WIND TURBINE ROLLING BEARING FAILURE PREDICTION BASED ON PSO-TRANSFORMER-BiLSTM MODELING

Zhenqiang ZHAO[1], Houxu PU[2], Jianhua DONG[3], Hong CHEN[4*]

The failure of wind turbine gearbox bearings can negatively impact grid-connection efficiency and may even result in severe accidents, causing substantial financial losses for wind farm operators. This study introduces a PSO-Transformer-BiLSTM model, which integrates the Transformer model, Bidirectional Long Short-Term Memory (BiLSTM) model, and Particle Swarm Optimization (PSO) algorithm to solve wind gearboxes bearing failure problem. In this model, the Transformer component is responsible for extracting signal features, while the BiLSTM model is applied for both signal feature extraction and signal timing prediction. The PSO algorithm is utilized to fine-tune the model's hyperparameters for optimal performance. The vibration fault dataset for the rolling bearings was obtained through a wind power fault simulation platform using components such as inner-ring-wear rolling bearings, acceleration sensors, and cloud vibration meters. Experimental results demonstrate that the PSO-Transformer-BiLSTM model achieves Root Mean Square Error (RMSE), R^2 values, and Mean Absolute Error (MAE) of 13.26, 0.93, and 10.37, respectively, on the faulty bearing vibration dataset. These results indicate an improvement in performance compared to individual Transformer and BiLSTM models, with the R^2 value increasing by 0.01 and 0.03, respectively. Additionally, when compared with other phase prediction models, such as CNN-LSTM-XGB, TCN-LSTM, and CNN-LSTM-SE, the R^2 coefficients of the PSO-Transformer-BiLSTM model surpass them by 0.12, 0.02, and 0.01, respectively. These findings confirm that the PSO-Transformer-BiLSTM model delivers reliable performance on faulty bearing datasets and offers valuable insights for diagnosing vibration faults in wind turbines.

Keywords: Wind turbine, Rolling bearing, Vibration signal, PSO-Transformer-BiLSTM

1. Introduction

Renewable energy is an imperative to address the environmental impact of fossil fuels [1]. By the end of June 2024, the total installed capacity of wind power

[1] Department of Electromechanics, Hami Vocational and Technical College, Hami, Xinjiang Uygur Autonomous Region, China
[2] Department of Electromechanics, Hami Vocational and Technical College, Hami, Xinjiang Uygur Autonomous Region, China
[3] Department of Electromechanics, Hami Vocational and Technical College, Hami, Xinjiang Uygur Autonomous Region, China
[4] * Prof., School of Mechanical and Power Engineering, Zhengzhou University, Zhengzhou, Henan, China, Corresponding author, e-mail: hongchen229@zzu.edu.cn

in China will have reached 46.7 million kilowatts, and it is expected that as of the end of 2025, the country's wind power installation capacity will reach about 53 million kilowatts. However, the reliability of wind turbines, which have been in harsh environmental conditions for a long time, is generally low, and domestic wind turbines have entered a period of frequent failures [2]. Bearing as a key bearing component in the gearbox, its failure will not only lead to an increase in downtime, but also may lead to more serious mechanical damage, resulting in huge economic losses and maintenance costs [3]. Currently, the bearing failure rate has been as high as 20% [4]. The formation of fatigue life prediction theory can also extend the life cycle of the product and save costs in the production industry [5]. If the bearings' health condition can be determined quickly and accurately, and timely predictions can be made before major failures occur, it is of great significance to increase turbine running time, improve operational reliability, and avoid major accidents and losses from unplanned downtime.

Vibration monitoring is extensively applied in fault diagnosis, offering high efficiency in detecting fan bearing issues through vibration signals, high sensitivity, simple data processing, high-precision diagnosis, real-time monitoring and data analysis. Based on vibration detection technology Pavithra et al [6] proposed a method for classifying machine faults using spectral image-based machine vibration signals and deep convolutional neural network representations, the model accuracy is 100% in the binary classification dataset of normal bearings and faulty bearings; Yan Zhao [7] proposed a fatigue life assessment process based on measured vibration data, combined with the SSI-UPC to the wind turbine tower self-oscillation frequency and damping ratio of the law to summarize the law; Jun Pi et al [8] combined with the bearing vibration fault simulation platform combined with machine learning to realize bearing fault diagnosis; Ahmed et al [9] used a plain Bayesian model to discriminate helical gear tooth wear faults in gear transmission achieving an accuracy of 93.9% and an area under the ROC curve of 99.1%.

Deep learning has the advantages of being adaptive, data-driven, and self-extracting features, among which the Transformer model [10] has a powerful sequence modeling capability and a self-attention mechanism, which can effectively capture complex time dependencies in vibration signals, thus improving the accuracy and efficiency of fault diagnosis. Ding et al [11] further developed the vanilla Transformer for sequence processing, building on the success of the original model, an end-to-end time-frequency Transformer model is introduced to extract meaningful abstractions from the time-frequency representation of vibration signals. The effectiveness of the proposed approach is validated through a case study on an experimental bearing dataset, showcasing its advantages over the benchmark model and other cutting-edge methods; Wen Jiangtao et al. [12] achieved a prediction of the remaining life of rolling bearings based on the improved Transformer model with a root mean square difference of 0.0941; Lin Tantao et al. [13] fused the faulty acoustic-vibration signals from the outer ring, inner ring, cage, and rolling element of a rolling bearing , using the Transformer model for diagnosis, with an accuracy of 89.51%.

Bearing vibration signals are sequential data that Bidirectional Long Short-Term Memory (BiLSTM) is commonly utilized for forecasting based on sequential data. The BiLSTM is an effective method for capturing long-range dependencies in time series data, automatically extracting fault characteristics, and improving the accuracy and reliability of the model through the bi-directional flow of information. Nacer et al [14] based on BiLSTM improved model using Case Western Reserve University bearing dataset tested the performance under 16 conditions and different loads, and the accuracy was as high as 99.96%; Guo et al [15] developed an ACNN-BiLSTM diagnosis of bearing faults model building on the foundation of the BiLSTM model with an average diagnostic accuracy as high as 99.79% on Jiangnan University bearing dataset; Sun et al [16] introduced a new model built on the enhanced DBO algorithm. They proposed a rolling bearing fault diagnosis framework that leverages VMD-CNN-BiLSTM, optimized through the CSADBO algorithm. The model achieves an average accuracy of 99.6% in diagnostics in regard to the bearing dataset from Case Western Reserve University; Guo et al [17] combine the MSCNN and BiLSTM to construct MSCNN-BiLSTM for real-time and stable monitoring of the rolling bearing operation status of a marine engine, and the experimental findings indicate that the proposed model is an effective fault diagnostic network model with strong generalization ability, which is able to accurately identify the different states of the rolling bearings at a faster rate.

The PSO operator can be used to automatically optimize the hyperparameters of deep learning models. Each particle corresponds to a set of potential parameters, and its performance is assessed using an objective function to progressively converge toward the optimal parameter combination. Song et al [18] addressed the challenges of insufficient utilization of the time-dependent bearing fault signalling features, the high cost of parameter tuning represents a significant challenge for many organisations, and the challenge of acquiring sufficient training data under a range of operational conditions. They proposed a CNN-BiLSTM model optimized with a PSO operator for hyper-parameter tuning, achieving a diagnostic accuracy of 99.44%. The diagnostic accuracy reaches 99.44%; Geethanjali et al [19] The PSO optimization method is used to optimize the ANN model for detecting normal, magnetizing inrush current, core overexcitation, internal fault and external fault states of the transformer, and the results demonstrate that the PSO-optimized ANN model is better in comparison with the BPN model; Bharti et al [20] developed a PSO-BiLSTM model, combining the PSO algorithm with a BiLSTM neural network, the objective is to forecast the short-term traffic flow. The model was compared with LSTM, ELM, GRU, WNN, MLP, and ARIMA, and the experimental results demonstrate that the PSO-BiLSTM model surpasses the others in both accuracy and stability.

Building on the aforementioned research, this paper focuses on the faulty rolling bearing vibration signals from the wind power fault simulation platform as the research subject. It develops the PSO-Transformer-BiLSTM model to predict faulty bearing signals, aiming to achieve effective bearing fault prediction and offer technical support for future fault diagnosis of rolling bearings.

2. Equipment and methods

2.1 Simulation of wind turbine gearbox bearing failure

The wind turbine gearbox output shaft condition detection platform built by this institute is shown in Fig. 1. The condition detection platform includes wind power transmission fault simulation experimental bench (Jiangsu Union Yiyou Measurement and Control Technology Co., Ltd.), 6212 inner ring wear bearing, PDES-G cloud diagnostic instrument (Zhengzhou Enptech Co., Ltd.) and EAG04-100 acceleration sensor (Zhengzhou Enptech Co., Ltd.). The wind power transmission fault simulation test bench with 6212 faulty bearings to simulate the bearing failure of the wind turbine gearbox output shaft, through the PDES-G cloud diagnostic instrument and EAG04-100 acceleration sensor for the faulty bearing vibration signal acquisition. Acceleration sensors are mounted on the positive X and positive Y axes of the planetary gearbox output shaft housing to collect bearing vibration signals from the gearbox output shaft.

The defect of the faulty bearing is the wear of the inner ring of the bearing, the data in the motor IF is 3093RPM, IF is 51.5Hz. the vibration signal (horizontal and vertical direction) is sampled at 25.6kHz, and the data is collected every 10 seconds for 0.1 seconds, i.e., 2560 samples are recorded, Fig. 2 shows the sample of the faulty bearing.

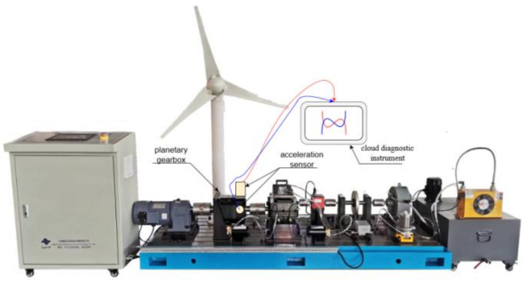

Fig. 1. Wind turbine gearbox output shaft condition detection platform

Fig. 2. Faulty bearing

2.2 Transformer

The Transformer model is fundamentally based on the self-attention mechanism, enabling the model to capture long-range dependencies by considering all elements in the sequence when processing each individual element. The

Transformer is composed of two primary components: the encoder and the decoder, both consisting of multiple identical layers. The encoder handles the input sequence, while the decoder leverages the encoder's output to produce the target sequence. Each layer within the encoder and decoder includes a multi-head attention mechanism and a feed-forward neural network, accompanied by residual connections and layer normalization to ensure stable gradient propagation. The structure of the Transformer model is illustrated in Fig. 3.

2.3 BiLSTM

BiLSTM is a recurrent neural network model built upon LSTM [21]. Unlike the traditional LSTM, the BiLSTM [22] model can handle both unidirectional dependencies and simultaneously consider information from both the past and future in time series data. The BiLSTM architecture comprises two principal components: two LSTM networks are employed, one forward and one backward. The forward LSTM processes the sequence data from left to right, while the backward LSTM traverses the same sequence from right to left. The outputs from both directions are combined to form the final output of the BiLSTM. This structure allows the model to leverage both historical and future context, enhancing its capability to capture complex patterns within sequential data.

As shown in Fig. 4, the forward layer performs computations step-by-step, storing the output of the forward hidden layer at each step. Afterward, the backward layer computes in reverse, saving the output from the backward hidden layer at each step. Subsequently, the outputs from the forward and backward layers are integrated to generate the final output.

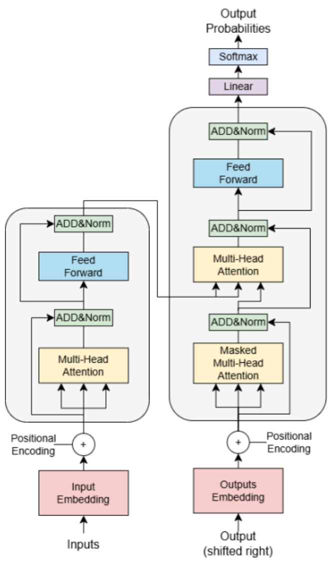

Fig. 3. Transformer framework

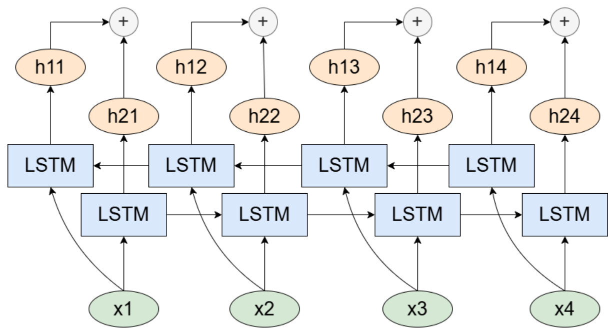

Fig. 4. BiLSTM model structure

2.4 PSO algorithm

PSO is a global search algorithm introduced by Kennedy and Eberhart, inspired by the foraging and evasive behaviors of swarming animals [23-24]. PSO offers several advantages, including fast convergence, a small number of parameters, high robustness, and simplicity, making it easy to implement. Due to these benefits, PSO is frequently employed for hyper-parameter optimization in other models [24].

The core of the PSO algorithm is to verify the fitness of each point location by moving the particle regularly many times in the solution space, so the particle needs to constantly update its position to explore more points while running.

The population position and velocity of the particle swarm are initialized as shown below:

$$X = lb + rand * (ub - lb) \tag{1}$$

$$V = V\min + rand * (V\max - V\min) \tag{2}$$

where X represents ub, lb represents the upper line position boundary of the particle, and Vmax, Vmin represent the velocity boundary of the particle, respectively.

Since the particles are affected by the individual inertial motion direction, the individual optimal motion direction and the group optimal motion direction, the velocity update is vectorially superimposed by all three, as shown in the following equation:

$$V^{i+1} = wV^{i} + C_1(P^{i}_{best} + X^{i}) + C_2(G^{i}_{best} - X^{i}) \tag{3}$$

Where, w, C_1, C_2 wind up denoting the weight coefficients and learning factors, respectively.

The i+1st position of the particle is affected by the ith position and the i+1st velocity with the following mathematical expression:

$$X^{i+1} = X^{i} + V^{i+1} \tag{4}$$

Where x^k denotes the position of the algorithm at the kth iteration, v^{k+1} denotes the displacement of the k+1st generation, and x^{k+1} i.e. the position of the k+1st generation is based on the position of the kth generation plus the displacement of the k+1st generation.

2.5 PSO-transformer-BiLSTM bearing diagnostic modeling

The Transformer's feature learning method obviates the necessity for manually-driven feature extraction, which is a hallmark of traditional approaches. This enables end-to-end information processing. Nevertheless, depending on a solitary model for time series prediction may not necessarily result in optimal outcomes. To address this, a Transformer-BiLSTM neural network based on one-dimensional data is put forth, integrating the Transformer's capacity for local feature extraction with the BiLSTM's aptitude for handling nonlinear temporal dependencies. This model is devised for the extraction and classification of one-dimensional bearing fault vibration signals.

In the Transformer-BiLSTM network, the choice of hyperparameters, including the learning rate, batch size and number of iterations, has a significant impact on the model's performance. Achieving optimal results for complex networks requires a substantial amount of tuning experience and a large number of experiments. To address this challenge, the particle swarm optimisation (PSO) algorithm is employed to optimise the model's hyperparameters, thereby reducing the time required for manual tuning. This approach ensures the discovery of the best training parameters, fully leveraging the model's diagnostic capabilities.

2.6 Training environment

The system configuration used is Windows 11, the processor is i7-12650H, 16GB RAM, 1T hard disk and Nvidia RTX4060 graphics card, and Matlab2023b software is used for PSO-Transformer-BiLSTM model building. Its hyperparameter settings are shown in Table 1.

Table 1

LSTM model parameters

hyperparameterization	parameter value
neuron setting	[16,16]
learning rate	0.001
Epoch	70
Batchsize	50
Training set: validation set: test set	8:1:1

2.7 Evaluation indicators

To evaluate the predictive performance of the model, four metrics were used: MAE, RMSE, R^2 coefficient, and Training Loss Value.

The calculation of MAE and RMSE is shown in Equation (6) and Equation (7). Where *RULt* is the real RUL value of the rolling bearing at the t^{th} moment; where $\overline{RUL_t}$ is the predicted RUL value of the rolling bearing at the tth moment; *n* is the number of samples in the test set. R^2 coefficient indicates the proportion of the variation explained by the model to the total variation, and its he value is expressed as a number between 0 and 1. A value that is closer to 1 is indicative of a greater degree of the characteristic in question, represents the better the model is trained on the bearing vibration dataset and the more accurate the state prediction is; the higher the value of the training loss indicates the worse the stability of the model, and the lower the value of the training loss indicates the better the stability of the model. The higher the training loss value, the worse the model stability, and the lower the training loss value, the better the model stability.

$$E_t = RUL_t - \overline{RUL_t} \tag{5}$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}\left|E_t\right| \tag{6}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(E_t\right)^2} \tag{7}$$

2.8 Dataset

The input data used in this study were rolling bearing vibration signals acquired via an EAG04-100 accelerometer.The data type is a one-dimensional time series vibration signal with a sampling frequency of 25.6 kHz, and 0.1 seconds of data are collected every 10 seconds, totaling 2560 sampling points. The signal direction is the positive X-axis and positive Y-axis of the sensor installed in the output shaft housing of the planetary gearbox, and the vibration signals in the horizontal and vertical directions are collected respectively.

Noise reduction and normalization were used for data processing. First, wavelet threshold denoising was used to eliminate high-frequency noise; then, Z-score standardization was performed on the raw signals; finally, the continuous signals were divided into segments with a length of 512 (with an overlap rate of 50%) and marked as “normal” or “faulty” according to the inner ring wear fault simulation platform. ".

3. Results and analysis

3.1 Prediction performance of PSO-Transformer-BiLSTM model for faulty bearing vibration signal

3.1.1 Training of faulty bearing detection models

Fig. 5(a) and (b) show the training RMSE and loss function plots of the PSO-Transformer-BiLSTM model for the faulty bearing vibration signals, respectively; from Fig. 5(a), it can be seen that the model shows a rapidly decreasing RMSE in the first 50 iterations, and the curve decreases slowly up to 200 Epochs, and this is followed by a steady phase in the RMSE. Fig. 5(b) shows that the loss function curve decreases rapidly from 0.496 to 0.087 in the first 50 iterations, and then the curve decreases slowly until it decreases to 0.035 at the 1400th Epoch, which indicates that the network model can obtain good detection without both underfitting and overfitting. The results of the network model.

3.1.2 Evaluation of the prediction performance of the faulty bearing vibration signal detection model

In the previous section, the faulty bearing signal dataset is divided into 8:1:1, and the validation and test sets of PSO-Transformer-BiLSTM are 10% of the dataset, respectively, and the validation and test sets will be used to calibrate the model and to test the effect of the improved model's performance on the faulty bearing signal dataset, respectively.

Table 2 demonstrates the MAE, RMSE and R^2 coefficients for the validation and prediction of the PSO-Transformer-BiLSTM model on the faulty bearing signal dataset. The MAE, RMSE and R^2 coefficients in the model validation stage are 10.13, 12.93 and 0.93, respectively, and the MAE, RMSE and R^2 coefficients in the model prediction stage are 10.37, 13.26 and 0.93, respectively, with the largest difference of 0.33 for the RMSE, the smallest gap of 0 for the R^2 coefficients, and the gap of 0.24 for the MAE, and the difference of the above data are all less than 1, especially the error of R^2 coefficient is less than 0.005. The experiments show that the PSO-Transformer-BiLSTM model is suitable for the fault prediction of wind turbine bearings.

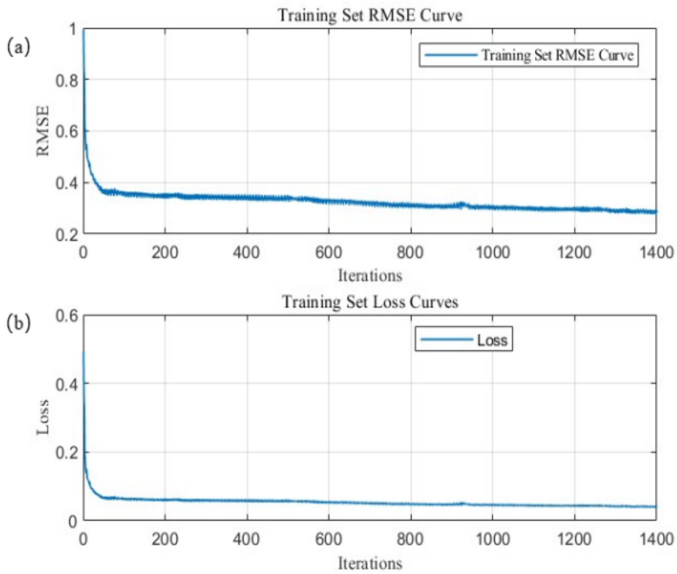

Fig. 5 PSO-Transformer-BiLSTM model RMSE plot
Note: Fig. 5(a) shows the RMSE curve of PSO-Transformer-BiLSTM model
Fig. 5(b) shows the PSO-Transformer-BiLSTM model training loss function plot

Table 2

Prediction results of PSO-Transformer-BiLSTM model for faulty bearing

model stage	MAE	RMSE	R^2 Coefficients
validate (a theory)	10.13	12.93	0.93
beta (software)	10.37	13.26	0.93

3.2 Ablation experiments

In this paper, ablation experiments are set up to verify the performance impact of each added module on PSO-Transformer-BiLSMT. The ablation experiment splits the PSO-Transformer-BiLSMT model into three classes, BiLSTM, Transformer-BiLSTM and PSO-Transformer-BiLSMT, and Table 3 shows the prediction set MAE, RMSE and R2 coefficients and the training set loss function corresponding to the above three classes of modeled faulty bearing vibration signals data sets. set loss function.

Table 3 demonstrates the MAE, RMSE and R2 coefficients of the prediction sets for the three models BiLSTM, Transformer-BiLSTM and PSO-Transformer-BiLSTM. The MAE, RMSE and R2 coefficients of Transformer are the lowest among the three models, which are 12.43, 15.50 and 0.90, respectively, while BiLSTM is the next lowest, with the MAE, RMSE and R2 coefficients of 10.71, 13.65 and 0.92, respectively. BiLSTM, which is not optimized, has a higher MAE, RMSE and R coefficients than that of Transformer. BiLSTM without any optimization instead performs better than Transformer-BiLSTM with Transformer module in the faulty bearing dataset. However, the PSO-Transformer-BiLSTM model optimized with PSO operator achieves better performance with the highest MAE, RMSE and R2 coefficients of 10.37, 13.26 and 0.93.

Although the improvement in the R^2 coefficient of the PSO-Transformer-BiLSTM model compared to the BiLSTM model is marginal (0.01), significant enhancements are observed in MAE and RMSE, with reductions of 0.34 and 0.39, respectively. These results indicate that the PSO-Transformer-BiLSTM model outperforms BiLSTM in both reliability for routine fault monitoring and capability to detect sudden abnormal signals. Furthermore, the PSO-Transformer-BiLSTM model demonstrates a 20s reduction in prediction time compared to the BiLSTM model, highlighting its efficiency in real-time applications.

Table 3

Prediction results of BiLSTM, Transformer-BiLSTM, PSO-Transformer-BiLSTM models for faulty bearings

Model Category	MAE	RMSE	R^2 Coefficients	Prediction Time
BiLSTM	10.71	13.65	0.92	641s
Transformer-BiLSTM	12.43	15.50	0.90	622s
PSO-Transformer-BiLSTM	10.37	13.26	0.93	621s

Fig. 6 shows the training loss function of BiLSTM, Transformer-BiLSTM and PSO-Transformer-BiLSTM models. From the figure, it can be seen that the training loss function of PSO-Transformer-BiLSTM model is the lowest compared to BiLSTM and Transformer-BiLSTM, which indicates that PSO-Transformer-BiLSTM model is the most stable. When the Epoch of BiLSTM model is 1400, its loss function is 0.38, and the loss function of Transformer-BiLSTM model at the 1400th Epoch is 0.46, which also explains the higher prediction performance of BiLSTM than Transformer-BiLSTM in terms of loss function.

3.3 Prediction results of different models for the faulty bearing dataset

To further evaluate the predictive performance of the PSO-Transformer-BiLSTM model on the faulty bearing dataset, a comparison was conducted against several other models, including the CNN-LSTM-XGB model, the TCN-LSTM model, the CNN-LSTM-SE model, and the PSO-Transformer-BiLSTM model itself.

Table 4 shows the performance results of four target detection network models for predicting the vibration signals of faulty bearings. As can be seen from the table, the MAE, RMSE and R2 coefficient of PSO-Transformer-BiLSTM have the best performance among the four models, which are 10.37, 13.26 and 0.93, respectively; and the worst performance model is CNN-LSTM-XGB, which has MAE, RMSE and R2 coefficients of 16.11, 17.26 and 0.81, and has a difference of 5.74, 4.00 and 0.12 with the evaluation index of PSO -Transformer-BiLSTM with MAE, RMSE and R coefficient of 16.11, 17.26 and 0.81, respectively, which are different from those of PSO-Transformer-BiLSTM with evaluation indexes of 5.74, 4.00 and 0.12, respectively. Overall, the PSO-Transformer-BiLSTM model proposed in this study has good results in the prediction of the faulty bearing vibration signal dataset in comparison with other models. with good results.

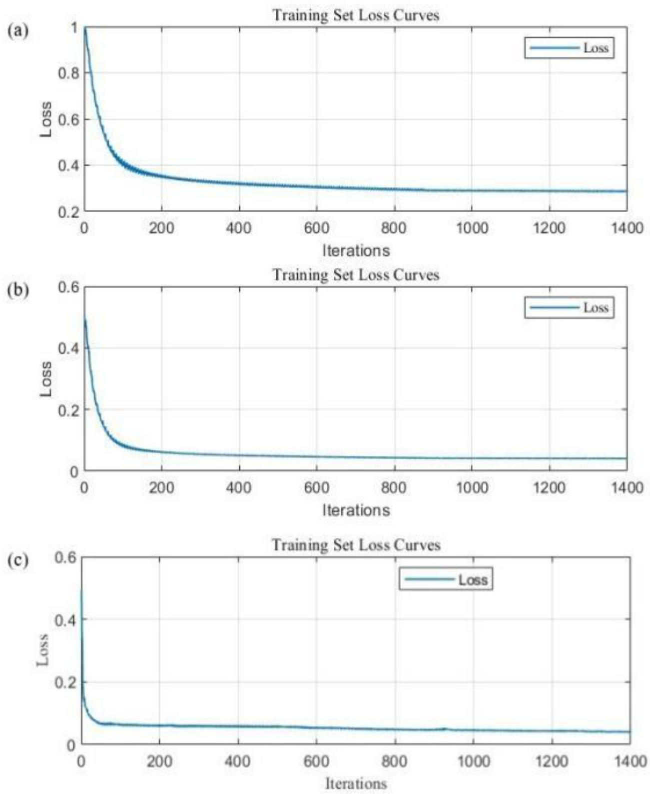

Fig. 6. Training loss function of BiLSTM, Transformer-BiLSTM, PSO-Transformer-BiLSTM models

Note: Fig. 6(a), (b) and (c) show the training loss function plots of BiLSTM, Transformer-BiLSTM, PSO-Transformer-BiLSTM respectively

Table 4

Comparison of model performance

Model Category	MAE	RMSE	R^2 Coefficients
CNN-LSTM-XGB	16.11	17.26	0.81
TCN-LSTM	11.36	16.52	0.91
CNN-LSTM-SE	10.85	13.76	0.92
PSO-Transformer-BiLSTM	10.37	13.26	0.93

4. Conclusion

In the context of big data, accurately estimating the life of rolling bearings remains a challenge. To address this issue, this study proposes a novel PSO-Transformer-BiLSTM prediction model integrating the Transformer, PSO operator, and BiLSTM for rolling bearing life prediction. The Transformer's attention module, featuring residual connections, effectively captures long-term dependencies within time series data. To further enhance the model's capacity to extract crucial features from long sequences and reduce the impact of noise interference, the PSO operator is incorporated to optimize the Transformer-BiLSTM backbone. The BiLSTM model is then employed to implement bearing failure prediction, enabling a more comprehensive understanding of the time series context and interdependencies.

The main contributions and findings of this study are summarized as follows:

We propose and implement a novel hybrid deep learning architecture that synergistically integrates the Transformer (for capturing complex long-range dependencies and feature interactions in vibration signals), the BiLSTM (for effectively modeling bidirectional temporal dynamics in bearing degradation), and the PSO algorithm (for efficiently optimizing critical hyperparameters like learning rate, batch size, and iterations). This integration overcomes the limitations of manual hyperparameter tuning and leverages the strengths of each component, leading to enhanced overall performance.

The proposed PSO-Transformer-BiLSTM model demonstrates superior predictive performance on a real-world wind turbine gearbox bearing vibration dataset collected from a dedicated fault simulation platform. Experimental results show significant improvements: achieving RMSE=13.26, MAE=10.37, and R^2 =0.93. These results indicate exceptional model fit and prediction accuracy, outperforming baseline models (Transformer, BiLSTM) and state-of-the-art hybrid models (CNN-LSTM-XGB, TCN-LSTM, CNN-LSTM-SE), particularly evidenced by the highest R^2 coefficient.

The model provides a robust and accurate tool for predicting incipient rolling bearing failures in wind turbines. By enabling timely fault detection, it offers significant practical value. This capability contributes directly to improving wind turbine operational reliability, reducing costly unplanned downtime, minimizing maintenance expenditures, preventing catastrophic failures, and ultimately enhancing grid-connection efficiency and economic benefits for wind farm operators.

This work empirically validates the effectiveness of employing the Particle Swarm Optimization (PSO) algorithm for hyperparameter tuning in complex hybrid deep learning models (specifically Transformer-BiLSTM) applied to bearing fault prognostics. The significant performance boost (e.g., R^2 increased by 0.03 compared to the non-optimized Transformer-BiLSTM) achieved by the PSO-optimized model underscores the critical importance and efficacy of automated hyperparameter optimization in this domain.

Notably, compared with the BiLSTM model, the PSO-Transformer-BiLSTM model increases the R^2 coefficient by 0.01, and compared with the Transformer-BiLSTM model, it shows a 0.03 improvement. These results provide an experimental basis for subsequent utilization of the PSO operator in model optimization and offer a theoretical foundation for further algorithm-oriented optimization research.

Despite the promising results achieved by the proposed hybrid prediction model in rolling bearing prediction, there are still limitations in terms of the interpretability of prediction results and the quantification of confidence levels. Future research should focus on exploring how to optimize the model to better fit the monitored degradation data, aiming to achieve more accurate and effective bearing life predictions.

Acknowledgements

This work was supported by the Special Program for Talent Development of the Organization Department of the CPC Committee of Xinjiang Uygur Autonomous Region (Project Title: Research on Key Technologies and System Construction for Intelligent Predictive Maintenance of Wind Turbines)

R E F E R E N C E S

[1] *Abohela, I., Hamza, N., Dudek, S.*, 2013. Effect of roof shape, wind direction, building height a nd urban configuration on the energy yield and positioning of roof mounted wind turbines. r enew. energy 50, 1106-1118. .

[2] *Xiao Xiaocong.* Research on intelligent monitoring method of main bearing operation status of large direct-drive wind turbine based on SCADA data. Hunan University of Science and Tec hnology, 2023.

[3] *Chen X, Cui Z, Shen J,* et al. Fault analysis and identification of motor bearing based on esmd a nd svm. UPB Scientific Bulletin, Series C: Electrical Engineering and Computer Science, 2 021, 83(4): 239-250.

[4] *Yuan Jingyi*. Research on fault diagnosis method of wind turbine gearbox based on online moni toring data. North China Electric Power University, 2018.

[5] *Ling L*. Residual life prediction for the spindle of retired machine based on fatigue damage mec hanism. UPB Scientific Bulletin, Series D: Mechanical Engineering, 2018, 80(02): 183-169.

[6] *Pavithra, R.*, and Prakash Ramachandran. "Deep convolution neural network for machine healt h monitoring using spectrograms of vibration signal and its EMD-intrinsic mode functions." Journal of Intelligent & Fuzzy Systems 44.6 (2023): 8827-8840.
[7] *Zhao Yan.* Research on vibration monitoring and dynamic response of wind turbine structure. Z hejiang University, 2018.
[8] *Pi Jun, Liu Peng, Hu Chao*. A rolling bearing fault diagnosis method based on GA-PNN networ k. Mechanical Design and Manufacturing, 1-5, 2024-09-08.
[9] *Abdulameer, Ahmed Ghazi*, et al. "Naïve Bayes algorithm for timely fault diagnosis in helical g ear transmissions using vibration signal analysis." International Journal on Interactive Desig n and Manufacturing (IJIDeM) (2024): 1-12.
[10] *Ding Y, Jia M, Miao Q,* et al. A novel time-frequency Transformer based on self-attention mec hanism and its application in fault diagnosis of rolling bearings. bearings. Mechanical Syste ms and Signal Processing, 2022, 168:108616.
[11] *Wen Jiangtao, Zhang Zhe*. Remaining life prediction method for rolling bearings based on imp roved Transformer. Journal of Yanshan University, 2024, 48(04):312-321.
[12] *Lin Tantao, Niu Qingbo, Ma Tianxu*, et al. Intelligent bearing sound-vibration fusion fault diag nosis based on Transformer. Bearing, 2023, (02):67-73.
[13] *Nacer S M, Nadia B, Abdelghani R*, et al. A novel method for bearing fault diagnosis based on BiLSTM neural networks. The International Journal of Advanced Manufacturing Technolo gy, 2023, 125(3): 1477-1492.
[14] *Guo Y, Mao J, Zhao M*. Rolling bearing fault diagnosis method based on attention CNN and B iLSTM network. Neural processing letters, 2023, 55(3): 3377-3410.
[15] *Sun, W., Wang Y, You X.* et al. Optimization of VMD-CNN-BiLSTM Rolling Bearing Fault Di agnosis Model Based on Improved DBO Algorithm. 2024.
[16] *Guo, Xiaocheng, Yang Yifei, He Zujun* et al. "Fault Diagnosis Method Based on MSCNN-BiLS TM for Marine Engine Rolling Bearing." 2024 3rd International Joint Conference on Inform ation and Communication Engineering (JCICE). IEEE, 2024.
[17] *Song B, Liu Y, Fang J*, et al. An optimized CNN-BiLSTM network for bearing fault diagnosis under multiple working conditions with limited training samples. Neurocomputing, 2024, 5 74: 127284.
[18] *Geethanjali M, Slochanal S M R, Bhavani R*. PSO trained ANN-based differential protection s cheme for power transformers. Neurocomputing, 2008, 71(4-6): 904-918.
[19] *Redhu, Poonam, and Kranti Kumar*. "Short-term traffic flow prediction based on optimized de ep learning neural network: PSO-Bi-LSTM." Physica A. Statistical Mechanics and its Appli cations 625 (2023): 129001.
[20] *Li B, Tang B, Deng L*, et al. Self-attention ConvLSTM and its application in RUL prediction o f rolling bearings. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 1-11.
[21] *Yin X H, Song R, Chen Z D,* et al. PSO-LSTM based construction schedule prediction method for shield tunneling. UPB Sci. Bull., Series C, 2024, 86(1):31-43.
[22] *Li Zhigang, Wang Xiangyang*. Research on collector pressure prediction based on improved P SO-ANFIS. Mechanical Design and Manufacturing, 2024, (09):1-5.
[23] *Chen J, Liu K, Luo X*, et al. A State-Migration Particle Swarm Optimizer for Adaptive Latent F actor Analysis of High-Dimensional and Incomplete Data. IEEE/CAA Journal of Automatic a Sinica, 2024, 11(11):2220-2235.
[24] *Yu Y, Kong L, Liu Y,* et al. PSO optimized SVM parameters for far infrared pedestrian detectio n. UPB Scientific Bulletin, Series C: Electrical Engineering and Computer Science, 2019, 8 1(4): 15-26.