.001, the batch size to 16, and the number of epochs to 100. The dataset was divided into a ratio of 4:1:5 for the training set, validation set, and testing set.

4.3. **Evaluation Metrics.** To assess the model's diagnostic capability for bearing faults, this study utilized evaluation metrics including accuracy, precision, recall, and F1 score. These metrics were calculated as follows: Accuracy (Acc): The ratio of correctly predicted instances to the total instances. Precision (Pre): The ratio of true positive predictions to the total positive predictions. Recall (Rec): The ratio of true positive predictions to the total actual positive instances. F1 Score ($F1$): A harmonic mean of precision and recall.

$$Acc = \frac{TP + TN}{TP + FP + FN + TN} \tag{13}$$

$$Pre = \frac{TP}{TP + FP} \tag{14}$$

$$Rec = \frac{TP}{TP + FN} \tag{15}$$

$$F1 = 2 \times \frac{Pre \times Rec}{Pre + Rec} \tag{16}$$

where TP, TN, FP and FN stand for the number of true positive, true negative, false positive and false negative samples, respectively. In the context of the experimental dataset with different fault types, each fault sample was treated as a positive sample, and the remaining types were considered negative samples.

4.4. **Experimental Results.** The MD used in the proposed model effectively reflected the degradation state during the bearing's operation, displaying a consistent overall trend. Therefore, the feature selection method in this study proved to be comprehensive, effective, and capable of representing degradation well. The constructed degradation indicator provided a clear representation of the bearing's gradual transition from a healthy state to failure. To address local oscillations during degradation, the EWMA method was applied to smooth the MD, eliminating the negative impact of oscillations on the indicator's effectiveness. Additionally, the hybrid deep learning model proposed in this study improved model structure and overcame the influence of outliers, eliminating the need for separate denoising. As shown in Figure 6, after the accuracy curve converged, the proposed model achieved average training and validation accuracies of 95.38% and 96.53%, respectively, on the experimental dataset. The model exhibited high accuracy on both the training and validation sets, indicating its ability to capture underlying data features. Furthermore, the proposed model showed no signs of overfitting, demonstrating robust resistance to overfitting.

Figure 7 represents the distribution of system reconstruction errors (mean absolute error, MAE) of the model on the training dataset. This distribution, depicting the error

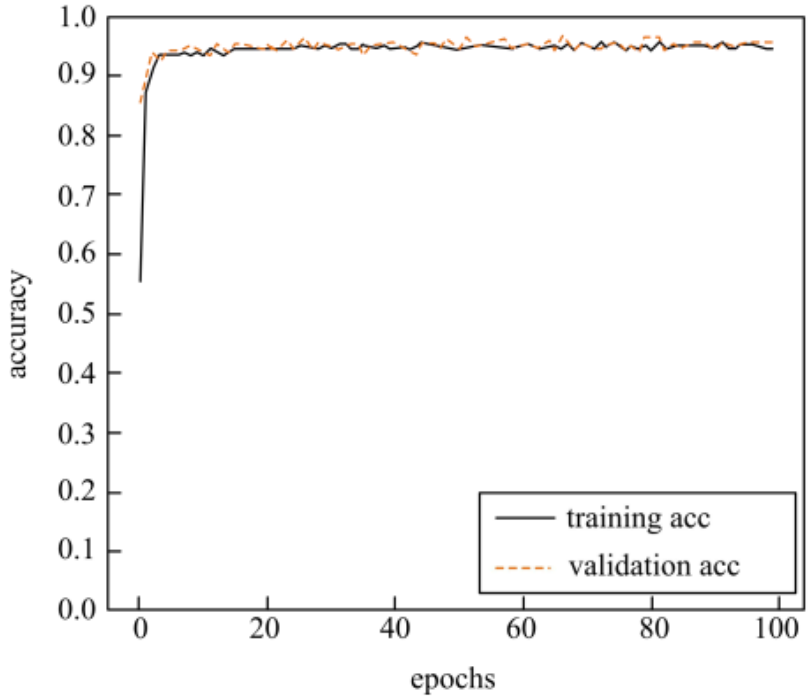

Figure 6. Accuracy results of the proposed model on the training and validation sets.

distribution of normal data, was used to set the model's error threshold. A threshold of $t = 0.06$ was chosen. MAE, compared to mean error, is advantageous because it avoids positive and negative errors canceling each other out, providing a better reflection of the actual prediction errors.

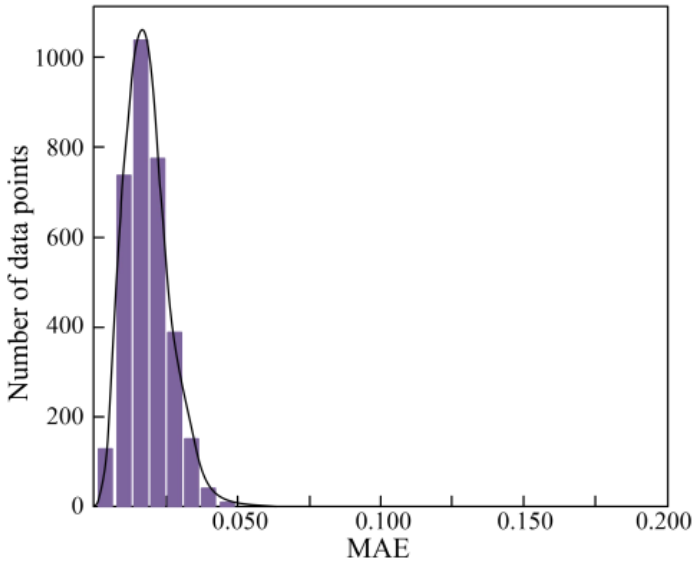

Figure 7. MAE error distribution of the proposed method on training set.

Furthermore, a comparative analysis was conducted to validate the performance of the proposed VAE-LSTM model against other models. The study compared the proposed model against AE, DAE, SDAE, VAE, and two additional methods presented in [22] and [23]. The method in [22] extracted features and constructed a HI index for remaining

useful life (RUL) prediction. And the model proposed in [23] used SAE for feature fusion followed by LSTM for RUL prediction. Experimental results are summarized in Table 1. The results demonstrate that the proposed model achieved superior training and validation accuracy compared to various AE-type models, resulting in an improvement of 3.5%, 2.4%, and 2.9% in accuracies and 0.6%, 0.6%, and 4.0% in F1 scores.

Table 1. Comparison results of different models on the experiment dataset. (%)

Models	*Acc*	*Pre*	*Rec*	*F*1
AE	83.15	83.27	81.77	82.51
DAE	87.67	87.72	85.43	86.56
SDAE	88.92	89.03	88.25	88.64
VAE	86.33	86.35	85.77	86.06
[22]	90.15	90.22	89.43	89.82
[23]	92.07	92.13	91.77	91.95
proposed method	96.77	96.82	95.68	96.24

The multi-objective feature selection algorithm based on AMS proposed in this study was compared to a single use of the LS algorithm [24] and 3σ algorithm [23]. The results given in Figure 8 indicated that the feature selection method presented in this study exhibited superior feature selection capabilities. For instance, when the feature dimension was 16, the LS algorithm achieved the best F1-Score of 93.27%, whereas the 3σ algorithm achieved the best F1-Score of 93.89% with 18 features. In comparison, the proposed algorithm selected 15 features with the best F1-Score of 96.24%. It is proved that the proposed method has superior dimensionality reduction effect.

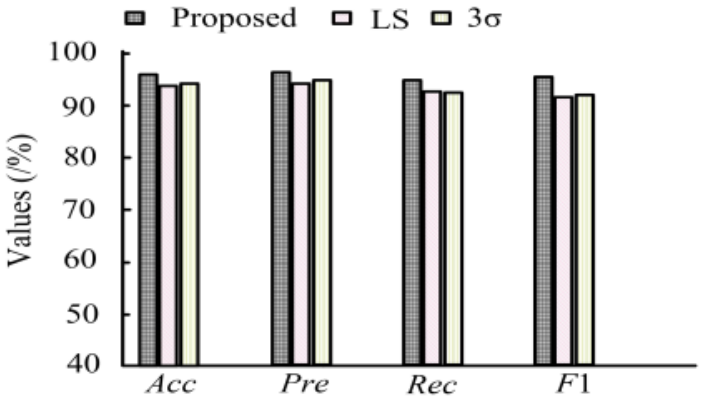

Figure 8. Comparison of different feature selection algorithms.

Feature selection impacts model prediction outcomes significantly. If all features are directly fused into the degradation indicator as input for the prediction model, it can lead to issues such as feature redundancy, low algorithm efficiency, and reduced prediction accuracy. Therefore, predictions were made using both the complete feature set without feature selection and the subset of features obtained through the AMS algorithm proposed in this study. Experimental results are presented in Table 2. A comparison indicates that models incorporating feature selection exhibit superior predictive performance. This demonstrates that feature selection effectively mitigates the introduction of excessive redundant variables, which can otherwise lead to decreased prediction accuracy.

Table 2. Performance of the proposed VAE-LSTM model with or without AMS-based feature selection.

Models	Acc	Pre	Rec	$F1$
with AMS	88.42	89.07	87.94	88.50
without AMS	96.77	96.82	95.68	96.24

5. **Conclusion.** In the context of rolling bearing degradation trend prediction, traditional feature selection methods suffer from strong human dependence, poor adaptability, low generalization capability, and an inability to fully extract performance degradation information from the data. This paper introduces a rolling bearing degradation trend prediction method based on AMS and integrated deep learning networks. Experimental results show that the multi-objective feature selection strategy based on AMS effectively reduces human dependence, enhances adaptability and generalization, and automatically acquires comprehensive, effective features with strong representation capabilities for bearing performance degradation. Furthermore, the paper presents the hybrid VAE-LSTM model. Experimental results demonstrate that the proposed VAE-LSTM model possesses a robust ability to automatically learn bearing degradation features while maintaining good generalization and resistance to overfitting. Satisfactory prediction results were obtained, with prediction accuracy surpassing that of various AE-type models and other methods. The system can provide early notifications to operators for maintenance intervention, thus preventing equipment losses due to component failures. It can be foreseen that the proposed VAE-LSTM model will find broader applications in industrial equipment monitoring. Additionally, this study utilized data from the entire lifecycle for research. However, in practical scenarios, obtaining complete lifecycle data can be challenging, and incomplete lifecycle data is more common. Therefore, future work will focus on equipment remaining useful life prediction based on incomplete life cycle data.

Data Availability. The data used to support the findings of this study are included within the article.

Conflicts of Interest. The author declares that there is no conflict of interest regarding the publication of this paper.

Funding Statement. State Grid Qinghai Province Power Company Technology Project: Project Name: Development and Application of Non-stop Transportation System and Cable Force Dynamic Monitoring and Control System for Transmission Line Freight Cableway, Project Number: 522804230001

REFERENCES

[1] W.-D. Wei, J.-S. Li, B. Chen, M. Wang, P.-F. Zhang, D. B. Guan, J. Meng, H.-Q. Qian, Y.-H. Cheng, C.-Q. Kang, K.-S. Feng, Q. Yang, N. Zhang, X. Liang, and J.-J. Xue, "Embodied greenhouse gas emissions from building China's large-scale power transmission infrastructure," *Nature Sustainability*, vol. 3, pp. 739–747, 2021.

[2] M.-J. Zhu, J.-H. Li, K.-Y. Xu, C. Liu, and F. Wang, "Research on parametric design and feature modeling of freight ropeway running car of transmission line," *Journal of Physics: Conference Series*, vol. 2292, p. 012013, 2022.

[3] C. Liu, J. Qin, Y.-J. He, G.-L. Sun, and L. Qiao, "Research on structural design and parametric modeling of trestle of material ropeway of transmission line," *Journal of Physics: Conference Series*, vol. 2160, p. 012071, 2022.

[4] L. B. Baral, J. J. Nakarmi, K. N. Poudyal, N. Karki, and D. Nalmpantis, "Gravity and muscle force operated surface ropeway: An efficient, cheap, and eco-friendly transport mode for mountainous countries," *The European Physical Journal Plus*, vol. 134, p. 55, 2019.
[5] A. V. Lagerev and I. A. Lagerev, "Optimal planning of the mobile cargo ropeway repair strategy," *International Journal of System Assurance Engineering and Management*, vol. 45, pp. 1–13, 2023.
[6] M. Cerrada, R.-V. Sánchez, C. Li, F. Pacheco, D. Cabrera, J. V. Oliveira, and R. E. Vasquez, "A review on data-driven fault severity assessment in rolling bearings," *Mechanical Systems and Signal Processing*, vol. 99, pp. 169–196, 2018.
[7] B. Peng, Y. Bi, B. Xue, M.-J. Zhang, and S.-T. Wan, "A survey on fault diagnosis of rolling bearings," *Algorithms*, vol. 15, p. 347, 2022.
[8] Y.-G. Xu, K. Zhang, C.-Y. Ma, L.-L. Cui, and W.-K. Tian, "Adaptive Kurtogram and its applications in rolling bearing fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 130, pp. 87–107, 2019.
[9] M. Hakim, A. A. Omran, A. N. Ahmed, M. A. Waily, and A. Abdellatif, "A systematic review of rolling bearing fault diagnoses based on deep learning and transfer learning: Taxonomy, overview, application, open challenges, weaknesses and recommendations," *Ain Shams Engineering Journal*, vol. 14, p. 101945, 2023.
[10] C. Malla and I. Panigrahi, "Review of condition monitoring of rolling element bearing using vibration analysis and other techniques," *Journal of Vibration Engineering & Technologies*, vol. 7, pp. 407–414, 2019.
[11] B. Pang, M. Nazari, and G. Tang, "Recursive variational mode extraction and its application in rolling bearing fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 165, p. 108321, 2022.
[12] D. Verstraete, A. Ferrada, E. L. Droguett, V. Meruane, and M. Modarres, "Deep learning enabled fault diagnosis using time-frequency image analysis of rolling element bearings," *Shock and Vibration*, vol. 2017, p. 5067651, 2017.
[13] S. Buchaiah and P. Shakya, "Bearing fault diagnosis and prognosis using data fusion-based feature extraction and feature selection," *Measurement*, vol. 188, p. 110506, 2022.
[14] M. A. Hall, "Correlation-based feature selection for discrete and numeric class machine learning," in *Proceedings of the Seventeenth International Conference on Machine Learning*, pp. 359–366, 2010.
[15] F. Yang, M. S. Habibullah, T.-Y. Zhang, Z. Xu, P. Lim, and S. Nadarajan, "Health index-based prognostics for remaining useful life predictions in electrical machines," *IEEE Transactions on Industrial Electronics*, vol. 63, pp. 2633–2644, 2016.
[16] B. Zhang, L. Zhang, and J. Xu, "Degradation feature selection for remaining useful life prediction of rolling element bearings," *Quality and Reliability Engineering International*, vol. 32, pp. 547–554, 2016.
[17] Z. Yang, X. Wang, and R. Yang, "Transfer learning-based rolling bearing fault diagnosis," in *10th Data Driven Control and Learning Systems Conference*, pp. 354–359, 2021.
[18] J.-M. Li, X.-F. Yao, X.-D. Wang, Q.-W. Yu, and Y.-G. Zhang, "Multiscale local features learning based on BP neural network for rolling bearing intelligent fault diagnosis," *Measurement*, vol. 153, p. 107419, 2020.
[19] C. Lu, Z.-Y. Wang, W.-L. Qin, and J. Ma, "Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification," *Signal Processing*, vol. 130, pp. 377–388, 2017.
[20] J. Yu, "A selective deep stacked denoising autoencoders ensemble with negative correlation learning for gearbox fault diagnosis," *Computers in Industry*, vol. 108, pp. 62–72, 2019.
[21] D.-F. Zhao, S.-L. Liu, D. Gu, X. Sun, L. Wang, Y. Wei, and H.-L. Zhang, "Enhanced data-driven fault diagnosis for machines with small and unbalanced data based on variational auto-encoder," *Measurement Science and Technology*, vol. 31, p. 035004, 2019.
[22] F. Yang, M. S. Habibullah, and Y. Shen, "Remaining useful life prediction of induction motors using nonlinear degradation of health index," *Mechanical Systems and Signal Processing*, vol. 148, p. 107183, 2021.
[23] T. Han, J. Pang, and A.-C. Tan, "Remaining useful life prediction of bearing based on stacked autoencoder and recurrent neural network," *Journal of Manufacturing Systems*, vol. 61, pp. 576–591, 2021.
[24] S.-K. Wang, Z.-Y. Zhang, P.-G. Wang, and Y.-R. Tian, "Failure warning of gearbox for wind turbine based on 3σ-median criterion and NSET," *Energy Reports*, vol. 7, pp. 1182–1197, 2021.

[25] R. Huang, W. Jiang, and G. Sun, "Manifold-based constraint Laplacian score for multi-label feature selection," *Pattern Recognition Letters*, vol. 112, pp. 346–352, 2018.
[26] N. A. Adegoke, A. N. H. Smith, M. J. Anderson, R. A. Sanusi, and M. D. M. Pawley, "Efficient homogeneously weighted moving average chart for monitoring process mean using an auxiliary variable," *IEEE Access*, vol. 7, pp. 94021–94032, 2019.
[27] D. Bank, N. Koenigstein, and R. Giryes, "Autoencoders," in *Machine Learning for Data Science Handbook: Data Mining and Knowledge Discovery Handbook*, pp. 353–374, 2023.
[28] Y. Yu, X.-S. Si, C.-H. Hu, and J.-X. Zhang, "A review of recurrent neural networks: LSTM cells and network architectures," *Neural Computation*, vol. 31, pp. 1235–1270, 2019.
[29] Y.-B. Li, X.-Z. Wang, S.-B. Si, and S.-Q. Huang, "Entropy based fault classification using the Case Western Reserve University data: A benchmark study," *IEEE Transactions on Reliability*, vol. 69, pp. 754–767, 2019.
[30] T.-Y. Wu, H.-N. Li, S. Kumari, and C.-M. Chen, "A spectral convolutional neural network model based on adaptive Fick's law for hyperspectral image classification," *Computers, Materials & Continua*, vol. 79, no. 1, pp. 19–46, 2024.
[31] Y.-R. Ma, Y.-J. Peng, and T.-Y. Wu, "Transfer learning model for false positive reduction in lymph node detection via sparse coding and deep learning," *Journal of Intelligent & Fuzzy Systems*, vol. 43, no. 2, pp. 2121–2133, 2022.