Proceedings of the 4th International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/2026.32843

Risk Level Classification and Fault Recovery Time Prediction of Industrial Internet of Things Devices Based on Transformer-BiGRU

Yuanxin Li

Zhongshan, China
1162088945@qq.com

Abstract. In the wave of Industry 4.0, industrial Internet of Things (iot) devices are increasingly widely used in manufacturing, energy, chemical and other fields. The scale and operational complexity of these devices are constantly rising, and their stable operation is crucial for production efficiency and safety. To make up for the shortcomings of existing algorithms in long time series feature extraction and complex association mining, this paper proposes the Transformer-BiGRU classification and regression algorithm, and first conducts correlation analysis and violin plot analysis. Experiments show that the core evaluation indicators of the classification algorithm have significant advantages. The accuracy rate, recall rate, and precision rate all reach 84%, 84%, and 86% respectively. The F1 value is 83%, all higher than all comparison machine learning algorithms. The AUC reaches 93%, although slightly lower than CatBoost's 94%, it is higher than other algorithms such as Random Forest. Strong generalization and category discrimination capabilities; The MSE of the regression algorithm is 16.598, the RMSE is 4.074, and the MAE is 2.615, all of which are the lowest. The MAPE is 51.468, which is in a relatively low range. The R^2 reaches 0.442, which is significantly better than the traditional algorithm. This algorithm integrates the global dependency capture capability of Transformer with the temporal feature extraction advantages of BiGRU, providing a reliable solution for the precise analysis and prediction of the operating status of industrial Internet of Things devices, which is of great significance for ensuring the efficiency and safety of industrial production.

Keywords: Industrial Internet of Things, Transformer, BiGRU

1. Introduction

Under the wave of Industry 4.0, industrial Internet of Things (iot) devices are increasingly widely used in manufacturing, energy, chemical engineering and other fields. The scale and operational complexity of these devices continue to rise, and their stable operation is directly related to production efficiency and safety [1]. However, when equipment is constantly in a high-load and multi-condition environment for a long time, it is prone to failure due to wear, aging or abnormal conditions [2].

DOI: 10.54254/2755-2721/2026.32843

Machine learning algorithms provide an effective technical path for the risk level classification and fault recovery time prediction of industrial Internet of Things (iot) devices [3]. In terms of risk level classification, traditional classification algorithms such as Random Forest and XGBoost can divide equipment risks into different levels by learning the correlation between equipment operation characteristics and historical faults, accurately identifying high-risk equipment and helping operation and maintenance personnel prioritize resource allocation. In terms of fault recovery time prediction, regression algorithms such as linear regression and LightGBM can quantitatively predict the recovery time after a fault based on equipment type, fault type and operating status, providing data support for the formulation of operation and maintenance plans [4]. In addition, for the time series characteristics of industrial data, algorithms such as recurrent neural networks can capture the temporal dependencies of the data and improve the accuracy of predictions [5].

To address the deficiencies of existing algorithms in long time series feature extraction and complex association mining, this paper proposes the Transformer-BiGRU classification algorithm and the Transformer-BiGRU regression algorithm. The self-attention mechanism of the Transformer module can effectively capture the global dependencies in long time series data and explore the potential correlations among different operational features.

2. Data sources

The dataset used in this article contains 729 complete equipment operation records, covering three types of features: equipment basic information, operation and maintenance data, and multi-dimensional sensor monitoring data. The basic information of the equipment includes the unique number type and operating duration. The sensor monitoring data involves key operating indicators such as temperature, vibration, pressure, voltage, current, rotational speed, environmental humidity and oil level. The operation and maintenance data includes daily operating duration, monthly maintenance frequency and the number of days since the last maintenance. Output the correlation heat maps of each variable, as shown in Figure 1.

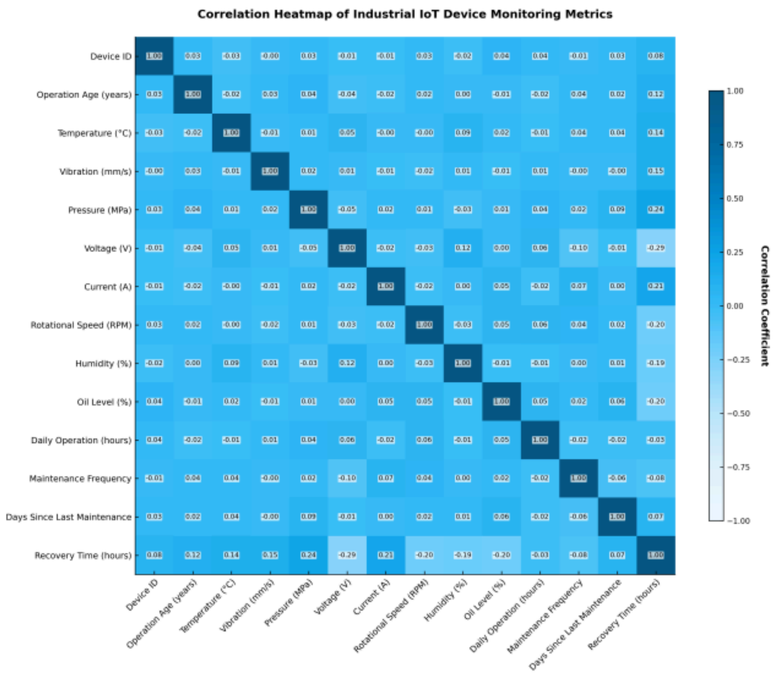

Figure 1. The correlation heat map

DOI: 10.54254/2755-2721/2026.32843

Output the violin plots of each variable and observe the distribution of the data, as shown in Figure 2.

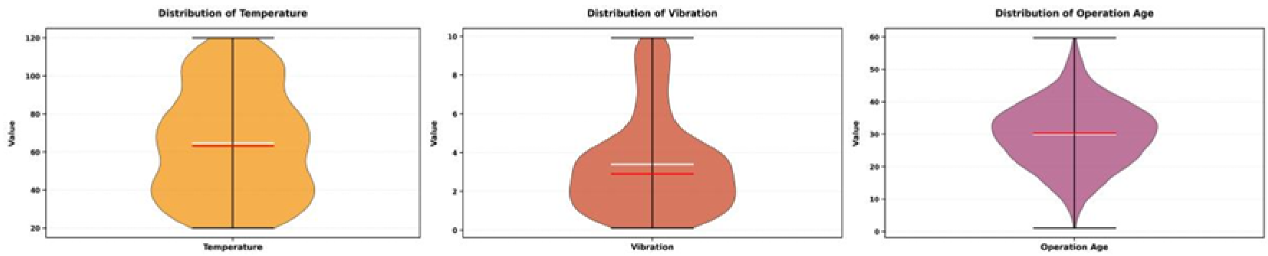

Figure 2. Violin diagrams of each variable

3. Method

3.1. Transformer

The Transformer algorithm is a deep learning model based on the self-attention mechanism. Its core lies in efficiently capturing global dependencies by calculating the association weights of each element in the input sequence with all other elements. It adopts an encoder-decoder structure and achieves parallel computing by means of multi-head attention mechanism and position encoding, breaking through the limitation of traditional recurrent neural networks that rely on temporal iteration. It shows significant efficiency advantages when processing long sequence data [6]. The network structure of the Transformer is shown in Figure 3.

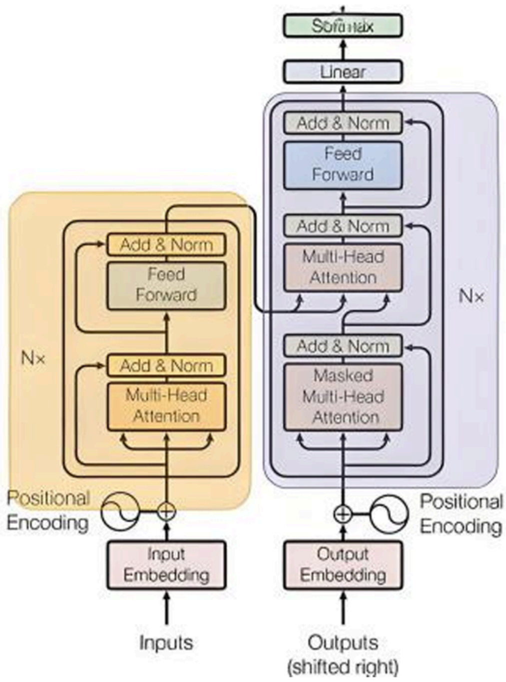

Figure 3. The network structure of the Transformer

3.2. BiGRU

BiGRU is a bidirectional extension form of the GRU algorithm. It captures the forward and reverse time series information of the sequence respectively through the forward GRU and backward GRU, and dynamically regulates the transmission and forgetting of information by using the update gate and reset gate, effectively alleviating the gradient vanishing problem of traditional RNNS. This algorithm is good at mining local temporal correlations and performs outstandingly in the context feature extraction of sequential data [7]. The network structure of BiGRU is shown in Figure 4.

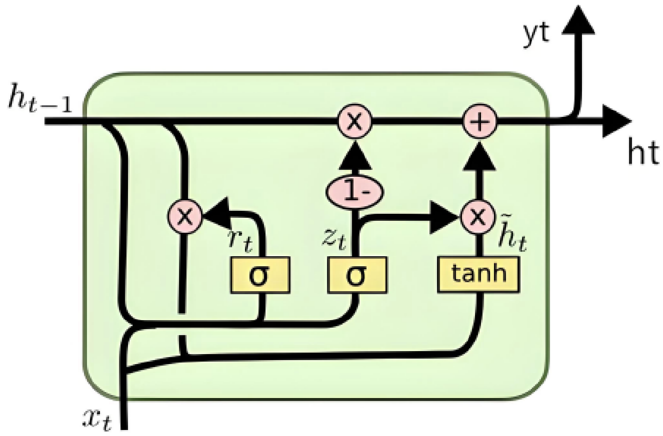

Figure 4. The network structure of BiGRU

3.3. Transformer-BiGRU

The Transformer-BigRU classification algorithm is a hybrid architecture that integrates the advantages of both models. The core idea is to capture long-distance dependent features at the global level through Transformer and then use BiGRU to mine local temporal correlation information to form complementary feature representations [8]. The process of this algorithm is as follows: The input data is processed through the multi-head attention mechanism of the Transformer to obtain the global feature vector, then input into the BiGRU layer for bidirectional temporal feature refinement, and finally output the category probability distribution through the fully connected layer and the softmax activation function [9].

The Transformer-BiGRU regression algorithm is a hybrid model designed for sequential data regression tasks. On the basis of retaining the core advantages of the fusion architecture, the output layer is replaced with a regression head suitable for continuous value prediction [10].

4. Result

The parameter Settings of the project are as follows: The ratio of the training set to the dataset is 0.7. The number of input channels is the feature dimension. The maximum position encoding is 256 times 2. The number of heads in the self-attention mechanism is 4. The number of key channels for each head is 32 and the total number of key channels is 128. The forward GRU layer has 6 units and the reverse GRU layer has 10 units. The Dropout rate is 0.01. The optimizer selects Adam, with a maximum number of training rounds of 200 and a batch size of 256. Each training session shuffled the dataset. The initial learning rate is 0.01, and the learning rate decline factor is 0.1. After 80 training sessions, the learning rate is adjusted to the initial value multiplied by the decline factor. The L2 regularization coefficient is 0.001, and the gradient clipping threshold is 10.

First, the risk levels of industrial Internet of Things (iot) devices were classified, and the results are shown in Table 1. Draw the bar comparison charts of each indicator as shown in Figure 5.

DOI: 10.54254/2755-2721/2026.32843

Table 1. The results of the comparative experiment

Model	Accuracy	Recall	Precision	F1	AUC
Decision tree	0.685	0.685	0.685	0.685	0.773
Random Forest	0.804	0.804	0.839	0.777	0.924
AdaBoost	0.79	0.79	0.847	0.776	0.913
GBDT	0.712	0.712	0.718	0.695	0.89
CatBoost	0.808	0.808	0.775	0.767	0.94
ExtraTrees	0.753	0.753	0.745	0.719	0.883
KNN	0.479	0.479	0.472	0.46	0.686
XGBoost	0.758	0.758	0.758	0.752	0.902
Transformer-BiGRU	0.836	0.836	0.863	0.832	0.927

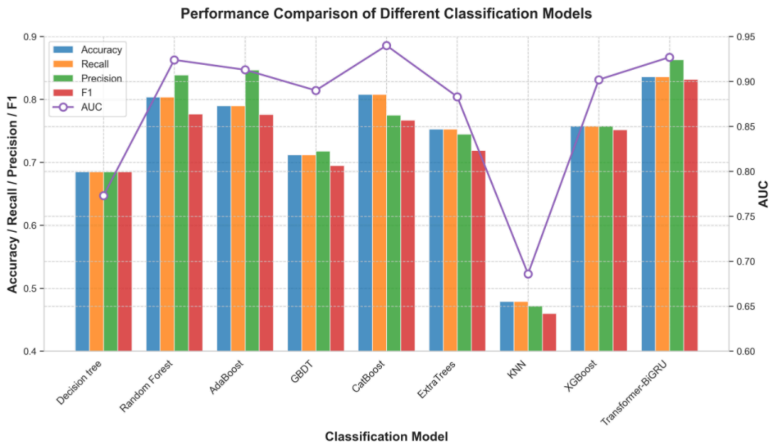

Figure 5. The bar comparison charts of each indicator

The experimental results show that the Transformer-BiGRU classification algorithm proposed in this paper has significant advantages in all core evaluation indexes: the accuracy, recall, accuracy and F1 value are 84%, 84%, 86% and 83%, respectively, which are superior to all comparative machine learning algorithms. Among them, the accuracy rate, recall rate and accuracy rate are 4, 4 and 9 percentage points higher than the second place CatBoost, respectively, and the F1 value is 5 percentage points higher than the second place Random Forest, which fully reflects the superiority of the fusion model in feature extraction and classification decision-making. In terms of AUC index, this model reaches 93%, slightly lower than CatBoost's 94%, but higher than Random Forest, AdaBoost, XGBoost and other algorithms, and has strong generalization and category discrimination ability. In the traditional algorithm, random forest and CatBoost perform better, with all indicators higher than 77% and AUC exceeding 92%; AdaBoost, XGBoost, etc. are at a medium level, with accuracy and F1 values between 75% – 79%; Decision tree and GBDT performed weakly, with indicators ranging from 69% to 72%; KNN performed the worst, with all indicators below 48% and an AUC of only 69%.

Secondly, the fault recovery time was predicted. The comparison results of each model are shown in Table 2, and the bar comparison charts of each indicator are shown in Figure 6.

DOI: 10.54254/2755-2721/2026.32843

Table 2. The results of the comparative experiment

Model	MSE	RMSE	MAE	MAPE	R^2
Decision tree	29.519	5.433	3.359	76.353	-0.121
Random Forest	22.336	4.726	3.218	58.419	0.301
AdaBoost	19.256	4.388	2.732	51.315	0.347
GBDT	23.096	4.806	2.997	58.183	0.376
CatBoost	25.814	5.081	3.125	64.166	0.235
ExtraTrees	26.503	5.148	3.361	57.782	0.289
KNN	36.724	6.06	4.319	86.211	0.024
XGBoost	25.78	5.077	3.315	53.753	0.343
Transformer-BiGRU	16.598	4.074	2.615	51.468	0.442

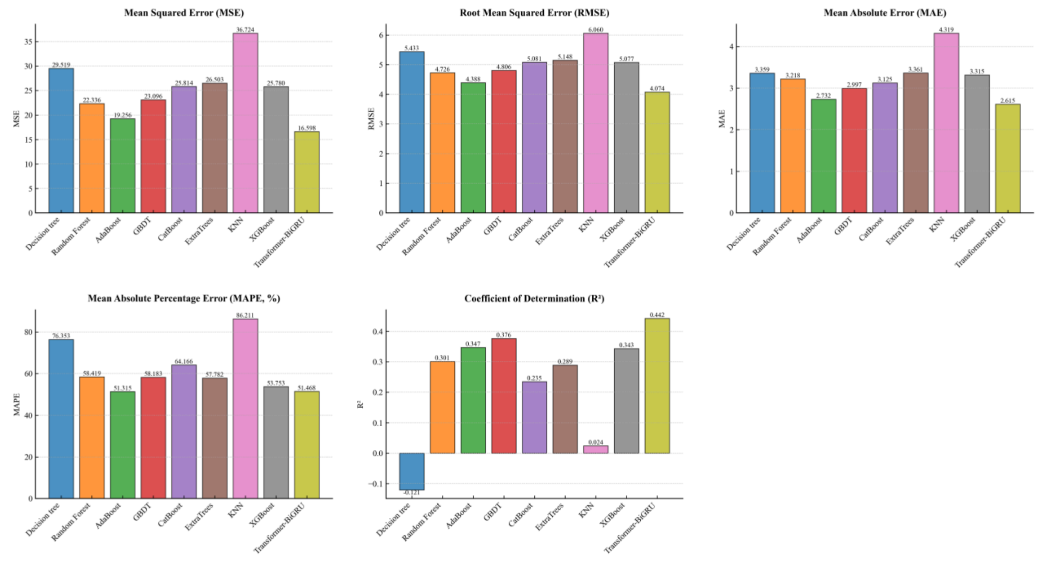

Figure 6. The bar comparison charts of each indicator

From the experimental results, the Transformer-BiGRU regression algorithm proposed in this paper shows significant advantages in all key evaluation indicators. Its MSE reaches 16.598, RMSE is 4.074, and MAE is as low as 2.615, all of which are the lowest levels among all comparison algorithms. It indicates that this algorithm performs the best in terms of prediction accuracy and error control. In terms of the MAPE indicator, although the 51.468 of Transformer-BiGRU is slightly higher than the 51.315 of AdaBoost, it is still in a relatively low range and much lower than the 76.353 of Decision tree and 86.211 of KNN, demonstrating strong predictive stability. In terms of model fitting ability, the R^2 value of Transformer-BiGRU reaches 0.442, which is significantly higher than 0.376 of GBDT, 0.347 of AdaBoost, 0.343 of XGBoost and 0.301 of Random Forest. It is superior to ExtraTrees' 0.289 CatBoost's 0.235 KNN's 0.024 and Decision tree's -0.121, indicating that this algorithm can capture the potential patterns in the data more fully.

DOI: 10.54254/2755-2721/2026.32843

5. Conclusion

To make up for the deficiencies of existing algorithms in long time series feature extraction and complex association mining, this paper proposes the Transformer-BiGRU classification algorithm and the Transformer-BiGRU regression algorithm, and first conducts correlation analysis and violin plot analysis. The classification algorithm has obvious advantages in core evaluation indicators, with an accuracy rate and recall rate of 84%, an precision rate of 86%, and an F1 value of 83%, both higher than other machine learning algorithms. The AUC reaches 93%, and it has strong generalization and category discrimination capabilities. The MSE, RMSE and MAE of the regression algorithm are all the lowest, the MAPE is in a relatively low range, the R^2 reaches 0.442, and the fitting ability far exceeds that of the traditional algorithm. This algorithm integrates the global dependency capture capability of Transformer with the temporal feature extraction advantages of BiGRU, offering superior performance. It provides a reliable solution for monitoring and predicting the stable operation of industrial Internet of Things devices and is of great significance for enhancing production efficiency and safety levels in related fields.

References

[1] Abbas, Ghulam, et al. "Cira-cyber intelligent risk assessment methodology for industrial internet of things based on machine learning." IEEE Access (2025).
[2] Alattas, Khalid A., and Abbas Mardani. "A novel extended Internet of things (IoT) Cybersecurity protection based on adaptive deep learning prediction for industrial manufacturing applications." Environment, Development and Sustainability 24.7 (2022): 9464-9480.
[3] Sun, Wen-Lin, Ying-Han Tang, and Yu-Lun Huang. "HiRAM: A hierarchical risk assessment model and its implementation for an industrial Internet of Things in the cloud." Software Testing, Verification and Reliability 33.5 (2023): e1847.
[4] Zolanvari, Maede, et al. "Machine learning-based network vulnerability analysis of industrial Internet of Things." IEEE internet of things journal 6.4 (2019): 6822-6834.
[5] Arico, Pietro, et al. "Passive BCI in operational environments: insights, recent advances, and future trends." IEEE Transactions on Biomedical Engineering 64.7 (2017): 1431-1436.
[6] Adaros Boye, Carolina, Paul Kearney, and Mark Josephs. "Cyber-risks in the industrial internet of things (IIoT): Towards a method for continuous assessment." International conference on information security. Cham: Springer International Publishing, 2018.
[7] Khalil, Ruhul Amin, et al. "Deep learning in the industrial internet of things: Potentials, challenges, and emerging applications." IEEE Internet of Things Journal 8.14 (2021): 11016-11040.
[8] Tsang, Yung Po, et al. "An Internet of Things (IoT)-based risk monitoring system for managing cold supply chain risks." Industrial Management & Data Systems 118.7 (2018): 1432-1462.
[9] Mouratidis, Haralambos, and Vasiliki Diamantopoulou. "A security analysis method for industrial Internet of Things." IEEE Transactions on Industrial Informatics 14.9 (2018): 4093-4100.
[10] Zhou, Sheng, et al. "Robust risk-sensitive task offloading for edge-enabled Industrial Internet of Things." IEEE Transactions on Consumer Electronics 70.1 (2023): 1403-1413.