Computing, Telecommunication and Control, 2026, Vol. 19, No. 2, Pp. 16–30.
Информатика, телекоммуникации и управление. 2026. Том 19, № 2. С. 16–30.

Research article
DOI: https://doi.org/10.18721/JCSTCS.19202
UDC 004.8

TFCA-GRU: A TEMPORAL AND FEATURE COOPERATIVE ATTENTION GRU FOR REMAINING USEFUL LIFE PREDICTION OF BEARINGS ACROSS CONDITIONS

C. Guo ✉, *V.V. Potekhin*

Peter the Great St. Petersburg Polytechnic University,
St. Petersburg, Russian Federation

✉ chenxiguo.academic@gmail.com

Abstract. Accurate prediction of the remaining useful life (RUL) of rolling bearings under different operating conditions is essential to enable proactive maintenance of industrial systems. To address the limitations of traditional recursive models in capturing complex time dependencies and distinguishing sensor contributions, we propose a temporal and feature cooperative attention-based gated recursive unit (TFCA-GRU). By integrating temporal and feature channel attention mechanisms, the model adaptively focuses on degradation-related time steps and key sensor features to build more robust RUL models from multidimensional time-frequency fusion data. Therefore, the TFCA-GRU enables sensitive extraction of critical features from long-term dependencies and local feature information. The validity of the proposed TFCA-GRU is verified using the bearing dataset from the XJTU-SY and the IEEE PHM 2012 Prognostic Challenge. This paper also provides the model's intrinsic interpretability through structured non-uniform attentional learning across time and feature dimensions. This reveals TFCA-GRU's selective attention to information time intervals and feature dimensions, which improves the transparency of the model. Experimental results show that the proposed TFCA-GRU performs better than some existing prediction methods in bearing RUL prediction across operating conditions.

Keywords: remaining useful life, deep learning, attention mechanism, gated recurrent unit, time-frequency feature fusion, model interpretability

Acknowledgements: The research was financially supported by the China Scholarship Council (CSC.202309810002).

Citation: Guo C., Potekhin V.V. TFCA-GRU: A temporal and feature cooperative attention GRU for remaining useful life prediction of bearings across conditions. Computing, Telecommunications and Control, 2026, Vol. 19, No. 2, Pp. 16–30. DOI: 10.18721/JCSTCS.19202

© Guo C., Potekhin V.V., 2026. Published by Peter the Great St. Petersburg Polytechnic University

Научная статья
DOI: https://doi.org/10.18721/JCSTCS.19202
УДК 004.8

TFCA-GRU: АЛГОРИТМ GRU С ВРЕМЕННЫМ И ХАРАКТЕРИСТИЧЕСКИМ КООПЕРАТИВНЫМ ВНИМАНИЕМ ДЛЯ ПРОГНОЗИРОВАНИЯ ОСТАТОЧНОГО СРОКА СЛУЖБЫ ПОДШИПНИКОВ В РАЗЛИЧНЫХ УСЛОВИЯХ

Ч. Го ✉ , *В.В. Потехин*

Санкт-Петербургский политехнический университет Петра Великого,
Санкт-Петербург, Российская Федерация

✉ chenxiguo.academic@gmail.com

Аннотация. Точное прогнозирование остаточного ресурса (Remaining Useful Life, RUL) подшипников качения в различных условиях эксплуатации является ключевым элементом для организации проактивного обслуживания промышленных систем. Для преодоления ограничений традиционных рекурсивных моделей, которые не всегда способны улавливать сложные временные зависимости и различать вклад отдельных датчиков, мы предлагаем рекуррентный блок с механизмом совместного внимания по временной и признаковой размерностям (TFCA-GRU). Механизм TFCA-GRU интегрирует внимание к временной последовательности и к каналам признаков, что позволяет модели адаптивно фокусироваться на ключевых этапах временной динамики, связанных с деградацией, а также на наиболее информативных сенсорных признаках. Это, в свою очередь, позволяет строить более устойчивые модели RUL на основе многомерных частотно-временных данных. Таким образом, TFCA-GRU обеспечивает более чувствительное извлечение критических признаков из долговременных зависимостей и локальной признаковой информации. Модель была верифицирована на наборах данных XJTU-SY и IEEE PHM 2012 Prognostic Challenge. В работе также подчеркивается внутренняя интерпретируемость модели, которая достигается за счет структурированного неравномерного обучения внимания по временной и признаковой размерностям. Это позволяет раскрыть избирательное внимание TFCA-GRU к информативным временным интервалам и признаковым размерностям, что в свою очередь повышает прозрачность модели. Экспериментальные результаты показывают, что предложенная модель TFCA-GRU превосходит некоторые существующие методы прогнозирования RUL подшипников в различных условиях эксплуатации, что делает ее перспективной для практического применения в промышленных системах.

Ключевые слова: остаточный ресурс, глубокое обучение, механизм внимания, рекуррентный блок, слияние временных и частотных характеристик, интерпретируемость модели

Финансирование: Исследование выполнено при финансовой поддержке Китайского государственного комитета по стипендиям (CSC.202309810002).

Для цитирования: Guo C., Potekhin V.V. TFCA-GRU: A temporal and feature cooperative attention GRU for remaining useful life prediction of bearings across conditions // Computing, Telecommunications and Control. 2026. Т. 19, № 2. С. 16–30. DOI: 10.18721/JCSTCS.19202

Introduction

Mechanical systems are the foundation of modern industry and accurate prediction of the Remaining Useful Life (RUL) of critical components is essential to ensure reliability and reduce maintenance costs [1, 2]. As a core element of Prognostics and Health Management (PHM), RUL prediction is particularly important for rolling bearings, which are prone to failure and widely used in rotating machinery. Effective RUL estimation enables timely maintenance, minimizes downtime and improves the efficiency and safety of the entire system [3].

© Го Ч., Потехин В.В., 2026. Издатель: Санкт-Петербургский политехнический университет Петра Великого

As the requirements for equipment reliability and operational safety continue to increase, various advanced techniques have been applied to the RUL prediction of rolling bearings. Existing methods are broadly categorized into model-based and data-driven approaches. Model-based approaches rely on physical or mathematical models to simulate degradation behavior, but their applicability in the real world is limited due to the complexity of the failure mechanisms and the need for extensive *a priori* knowledge. In contrast, data-driven approaches utilize sensor data and machine learning algorithms such as support vector regression, random forests and neural networks to learn degradation patterns without the need for explicit physical models. These methods offer greater adaptability and scalability for industrial applications, making them more suitable for complex nonlinear systems [4, 5].

With the rapid development of AI, Deep Learning (DL) techniques hold great promise in the field of PHM, especially in the RUL prediction of rolling bearing [6]. These methods, including Convolutional Neural Network (CNN), Recurrent Neural Network (RNN) and Gated Recurrent Unit (GRU), can learn degradation patterns directly from raw sensor signals without the need for manual feature extraction. Various strategies have been proposed to improve prediction accuracy and robustness [7]. For example, in [8], a multiscale variational Gaussian self-attention (MVGA) mechanism was developed, that is integrated with a Depth-Wise Sparse Attention Long Short-Term Memory (LSTD) to enhance temporal modelling and mitigate feature confounding through depth propagation. In [9], the focus is on the construction of effective Health Indicators (HIs) via an unsupervised approach by employing a multiscale convolutional LSTM auto-encoder that combines support vector regression and CNN-BiLSTM networks, resulting in improved RUL prediction without the need for manual preconditioning. In [10], the problem of uncertainty modelling and noisy signal interference with the Deep Residual Temporal-Spatial Network (DRTS-Net) was addressed, which integrates attention and soft thresholding to enhance interpretability and probabilistic health estimation.

GRU networks are able to capture long-term temporal dependencies while reducing computational complexity, and thus have become a widely adopted architecture in time series modelling. GRUs are more compact, converge faster, and have a smaller number of parameters than traditional LSTM networks, making them well suited for industrial applications that require real-time and resource-efficient RUL prediction. In recent years, various GRU-based models have been developed to improve the performance of bearing RUL estimation. For example, in [11], a Multiscale Efficient CNN (MSECNN) is proposed, incorporating bidirectional GRUs to exploit spatial and temporal degradation features. They implemented image-based feature extraction using the Gramian Angular Difference Field (GADF) transform to improve the prediction performance under stable conditions. Similarly, in [12], a multi-scale gated CNN-GRU architecture is introduced and used a transformer encoder to improve long-range dependency modelling. In [13], stacked GRUs integrating Dynamic Time Warping (DTW) based attentional mechanisms and Bayesian inference are employed to achieve prediction accuracy and uncertainty estimation. Together, these approaches demonstrate the versatility and potential of GRU-based frameworks for RUL prediction tasks.

Although effective, existing GRU-based models face several significant limitations. Firstly, many models rely on signal transformation or manual feature extraction (e.g., GADF coding or engineering health metrics), which can introduce information loss and increase preprocessing complexity. Secondly, while GRUs can capture temporal dependencies, they lack the intrinsic ability to distinguish which time steps or sensor channels are most informative about degradation progress. While attentional mechanisms have been employed to address this issue, most existing implementations deal with the time and feature dimensions separately or asymmetrically, failing to jointly model their interactions. Moreover, architectures that combine GRUs with complex modules such as transformers or Bayesian estimators typically incur significant computational overheads, which limits their utility in real-world deployments under different operating conditions.

To address these challenges, we propose a novel architecture – Temporal and Feature Cooperative Attention GRU (TFCA-GRU) – which integrates attention mechanisms along the time and feature axes into a unified, lightweight GRU-based framework. This design enables the model to dynamically emphasize critical time steps and degradation-sensitive sensor dimensions, thereby improving prediction accuracy and interpretability under different operating conditions without sacrificing computational efficiency. The main contributions of this work are as follows:

1. A novel time and feature cooperative attention-based GRU framework (TFCA-GRU) is proposed for predicting the RUL of bearings under different operating conditions. The model effectively captures long-range time dependencies and critical sensor interactions through joint attention modelling.

2. A Multi-Head Temporal Attention Mechanism (MTAM) is designed to dynamically highlight the time steps associated with degradation, thereby improving the model's ability to identify critical stages in the degradation process. In addition, a feature-focused module is introduced for adaptively re-weighting input features, thus allowing the model to focus on condition-sensitive signals while maintaining a lightweight structure.

3. Experiments conducted on the XJTU-SY and IEEE PHM 2012 Prognostic Challenge datasets demonstrate the effectiveness and generality of the proposed method. The model shows good prediction accuracy and robustness across operating conditions.

4. The interpretability of the proposed model is verified through attention visualization, which demonstrates its ability to selectively focus on the information time interval and feature dimensions, thus improving the transparency of the model.

Proposed methods

TFCA-GRU framework for bearing RUL

The overall system framework of rolling bearing RUL based on TFCA-GRU is shown in Fig. 1. The system contains four main components, which are: equipment data acquisition, data preprocessing, TFCA-GRU model, and RUL prediction.

GRUs

LSTM networks were introduced to alleviate the gradient vanishing and gradient explosion problems faced by traditional RNNs when processing long sequence-dependent information. LSTMs enhance the capture of long dependencies by designing forgetting gates, input gates, and output gates to selectively store and update time-series information. However, with the increasing demand for network efficiency and model simplicity, GRU has emerged as an improved version of LSTM.

GRU achieves efficient management of information flow by combining the forgetting and input gates of LSTM into a single update gate and introducing a reset gate. This improvement not only reduces the number of network parameters, but also reduces the model complexity and effectively avoids the risk of overfitting. Meanwhile, GRU maintains the sensitivity to information dependence on long series modelling, which further improves the performance of time series prediction.

The core structure of GRU consists of four main states:

1) reset gate r_t,
2) update gate z_t,
3) candidate hidden state $\tilde{h}_t$,
4) output hidden state h_t.

The parameter update equations of GRU model are as follows:

$$r_t = \sigma\left(W_r X_t + U_r h_{t-1} + b_r\right); \tag{1}$$

$$z_t = \sigma\left(W_z X_t + U_z h_{t-1} + b_z\right); \tag{2}$$

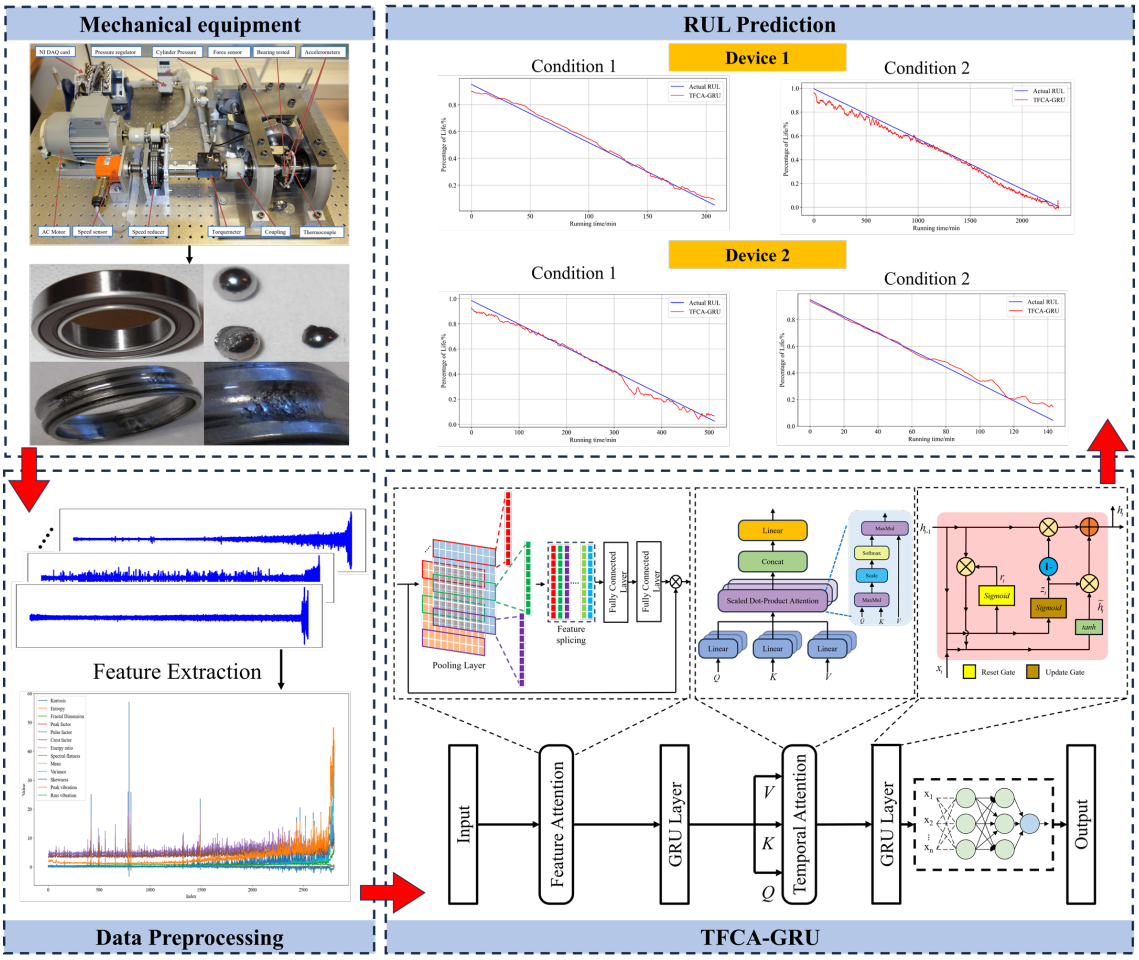


Fig. 1. System framework of the rolling bearing RUL based on TFCA-GRU

$$h_t^* = \tanh\left(WX_t + r_t \odot Uh_{t-1} + b\right); \tag{3}$$

$$h_t = \left(1 - z\right) \odot h_t^* + z_t \odot h_{t-1}, \tag{4}$$

where $\sigma \in (0,1)$ is the activation function, $\tanh \in (-1,1)$ is the hyperbolic tangent function, W and U are the weight matrices, b is the bias term, and $\odot$ is the element-level multiplication.

MTAM

MTAM aims to simulate human's ability to selectively pay attention to time-series information, so that the model can dynamically allocate attention weights, focus on important time steps, and enhance the ability to capture long-time dependencies. Traditional RNNs are limited to a fixed time window, making it difficult to adequately model long-term dependent information. With the introduction of the temporal attention mechanism, however, the model is no longer limited to sequential traversal, but models global associations for all time steps, which substantially improves the flexibility and accuracy of information capture.

The core idea of the temporal attention mechanism designed in this paper is to map the input sequence to three vectors: query, key and value, – and calculate the attention weights through the interaction between these three to determine the degree of attention of the model to different time steps [14].

Suppose the input time sequence is $X = [x_1, x_2, ..., x_T]$, where $x_t \in \mathbb{R}^d$ is the feature vector of the t-th time step, T is the length of the sequence, and d is the feature dimension. In the temporal attention mechanism, the input is first linearly mapped to generate a query vector, a key vector and a value vector. The transformation process of the three vectors is as follows:

$$Q = XW_Q + b_Q; \tag{5}$$

$$K = XW_K + b_K; \tag{6}$$

$$V = XW_V + b_V, \tag{7}$$

where W_Q, W_K and W_V are the weight matrices after linear mapping, b_Q, b_K and b_V are the bias vectors.

The model then calculates the similarity weights of the query and key for each time step using scaled dot product, and the attention weight matrix A is weighted and summed with the value vector V to obtain the final attention representation A:

$$A = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)\cdot V, \tag{8}$$

where A is the attention weight of each time step to the other time steps, and $\sqrt{d}$ is a scaling factor to prevent the gradient from disappearing.

In order to enhance the model's ability to capture multi-level attention subspace information, transformer introduces the multi-head attention mechanism. The advantage of the latter is that it allows model to parallelize the modeling of elements in the sequence from multiple perspectives, and different attention heads can focus on different feature patterns of the input sequence, thus mining the contextual information in a more comprehensive way. The computational process can be represented as follows:

$$\text{head}_j = \text{Attention}\left(QW_j^Q, KW_j^K, VW_j^V\right); \tag{9}$$

$$\text{MultiHead}\left(Q,K,V\right) = \text{Concat}\left(\text{head}_1, \ldots, \text{head}_h\right)W^H, \tag{10}$$

where $j = 1, 2, ..., h$ and j is the number of heads.

Feature attention mechanism

In multidimensional time series data, there are often important correlations and differences between multiple features. Traditional time series modelling approaches pay fixed attention to individual features, making it difficult to capture the important impact of key features on the forecasting task. In order to better portray these differences, this paper introduces the Feature Attention Mechanism (FAM), which aims to assign higher weights to important features to enhance the model's ability to perceive critical information.

Suppose the input time series is denoted as $X = [x_1, x_2, ..., x_T]$. For any time-step t, the FAM first performs average pooling along the time dimension for all time-steps to capture the global feature performance. The computational expression for this process is:

$$\overline{X} = \frac{1}{T}\sum_{t=1}^{T} X_t, \tag{11}$$

where $\overline{X} \in \mathbb{R}^{B \times F}$ is the global average representation of each feature over the whole time series. In this paper, we extract a total of 13 dimensional features, so F is 13. Then, the feature dimensions are compressed by nonlinear mapping and the intermediate feature Z is reconstructed using feedforward network to obtain the attention weight A. The computation process is as follows:

$$Z = \mathrm{ReLU}\left(W_1 \overline{X} + b_1\right); \tag{12}$$

$$A = \sigma\left(W_2 Z + b_2\right), \tag{13}$$

where W_1 and W_2 are the weight matrices, b_1 and b_2 are the bias terms, and σ is the Sigmoid activation function.

Finally, in order to be consistent with the dimension of the input vectors, the attention weights A are dimensionally extended and the attention weighting operation is performed with the original features, and the output of the FAM is computed as follows:

$$X' = X \odot A', \tag{14}$$

where A' is the expanded attention weight, $\odot$ is the elemental multiplication, and X' is the feature-weighted output.

Prediction process of rolling bearing RUL based on TFCA-GRU

Step 1: The vibration signal is collected by industrial equipment to obtain the original time series data.

Step 2: Based on the feature extraction methods in time and frequency domains, the vibration signals are analyzed and converted into multi-dimensional features to generate the time-frequency domain fusion feature matrix.

Step 3: The processed data set is divided into training set and test set, and the training set is input into the constructed TFCA-GRU network model for multiple iterations optimization, to gradually learn the deterioration law of the equipment and generate the final RUL prediction model.

Step 4: Input the test set into the trained model to complete the accurate prediction of the RUL of rolling bearings.

Experiments and results analysis

The model modelling and experimental tests in this paper are developed based on the PyTorch DL framework using Python version 3.11. The development environment is PyCharm 2023. The experimental hardware configuration is Intel Xeon(R) Platinum 8352 CPU and NVIDIA GeForce RTX 4090 GPU.

Dataset and data preprocessing

In this article, the IEEE PHM 2012 Prognostic Challenge dataset [15] and the XJTU-SY dataset [16] will be used for experimental testing. The working conditions of the dataset and the division of the dataset are shown in Table 1.

The IEEE PHM 2012 Prognostic Challenge dataset was collected from the PRONOSTIA platform, which has been widely used in many equipment residual life prediction studies. The platform consists of a rotary drive system, radial loading device and measurement system, and is capable of simulating the degradation process of rolling bearings under real operating conditions. The vibration signals are collected at a sampling frequency of 25.6 kHz, and data segments of 0.1 second duration are recorded every 10 seconds, which is equivalent to 2560 sampling points per segment. When the amplitude of the monitored vibration exceeds 20 G, the experiment is judged as bearing failure and terminated. In this paper, the training set is used in strict accordance with the official classification scheme, and the rest of the bearing data is used as the test set.

Table 1

Running conditions and experimental setup for the dataset

Datasets	Operating conditions	Training dataset	Test dataset
IEEE PHM 2012 Prognostic Challenge	1800 rpm, 4 kN	Bearing1_1 Bearing1_2	Bearing1_3, Bearing1_4, Bearing1_5, Bearing1_6, Bearing1_7
	1650 rpm, 4.2 kN	Bearing2_1 Bearing2_2	Bearing2_3, Bearing2_4, Bearing2_5, Bearing2_6, Bearing2_7
XJTU-SY	2100 rpm, 12 kN	Bearing1_1 Bearing1_2	Bearing1_3, Bearing1_4, Bearing1_5
	2250rpm, 11 kN	Bearing2_1 Bearing2_2	Bearing2_3, Bearing2_4, Bearing2_5

The XJTU-SY dataset is based on a highly integrated bearing life test platform, which includes a motor system for controlling the rotational speed, a hydraulic control device for loading, and an array of multi-channel vibration sensors for real-time acquisition of vibration signals during operation. The platform supports flexible adjustment of radial load and rotational speed to simulate bearing degradation under different operating conditions. The data is collected at a sampling frequency of 25.6 kHz, with complete samples taken every 1.28 seconds, each containing 32768 consecutive data points, providing high-resolution, long-period vibration sequences for subsequent failure modelling and remaining life prediction.

Furthermore, the 13 time-domain and frequency-domain features were extracted from the vibration signals of the dataset species. The time-domain features included kurtosis, mean, variance, skewness, peak vibration, root-mean-square vibration, pulse factor, crest factor, and peak factor, while the frequency-domain features covered entropy, fractal dimension, energy ratio, and spectral flatness.

Assessment metrics and comparison algorithms

In the field of RUL prediction, the following three methods are commonly used to assess the prediction performance of the model [17, 18]:

1) Root Mean Square Error (RMSE),
2) Mean Absolute Error (MAE),
3) scoring function (Score) and R-Square (R^2) [19].

The three assessments are calculated as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{y_i} - y_i\right)^2}; \tag{15}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|\widehat{y_i} - y_i\right|; \tag{16}$$

$$\mathrm{Score} = \begin{cases} \sum_{i=1}^{N_2}\left(e^{-\frac{R^i_{\mathrm{pre}} - R^i_{\mathrm{true}}}{13}} - 1\right), & R^i_{\mathrm{pre}} < R^i_{\mathrm{true}} \\ \sum_{i=1}^{N_2}\left(e^{-\frac{R^i_{\mathrm{pre}} - R^i_{\mathrm{true}}}{10}} - 1\right), & R^i_{\mathrm{pre}} \geq R^i_{\mathrm{true}} \end{cases}; \tag{17}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(\widehat{y_i} - y_i\right)^2}{\sum_{i=1}^{n}\left(\overline{y_i} - y_i\right)^2}, \tag{18}$$

where n is the number of samples, y_i and $\widehat{y_i}$ are the actual and predicted RUL values at the i-th moment, $\overline{y}$ is the actual RUL mean, and N_2 is the total number of test samples. Generally, lower values of RMSE and MAE indicate better performance of the model. The closer the value of R^2 is to 1 indicates higher prediction accuracy of the model.

In order to demonstrate the superiority of the proposed TFCA-GRU, five representative baseline RUL prediction methods are used in this paper to compare the RUL prediction effectiveness, as well as to compare the evaluation data with the current novel RUL prediction methods. The baseline methods include CNN, LSTM, CNN-LSTM, CNN-Transformer, and the novel methods include MSCNN [20], TCN-SA [21], DCTCN [22], DSCN [23].

Experiment 1: PHM 2012 dataset

The RUL prediction results of the proposed TFCA-GRU with the baseline model are illustrated in Fig. 2. Firstly, from the prediction curves, the proposed model can track the actual RUL degradation trend stably under all four operating conditions, and shows good dynamic response capability in both the initial and termination stages of degradation. In contrast, the baseline approach exhibits different degrees of drift and prediction lag under several scenarios, which is difficult to accurately reflect the system degradation evolution. In the box-and-line plot, TPCA-GRU shows the smallest median prediction error and small distribution variance in all scenarios, and the fewest outliers, which reflects its advantages of prediction stability and robustness.

The evaluation data of the proposed TFCA-GRU with the novel models on the PHM 2012 dataset are shown in Table 2. From the overall trend, TFCA-GRU significantly outperforms the other four models in terms of MAE and RMSE metrics under all data scenarios, showing strong consistency and robustness. In addition, a side-by-side comparison between the models shows that although the SCTCN, DSCN and TCN-SA models perform the best MAE and RMSE metrics in multiple scenarios, they exhibit unstable prediction and large error fluctuations in cross-case conditions.

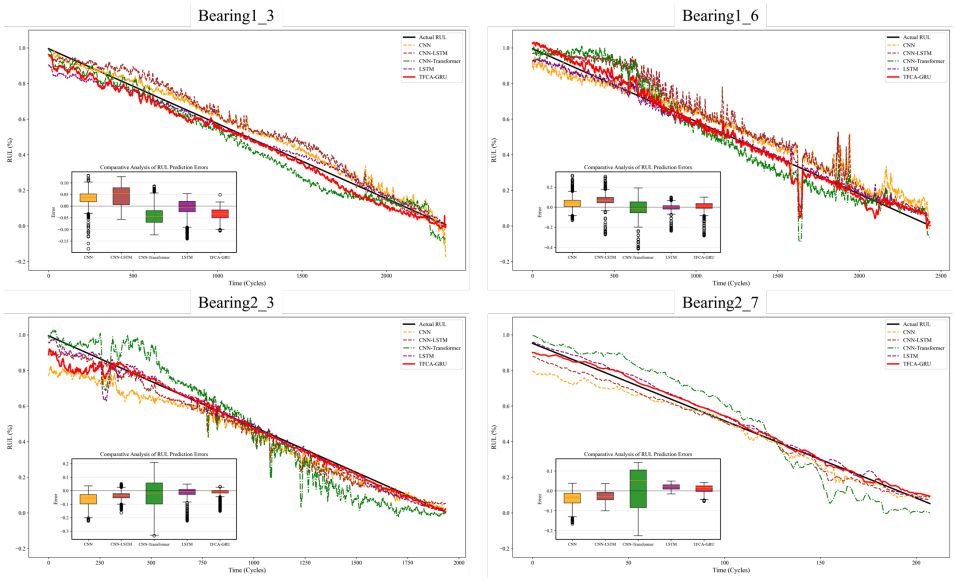


Fig. 2. Comparison of the prediction results for the four working conditions in the PHM 2012 dataset

Table 2

Performance comparison on PHM 2012 dataset

	MSCNN		TCN-SA		SCTCN		DSCN		TFCA-GRU	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
1–3	0.078	0.096	0.066	0.085	0.098	0.121	0.071	0.085	**0.035**	**0.041**
1–4	**0.015**	0.180	0.103	0.139	0.130	0.149	0.084	0.108	0.037	**0.045**
1–5	0.089	0.151	0.086	0.126	0.163	0.222	0.087	0.146	**0.051**	**0.040**
1–6	0.105	0.149	0.084	0.116	0.178	0.221	0.084	0.112	**0.030**	**0.042**
1–7	0.106	0.147	0.089	0.125	0.084	0.109	0.089	0.121	**0.044**	**0.057**
2–3	0.269	0.325	0.191	0.232	0.110	0.216	0.191	0.238	**0.024**	**0.042**
2–4	0.185	0.226	0.091	0.129	0.117	0.145	0.257	0.298	**0.021**	**0.030**
2–5	0.137	0.176	0.192	0.222	0.189	0.221	0.131	0.171	**0.034**	**0.074**
2–6	0.103	0.126	0.08	0.098	0.092	0.115	0.101	0.130	**0.031**	**0.047**
2–7	0.241	0.351	0.196	0.230	0.167	0.195	0.216	0.255	**0.019**	**0.022**

Experiment 2: XJTU-SY dataset

The RUL prediction result plots of the proposed TFCA-GRU with the baseline model are illustrated in Fig. 3. From the time-series prediction curves, traditional deep models such as CNN and LSTM have obvious lagging or advancing phenomena, especially in the early and late stages of degradation, with the largest prediction deviation. The proposed model is able to closely track the degradation trend of the real RUL decreasing over time, especially in Bearing1_3 and Bearing2_5, where its prediction curves almost overlap with the real curves, indicating its strong ability to capture the degradation features. As seen from the boxplots, the proposed model shows the smallest median error, narrower error intervals, and very few outliers in all four subplots, which indicates that its predictions are not only accurate but also stable.

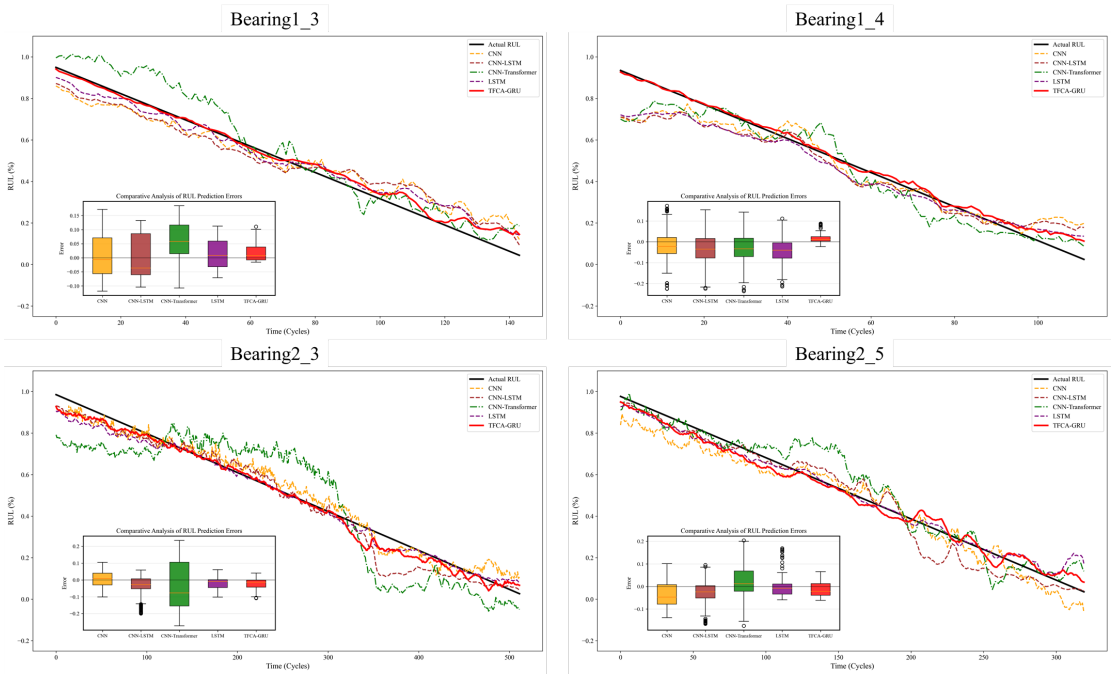


Fig. 3. Comparison of the prediction results of the four operating conditions on XJTU-SY dataset

The evaluation data of the proposed TFCA-GRU and the new model on the XJTU-SY dataset are presented in Table 3. It can be seen that the MAE and RMSE values of TFCA-GRU are significantly better than the other models in both working conditions. Compared with the other models, TCN-SA and DSCN, although better in some indicators, are difficult to maintain a stable advantage under multiple operating conditions.

Table 3

Performance comparison on XJTU-SY dataset

	MSCNN		TCN-SA		SCTCN		DSCN		TFCA-GRU	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
1–3	0.227	0.262	0.120	0.139	0.114	0.133	0.092	**0.104**	**0.083**	0.110
2–3	0.086	0.104	0.092	0.118	0.160	0.191	0.094	0.118	**0.063**	**0.086**

Ablation experiment

The ablation models are set up as GRU, TA-GRU with the introduction of temporal attention, FA-GRU with the introduction of feature attention, and TFCA-GRU.

The prediction curves of the ablation model in XJTU-SY Bearing2_5 are illustrated in Fig. 4. It can be seen that the prediction curve of TFCA-GRU basically coincides with the real RUL, which fully demonstrates the effectiveness of the joint time-feature modelling. Compared with other models, TFCA-GRU is able to extract the key dependencies in the degraded signals more comprehensively, thus achieving higher accuracy and more robust life prediction.

The comparison of the comprehensive evaluation of the models is demonstrated in Table 4. GRU performs relatively poorly on all metrics, showing its limited predictive performance when no attention mechanism is introduced. In contrast, TFCA-GRU has the best overall performance on all metrics. Therefore, this justifies TFCA-GRU.

Table 4

Performance comparison of ablation models

Ablation models	R2	RMSE	MAE	Score
GRU	0.954	0.058	0.148	1.404
TA-GRU	0.987	0.037	0.033	0.885
FA-GRU	0.971	0.046	0.037	1.109
TFCA-GRU	**0.991**	**0.031**	**0.026**	**0.779**

Model interpretability analysis

First, we compare and analyze the effect of its temporal attention on the model's time-awareness ability, and Fig. 5 demonstrates the heat map of attention in the temporal dimension. TFCA-GRU presents obvious non-uniformity, and the model is able to dynamically distribute the degree of attention at different time steps, which significantly enhances the ability of identifying the key degradation phases; whereas the GRU model has a more uniform distribution of attention, and lacks the ability of efficiently capturing the important timing information is effectively captured.

Meanwhile, Fig. 6 demonstrates the distribution of the model's attentional weights on feature dimensions. GRU has a relatively even distribution of weights across feature dimensions, and the model lacks differentiation of the importance of input features. TFCA-GRU attentional distributions show significant differences, with some feature dimensions (e.g., indexes 0, 10, etc.) obtaining higher

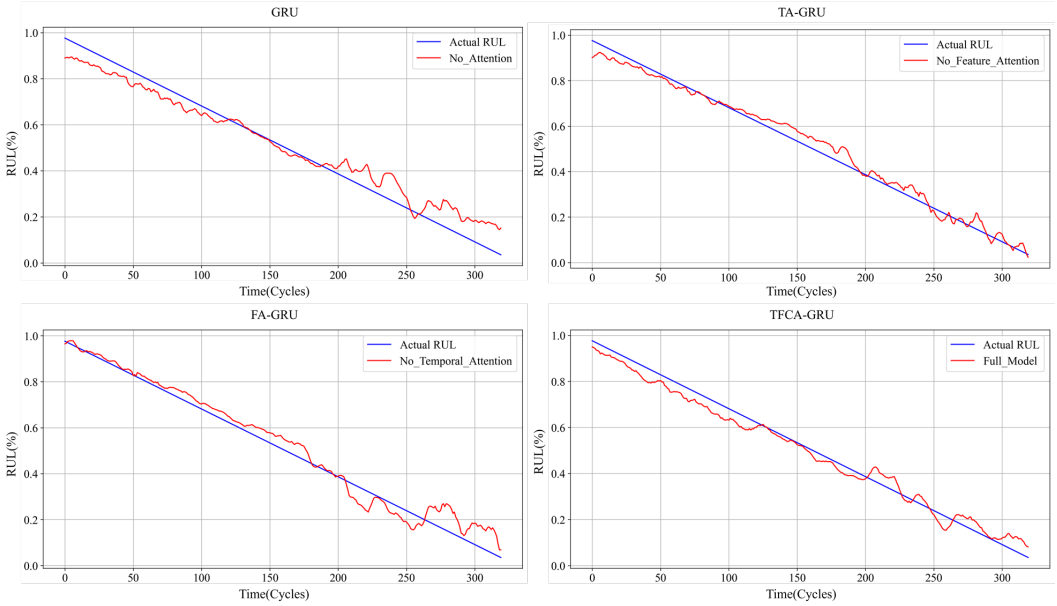


Fig. 4. Comparison of predictions from ablation models

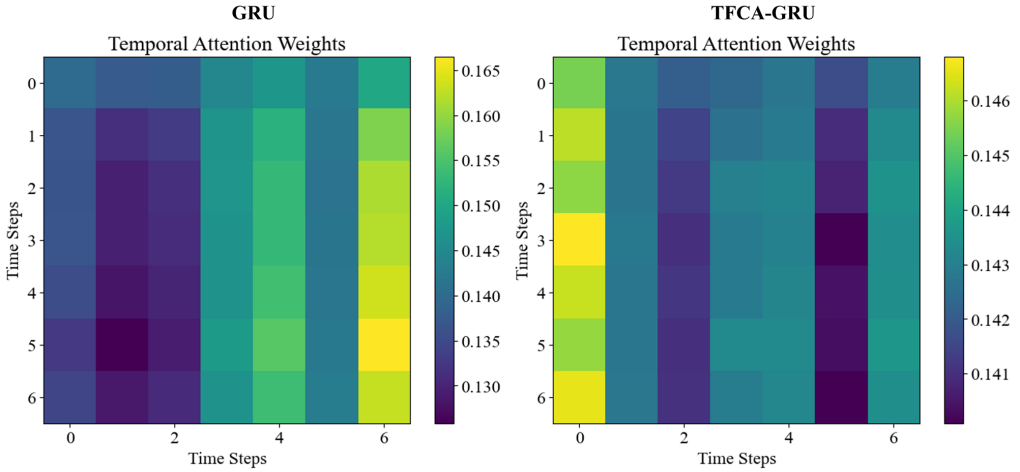


Fig. 5. Heat map of temporal attention

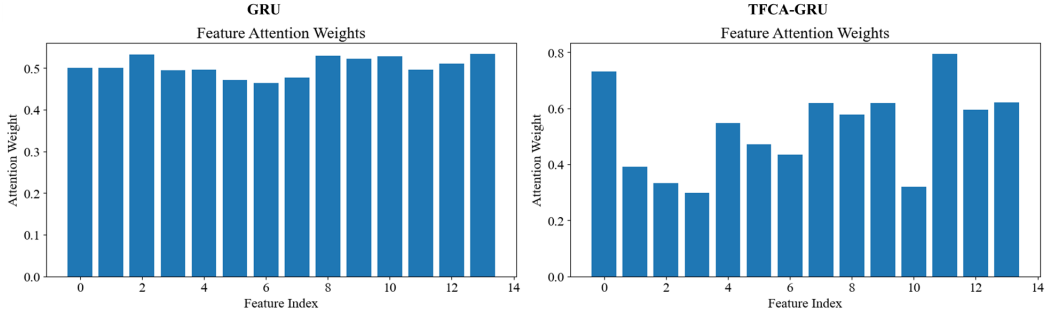


Fig. 6. Distribution of feature attention weights

weights, suggesting that the model is capable of adaptively identifying and reinforcing the features that are most relevant to the trend of degradation.

In this study, we proposed TFCA-GRU, a temporal and feature cooperative attention-based gated recurrent unit framework for the prediction of the rolling bearings RUL under variable operating conditions. The model addresses two core challenges in industrial prognostics: the difficulty of capturing long-term temporal dependencies and the inability of conventional models to distinguish between informative and irrelevant sensor features.

Conclusions and further work

By incorporating MTAM and FAM into the GRU architecture, TFCA-GRU is capable of dynamically focusing on degradation-critical time steps and sensor dimensions. This joint attention design enhances the model's ability to learn representative patterns from multidimensional time-frequency fusion data while preserving computational efficiency. Unlike traditional GRU-based or hybrid deep models that rely heavily on handcrafted preprocessing or suffer from scalability issues, TFCA-GRU offers a unified and interpretable solution.

Comprehensive experiments on two public datasets – IEEE PHM 2012 Prognostic Challenge and XJTU-SY – demonstrated that TFCA-GRU consistently outperforms baseline and state-of-the-art models in terms of prediction accuracy, robustness, and generalization across working conditions. Ablation studies confirmed the effectiveness of each attention component, and visualization of attention distributions further illustrated the model's capacity to highlight key temporal intervals and dominant features, thereby improving transparency and trustworthiness in critical applications.

Overall, TFCA-GRU provides a practical and extensible framework for data-driven RUL prediction under complex and non-stationary degradation scenarios. Future research will focus on further reducing the model's computational overhead, exploring domain adaptation techniques to improve cross-machine transferability, and extending the framework to other types of mechanical components and multi-sensor fusion settings.

REFERENCES

1. **Sai Van Cuong, Shcherbakov M.V.** Architecture of predictive maintenance system of complex multiobject systems in Industry 4.0 concept. *Software & Systems*, 2020, Vol. 33, No. 2, Pp. 186–194. DOI: 10.15827/0236-235X.130.186-194

2. **Cui L., Shen Q., Xiao Y., Liu D., Wang H.** Sparse graph structure fusion convolutional network for machinery remaining useful life prediction. *Reliability Engineering & System Safety*, 2025, Vol. 254A, Art. no. 110592. DOI: 10.1016/j.ress.2024.110592

3. **Dli M.I., Puchkov A.Yu., Lobaneva E.I.** Method for estimating the time of useful use of equipment based on neural networks. *Bulletin of the Saint Petersburg State Institute of Technology* (*Technical University*), 2021, Vol. 85, No. 59, Pp. 107–112. DOI: 10.36807/1998-9849-2021-59-85-107-112

4. **Zhang J., Tian J., Yan P., Wu S., Luo H., Yin S.** Multi-hop graph pooling adversarial network for cross-domain remaining useful life prediction: A distributed federated learning perspective. *Reliability Engineering & System Safety*, 2024, Vol. 244, Art. no. 109950. DOI: 10.1016/j.ress.2024.109950

5. **Xu Z., Li C., Yang Y.** Fault diagnosis of rolling bearings using an Improved Multi-Scale Convolutional Neural Network with Feature Attention mechanism. *ISA Transactions*, 2021, Vol. 110, Pp. 379–393. DOI: 10.1016/j.isatra.2020.10.054

6. **Liu H., Sun Y., Ding W., Wu H., Zhang H.** Enhancing non-stationary feature learning for remaining useful life prediction of aero-engine under multiple operating conditions. *Measurement*, 2024, Vol. 227, Art. no. 114242. DOI: 10.1016/j.measurement.2024.114242

7. **Yin Y., Tian J., Liu X.** Remaining useful life prediction based on parallel multi-scale feature fusion network. *Journal of Intelligent Manufacturing*, 2024, Vol. 36, Pp. 3111–3127. DOI: 10.1007/s10845-024-02399-y

8. **Fan W., Chang Z., Xu S., Fan Z., Yuan Y.** RUL Prediction of Rolling Bearings with MVGA-DSAL: A Multiscale Variant Gaussian Attention Model with Depth-Wise Sparse Attention LSTM. *IEEE Transactions on Instrumentation and Measurement*, 2025, Vol. 74, Art no. 3523111. DOI: 10.1109/TIM.2025.3551998

9. **Qu Y., Fu S., Yong M., Tian J., Lv Z., Li R.** Health Indicator Construction and Remaining Useful Life Prediction Based on MSC-LSTM-AE Model for Working Bearings. *IEEE Sensors Journal*, 2025, Vol. 25, No. 9, Pp. 15525–15535. DOI: 10.1109/JSEN.2025.3548675

10. **Yang T., Li G., Huo J., Li X., Zhou X., Han Q.** The DRTS-Net Framework: Enhanced Probabilistic Health Perception for Wind Turbine Bearings Under Complex Operating Conditions with Improved Interpretability. *IEEE Transactions on Instrumentation and Measurement*, 2025, Vol. 74, Art no. 3517412. DOI: 10.1109/TIM.2025.3545504

11. **Ma P., Li G., Zhang H., Wang C., Li X.** Prediction of Remaining Useful Life of Rolling Bearings Based on Multiscale Efficient Channel Attention CNN and Bidirectional GRU. *IEEE Transactions on Instrumentation and Measurement*, 2024, Vol. 73, Art no. 2508413. DOI: 10.1109/TIM.2023.3347787

12. **Guo D., Cao Z., Fu H., Li Z.** Remaining Useful Life Estimation for Rolling Bearings Using MSGCNN-TR. *IEEE Sensors Journal*, 2022, Vol. 22, No. 24, Pp. 24333–24343. DOI: 10.1109/JSEN.2022.3221753

13. **Que Z., Jin X., Xu Z.** Remaining Useful Life Prediction for Bearings Based on a Gated Recurrent Unit. *IEEE Transactions on Instrumentation and Measurement*, 2021, Vol. 70, Art no. 3511411. DOI: 10.1109/TIM.2021.3054025

14. **Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser L., Polosukhin I.** Attention is all you need. *arXiv:1706.03762*, 2017. DOI: 10.48550/arXiv.1706.03762

15. **Nectoux P., Gouriveau R., Medjaher K., Ramasso E., Chebel-Morello B. et al.** PRONOSTIA: An experimental platform for bearings accelerated degradation tests. *IEEE International Conference on Prognostics and Health Management* (*PHM'12*), 2012, pp. 1–8.

16. **Wang B., Lei Y., Li N., Li N.** A hybrid prognostics approach for estimating remaining useful life of rolling element bearings. *IEEE Transactions on Reliability*, 2020, Vol. 69, No. 1, Pp. 401–412. DOI: 10.1109/TR.2018.2882682

17. **Wang T., Liu H., Guo D., Sun X.-M.** Continual residual reservoir computing for remaining useful life prediction. *IEEE Transactions on Industrial Informatics*, 2024, Vol. 20, No. 1, Pp. 931–940. DOI: 10.1109/TII.2023.3271661

18. **Qin Y., Chen D., Xiang S., Zhu C.** Gated Dual Attention Unit Neural Networks for Remaining Useful Life Prediction of Rolling Bearings. *IEEE Transactions on Industrial Informatics*, 2021, Vol. 17, No. 9, Pp. 6438–6447. DOI: 10.1109/TII.2020.2999442

19. **Babu G.S., Zhao P., Li X.-L.** Deep Convolutional Neural Network Based Regression Approach for Estimation of Remaining Useful Life. *Database Systems for Advanced Applications* (*DASFAA 2016*), 2016, Vol. 9642, Pp. 214–228. DOI: 10.1007/978-3-319-32025-0_14

20. **Zhu J., Chen N., Peng W.** Estimation of Bearing Remaining Useful Life Based on Multiscale Convolutional Neural Network. *IEEE Transactions on Industrial Electronics*, 2019, Vol. 66, No. 4, Pp. 3208–3216. DOI: 10.1109/TIE.2018.2844856

21. **Wang Y., Deng L., Zheng L., Gao R.X.** Temporal convolutional network with soft thresholding and attention mechanism for machinery prognostics. *Journal of Manufacturing Systems*, 2021, Vol. 60, Pp. 512–526. DOI: 10.1016/j.jmsy.2021.07.008

22. **Zou H., Zhu J.** DCTCN: Deep Complex Temporal Convolutional Network for Real Time Speech Enhancement. *2021 11th International Conference on Intelligent Control and Information Processing* (*ICICIP*), 2021, Pp. 112–118. DOI: 10.1109/ICICIP53388.2021.9642159

23. **Hu K., Yan A.** Multitarget Robust Deep Stochastic Configuration Network Parameter Modeling Method. *2024 6th International Conference on Industrial Artificial Intelligence* (*IAI*), 2024, Pp. 1–5. DOI: 10.1109/IAI63275.2024.10729968

INFORMATION ABOUT AUTHORS / СВЕДЕНИЯ ОБ АВТОРАХ

Guo Chenxi
Го Чэньси
E-mail: chenxiguo.academic@gmail.com
ORCID: https://orcid.org/0000-0002-1497-2893

Vyacheslav V. Potekhin
Потехин Вячеслав Витальевич
E-mail: Slava.Potekhin@spbstu.ru
ORCID: https://orcid.org/0000-0001-9850-9558

Submitted: 24.05.2025; Approved: 25.05.2026; Accepted: 03.06.2026.
Поступила: 24.05.2025; Одобрена: 25.05.2026; Принята: 03.06.2026.