International Journal of Engineering Inventions
e-ISSN: 2278-7461, p-ISSN: 2319-6491
Volume 15, Issue 3 [March 2026] PP: 134-138

An Improved Grey Wolf Optimizer Optimized GRU Network for Gear Remaining Useful Life Prediction

Bo Sun*, Xiangqin Wu, Lixin Zhang, Chengyang Deng
School of Mechanical and Vehicle Engineering, Changchun University, Changchun, 130022, Changchun Province, People's Republic of China

Abstract

Accurate remaining useful life (RUL) prediction of gears is critical to ensure stable operation of mechanical equipment and reduce economic losses caused by unplanned shutdowns. The gated recurrent unit (GRU) is widely used in gear RUL prediction for its excellent time series processing performance, while its prediction accuracy heavily depends on hyperparameter selection. The traditional grey wolf optimizer (GWO), commonly used for GRU hyperparameter optimization, suffers from slow convergence and local optimum stagnation. To address these issues, this paper proposes an improved grey wolf optimizer (SCGWO) for GRU hyperparameter optimization, with three core improvements: Sine chaotic mapping for population initialization, a sinusoidal attenuation nonlinear convergence factor, and a cuckoo search algorithm (CSA) based position update strategy. Experimental results on the gear life cycle dataset show that the proposed SCGWO converges within 3 iterations, significantly outperforming the 8 iterations required by traditional GWO. Meanwhile, the SCGWO-GRU model reduces MAE and RMSE by 2. 62% and 4. 02% respectively compared with the traditional GWO-GRU model, achieving a determination coefficient R^2 *of 0. 9834, with both faster convergence and higher prediction accuracy.*

Key words*: Gear RUL Prediction; Grey Wolf Optimizer; Gated Recurrent Unit; Hyperparameter Optimization; Cuckoo Search Algorithm*

Date of Submission: 15-03-2026 Date of acceptance: 31-03-2026

I. Introduction

With the development of intelligent manufacturing, fault prediction and health management (PHM) has become a core technology for safe operation of mechanical equipment. As a key task of PHM, remaining useful life (RUL) prediction provides scientific guidance for equipment maintenance by quantifying the remaining normal working time of components before failure[1]. Gears, as core transmission components, are prone to fatigue and fracture under heavy load and high-speed conditions, and gear failures account for about 10% of total mechanical equipment failures, leading to huge economic losses and safety risks[2]. Therefore, accurate gear RUL prediction has become a research hotspot in the mechanical PHM field.

Data-driven RUL prediction methods have been widely applied for avoiding complex failure mechanism modeling, relying only on equipment operation data to achieve high-precision prediction[3]. Among them, the gated recurrent unit (GRU), a simplified variant of long short-term memory (LSTM), reduces parameters by nearly 1/3 compared with LSTM while maintaining excellent long-term dependency capture ability for time series, making it suitable for gear RUL prediction[4]. However, GRU's prediction performance is highly dependent on key hyperparameters (batch size, learning rate, hidden layer number and unit number), and traditional manual tuning is inefficient and hard to reach the optimal configuration.

The grey wolf optimizer (GWO), a swarm intelligence algorithm with simple structure and good stability, is commonly used for hyperparameter optimization[5]. Nevertheless, traditional GWO has obvious defects: uneven population distribution from random initialization, inflexible balance between global exploration and local exploitation from linear convergence factor, and easy local optimum stagnation from single position update strategy[6]. These defects limit its optimization efficiency for GRU hyperparameters and further affect gear RUL prediction performance.

To solve the above problems, this paper proposes an improved grey wolf optimizer (SCGWO) with three targeted improvement strategies, and constructs a SCGWO-GRU model for gear RUL prediction. The convergence performance and prediction accuracy of the proposed model are verified through public gear life cycle dataset.

II. Theoretical Background

2. 1 Gated Recurrent Unit (GRU)

GRU solves the gradient vanishing and explosion problems of traditional recurrent neural network (RNN) through two gate structures: update gate Z_t and reset gate R_t. It takes the input X_t at time t and the previous hidden state H_{t-1} as inputs, and outputs the current hidden state H_t. The core calculation formulas are as follows:
Reset gate and update gate calculation:

$$R_t = \sigma(X_t W_{xr} + H_{t-1} W_{hr} + b_r) \quad (1)$$
$$Z_t = \sigma(X_t W_{xz} + H_{t-1} W_{hz} + b_z) \quad (2)$$

Candidate hidden state calculation:

$$\widetilde{H}_t = \tanh(X_t W_{xh} + (R_t \odot H_{t-1}) W_{hh} + b_h) \quad (3)$$

Hidden state update:

$$H_t = (1 - Z_t) \odot H_{t-1} + Z_t \odot \widetilde{H}_t \quad (4)$$

Where σ is the sigmoid activation function, $\odot$ represents Hadamard product, W and b are weight parameters and bias terms respectively. The reset gate controls the retention of historical state information, and the update gate balances the weight of historical and new information, enabling GRU to effectively capture the long-term dependency of gear degradation time series.

2. 2 Traditional Grey Wolf Optimizer (GWO)

GWO simulates the social hierarchy and hunting behavior of grey wolf populations, dividing individuals into α (optimal solution), β (suboptimal), δ (third optimal) and ω wolves. The hunting process includes encircling and attacking prey, with ω wolves' position updated under the guidance of α, β and δ wolves.
The core mathematical model of prey encircling is:

$$D = \left|C \cdot X_p(t) - X(t)\right|, X(t+1) = X_p(t) - A \cdot D \quad (5)$$
$$A = 2a \cdot r_1 - a, C = 2 \cdot r_2 \quad (6)$$
$$a = 2 - 2 \cdot \frac{t}{T_{max}} \quad (7)$$

Where t is the current iteration, T_{max} is the maximum iteration, X_p and X are the position of prey and grey wolf, A and C are coefficient vectors, $r_1, r_2 \in [0, 1]$, and a is the linear convergence factor decreasing from 2 to 0.

The position update guided by the top three optimal wolves is:

$$\begin{cases} D_\alpha = |C_1 \cdot X_\alpha - X(t)|, D_\beta = \left|C_2 \cdot X_\beta - X(t)\right|, D_\delta = |C_3 \cdot X_\delta - X(t)| \\ X_1 = X_\alpha - A_1 \cdot D_\alpha, X_2 = X_\beta - A_2 \cdot D_\beta, X_3 = X_\delta - A_3 \cdot D_\delta \\ X(t+1) = \frac{X_1 + X_2 + X_3}{3} \end{cases} \quad (8)$$

when $|A| < 1$, it carries out local exploitation. The linear convergence factor, random initialization and single position update strategy of traditional GWO lead to slow convergence and easy local optimum, which need to be improved.

III. Improved SCGWO Algorithm and SCGWO-GRU Prediction Model

3. 1 Improvement Strategies for GWO

Three targeted improvement strategies are proposed to solve the defects of traditional GWO, constructing the improved SCGWO algorithm.

3. 1. 1 Sine Chaotic Mapping for Population Initialization

Random initialization leads to uneven population distribution and low search efficiency. Sine chaotic mapping, with better ergodicity and uniform distribution than Logistic mapping, is used to initialize the grey wolf population, with the formula:

$$x_{k+1} = 2\sin(\pi x_k)\,, x_k \in (0, 1) \quad (9)$$

The chaotic sequence generated by formula (9) is mapped to the hyperparameter search space to obtain the initial population position, which improves population diversity and lays a foundation for accelerating convergence.

3. 1. 2 Nonlinear Convergence Factor Based on Sinusoidal Attenuation

The linear convergence factor cannot flexibly balance global and local search capabilities. A sinusoidal attenuation nonlinear convergence factor is proposed:

$$a = 2 \cdot \cos\left(\frac{\pi}{2} \cdot \frac{t}{T_{max}}\right) \quad (10)$$

In the early iteration, a decreases slowly to maintain a large value and enhance global exploration; in the late iteration, a decreases rapidly to strengthen local exploitation, realizing adaptive balance of search capabilities.

3. 1. 3 CSA-Based Position Update Strategy

Traditional GWO is easy to fall into local optimum due to single position update mechanism. The Levy flight mechanism of cuckoo search algorithm (CSA) is introduced to improve the position update strategy, with the new formula:

$$X(t+1) = X_{best}(t) + \alpha \oplus Levy(\lambda) \quad (11)$$

Where $X_{best}(t)$ is the current optimal position, $\alpha > 0$ is the step control factor, and $Levy(\lambda)$is the random step generated by Mantegna algorithm, which alternates short-distance search and long-distance jump to expand the search range and help the algorithm jump out of local optimum.

3. 2 Implementation Steps of SCGWO Algorithm

1. Initialize algorithm parameters: population size $N = 10$, maximum iteration$T_{max} = 30$, and search range of GRU hyperparameters.
2. Generate initial grey wolf population through Sine chaotic mapping.
3. Calculate individual fitness value, select and record the positions of $\alpha, \beta and \delta$ wolves.
4. Update the nonlinear convergence factor a and coefficient vectors A and C.
5. Update the position of grey wolf individuals through the CSA-based strategy.
6. Recalculate fitness values and update the positions of α, β and δ wolves.
7. Judge whether the termination condition is met. If yes, output the position of α wolf as the optimal hyperparameter solution; otherwise, return to step 4 to continue iteration.

3. 3 Gear RUL Prediction Process of SCGWO-GRU Model

1. **Data preprocessing**: Extract 13 time-domain and 4 frequency-domain features from the original gear vibration signal to construct a 17-dimensional feature dataset; use ISOMAP algorithm to reduce the dimension to 1, normalize the data to [0, 1], and construct RUL labels for samples.
2. **Dataset division**: Process the time series data with sliding window method (SWM), and divide the dataset into training set (first 530 samples, 90%) and test set (last 70 samples, 10%).
3. **Hyperparameter optimization**: Take the MAE of GRU on the training set as the fitness function, and obtain the optimal hyperparameter configuration through SCGWO iterative optimization.

4. **Model training and prediction**: Train the GRU model with the optimal hyperparameters, input the test set to complete gear RUL prediction, and use MAE, RMSE and R^2 to evaluate the model performance.

IV. Experimental Verification and Results Analysis

4. 1 Experimental Setup

The experiment adopts the public gear life cycle dataset[8] as Figure 1 shown. The test gear material is 20CrMnMo, with a speed of 1000r/min, torque of 1300N·m, vibration signal sampling frequency of 50kHz, and a total of 600 samples of 600-minute degradation data. The experimental environment is Intel Core i5-12500H processor, NVIDIA RTX3050 GPU, Python 3. 9 and PyTorch 1. 13 framework.
The hyperparameter search range of GRU is set as: batch size [30, 90], learning rate [0. 0005, 0. 01], hidden layer units [30, 300], hidden layer number[1, 4]. The parameters of SCGWO and traditional GWO are set consistently: population size 10, maximum iteration 30, optimization dimension 4.

4. 2 Convergence Performance Comparison

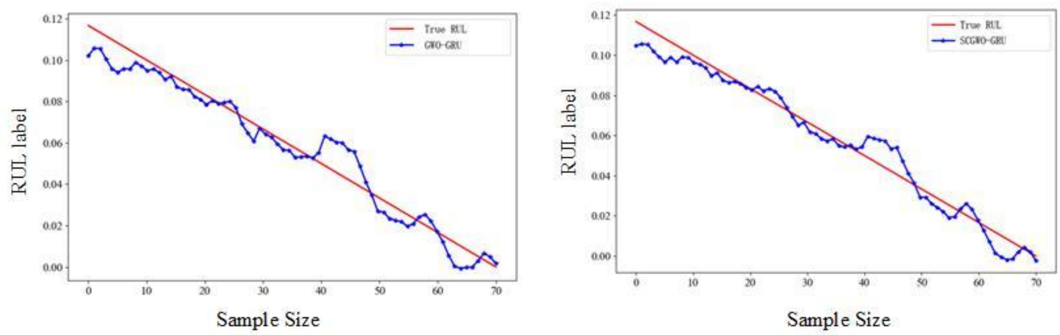
The convergence curves of SCGWO and traditional GWO are shown in Table 1. The traditional GWO converges after 8 iterations, while the improved SCGWO reaches convergence only after 3 iterations, with the convergence speed increased by 62. 5%, and the final convergence fitness value of SCGWO is lower than that of traditional GWO. The results prove that the proposed improvement strategies significantly accelerate the convergence speed and enhance the global search ability of the algorithm.

Table 1 Convergence performance comparison of optimization algorithms

Algorithm	**Convergence Iteration**	**Final Convergence Fitness Value**
raditional GWO	8	0. 00447
Improved SCGWO	3	0. 00435

4. 3 RUL Prediction Performance Comparison

To further compare the prediction accuracy of the improved GWO algorithm and the traditional GWO algorithm in optimizing the GRU model for gear remaining useful life prediction, the gear set 1 dataset was used for validation. The optimal hyperparameter combination corresponding to the position of the best grey wolf individual after the convergence of the improved GWO algorithm is as follows: batch size = 30, learning rate = 0.0005, number of hidden units = 95, and number of hidden layers = 4; the hyperparameter combination corresponding to the traditional GWO algorithm is: batch size = 30, learning rate = 0.00073, number of hidden units = 33, and number of hidden layers = 3. Figure 1 shows the gear life prediction curves of the GRU hyperparameters optimized by the two algorithms.



a) GWO-GRU b）SCGWO-GRU

Figure 1 Life prediction curves of the two algorithms

As shown in the figure, the prediction curves of both models exhibit small fluctuations and decrease consistently with the trend of the true RUL values with minor variations. The model evaluation metrics of the two algorithms are as follows: for the GWO-GRU algorithm, the MAE is 4.537×10^{-3}, the RMSE is 5.965×10^{-3}, and the R^2 is 0.9802; for the improved GWO-GRU algorithm, the MAE is 4.418×10^{-3}, the RMSE is

5.726×10^{-3}, and the R^2 is 0.9834. Among them, the MAE and RMSE of the improved GWO-GRU model are reduced by 2.62% and 4.02% compared with the GWO-GRU model, indicating a certain improvement in the prediction accuracy of the improved GWO-GRU model.
Overall, in this dataset, the improved grey wolf optimizer not only achieves a significant acceleration in convergence speed but also a synchronous improvement in prediction accuracy compared with the traditional GWO algorithm.

V. Conclusion

Aiming at the slow convergence of traditional GWO in GRU hyperparameter optimization for gear RUL prediction, this paper proposes an improved SCGWO algorithm with three targeted improvement strategies, and constructs a SCGWO-GRU gear RUL prediction model. Experimental results show that the SCGWO algorithm converges within 3 iterations, with a 62. 5% increase in convergence speed compared with traditional GWO; meanwhile, the SCGWO-GRU model reduces MAE and RMSE by 2. 62% and 4. 02% respectively, achieving higher prediction accuracy. The proposed model effectively solves the problems of slow convergence of traditional GWO and low prediction accuracy caused by manual hyperparameter tuning, and provides an efficient method for gear RUL prediction. In future work, we will further explore gear RUL prediction under variable working conditions and multi-sensor data fusion to improve the generalization ability of the model.

Funding
This work was supported by the Research Project of Jilin Provincial Department of Science and Technology (20240602073RC).

References

[1] Lei Y, Li N, Guo L, et al. Machinery health prognostics: A systematic review from data acquisition to RUL prediction[J]. Mechanical Systems and Signal Processing, 2018, 104: 799-834.
[2] Qu L S, He Z J. Mechanical Fault Diagnostics[M]. Shanghai: Shanghai Science and Technology Press, 1986.
[3] Si X S, Wang W, Hu C H, et al. Remaining useful life estimation–a review on the statistical data driven approaches[J]. European Journal of Operational Research, 2011, 213(1): 1-14.
[4] Cho K, Van Merriënboer B, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[J]. arXiv preprint arXiv: 1406. 1078, 2014.
[5] Mirjalili S, Mirjalili S M, Lewis A. Grey wolf optimizer[J]. Advances in Engineering Software, 2014, 69: 46-61.
[6] Mittal N, Singh U, Sohi B S. Modified grey wolf optimizer for global engineering optimization[J]. Applied Computational Intelligence and Soft Computing, 2016, 2016: 7950348.
[7] Yang X S, Deb S. Cuckoo search via Lévy flights[C]//2009 World Congress on Nature & Biologically Inspired Computing. IEEE, 2009: 210-214.
[8] Qin Y, Yang J, Zhou J, et al. Dynamic weighted federated remaining useful life prediction approach for rotating machinery[J]. Mechanical Systems and Signal Processing, 2023, 202: 110688.