

Article citation info:

Shi L, Li X, Yao W, Liang J, Zhao Y, Liu Y, Reliability Prediction Method for Onboard ATP Based on Optimized Empirical Mode Decomposition, *Eksploracja i Niezawodność – Maintenance and Reliability* 2026: 28(2) <http://10.17531/ein/213708>

Reliability Prediction Method for Onboard ATP Based on Optimized Empirical Mode Decomposition

Indexed by:



Lei Shi^a, Xinran Li^{a,*}, Wenjun Yao^a, Junyi Liang^a, Yuanyuan Zhao^a, Yazhi Liu^a

^a Lanzhou Jiaotong University, China

Highlights

- Time series decomposition is applied to predict ATP failure rates.
- The proposed model demonstrates robust performance with small training datasets.
- A hybrid WLSSVM-temporal decomposition framework is proposed.

Abstract

Automatic Train Protection (ATP) system reliability significantly impacts rail safety and maintenance efficiency, but current strategies lack data-driven spare parts optimization. We propose a hybrid failure rate prediction model combining time-varying filtered empirical mode decomposition (TVFEMD) and machine learning. First, ATP operational data undergoes interval segmentation and zero-value preprocessing. Next, grey wolf optimization (GWO) adaptively tunes TVFEMD parameters to decompose failure rate series into intrinsic mode functions (IMFs). Each IMF is independently predicted via weighted least squares support vector machine (WLSSVM), with final outputs aggregated through superposition. Validated using real ATP system data, the model achieves 0.0028 MAE, 0.0066 RMSE, and 0.0139 MAPE for 10-minute interval predictions with zero-inflated data, surpassing baseline methods. Results confirm its effectiveness for ATP failure rate forecasting.

Keywords

time series, ATP, machine learning, fault prediction, TVFEMD

This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>)

1. Introduction

As high-speed railways develop rapidly, China's train operation control system (CTCS) has become a key technology to ensure traffic safety, of which CTCS-3 level has been widely used as the core system for high-speed railways of 300 km/h and above. Fig. 1 shows a schematic of the overall structure of the CTCS-3 and ATP system. Functioning as the essential safety-critical component of CTCS-3, the ATP delivers continuous train state monitoring and protection functionalities, with its reliability representing the primary safeguard for railway operational safety [1-2]. However, the long-term operation of the ATP system is susceptible to environmental conditions like elevated

temperatures, vibration, electro-magnetic interference, etc. and ageing of the equipment, with the risk of malfunction always present [3]. Once an abnormality occurs in the ATP system, it will lead to the interruption of train operation and may even cause a safety accident. Therefore, accurate prediction of the ATP system's failure rate is a significant research topic for high-speed rail to ensure safe operation.

The on-board ATP system, being vital for train operation safety, requires scientifically grounded testing and maintenance regimes [4]. At present, the maintenance management of on-board ATP equipment usually adopts the strategy of regular

(*) Corresponding author.
E-mail addresses:

L. Shi (ORCID: 0009-0000-1742-0245) shil1215@foxmail.com, X. Li (ORCID: 0009-0007-1416-6473) lxr12231704@163.com, W. Yao (ORCID: 0009-0006-4351-6244) 1433997571@qq.com, J. Liang (ORCID: 0009-0004-9086-0131) ljiywf2021@126.com, Y. Zhao (ORCID: 0009-0009-5600-6835) yuan1841973853@163.com, Y. Liu (ORCID: 0009-0009-7741-1151) liyazhil@163.com

maintenance or ‘after-action maintenance’ based on experience
5. Regular maintenance ignores individual differences in equipment and can lead to over- or under-maintenance, while

the ‘after-the-fact maintenance’ model can lead to operational disruptions due to the lack of an early warning mechanism for failures [6-7].

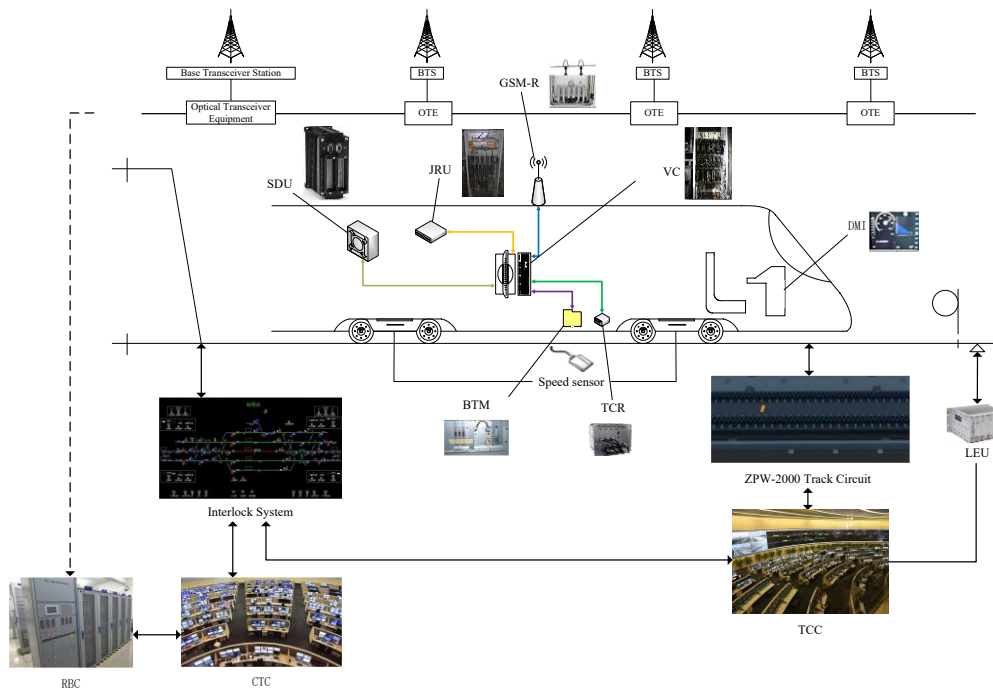


Fig. 1. Overall structure schematic of CTCS-3+ATP for the system.

Currently, data-driven prediction technology has been widely used in predictive maintenance, demand prediction, quality prediction, etc. 8. Data-driven ATP system failure prediction can accurately predict the probability of failure and achieve preventive maintenance by extracting key information and analysing trends in the timing data in its operation log, thus ensuring the efficiency and safety of system operation 9.

However, ATP system failure rate prediction faces serious challenges, mainly stemming from the significant operating data with nonlinear and nonstationary characteristics, and by environmental conditions, operating conditions and equipment status and other multiple factors 10. A single traditional method has limitations in dealing with such complex data, and there is an urgent need to develop more accurate and efficient failure rate prediction models 11.

Existing research on failure rate time series prediction for ATP systems remains scarce. This study consequently develops a novel hybrid methodology integrating Grey Wolf Optimizer (GWO)-based parameter optimisation, Time-Varying Filtered Empirical Mode Decomposition (TVFEMD) for data decomposition, and Weighted Least Squares Support Vector Machine (WLSSVM) for predictive modelling. Depending on

the results of the model prediction, the allocation of spare parts reserves and maintenance resources can be dynamically optimised. Increasing reserves before the failure rate rises and rationally allocating resources when it falls improves maintenance efficiency and reduces wastage.

2. Related Research

The failure rate prediction is a critical tool for the maintenance and operation of ATP system in high-speed railways, and a considerable number of studies have been conducted to propose related prediction methods, such as Autoregressive Integrated Moving Average (ARIMA) and Exponential Smoothing Method (ESM) [12-13]. Yet, the ARIMA model has significant limitations in dealing with the nonlinear dynamics and nonstationary features of the ATP system time series of failure rates due to its inherent assumptions of linearity and smoothness 14. In addition, although neural networks can capture the nonlinear features of data, their training process is susceptible to overfitting [15-18]. In order to overcome these limitations, time-frequency analysis methods have gradually gained attention.

Empirical Mode Decomposition (EMD) operates as a fully adaptive signal processing technique that analyzes complex

waveforms by separating them into Intrinsic Mode Functions (IMFs) and residual elements through an empirical, data-driven approach [19-23]. Although EMD has achieved significant application results in many fields, it still has some problems [24-29]. In order to improve the limitations of traditional EMD, Li et al. proposed a time-varying filter empirical modal decomposition (TVFEMD) method. Through adaptive time-varying filtering, this method achieves concurrent modal aliasing reduction and significant improvements in both decomposition accuracy and frequency band separation effectiveness [30]. Ma et al. used TVFEMD method to decompose millimetre-wave radar signals and successfully extracted heart rate and respiratory signals by reconstructing IMF components [31]. Zang Xu et al. combined TVFEMD with Sparrow Search Algorithm (SSA) optimised Least Squares Support Vector Machine (LSSVM) in the diagnosis of latent mechanical faults inside the transformer, which significantly improved the diagnostic accuracy [32]. Zhiyu Deng et al. proposed a time-varying filter empirical modal quadratic decomposition (TVFEMDII)-improved Deep Hybrid Kernel Extreme Learning Machine (DHKELM) hybrid model, which effectively extracts the implicit information in the complex components through the quadratic decomposition strategy [33]. TAN et al. Processing Input Stacking Integration Algorithm Parameters via Time-Varying Empirical Modal Decomposition to Effectively Improve Machine Learning Models' Sensitivity to Parameters [34]. Yufeng Yin establishes a combined TVFEMD-WOA-LSTM-ARMAX prediction model, which uses optimised LSTM and Autoregressive Moving Average with Extra Input (ARMAX) to predict the high and low frequency subsequences decomposed by TVFEMD, respectively, with improved prediction results [35]. Yufeng Yin establishes a combined TVFEMD-WOA-LSTM-ARMAX prediction model, and uses the optimised LSTM and Autoregressive Moving Average with Extra Input (ARMAX) to improve the prediction effect of the high and low frequency subsequences respectively decomposed by TVFEMD [35].

The TVFEMD method, while capable of decomposing failure rate time series into IMFs, necessitates coupling with prediction models for comprehensive component analysis.. Support Vector Machine (SVM) shows advantages in the field of fault prediction by virtue of its small-sample processing

capability and maximum interval regression characteristics [36-37]. However, the traditional SVM has some restrictions in addressing nonlinear problems. To solve this problem, LSSVM reduces the computational complexity and improves the training efficiency by introducing a quadratic cost function, but there is still the problem of information loss [38-43]. Suykens et al. proposed WLSSVM, which has achieved good application results in areas such as fault prediction and diagnosis due to its capability of excellent small sample processing [44]. In recent years, more and more studies have begun to combine WLSSVM with other methods to construct more accurate hybrid prediction models. R. Wang et al. combined the integrated empirical modal decomposition Integrated Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (ICEEMDAN) with WLSSVM to predict the decomposed photovoltaic power using the improved WLSSVM algorithm of the slime mould algorithm [45]. In order to further optimise the prediction performance of WLSSVM, Chen Kun et al. improved the Grey Wolf Optimisation (GWO) algorithm, which was used for the optimisation of penalty coefficients and kernel parameters of the WLSSVM model [46]. Lei Xiaoxing used the multivariate universe algorithm to optimise the regularisation parameters and kernel function parameters of the WLSSVM, and verified that the improved WLSSVM was better than the LSSVM in predicting small samples [47].

This paper combines the advantages of TVFEMD and WLSSVM, and introduces GWO to optimise the parameter selection problem of TVFEMD, and proposes the GWO-TVFEMD-WLSSVM prediction model, and discusses the effectiveness and applicability of the model.

3. Methodology

3.1. Optimisation parameters of the grey wolf algorithm

GWO simulates the social hierarchy and hunting mechanism of wolf packs and achieves parameter optimisation through global exploration and local exploitation. Let each grey wolf individual represent a set of candidate $[\xi, n]$ parameter combinations.

The grey wolf algorithm's population is divided into four classes, a, b, c and d, where a, b and c wolves simulate the process of tracking, encircling and striking prey by wolves, and d wolves are guided by these three to perform search and update operations [48].

The GWO algorithm models three main behaviors:

The wolf encirclement process:

$$\vec{D} = |\vec{C} \cdot \vec{X}_p(t) - \vec{X}(t)| \quad (1)$$

$$\vec{X}(t+1) = \vec{X}_p(t) - \vec{A} \cdot \vec{D} \quad (2)$$

$$\vec{A}(t) = 2\vec{a}(t) \cdot \vec{r}_1 - \vec{a}(t) \quad (3)$$

$$\vec{C} = 2\vec{r}_2 \quad (4)$$

$$a(t) = a \frac{t}{T} \min_{max} \quad (5)$$

$\vec{D}$ denotes the distance from other grey wolves to the leader grey wolf, $\vec{X}_p(t)$ is the position of d-wolf in the current pack, $\vec{X}(t+1)$ is the updated position of d-wolf, t is the current number of iterations, and $\vec{A}$ and $\vec{C}$ are the coefficient vectors. $\vec{a}$ is the convergence factor, which decreases linearly from a_{max} to a_{min} with the number of iterations, T is the maximum number of iterations, and $\vec{r}_1$ and $\vec{r}_2$ are random vectors between $[0,1]$.

The hunting process:

$$\begin{cases} \vec{D}_a = |\vec{C}_1 \cdot \vec{X}_a - \vec{X}(t)| \\ \vec{D}_b = |\vec{C}_2 \cdot \vec{X}_b - \vec{X}(t)| \\ \vec{D}_c = |\vec{C}_3 \cdot \vec{X}_c - \vec{X}(t)| \end{cases} \quad (6)$$

$$\begin{cases} \vec{X}_1 = \vec{X}_a - \vec{A}_1 \cdot \vec{D}_a \\ \vec{X}_2 = \vec{X}_b - \vec{A}_2 \cdot \vec{D}_b \\ \vec{X}_3 = \vec{X}_c - \vec{A}_3 \cdot \vec{D}_c \end{cases} \quad (7)$$

$$\vec{X}(t+1) = \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3} \quad (8)$$

where $\vec{X}_a$, $\vec{X}_b$ and $\vec{X}_c$ are the positions of wolves a, b and c, respectively, $\vec{D}_a$, $\vec{D}_b$ and $\vec{D}_c$ are the distances between wolves a, b and c and wolf d, respectively, and $\vec{X}_1$, $\vec{X}_2$ and $\vec{X}_3$ are the

lengths and directions of the steps taken by grey wolf d towards wolves a, b and c, respectively.

The attack process is mainly affected by $\vec{A}$, and the size of $|\vec{A}|$ determines the attack strategy of the grey wolves. When $|\vec{A}| > 1$, the rest of the grey wolves will move away from the leader grey wolf and expand the search range; when $|\vec{A}| \leq 1$, the pack will gradually shrink to encircle the prey and gradually approach the optimal solution.

GWO assumes that the optimal individuals in the population (i.e., a, b, and c) adequately reflect the high-quality solutions in the search space, and uses these three to guide the updating direction of the other individuals so as to achieve the global optimum or near-optimum. During the iteration process, GWO controls the search breadth and accuracy through random coefficients to reduce the risk of falling into the local optimum.

While the Grey Wolf Optimizer (GWO) is originally inspired by the social hierarchy and collaborative hunting strategies of grey wolves in nature, its application in this study is purely algorithmic rather than biological. There is no direct physical or functional correspondence between the hunting behavior of wolves and ATP. Instead, GWO is employed as a global optimization tool to tune key hyperparameters within the TVFEMD decomposition process.

During the iterative optimisation process, the grey wolf algorithm eventually obtains the combination of parameters that minimises the envelope entropy by renewing the positions of the individuals, schematically shown in Fig. 2.

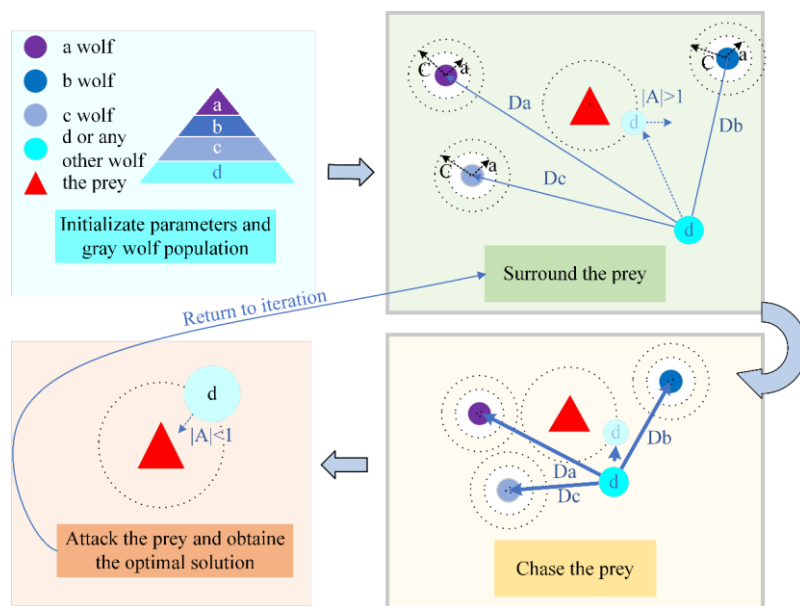


Fig. 2. Schematic flow of the grey wolf algorithm.

In ATP failure rate prediction, the GWO-optimized parameters significantly enhance the TVFEMD decomposition quality. Comparative experiments show that this adaptive optimization improves prediction accuracy and stability over fixed-parameter variants. Compared with other traditional optimization algorithms, GWO offers advantages such as a simple mechanism and fewer decisive parameters. By simulating the optimal individual mechanism of wolf-pack hunting behavior with relatively low parameter requirements, it achieves a balance between global exploration and local exploitation, outperforming Particle Swarm Optimization (PSO) and Genetic Algorithm (GA). These characteristics enable GWO to avoid entrapment in local extrema when addressing nonlinear and nonstationary time series optimization, thereby ensuring both decomposition accuracy and computational efficiency while meeting real-time application requirements.

In the context of ATP failure rate prediction, the underlying time series data are highly nonlinear, nonstationary, and often corrupted with noise. To improve the interpretability and predictive utility of such data, the TVFEMD method is introduced to extract multiscale features from the original signal. However, the decomposition performance of TVFEMD is sensitive to parameter selection, and manual tuning is prone to suboptimal configurations. The GWO algorithm solves this challenge by adaptively searching for the optimal parameter pair, specifically, the filter bandwidth threshold ξ and the filter order n , to minimise the envelope entropy and thus improve the quality of the decomposed components [49].

Envelope entropy is a metric for signal complexity based on information entropy, indicating the sparsity properties of the signal [50]. In the GWO optimisation process, the envelope entropy is adopted as the objective feature to find the minimal local entropy as the optimisation goal, thus guiding the parameter search process [51]. Literature [51] has proposed the use of envelope of entropy to measure the decomposition effectiveness of an algorithm, and the envelope entropy H can be expressed as:

$$H = -\sum_{i=1}^N P_i \log P_i \quad (9)$$

N is the aggregate amount of IMFs obtained by disaggregation and P_i is the normalised energy of the i -th component of the signal, defined as:

$$P_i = \frac{E_i}{\sum_{j=1}^N E_j} \quad (10)$$

E_i denotes the energy of the i -th component of the signal, which is calculated as:

$$E_i = \int_{t_0}^{t_1} |IMF_i(t)|^2 dt \quad (11)$$

$IMF_i(t)$ denotes the i -th eigenmode function, t_0 and t_1 are the time ranges of the signal.

Through the above process, the optimised TVFEMD method can effectively decompose the failure rate time series into several IMFs with different time scales and frequency characteristics.

3.2. Principle of TVFEMD

In this paper, the input data for TVFEMD decomposition refers to the failure rate time series derived from ATP system operation logs. This signal captures the temporal evolution of fault frequency within fixed time intervals and serves as the basis for multiscale decomposition and prediction. The specific construction and preprocessing procedures for this sequence are detailed in Section 4.1.

TVFEMD optimises the signal decomposition process by constructing a time-varying filter (TVF) through a B-spline approximation to obtain better noise immunity and mode separation [52]. The flowchart is shown in Fig. 3.

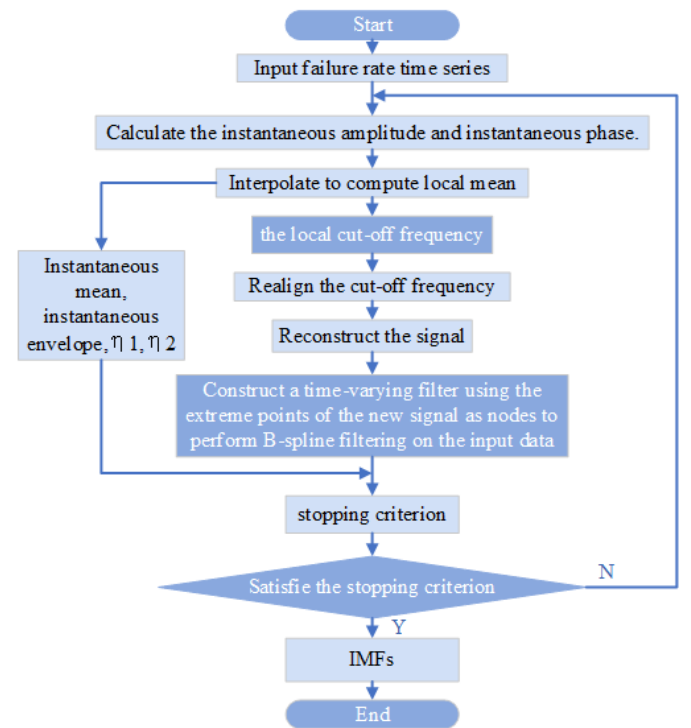


Fig. 3. Flowchart of TVFEMD.

The main steps in TVFEMD are:

1) Calculate the local cut-off frequency

The Hilbert transform of an N-component signal $X(t)$ yields its analytic signal $Z(t)$, $Z(t)$ which can be expressed as a combination of two signals as:

$$Z(t) = X(t) + j \cdot H(X(t)) = A(t)e^{j\varphi_1(t)} = a_1(t)e^{j\varphi_1(t)} + a_2(t)e^{j\varphi_2(t)} \quad (12)$$

where $H(X(t))$ is the Hilbert transform of $X(t)$, $A(t)$ denotes the instantaneous amplitude, and $\varphi(t)$ denotes the instantaneous phase. a_1 and a_2 are the amplitudes of the first and second components respectively, and $\varphi_1(t)$ and $\varphi_2(t)$ are the phases of the first and second components respectively.

Specifically, the instantaneous amplitude $A(t)$ reflects the energy magnitude or envelope strength of the signal at a given moment in time, defined as:

$$A(t) = \sqrt{X(t)^2 + H(X(t))^2} \quad (13)$$

The instantaneous phase $\varphi(t)$ then denotes the phase evolution at that point, which is used to further calculate the instantaneous frequency, defined as:

$$\varphi(t) = \arctan\left(\frac{H(X(t))}{X(t)}\right) \quad (14)$$

When $N = 2$, the instantaneous amplitude and instantaneous phase of $X(t)$ can be expressed as:

$$A^2(t) = a_1^2(t) + a_2^2(t) + 2a_1(t)a_2(t)\cos[\varphi_1(t) - \varphi_2(t)] \quad (15)$$

$$\begin{aligned} \varphi'(t) = \frac{1}{A^2(t)} & (\varphi_1'(t)(a_1^2(t) + a_1(t)a_2(t)\cos[\varphi_1(t) - \varphi_2(t)]) \\ & + \varphi_2'(t)(a_2^2(t) + a_1(t)a_2(t)\cos[\varphi_1(t) - \varphi_2(t)])) \\ & + \frac{1}{A^2(t)} (a_1'(t)a_2(t)\sin[\varphi_1(t) - \varphi_2(t)] \\ & - a_2'(t)a_2(t)\sin[\varphi_1(t) - \varphi_2(t)]) \end{aligned} \quad (16)$$

The TVF separates the similar components within the signal by means of a local cut-off frequency, which can be approximated as the bisecting frequency $\varphi'_{bis}(t)$ of $\varphi_1(t)$ and $\varphi_2(t)$.

$$\varphi'_{bis}(t) = \frac{\varphi_1'(t) + \varphi_2'(t)}{2} = \frac{\eta_2(t) - \eta_1(t)}{4a_1(t)a_2(t)} \quad (16)$$

Solve for the values of $a_1(t)$ and $a_2(t)$, let:

$$\beta_1(t) = |a_1(t) - a_2(t)| \quad (17)$$

$$\beta_2(t) = a_1(t) + a_2(t) \quad (18)$$

$\beta_1(t)$ and $\beta_2(t)$ are obtained by interpolating the local extrema estimates of $A(t)$. Setting $a_1(t) \geq a_2(t)$, the values of $a_1(t)$ and $a_2(t)$ can be solved for. Let:

$$\begin{aligned} \eta_1(t) = \varphi_1'(t)[a_1^2(t) - a_1(t)a_2(t)] + \varphi_2'(t)[a_2^2(t) - \\ a_1(t)a_2(t)] \end{aligned} \quad (19)$$

$$\eta_2(t) = \varphi_1'(t)[a_1^2(t) + a_1(t)a_2(t)] + \varphi_2'(t)[a_2^2(t) + a_1(t)a_2(t)] \quad (20)$$

$\eta_1(t)$ and $\eta_2(t)$ can be obtained by interpolating the local extremes of $\varphi_1'(t)A^2(t)$. Solving for the values of $\eta_1(t)$ and $\eta_2(t)$, $a_1(t)$ and $a_2(t)$, and substituting into Eq. (16) gives the local cut-off frequency of the signal.

2) B-spline approximate filtering of the input signal

Let $\beta^n(t)$ be the n th order B-spline order and m be the section sequence step, the signal in this B-spline space is defined as:

$$g_m^n(t) = \sum_{k=-\infty}^{\infty} c(k)\beta^n\left(\frac{t}{m} - k\right) \quad (21)$$

Where $c(k)$ is the B-spline coefficient. Given m and n , determine $c(k)$ such that the approximation error is minimised and substitute into Eq. (21).

$$g_m^n(t) = [p_m^n * x]_{lm} * b_m^n(t) \quad (22)$$

Constructing the new signal $h(t)$ with a local cut-off frequency and using the extreme points of $h(t)$ as nodes, a B-spline approximation filter is used on the input signal $X(t)$, which results in a filter of $m(t)$.

$$h(t) = \cos[\int \varphi'_{bis}(t) dt] \quad (23)$$

3) Verify that the remnant signal satisfies the stopping criterion

The weighted average of instantaneous frequencies (WAIF) of each component can be defined as:

$$\varphi_{avg}(t) = \frac{a_1^2(t)\varphi_1'(t) + a_2^2(t)\varphi_2'(t)}{a_1^2(t) + a_2^2(t)} \quad (24)$$

According to Loughlin instantaneous bandwidth, the instantaneous bandwidth is defined as the standard deviation of the WAIF, then the two-component instantaneous bandwidth can be expressed as:

$$B_{Loughlin}(t) = \sqrt{\frac{a_1'^2(t) + a_2'^2(t)}{a_1^2(t) + a_2^2(t)} + \frac{a_1^2(t)a_2^2(t)(\varphi_1'(t) - \varphi_2'(t))^2}{(a_1^2(t) + a_2^2(t))^2}} \quad (25)$$

When $B_{Loughlin}(t)$ is small enough, the signal can be considered as an IMF. Define $\theta(t)$ to measure the extent to which $B_{Loughlin}(t)$ deviates from the WAIF, which is the stopping criterion. Given a threshold ξ , the signal is considered to be IMF when $\theta(t) \leq \xi$, otherwise set $X(t) = X(t) - m(t)$, and repeat the screening process.

$$\theta(t) = \frac{\varphi_{avg}(t)}{B_{Loughlin}(t)} \quad (26)$$

3.3. WLSSVM forecasts

Literature [53] used the ε -SVR model for prediction and demonstrated that the ATP failure rate time series has volatility and randomness, and some data points may contain richer fault characteristics or higher information weights. In order to adapt to the non-stationarity and local volatility characteristics of ATP fault rate data, this paper introduces the WLSSVM, and constructs a sample weighting mechanism on the basis of LSSVM to optimise the model performance. The objective function of LSSVM is given in Eq. (28).

$$J(w, \xi) = \frac{1}{2} \|w\|^2 + \frac{\gamma}{2} \sum_{i=1}^N \xi_i^2 \quad (27)$$

Where γ is the penalty coefficient and all training samples share the same penalty coefficient. WLSSVM introduces a weight matrix in front of the error term to adjust its influence in the error term. This feature makes WLSSVM more suitable for complex time series with strong volatility and non-stationarity such as the ATP failure rate series. The objective function of WLSSVM can be expressed as follows:

$$J(w, \xi) = \frac{1}{2} \|w\|^2 + \frac{\gamma}{2} \sum_{i=1}^N W_i \xi_i^2 \quad (28)$$

w is the weight vector. The weight vector w is not defined in the WLSSVM model by an explicit function, but is obtained by solving the optimisation problem indirectly by the Lagrangian dyadic method, which means the set of directional coefficients in the eigenspace of the kernel function mapped to the regression function with the largest contribution to the regression function. ξ_i is the prediction error, W_i is the weight factor of the samples, which is used to adjust the contribution of each sample in the error term and can be expressed as:

$$W_i = \frac{1}{(y_i - \hat{y}_i)^2} \quad (29)$$

y_i is the true value of sample i and $\hat{y}_i$ is the model prediction. N is the total number of samples. Set $i = 1, 2, \dots, N$, the constraints are:

$$y_i = w^T \phi(x_i) + b + \xi_i \quad (30)$$

b is the bias term. $\phi(x_i)$ is the mapping function that nonlinearly maps the input sample x_i to the high-dimensional feature space, and y_i is the actual output value of the sample. To remove the constraints, the Lagrange multiplier method is used to construct the Lagrange function:

$$L(w, b, \xi, \alpha) = \frac{1}{2} \|w\|^2 + \frac{\gamma}{2} \sum_{i=1}^N W_i \xi_i^2 -$$

$$\sum_{i=1}^N \alpha_i [w^T \phi(x_i) + b + \xi_i - y_i] \quad (31)$$

Where $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_N]^T$ is the Lagrange multiplier.

The Karush-Kuhn-Tucker (KKT) condition for the optimisation problem is obtained by taking the partial derivatives of each variable w, b, ξ and α and setting them to zero. Substituting the KKT condition into the constraints, the weight vector w can be expressed as a weighted linear combination of the projections of the training samples in the feature space.

$$w = \sum_{i=1}^N \alpha_i \mu_i \phi(x_i) \quad (32)$$

By eliminating the original variables w, ξ_i and b , the matrix represents the constraints for all the samples, obtaining the joint equation.

$$\alpha = (\Omega + W^{-1})^{-1} y \quad (33)$$

$\Omega_{ij} = K(x_i, x_j)$ is the kernel matrix, $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$ is the kernel function, which is used to compute the similarity between the input samples; W is the weight diagonal matrix, and $W = \text{diag}(W_1, W_2, \dots, W_N)$, $y = [y_1, y_2, \dots, y_N]^T$ are the target value vectors of the samples.

By solving the system of equations to get α and b

$$\begin{bmatrix} 0 & 1^T \\ 1 & \Omega + W^{-1} \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \quad (34)$$

The final WLSSVM prediction function can be expressed as follows.

$$f(x) = \sum_{i=1}^N \alpha_i K(x, x_i) + b \quad (35)$$

$K(x, x_i)$ is the kernel function, which represents the mapping relationship between the input samples and the training samples. In this paper, we use the RBF kernel function, which is denoted as:

$$K(x, x_i) = \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right) \quad (36)$$

Where σ is the nuclear parameter.

3.4. GWO-TVFEMD-WLSSVM model

In this paper, GWO is used to optimise the parameters of TVFEMD, and then the optimised TVFEMD algorithm is employed to decimate the failure rate time series to establish the GWO-TVFEMD-WLSSVM failure rate prediction model, with the following main steps, and the flowchart is shown in Fig. 4.

1) Input failure rate time series. Preprocess the ATP failure rate data and classify the training set and test set.

2) Setting initial parameters and initialising grey wolf populations. Set the relevant parameters of the GWO algorithm

and the TVFEMD algorithm, and use the parameters of TVFEMD as the optimisation-seeking target.

3) Perform parameter optimisation search. Fitness values for each individual grey wolf decomposition result are calculated, and the optimal parameters are searched iteratively to finally obtain the optimal TVFEMD parameters.

4) TVFEMD decomposition using optimal parameters. The time series data of failure rates is subjected to decomposition

via the optimized TVFEMD.

5) A WLSSVM prediction model was built separately for each component. For each IMF component and residual term, build a separate WLSSVM model for prediction.

6) Get the prediction results. The predictions of all the sub-sequences were integrated to obtain the final estimates of the failure rate data. The predictive effectiveness of the model was also estimated using metrics including RMSE, MAE and MAPE.

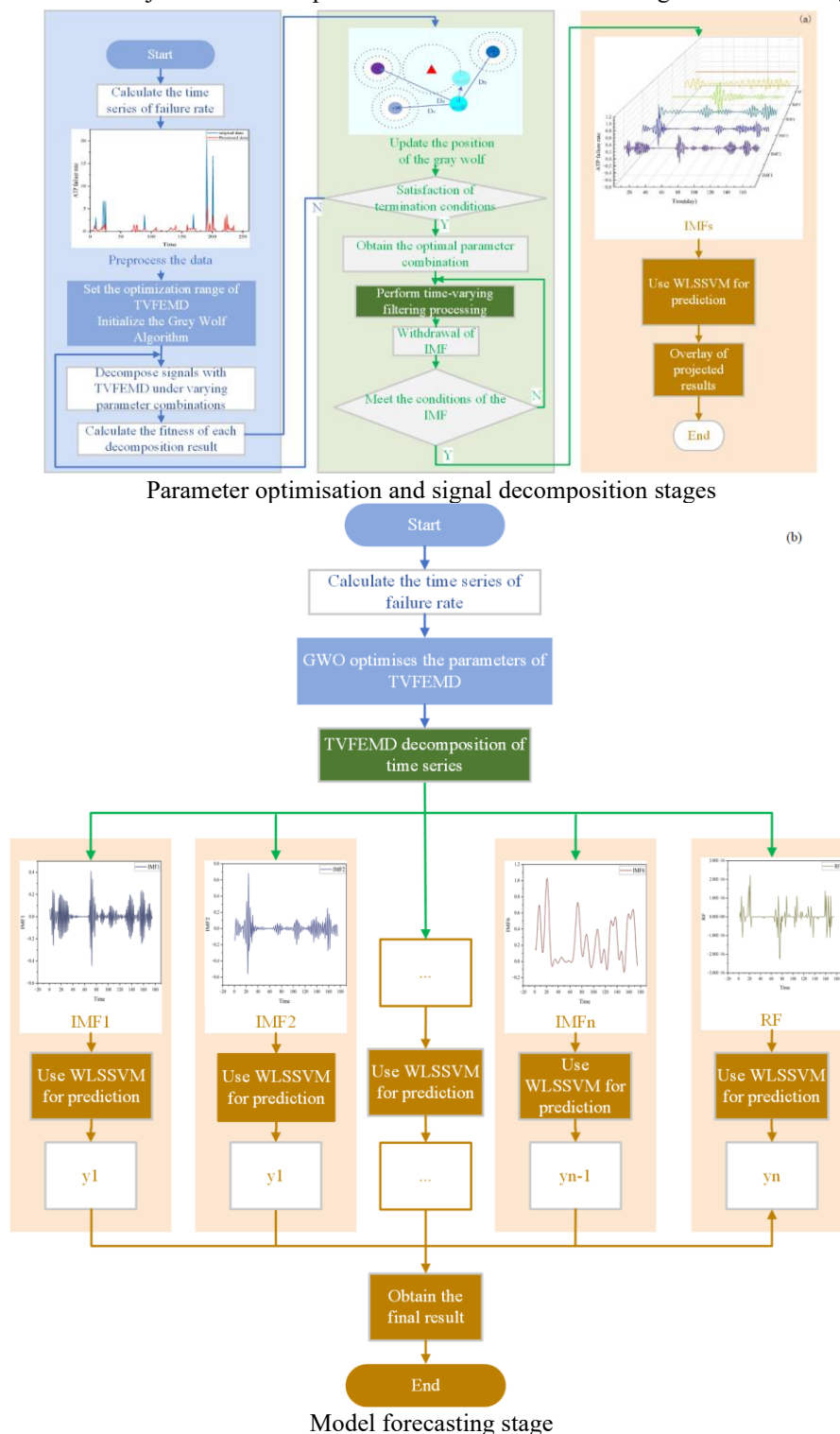


Fig. 4. Flowchart of the GWO-TVFEMD-WLSSVM model.

While previous studies such as 53 employed chaos-based modeling strategies for ATP failure rate series, assuming the presence of strong deterministic nonlinear dynamics, our method follows a data-driven decomposition–prediction paradigm. Preliminary nonlinear analysis on our dataset revealed only weak low-dimensional structure, and no consistent evidence of chaotic behavior. Therefore, no explicit chaotic dynamics model is constructed in this paper, but these complex time-evolutionary properties provide important theoretical insights in the model selection process. Given the stochastic, noisy, and nonstationary characteristics of real-world ATP fault data, the integration of TVFEMD and WLSSVM provides a more flexible and robust modeling alternative without relying on strict dynamical assumptions.

4. Case study

4.1. data processing

The figures in this article are derived from the ATP system's operating logs. These log texts record information about faults that occur during the actual operation of the ATP system. Based on the semantic characteristics of the ATP system operation log, ‘failure’ is defined as events that can be captured by the alarm log and have engineering impact, such as communication anomalies, module failures, hardware fluctuations, and so on, that occur during the operation of the system. To automatically extract and analyse the operational data of the on-board ATP system, the first step is to process the text. Keywords are extracted from the existing fault case database to establish a fault case thesaurus as shown in

Table 1. Thesaurus of some fault cases.

Fault Alarm Statements	Clarification	Keywords
VDX EBR1 port switched to invalid	Safety input/output unit (VDX) EBR1 port switched to invalid	invalid
VDX EBR1 port switched to valid	VDX EBR1 port switched to active	valid
BI-H VDX1:IN1 I/O failed	It is recommended to refer to the corresponding fault case processing	failed
BI-H EBR1 state wrong	BI-H EBR1 status error	wrong
BI-H EBR1 feedback timeout	BI-H EBR1 feedback timeout. EBR1 relay faulty, consider replacing EBR1 relay	feedback timeout
C018E000C016E03CA001	Abnormalities detected in tests performed by the VDX unit on the FS output port at 32ms intervals	C018

Regular expressions are used to match the text of running logs and filter the entries containing keywords from the log files. Based on the existing case base, the same fault is usually accompanied by multiple alarm messages, and the time intervals of these alarm messages usually do not exceed 10 minutes. Therefore, in this paper, fault alarm statements are classified and organised in 5-minute and 10-minute intervals. The failure rate is calculated according to Eq. (38), where f and t_i are the number of failures and the working hours in the i th day, respectively.

$$\lambda = \frac{\sum_{i=1}^N f_i}{\sum_{i=1}^N t_i}$$

(37)

The sliding window approach combined with the Z-Score criterion is used in this paper to identify outliers, when the absolute value of the Z-Score at a point exceeds 2, it is considered as an outlier and is replaced with the rolling mean at that location, and the optimal window size is defined by cross-validation. Model validation employs a conventional 70% training and 30% testing partition, maintaining rigorous separation between development and evaluation datasets to guarantee unbiased assessment of generalization characteristics.

This paper adopts a time series-based modelling approach. Different from the traditional reliability theory, which often models ‘failure’ as a system state crossing a failure surface in a multidimensional covariate space (e.g., stress-strength model), the constructed time series of failure rate essentially reflects the

dynamic evolution of the system state in the time dimension. Although the failure boundaries are not explicitly constructed, the multi-scale decomposition of the signals by TVFEMD can extract the intrinsic behavioural characteristics of the system under different degradation modes, thus forming a data-driven equivalent expression of the idea of ‘crossing the failure surface’, which is suitable for engineering scenarios lacking multi-dimensional physical characterization.

4.2. Data source

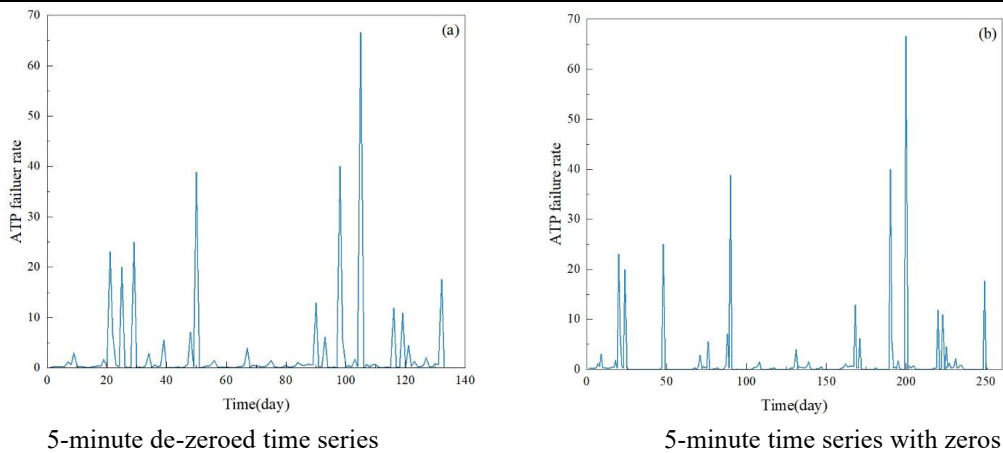
To validate the accuracy of the proposed GWO-TVFEMD-WLSSVM model, the operational data of the on-board ATP system of a road section from December 2021 to July 2022 is selected as a case study, and the number of faults is extracted and calculated. The failure rate is calculated in terms of days, and two time series of failure rate by date are generated. Each data point in the sequence represents the failure frequency within a fixed time window (e.g., 5 or 10 minutes), forming a univariate time series with temporal ordering and nonstationary characteristics.

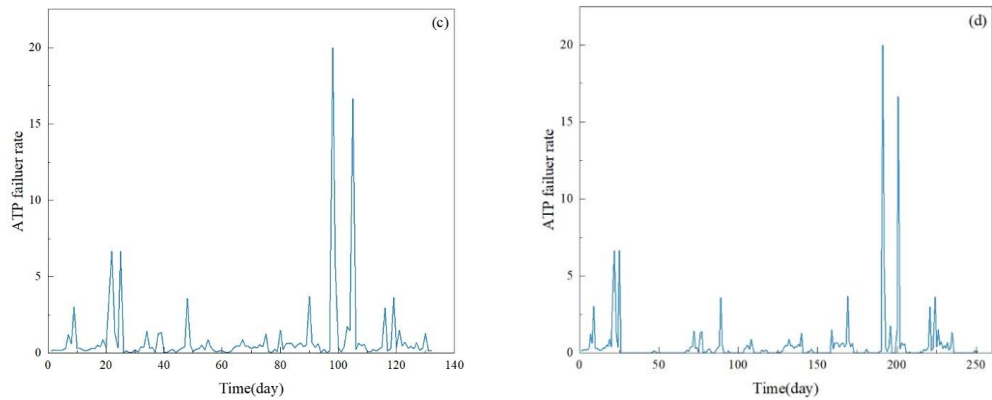
In order to explore the effect of zero values on prediction, the non-zero sequence after removing zero values was further constructed so as to obtain two sets of data, the original sequence (251 points) and the non-zero sequence (186 points), which represent the operational characteristics of the actual system and the changes in the system dynamics during the fault event, respectively, and are used for comparative analyses. As can be seen from the figure, the failure rate data under the 5-minute window is more volatile and sparse, especially in the presence of a high number of zeros, and the dynamics of the data change more frequently. This makes the model need to cope with more complex short-term fluctuations, and thus the prediction results under the 5-minute window show greater volatility. The data under the 10-minute window is time-aggregated, and the effects of noise and zero values are effectively suppressed, resulting in a smoother prediction curve that shows a more stable trend. To comprehensively characterise the time series dataset, the descriptive statistical analysis table is employed in this paper to show the overall characteristics of the data, as illustrated in

.By comparing statistics such as maximum, minimum, mean, standard deviation, skewness and kurtosis, it can be observed that there are significant differences between the various data sets.

Table 2. Table of descriptive statistics analysed for the four datasets.

	Max	Min	Mean	Std	Skewness	Kurtosis
5-minute de-zeroed time series	66.667	0.043	2.684	8.340	5.122	33.147
5-minute time series with zeros	66.667	0	1.422	6.207	7.130	62.416
10-minute de-zeroed time series	20	0.046	0.961	2.449	5.897	41.703
10-minute time series with zeros	20	0	0.505	1.837	8.003	76.425





10-minute de-zeroed time series

10-minute time series with zeros

Fig. 5. Time series of failure rates for different treatments.

The zero-valued data series show stronger sparsity and imbalance, especially in the 10-minute time granularity, its skewness and kurtosis reach 8.003 and 76.425, respectively, indicating that most of the time period of the failure rate is zero, and there are a few time periods with very high outliers. In contrast, after the de-zeroing process, the mean and standard deviation of the data are significantly improved, and the information density is enhanced, providing a clearer signal representation for data-driven modelling.

Furthermore, the significant differences in skewness, kurtosis and standard deviation of the different datasets in Table 2 provide theoretical support for the selection of modelling approaches. The highly skewed and heavy-tailed distributions reflect the abrupt and clustered nature of failure events over time, thereby reinforcing the suitability of TVFEMD for isolating modal components with distinct temporal-frequency characteristics. In parallel, the increased standard deviation and the presence of extreme values indicate higher information concentration at certain time points, justifying the incorporation of WLSSVM with a sample-weighting mechanism to enhance the model's sensitivity to critical instances and improve both predictive accuracy and robustness.

4.3. Determining model parameters and building model predictions

TVFEMD decomposition is used for the time series of this paper. The number of iterations for the grey wolf algorithm is chosen to be 100 and the population size to be 20. To avoid overdecomposition, reconstruction errors are used to identify the optimal numbers of IMFs. The optimal number of Intrinsic Mode Functions is determined through a reconstruction-based validation approach, where decomposed IMFs and residual components are combined to reconstruct the original signal, followed by error quantification between the reconstructed and original signals. The reconstruction errors corresponding to each maximum number of IMFs are shown in Table 3. Setting the optimal number of IMFs as 6. The fault rate time series were decomposed based on the TVFEMD method, and the WLSSVM model was built for the components to predict them respectively, and Fig. 6 shows the results of GWO-TVFEMD decomposition of the four time series datasets.

Table 3. Reconstruction errors for the optimal number of IMFs.

Number of IMFs	5-minute de-zeroed	5-minute with zero value	10-minute de-zeroed	10-minute with zero value
5	0.000020	0.000023	0.000009	0.000012
6	0.000020	0.000019	0.000009	0.000012
7	0.000021	0.000023	0.000009	0.000012
8	0.000021	0.000027	0.000010	0.000012
9	0.000023	0.000018	0.000013	0.000012
10	0.000020	0.000025	0.000011	0.000012
15	0.000022	0.000025	0.000012	0.000018
20	0.000024	0.000025	0.000012	0.000018

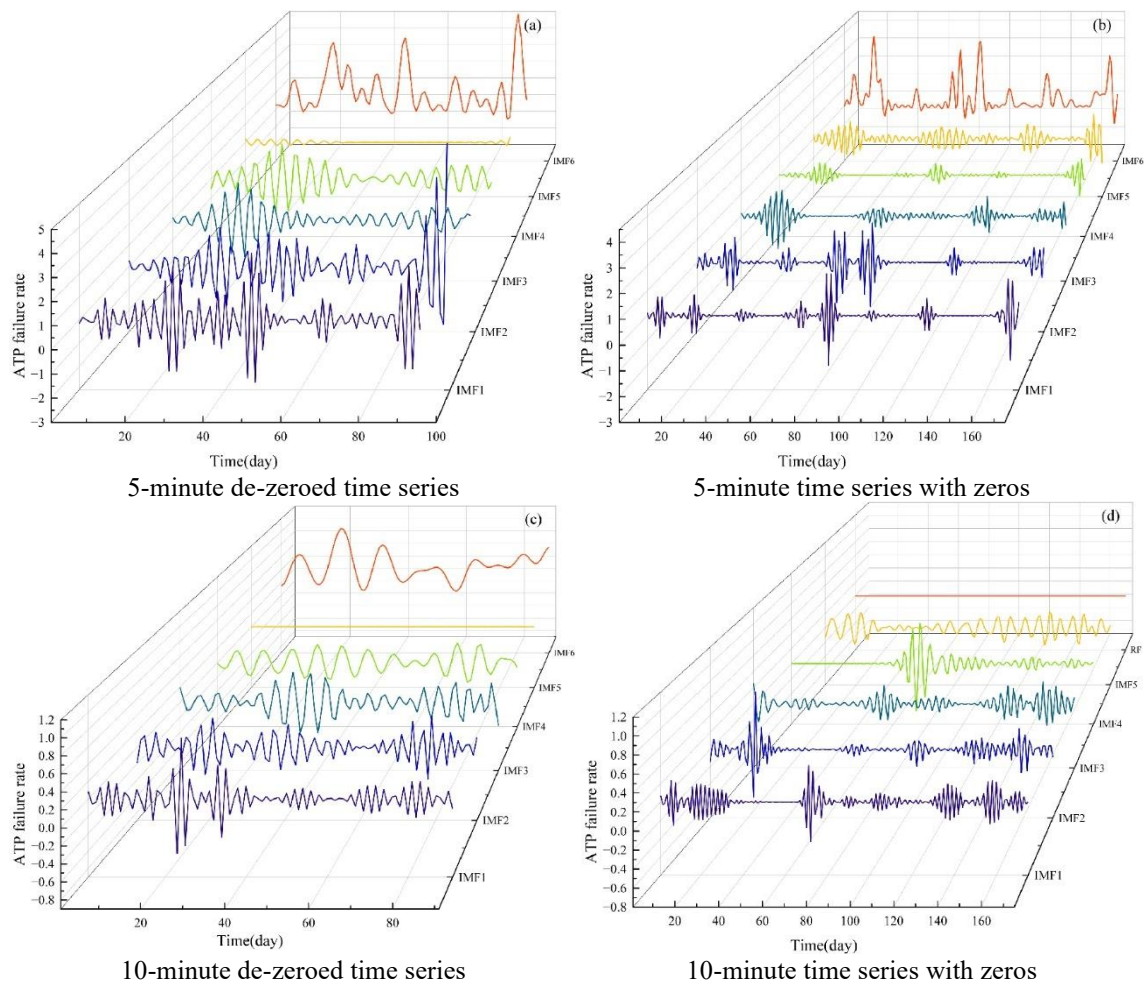


Fig. 6. Decomposition results for different time series.

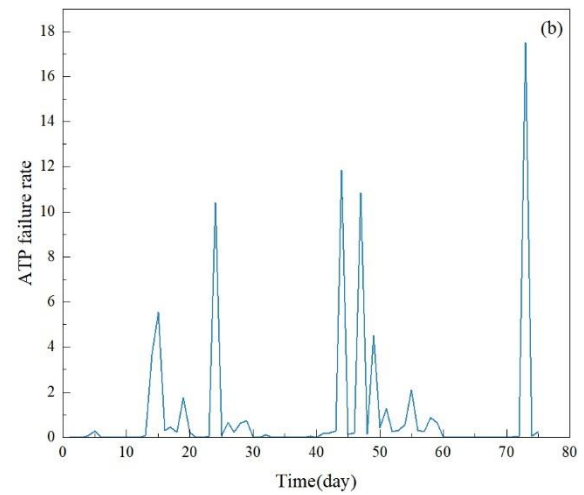
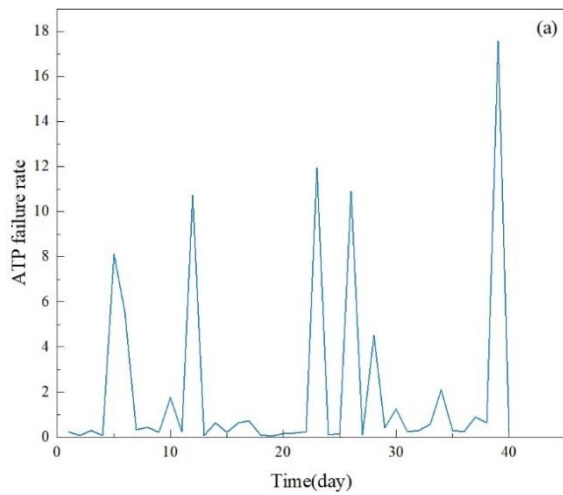
Before prediction, lag features are first created 54. The lag order determination employs both Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), with an optimization range spanning 1 to 20. When optimal AIC and BIC values indicate different lag orders, the median value is selected.

presents the component-specific optimal lag orders, which are subsequently used to generate corresponding lag features for each decomposed subsequence. These subsequences serve as labels for the WLSSVM model, while their associated lag features constitute the model inputs. Through grid search methodology, optimal combinations of parameters are identified

Table 4. Optimal lag order for each time series component.

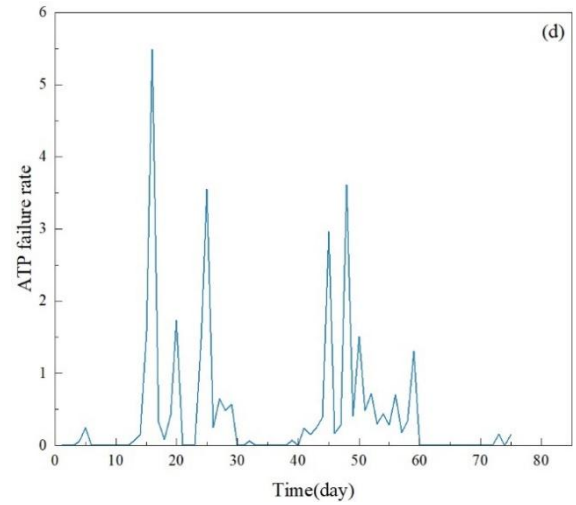
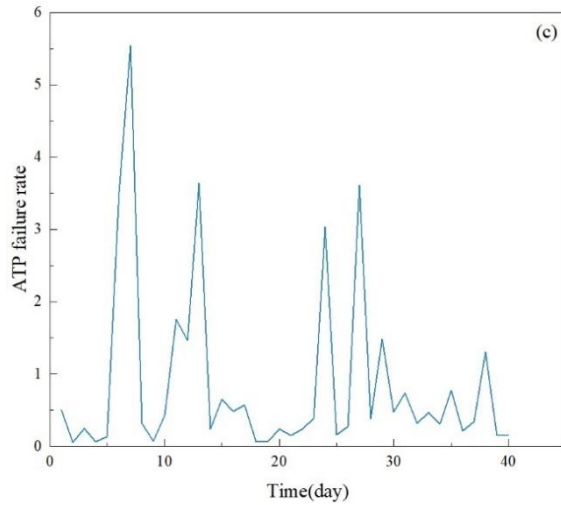
	AIC/BIC	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6	RF
5-minute time series with zeros	Optimal AIC Corresponding Order	20	16	20	20	20	10	1
	Optimal BIC Corresponding Order	15	16	20	17	20	10	1
10-minute time series with zeros	Optimal AIC Corresponding Order	15	20	19	18	14	10	1
	Optimal BIC Corresponding Order	15	20	14	16	12	10	1
5-minute de-zeroed time series	Optimal AIC Corresponding Order	16	17	20	18	16	7	1
	Optimal BIC Corresponding Order	16	17	20	15	16	7	1
10-minute de-zeroed time series	Optimal AIC Corresponding Order	13	18	12	10	6	5	1
	Optimal BIC Corresponding Order	13	13	11	6	6	5	1

for each IMF component, with the regularization parameter C ranging from 0.1 to 10000 and the kernel function parameter $K1$ varying between 0.1 and 10. The final prediction results, obtained by aggregating all component predictions, are visualized in Fig. 7.



5-minute de-zeroed fault time series results

5-minute time series results with zero values



10-minute de-zeroed fault time series results

10-minute time series results with zero values

Fig. 7. Prediction results for different time series.

4.4. Assessment of indicators

In this paper, MAE, RMSE and MAPE are chosen to evaluate the model. For each of the four time series, their corresponding errors are computed and summarized in

Table 5. Prediction errors for the four time series.

	5-minute with zeros	10-minute with zeros	5-minute de-zeroed fault	10-minute de-zeroed fault
MAE	0.0097	0.0033	0.0050	0.0120
RMSE	0.0259	0.0094	0.0181	0.0257
MAPE	0.0452	0.0167	0.0130	0.0303

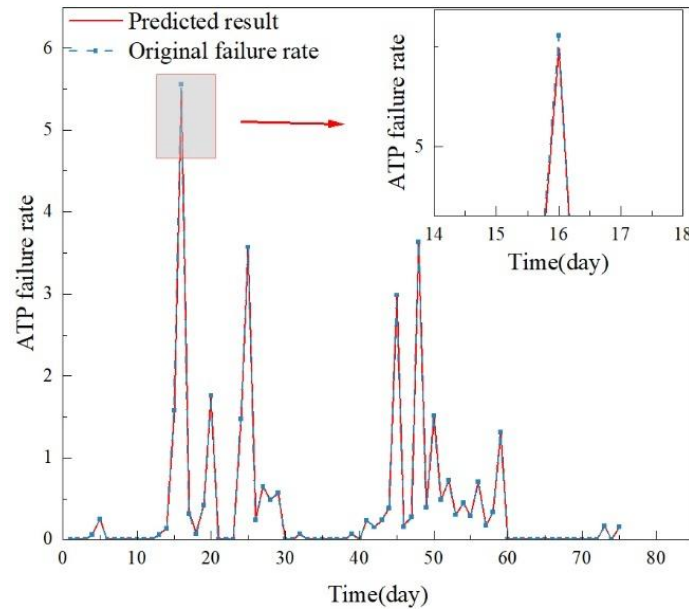


Fig. 8. Failure rate prediction results.

As can be seen from the table, the 10-minute interval dataset is able to demonstrate a more stable prediction performance with the smallest overall error when the zero value is included, and although the MAPE is slightly increased by the zero value, it has the smallest absolute error, suggesting that the prediction is more superior at this time interval. The zero value represents the health state of the system, which can help the model better capture the dynamics of the system during stable operation and fault occurrence, and avoid overlooking the fault-free periods. By retaining the zero value, the failure prediction accuracy of the model can be improved.

The final prediction is shown in Fig. 8.

In order to further enhance the interpretability of the prediction results and the credibility of the model, this paper evaluates the prediction stability of the WLSSVM model by means of the confidence interval estimation method based on residual analysis, and the confidence band prediction curves of the prediction results are shown in Fig. 9. The prediction residuals of the training set containing zero-value data at 10-minute intervals are analysed, and based on their distributional characteristics, the normal distribution is used to fit the error, and the fluctuation range of the error is estimated by the standard deviation. Finally, based on the nature of normal distribution, this paper constructs the corresponding 95% confidence interval for each prediction point.

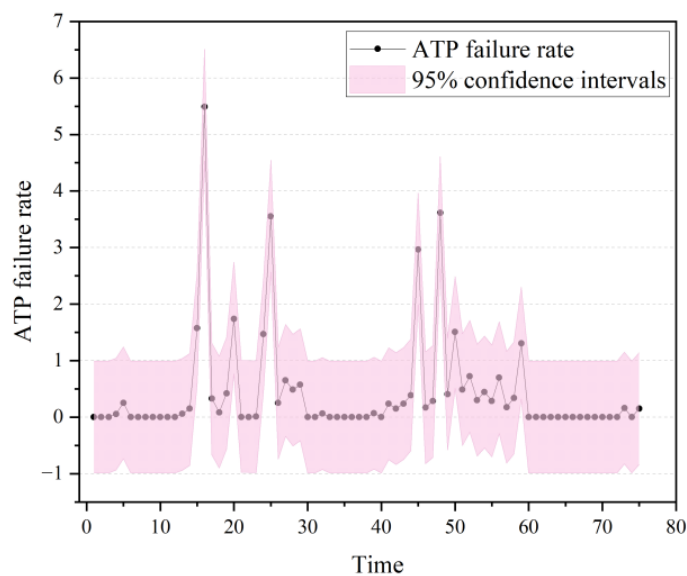


Fig. 9. Confidence band prediction curves for predicted results.

4.5. Ablation experiment

To further evaluate the role of the Grey Wolf algorithm in TVFEMD parameter selection, this paper predicts the candidate solutions generated during the real optimisation process as a comparison set of fixed parameters input into TVFEMD. Fig. 10 shows the MAE, RMSE and MAPE for each parameter combination $[\xi,n]$, where $[0.210,7]$ is the optimal solution. Comparative analysis reveals that while suboptimal parameter sets originate from authentic GWO optimization, their predictive efficacy remains consistently below that of the optimal configuration. This finding underscores the grey wolf algorithm’s significance in identifying high-performance TVFEMD parameters.

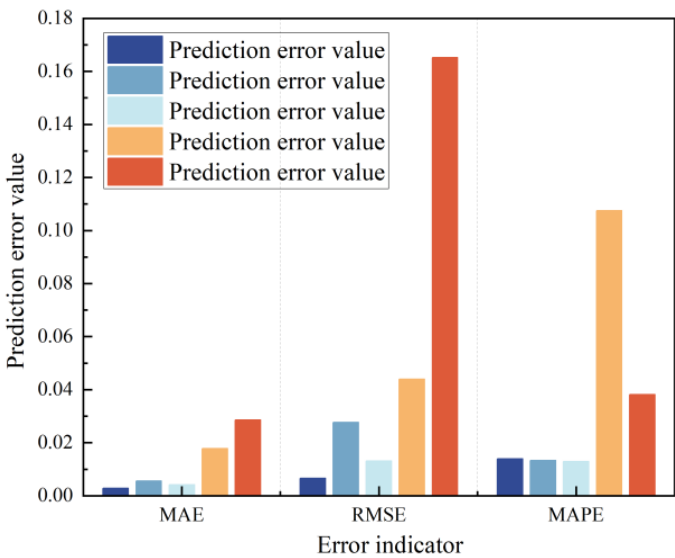


Fig. 10. Error assessment metrics for the prediction results of each parameter combination.

Table 6 presents the runtime comparison between the TVFEMD and GWO-TVFEMD algorithms. As shown in the table, the training phase of the GWO-TVFEMD requires approximately 354.6 seconds, which is longer than the fixed-parameter TVFEMD algorithm. However, during the testing phase, the runtime of the GWO-TVFEMD algorithm increases only slightly compared with TVFEMD, indicating that the additional computational cost is confined to the training stage and that GWO optimization does not significantly affect the efficiency of practical applications.

Table 6. Runtime comparison between TVFEMD and GWO-TVFEMD.

	Training time(s)	Testing time(s)	Overall time(s)
TVFEMD	8.622	0.017	8.642
GWO-TVFEMD	354.554	0.042	355.453

To further validate the predictive ability of the models in order to more comprehensively assess their advantages in failure rate time series prediction, four benchmark models are selected for comparative experiments and four sets of ablation experiments are designed to validate them for different key factors. These experiments, all based on time series with 10-minute intervals and containing zero values, systematically evaluate the effectiveness of the TVFEMD method, the value of parameter optimisation, and the advantages of the WLSSVM in failure rate prediction.

presents a comparative performance assessment between the standalone model and ablation variants when applied to zero-value-containing 10-minute test sequences. As shown in the table, although the WLSSVM model outperforms other single models in terms of predictive performance, its prediction error remains higher than that of the proposed model due to the

absence of TVFEMD-based data processing. Fig. 11 displays boxplots contrasting prediction outcomes from the reference model and our GTW approach, with GTW denoting the integrated GWO-TVFEMD-WLSSVM methodology. The GTW model performs significantly better on the test set compared to other commonly used prediction models. The

TVFEMD-WLSSVM model constructed with fixed parameters from the results of multiple GWO runs.
in the figure uses a non-optimal combination randomly selected

Table 7. Indicators for assessing the results of the models.

	Model	MAE	RMSE	MAPE(%)
10-minute time series with zero values	SVR	0.3562	0.9468	0.6310
	LSTM	0.4443	0.9811	1.1858
	ARIMA	0.4558	0.9674	0.7954
	XGBoost	0.3682	0.9271	0.9087
	WLSSVM	0.2330	0.8093	0.2671
	EMD-WLSSVM	0.0886	0.1006	0.3500
	TVFEMD-WLSSVM	0.0286	0.1652	0.0382
	GWO-TVFEMD-SVR	0.2525	0.3676	1.1974
	GWO-TVFEMD-New signal	0.4165	0.9788	0.6132
	GWO-TVFEMD-WLSSVM	0.0028	0.0066	0.0139

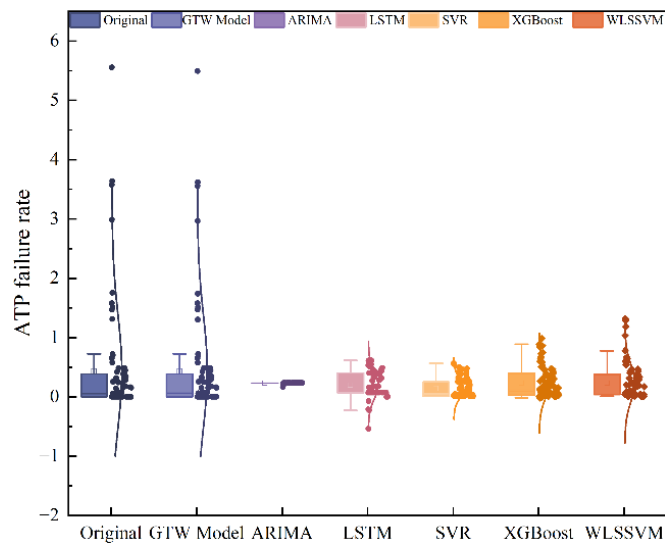


Fig. 11. Comparison of single model prediction results.

Fig. 12 shows the line graphs of the prediction results of each ablation experiment, and according to Fig. 12 and , it can be seen that the present model performs optimally on data.
the 10-minute zero-value-containing dataset, and the MAE is reduced by 0.2497 and 0.4137 compared with that of the GWO-TVFEMD-SVR model and the GWO-TVFEMD-New signal model, respectively, which is closer to the real value; The TVFEMD-WLSSVM model with optimised parameters of the Grey Wolf algorithm is reduced by 96.03%, and the MAPE is reduced by 33.61 percentage points compared to the EMD-WLSSVM model, with greater reliability. The prediction results of the GWO-TVFEMD-New signal model fluctuate less numerically and show a tendency to approximate a horizontal line, which indicates that it fails to adequately capture the information about the differences between the frequency components, and verifies the advantages of modality-independent modelling in dealing with complex non-stationary

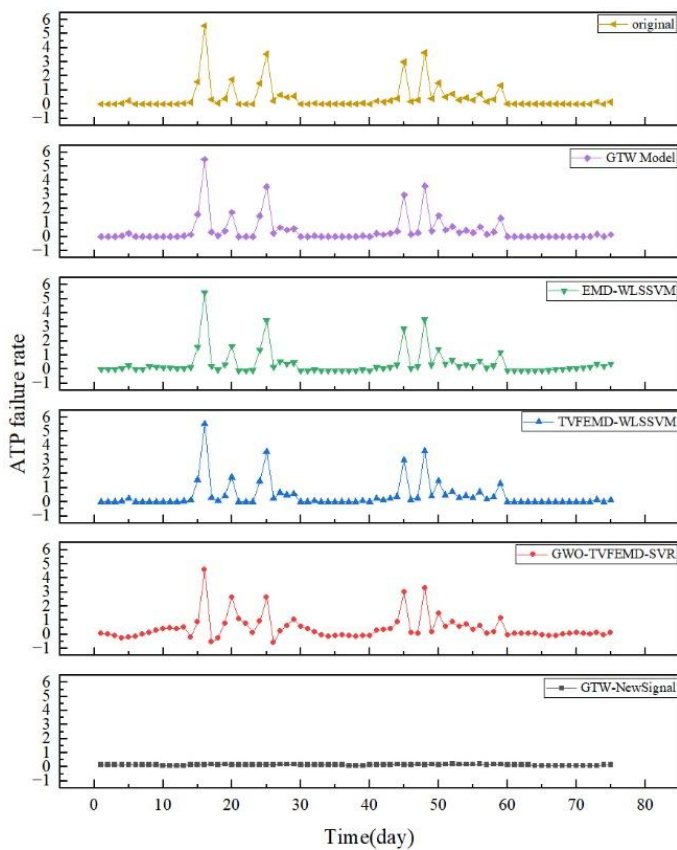


Fig. 12. Plot of prediction results for different combinations of models on a 10-minute test set containing zero values.

5. Conclusion

1) The TVFEMD decomposition method significantly improves the accuracy and stability of predictions. Currently

commonly used single prediction models are often unsatisfactory in dealing with such non-smooth and non-linear time series. TVFEMD decomposition of the failure rate time series transforms the original series into a set of relatively smooth subsequences. Combining the advantages of signal decomposition techniques and nonlinear modelling enables WLSSVM to effectively capture the complex patterns of fault rate changes and reduce the impact of non-smoothness and strong nonlinearity on modelling accuracy.

2) The developed GWO-TVFEMD-WLSSVM hybrid model outperforms traditional prediction methods across multiple evaluation metrics. Empirical results indicate consistently high prediction accuracy throughout various operational periods, demonstrating close alignment between predicted and observed failure rate patterns. This integrated approach proves particularly effective for ATP system failure rate estimation, establishing a reliable methodological framework for similar reliability engineering applications.

3) The present study's validation is confined to ATP system failure rate time series, which inherently limits the model's generalizability across different operational contexts. To advance model performance, future investigations will integrate environmental and operational status variables, thereby building more robust predictive frameworks with superior accuracy and adaptation capacity.

References

1. Shi Z, Li Y, Guo J, et al. Testing Modeling Method for Engineering Testing of High-Speed Railway Signaling System. *Journal of Southwest Jiaotong University*, 2024; 59(5): 1023-1033, <https://doi.org/10.3969/j.issn.0258-2724.20220674>.
2. Shangguan W, Hu F, Yuan M, et al. Reliability Analysis Method for On-board Equipment of Train Control System Based on Resilience Effect. *Journal of the China Railway Society*, 2018; 40(6): 75-82, <https://doi.org/10.3969/j.issn.1001-8360.2018.06.010>.
3. Zheng S, Cheng J, Sun D. ATP On-board Equipment Fault Classification and Management Method Research. *Railway Signalling & Communication*, 2013; 49(7): 1-4, <https://doi.org/10.13879/j.issn1000-7458.2013.07.009>.
4. Deng L, Liu J, Liu M, Cao Y, Wen C, Deng X. Remaining Useful Life Prediction Based on Cross-Temporal Dynamic Graph Convolutional Network. *Eksplotacja i Niezawodność - Maintenance and Reliability*, 2025, <https://doi.org/10.17531/ein/203249>.
5. Kang R. Intelligent Operations for ATP Systems of High-speed Railway Based on Machine Learning [D]. Beijing Jiaotong University, 2022, <https://doi.org/10.26944/d.cnki.gbjfu.2022.000050>.
6. Balakrishnan K, Mani I, Sankaran D. Predicting the overall equipment efficiency of core drill rigs in mining using ANN and improving it using MCDM. *Eksplotacja i Niezawodność - Maintenance and Reliability*, 2023; 25(3), <https://doi.org/10.17531/ein/169581>.
7. Lv Q, He G, Cheng J. Research on Operation and Maintenance Management of ATP On-board Equipment Based on Information Technology. *Railway Signalling & Communication*, 2014; 50(1): 4-7, <https://doi.org/10.13879/j.issn1000-7458.2014.01.002>.
8. Yuan Y, Zhang Y, Ding H. Research on Key Technology of Industrial Artificial Intelligence and Its Application in Predictive Maintenance. *Acta Automatica Sinica*, 2020; 46(10): 2013-2030, <https://doi.org/10.16383/j.aas.c200333>.
9. Yang B, Liang X, Xu S, et al. A time-series based deep survival analysis model for failure prediction in urban infrastructure systems.

- Engineering Applications of Artificial Intelligence, 2024; 136(Part A): 15, <https://doi.org/10.1016/j.engappai.2024.108876>.
10. Li W, Ma Y, Nie P, et al. Grey GM(1,N) and PLS Combined Model for Failure Rate Prediction of Airborne Equipment. *Fire Control & Command Control*, 2022; 47(8): 98-102, <https://doi.org/10.3969/j.issn.1002-0640.2022.08.016>.
 11. Xu D, Zhou C, Guan C, et al. Research on the Equipment Failure Rate Prediction Method Based on ARMA-BP Combined Model. *Fire Control & Command Control*, 2021; 46(1): 83-87, <https://doi.org/10.3969/j.issn.1002-0640.2021.01.015>.
 12. Kang J, Lu Y, Zhao B, et al. Remaining useful life prediction of cylinder liner based on nonlinear degradation model. *Eksplotacja i Niezawodność*, 2022; 24(1), <https://doi.org/10.17531/ein.2022.1.8>.
 13. Li R, Kang R. Research on failure rate forecasting method based on ARMA model. *Systems Engineering and Electronics*, 2008; (8): 1588-1591, <https://doi.org/10.3321/j.issn:1001-506X.2008.08.047>.
 14. Sun Y, Li C, Wei Y, et al. Fault rate prediction method of circuit breaker based on hybrid learning method. *Mechanical & Electrical Technique of Hydropower Station*, 2023; 46(12): 117-120, <https://doi.org/10.13599/j.cnki.11-5130.2023.12.038>.
 15. Wang H, Zhao X, Guo Q, et al. A novel hybrid model by integrating TCN with TVFEMD and permutation entropy for monthly non-stationary runoff prediction. *Scientific Reports*, 2024; 14(1), <https://doi.org/10.1038/s41598-024-81574-w>.
 16. Jinyong G, Sheng L, Junshan T, et al. Traffic Flow Prediction for Highway Based on Multi-Task Spatiotemporal Graph Network. *Transportation Safety and Environment*, 2025, <https://doi.org/10.1093/tse/tdaf003>.
 17. Wang C, Lin H, Yang M, et al. Ultra-short-term wind farm cluster interval power prediction based on cluster division and MQ-WaveNet-MSA. *Electric Power Systems Research*, 2025; 244, <https://doi.org/10.1016/j.epr.2025.111557>.
 18. Zhang J, Sun T, Guo X, et al. Short-term photovoltaic power prediction with CPO-BILSTM based on quadratic decomposition. *Electric Power Systems Research*, 2025; 243, <https://doi.org/10.1016/j.epr.2025.111511>.
 19. Huang N E, Shen Z, Long S R, et al. The empirical mode decomposition method and the Hilbert spectrum for non-stationary time series analysis. *Proceedings of the Royal Society A*, 1998, <https://doi.org/10.1098/rspa.1998.0193>.
 20. Poongadan S, Lineesh M C. A hybrid deep learning network for non-linear time series prediction. *Journal of Mathematics and Computer Science*, 2022; 12: Article ID 158, <https://doi.org/10.28919/jmcs/7398>.
 21. Flandrin P, Rilling G, Goncalves P. Empirical mode decomposition as a filter bank. *IEEE Signal Processing Letters*, 2004; 11(2): 112-114, <https://doi.org/10.1109/LSP.2003.821662>.
 22. Xu T, Lu C, Wang H, et al. Fault rate prediction method based on relevant vector EMD and GMDH reconstruction. *Journal of Vibration, Measurement & Diagnosis*, 2018; 38(6): 1275-1285+1300, <https://doi.org/10.16450/j.cnki.issn.1004-6801.2018.06.030>.
 23. Hu Y, Li F, Yang S. Failure rate prediction and early warning monitoring based on flight test data. *Journal of Ordnance Equipment Engineering*, 2020; 41(7): 224-227, <https://doi.org/10.11809/bqzbgcxb2020.07.044>.
 24. Wu Z, Huang N E. Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Advances in Adaptive Data Analysis*, 2011; 1(1), <https://doi.org/10.1142/S1793536909000047>.
 25. Zhang J, Zhang J, Zhong M, et al. Weak fault diagnosis of wind turbine bearings based on PSO-VMD-MCKD method. *Journal of Vibration, Measurement & Diagnosis*, 2020; 40(2): 287-296+418, <https://doi.org/10.16450/j.cnki.issn.1004-6801.2020.02.011>.
 26. Li N, Yang D, Wang Z, et al. Chatter online monitoring method for roll grinder using TVFEMD and instantaneous energy ratio. *Journal of Electronic Measurement and Instrumentation*, 2024; 38(1): 228-236, <https://doi.org/10.13382/j.jemi.B2306864>.
 27. Li S, Xin J, Jiang L, et al. A denoising method for bridge monitoring data based on TVFEMD-IMF energy entropy increment. *Journal of Vibration, Measurement & Diagnosis*, 2024; 44(1): 178-185+206, <https://doi.org/10.16450/j.cnki.issn.1004-6801.2024.01.027>.
 28. Wang D, Li X, Huang Y, et al. Short-term wind speed prediction based on time-varying filtered empirical modal decomposition with spatio-temporal correlation error correction. *Journal of North China Electric Power University (Natural Science Edition)*, 2025, <https://doi.org/10.1109/CCDC62350.2024.10587719>.
 29. Zhang N, Zhu Y, Sun N, et al. Vibration trend prediction of hydropower units based on OVMD-TVFEMD secondary decomposition and HPO-ELM. *Water Resources and Power*, 2023; 41(10): 204-207+199, <https://doi.org/10.20040/j.cnki.1000-7709.2023.20222110>.
 30. Heng A, Zhi B, Wei C. A time varying filter approach for empirical mode decomposition. *Signal Processing*, 2017, <https://doi.org/10.1016/j.sigpro.2017.03.019>.
 31. Ma X, Lu X, Zhang Z, et al. A non-contact HRV estimation method based on TVF-EMD. *Journal of Data Acquisition and Processing*,

- 2024; 39(4): 1009-1019, <https://doi.org/10.16337/j.1004-9037.2024.04.019>.
32. Zang X, Zhang T, Shao X, et al. A transformer internal mechanical fault diagnosis method based on TVFEMD and SSA-LSSVM. *Electric Machines & Control Application*, 2023; 50(9): 49-56, <https://doi.org/10.12177/emca.2023.106>.
 33. Deng Z, Xie J, Cui D. Daily sediment concentration prediction based on TVFEMDII-Ten Fish Swarm Algorithm-DHKELM model. *China Rural Water and Hydropower*, 2025, <https://doi.org/10.12396/znsd.241017>.
 34. Tan B, Wang D, Shi J, et al. Temperature field prediction of steel-concrete composite decks using TVFEMD-stacking ensemble algorithm. *Journal of Zhejiang University-Science A (Applied Physics & Engineering)*, 2024; 25(9): 732-748, <https://doi.org/10.1631/jzus.A2300441>.
 35. Yin Y. Research on short-term load forecasting of microgrid based on TVFEMD-WOA-LSTM-ARMAX [D]. Wuhan University of Science and Technology, 2024, <https://doi.org/10.1007/s42341-023-00506-z>
 36. Wang X, Shang H, Li N. Research on advanced manufacturing process monitoring and fault prediction method based on machine learning. *Automation and Machine Learning*, 2023; 4: 87-92, <https://doi.org/10.23977/autml.2023.040311>.
 37. Liang J, Yin L, Xin Y, et al. Short-term photovoltaic power prediction based on CEEMDAN-PE and BiLSTM neural network. *Electric Power Systems Research*, 2025; 246: 111706, <https://doi.org/10.1016/j.epsr.2025.111706>.
 38. Cortes C, Vapnik V. Support-vector networks. *Machine Learning*, 1995; 20(3): 273-297, <https://doi.org/10.1023/A:1022627411411>.
 39. Suykens J A K, Vandewalle J. Least squares support vector machine classifiers. *Neural Processing Letters*, 1999; 9(3): 293-300, <https://doi.org/10.1023/A:1018628609742>.
 40. Xiong R, Xiao C, Li Y. Electric drive system of shovel forecasting based on wavelet and least square support vector machines. *Journal of China Jiliang University*, 2008; (1): 51-55.
 41. Cui J, Li Q. Application of improved WLSSVM model in the prediction of gasoline dry point at the top of atmospheric tower. *Journal of Liaoning Petrochemical University*, 2023; 43(1): 67-72, <https://doi.org/10.12422/j.issn.1672-6952.2023.01.012>.
 42. Zhang X, Zhang Y. Prediction of external corrosion rate of buried pipelines based on KPCA-ALO-WLSSVM. *Journal of Safety and Environment*, 2022; 22(4): 1804-1812, <https://doi.org/10.13637/j.issn.1009-6094.2021.0275>.
 43. Chang X, Yang S, Yang D. Temperature modeling of PVC stripping process using weighted least square support vector machine. *Plastics Science and Technology*, 2020; 48(4): 74-77, <https://doi.org/10.15925/j.cnki.issn1005-3360.2020.04.019>.
 44. Suykens J A K, Brabanter J D, Lukas L, et al. Weighted least squares support vector machines: robustness and sparse approximation. *Neurocomputing*, 2002; 48(1-4): 85-105, [https://doi.org/10.1016/S0925-2312\(01\)00644-0](https://doi.org/10.1016/S0925-2312(01)00644-0).
 45. Wang R, Wang Y, Lu J. Prediction of photovoltaic power generation based on ICEEMDAN-DTW and ISMA-WLSSVM. *Journal of Engineering for Thermal Energy and Power*, 2023; 38(9): 131-140, <https://doi.org/10.16146/j.cnki.rndlgc.2023.09.016>.
 46. Chen K, Ding M, Liu J, et al. Short term wind power prediction based on WLSSVM optimized by improved grey wolf optimization algorithm. *Inner Mongolia Electric Power*, 2024; 42(2): 1-7, <https://doi.org/10.19929/j.cnki.nmgdljs.2024.0017>.
 47. Lei X. Multiverse optimization algorithm optimizes network traffic prediction of weighted least squares support vector machine. *Automation & Instrumentation*, 2023; (5): 51-55, <https://doi.org/10.14016/j.cnki.1001-9227.2023.05.051>.
 48. Badar A Q H. Grey wolf optimizer. *Advances in Engineering Software*, 2014; 69: 46-61, <https://doi.org/10.1016/j.advengsoft.2013.12.007>.
 49. Cao W. Fault diagnosis of tooth surface spalling of planetary gearbox based on GWO-TVF-EMD method [D]. Xi'an Technological University, 2021, <https://doi.org/10.13224/j.cnki.jasp.20210226>.
 50. Zhao S, Wei Z, Guo S, et al. Method for removing multi-frequency vibration information from 3D laser-scanned pavement point clouds. *Laser & Optoelectronics Progress*, 2023; (20): 060, <https://doi.org/10.3788/LOP223266>.
 51. Tang G, Wang X. Parameter optimized variational mode decomposition method with application to incipient fault diagnosis of rolling bearing. *Journal of Xi'an Jiaotong University*, 2015; 49(5): 73-81, <https://doi.org/10.7652/xjtub201505012>.
 52. Unser M. Splines: a perfect fit for signal and image processing. *IEEE Signal Processing Magazine*, 1999; 16(6): 22-38, <https://doi.org/10.1109/79.799930>.
 53. Kang R, Wang J, Cheng J, et al. Prediction of failure rate of ATP on-board equipment based on chaos theory. *Journal of the China Railway Society*, 2016; 38(9): 59-65, <https://doi.org/10.3969/j.issn.1001-8360.2016.09.009>.
 54. Fernandez Martinez R, Alberdi R, Fernandez E, et al. Improvement of transmission line ampacity utilization via machine learning-based dynamic line rating prediction. *Electric Power Systems Research*, 2024; 236: 110931, <https://doi.org/10.1016/j.epsr.2024.110931>.

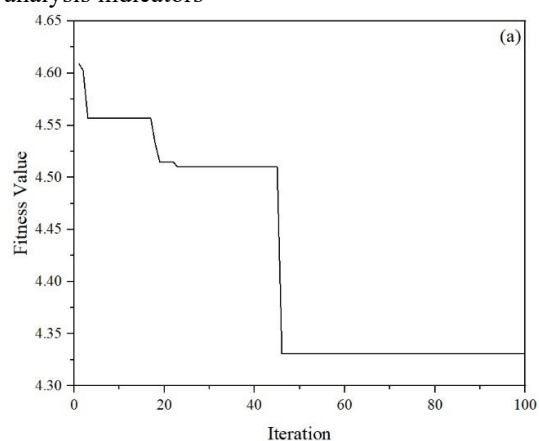
Appendix A

Thesaurus of fault cases

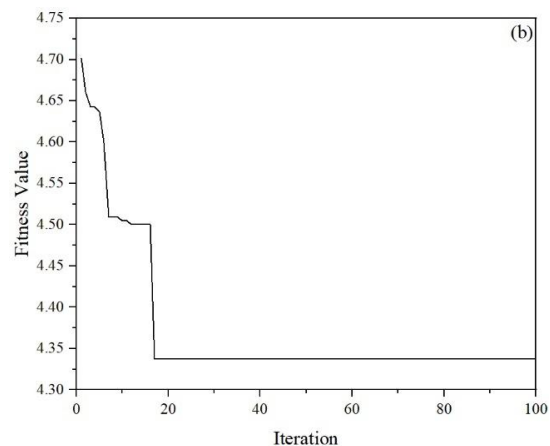
Ordinal number	Fault Alarm Statements	Clarification	Keywords
1	VDX EBR1 port switched to invalid	Safety input/output unit (VDX) EBR1 port switched to ineffective	invalid
2	VDX EBR1 port switched to valid	VDX EBR1 port switched to effective	invalid
3	VDX RBR port switched to invalid	VDX RBR port switched to ineffective	invalid
4	VDX EBF port switched to invalid	VDX EBF port switched to ineffective	invalid
5	VDX bypass port switched to invalid	VDX bypass port switched to ineffective	invalid
6	VDX EBR1 port switched to valid	VDX EBR1 port switched to active	valid
7	VDX EBR2 port switched to valid	VDX EBR2 port switched to active	valid
8	VDX RBR port switched to valid	VDX RBR port switched to active	valid
9	VDX EBF port switched to valid	VDX EBF port switched to active	valid
10	VDX bypass port switched to valid	VDX bypass port switches to active	valid
11	BI-H VDX1:IN1 I/O failed	It is recommended to refer to the corresponding fault case processing	failed
12	BI-H VDX2:IN1 I/O failed	It is recommended to refer to the corresponding fault case processing	failed
13	BI-H VDX1:IN2 I/O failed	It is recommended to refer to the corresponding fault case processing	failed
14	BI-H VDX2:IN2 I/O failed	It is recommended to refer to the corresponding fault case processing	failed
15	BI-H VDX1:IN3 I/O failed	It is recommended to refer to the corresponding fault case processing	failed
16	BI-H VDX2:IN3 I/O failed	It is recommended to refer to the corresponding fault case processing	failed
17	BI-H RBR state wrong	BI-H RBR status error	wrong
18	BI-H EBR1 state wrong	BI-H EBR1 status error	wrong
19	BI-H EBR2 state wrong	BI-H EBR2 status error	wrong
20	BI-H BPR state wrong	BI-H BPR status error	wrong
21	BI-H EBFB state wrong	BI-H EBFB status error	wrong
22	BI-H RBR feedback timeout	BI-H RBR feedback timeout.RB relay fault.	timeout
23	BI-H EBR1 feedback timeout	BI-H EBR1 feedback timeout.EBR1 relay failure.	timeout
24	BI-H EBR2 feedback timeout	BI-H EBR2 feedback timeout.EBR2 relay failure.	timeout
25	BI-H BPR feedback timeout	BI-H BPR feedback timeout.BP relay failure.	timeout
26	BI-H EBFR feedback timeout	BI-H EBFB feedback timeout.EBFB relay failure.	timeout
27	[STM] STM in failure state	STM failure	failure
28	[STM] STM in DA state failed	STM disabled in DA state	failed
29	direction control failure	Direction control malfunction	failure
30	DirCtrlPos err	Abnormal status of the driver's seat	err
31	Cab status fail	Abnormal vehicle status	fail
32	[CAB] REPORT_CAB_NOT_STANDSTILL_DETECTED	Full shutdown not detected	NOT
33	Level changed to L2, orderby=0	Level has been changed to L2, Sort = 0	orderby=0
34	Braketest failed in step 11	Brake Test Step 11 Failure	failed
35	Wrong feedback. Timeout expires 66523	Feedback timeout	wrong
36	Time 64623 BI-H EBFR state wrong	BI-H EBFR status error	wrong
37	Timeout Lifesign *	life-signal abnormality	Timeout
38	Brake Test step 0 failed	Unable to initiate brake test	failed
39	C018	An exception was detected in the test performed by the VDX unit on the FS output port every 32ms.	C018
40	C019	Abnormalities detected in tests performed by the VDX unit on the FS output port at 5 second intervals	C019
41	C002	VDX Unit HR2 Port Test Failure	C002
42	0028	Inconsistent VDX unit A/B code comparisons	0028
43	C018E000C016E03CA001	Abnormalities detected in tests performed by the	C018

Appendix B

Convergence analysis indicators

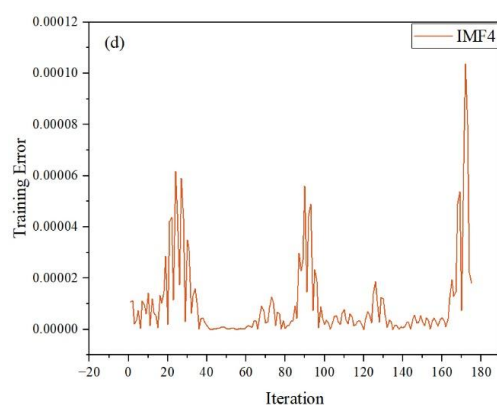
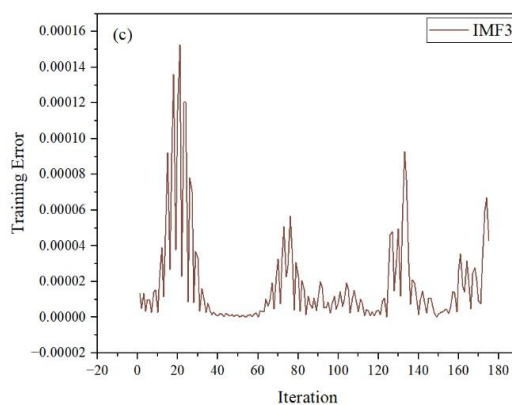
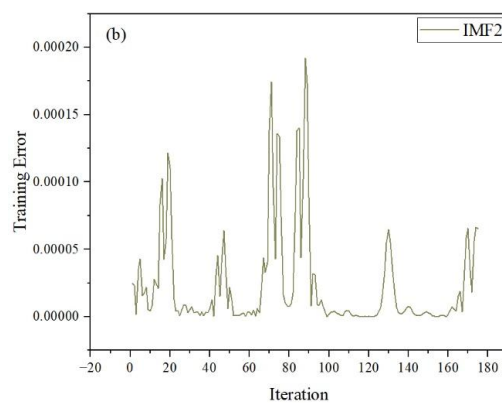
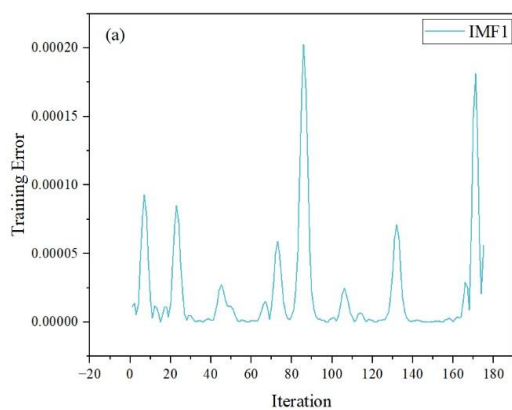


5-minute interval data set



10-minute interval data set

Figure 1. Algorithmic convergence curves for GWO.



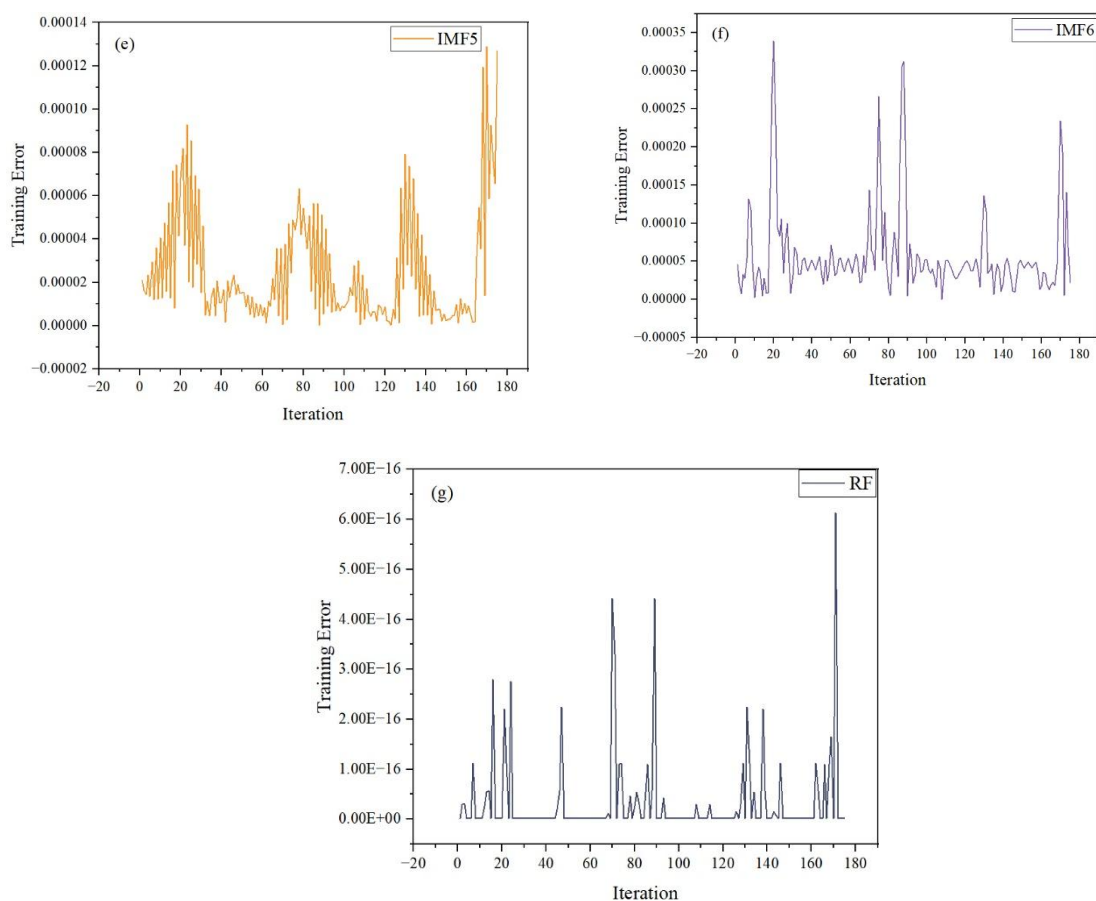
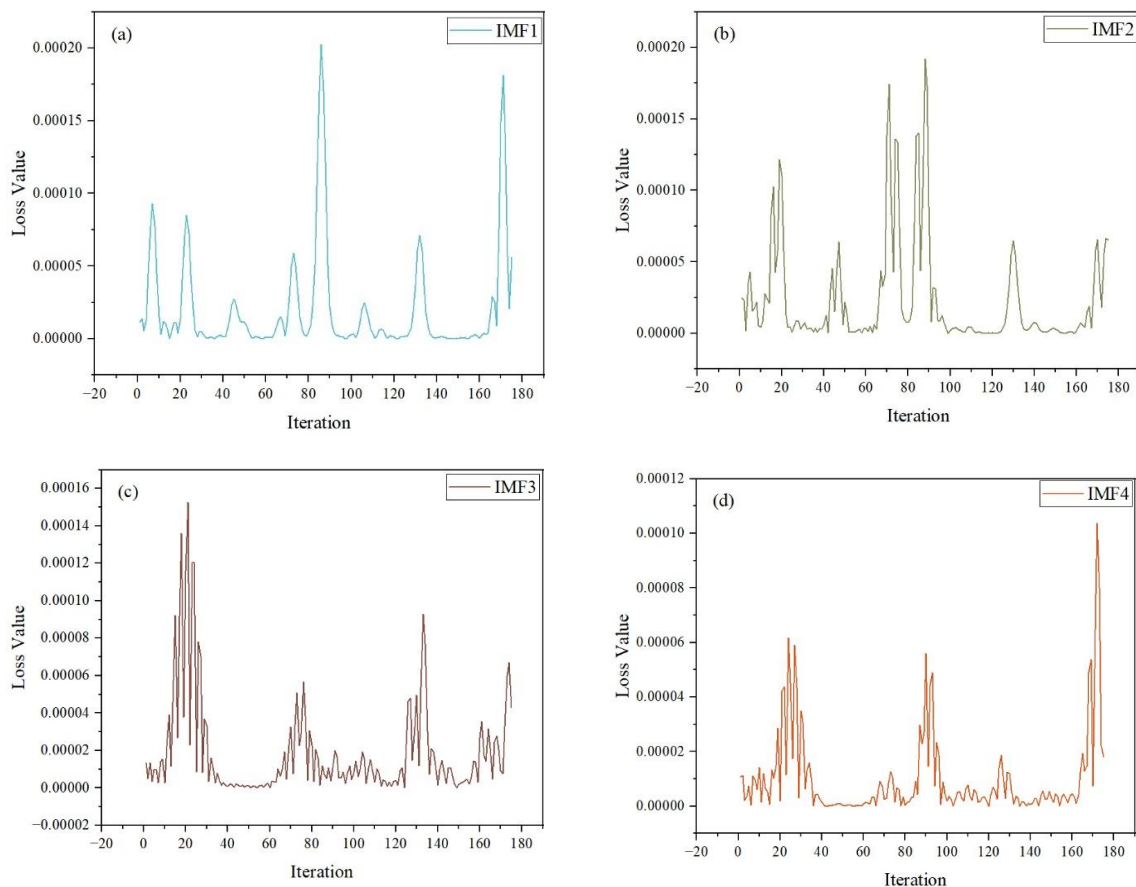


Figure 2. Training loss convergence curves for the WLSSVM on the 5-minute dataset.



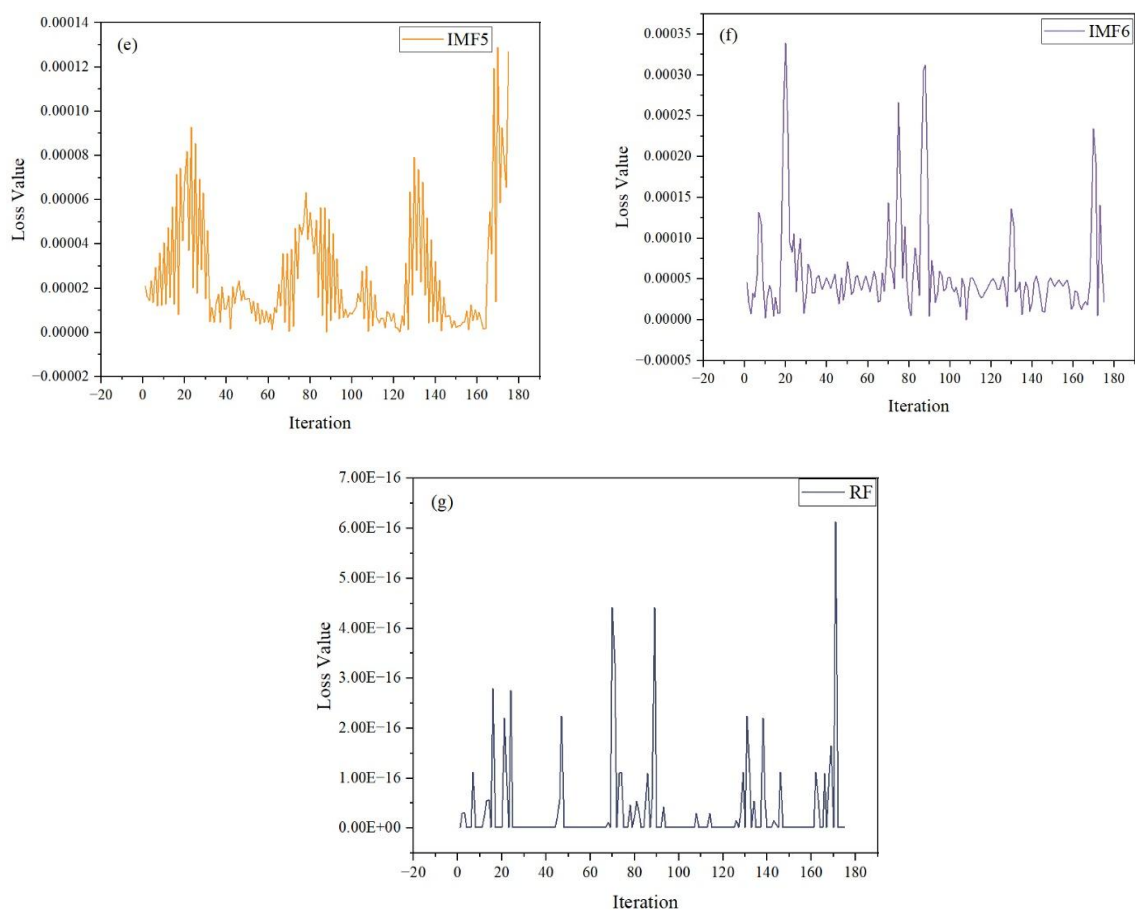
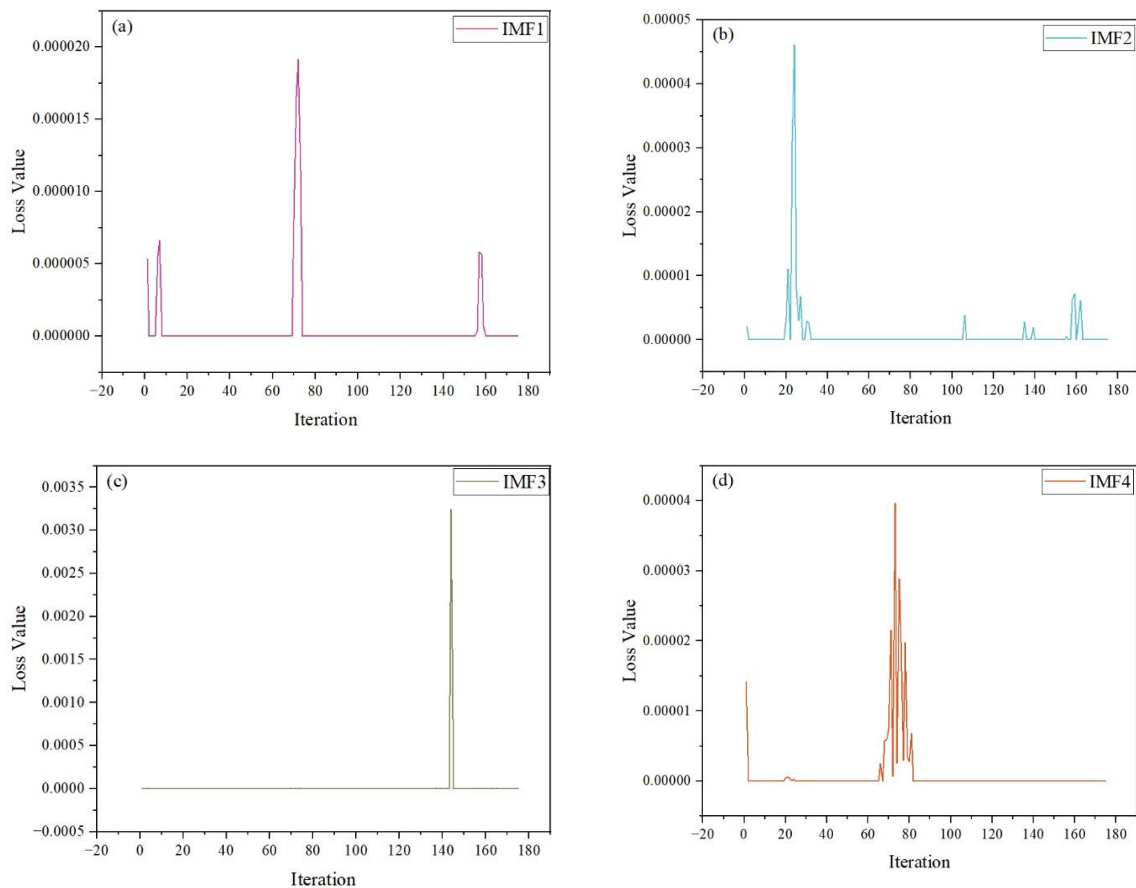


Figure 3. Trend plot of training error for WLSSVM on 5-minute dataset.



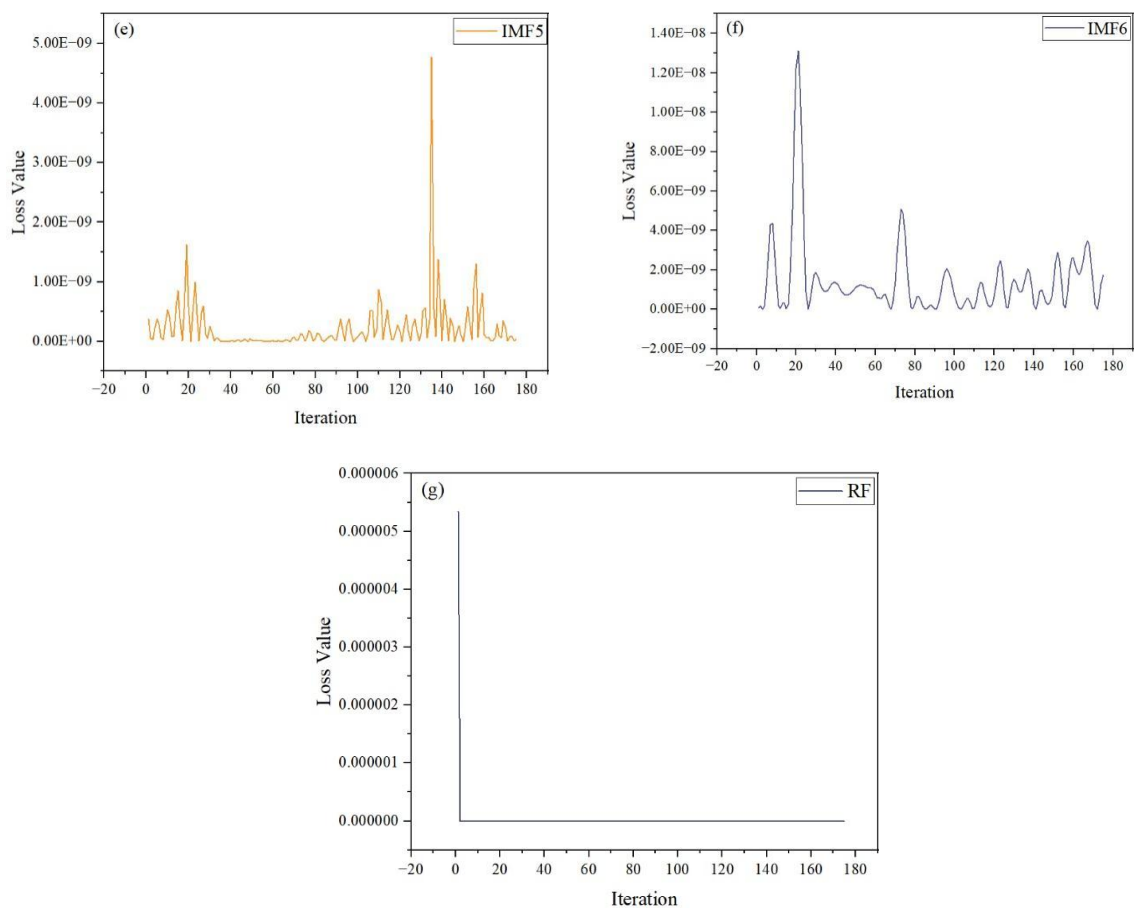
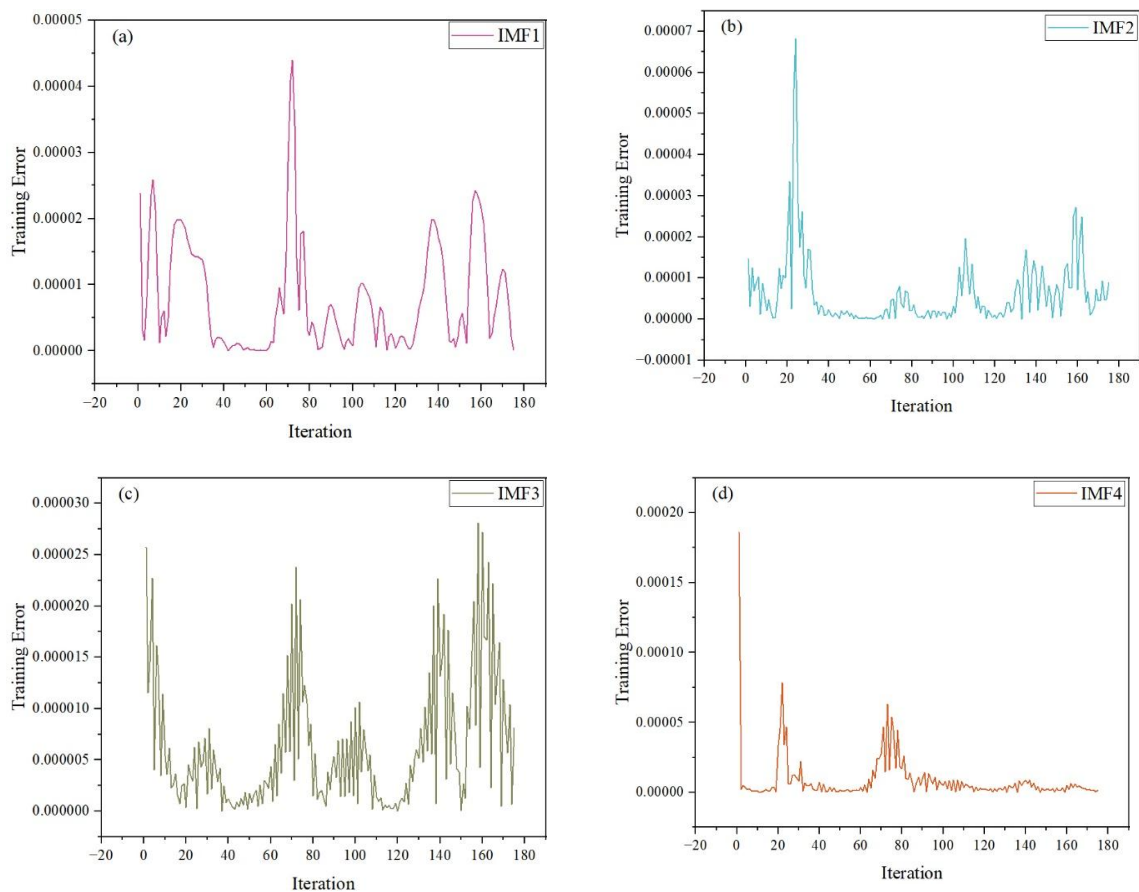


Figure 4. Training loss convergence curves for the WLSSVM on the 10-minute dataset.



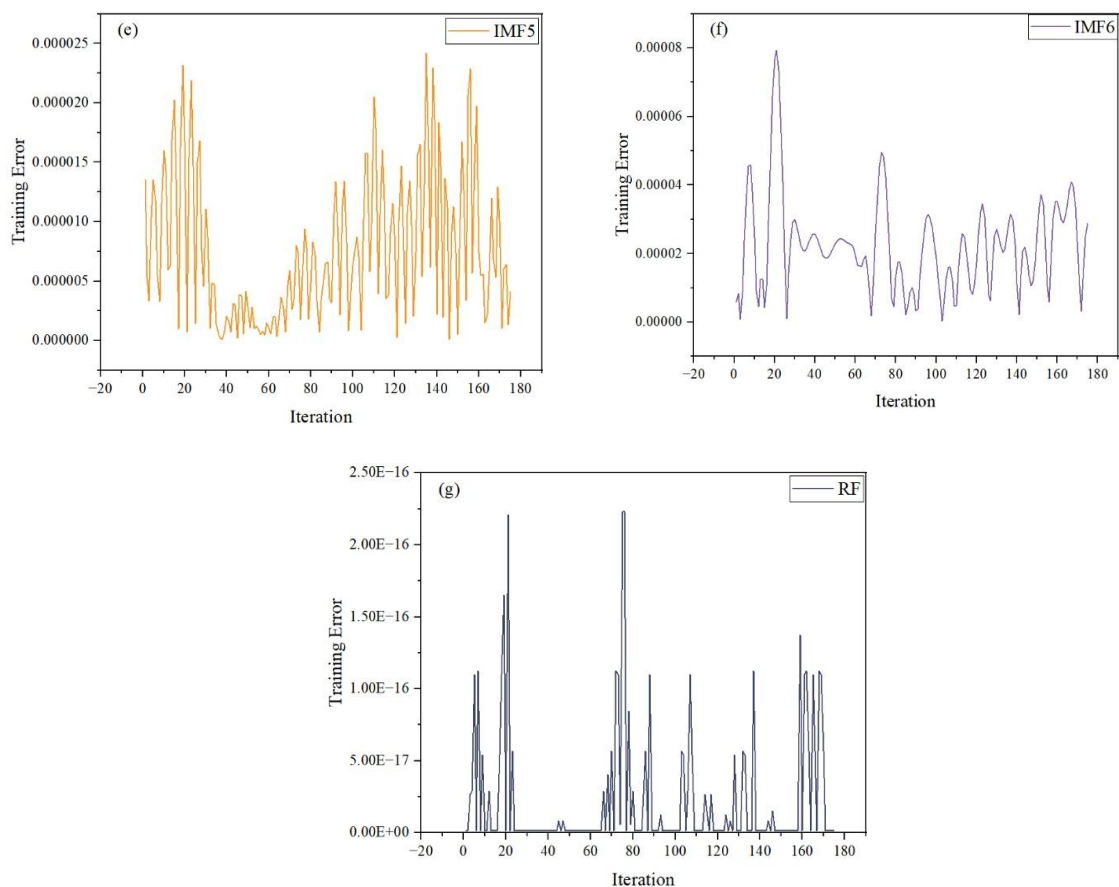


Figure 5. Trend plot of training error for WLSSVM on 10-minute dataset.