ISSN 1213–2489
e-ISSN 2787–9402

DOI: 10.21062/mft.2025.051 http://www.journalmt.com

Research on Stator Thermal Fault Detection of Steam Turbine Generator Based on Improved Transformer and Gaussian Mixture Model

Fengqin Cheng (0009-0003-3824-2945), Weinan Zheng (0009-0007-2215-2610)*
College of Electrical and Mechanical Engineering, Jilin University of Architecture and Technology, Jilin, 130114, China
*zhengwan152@163.com

This study proposes a multi-stage intelligent diagnostic approach integrating Physics-Guided Normalization (LPGN), enhanced Transformer networks, and Gaussian Mixture Models (GMM) for thermal fault detection in turbine generator stators. The methodology sequentially performs the following steps: (1) enhances localized anomaly features in temperature data through LPGN, (2) efficiently extracts temporal patterns via the optimized Transformer architecture, and (3) achieves unsupervised fault classification using GMM. Experimental results demonstrate the proposed method's superiority over conventional ARIMA and LSTM models across multiple evaluation metrics, exhibiting a lower RMSE and a higher detection accuracy. Ablation studies further validate the individual contributions of each component to performance improvement. This solution provides an efficient and reliable framework for intelligent thermal monitoring in large rotating electrical machinery.

Keywords: Stator Thermal Fault, Transformer, Gaussian Mixture Model, Fault Diagnosis

1 Introduction

In a steam turbine generator, the rotor plays a critical role. Connected to the steam turbine, it rotates at high speed under the turbine's drive, generating a powerful rotating magnetic field in its windings [1]. This magnetic field interacts with the stator windings and, based on the principle of electromagnetic induction, induces alternating current in the stator. Thus, the rotor is one of the core components responsible for energy conversion in the generator, and its performance directly impacts the generator's efficiency and stability [2-4].

Thermal fault detection in turbine generator stators is of great significance for ensuring the long-term safe and stable operation of power units. As the primary power-generating component of a generator, the stator's windings and core endure substantial thermal loads during operation [5]. Undetected issues such as cooling system failures, insulation aging, or localized overloads can easily lead to overheating faults, potentially resulting in severe consequences such as insulation breakdown or equipment burnout.

However, stator thermal fault detection presents several challenges: the concealed nature of fault locations, uneven heat distribution, and complex operating environments make it difficult for conventional detection methods to achieve precise localization and real-time early warning. Therefore, developing highly sensitive and high-resolution temperature monitoring technologies, along with intelligent diagnostic methods, is key to improving the accuracy and efficiency of stator thermal fault detection [6-7].

In summary, existing thermal fault detection methods for turbine generator stators still exhibit limitations in handling complex temporal relationships, identifying localized temperature rise anomalies, and modeling feature distributions, making it challenging to meet the requirements for high-precision and robust fault early-warning under actual operating conditions. To address these challenges, this study proposes a multi-stage intelligent detection approach that integrates physics-guided normalization, an enhanced Transformer network architecture, and Gaussian Mixture Models. This integrated methodology aims to comprehensively improve the model's capabilities in perceiving, modeling, and identifying stator thermal faults, thereby achieving more accurate and stable thermal anomaly prediction and diagnosis.

2 Related Work

Currently, thermal fault detection in turbine generator stators has become an important research direction in power equipment condition monitoring and intelligent operation and maintenance [8-9]. Research efforts primarily focus on temperature sensing technologies, thermal field modeling and simulation, fault feature extraction, and intelligent diagnostic methods. Conventional techniques such as thermistors, thermocouples, and infrared imaging have been widely adopted for field monitoring, yet they still exhibit certain limitations in real-time performance, resolution, and anti-interference capability [10-12].

In recent years, emerging technologies like Fiber Bragg Grating (FBG) sensing and distributed temperature sensing (DTS) have gained traction, offering higher sensitivity and spatially distributed monitoring capabilities [13]. Meanwhile, the integration of artificial intelligence and big data analytics has enabled intelligent fault identification and trend prediction, making it a key research focus.

Eyha et al. [14] propose a non-invasive diagnostic method for detecting inter-turn short circuit (ITSC) and dynamic eccentricity (DE) faults in excitation windings. This method, based on advanced analysis techniques of stray magnetic field signals, can be implemented during generator operation, enabling accurate monitoring of the rotor excitation winding condition. It also generates characteristic diagnostic maps that facilitate the identification of potential ITSC faults. Michalski et al. [15] develop a hybrid fault detection and diagnosis approach that incorporates unsupervised machine learning techniques, aiming to identify faults before generator shutdown. By analyzing historical operational data from hydropower stations, the method reveals patterns in fault development, thereby aiding in understanding the evolution of fault mechanisms and their potential consequences. El Idrissi et al. [16] propose a non-destructive testing (NDT) method for diagnosing stator imbalance defects in induction motors (IM), based on thermal imaging and machine learning. This technique evaluates the thermal distribution state within regions of interest (ROI) by analyzing the temperature distribution on the stator's outer surface, combined with multiple statistical indicators, thus providing criteria to determine whether the thermal distribution is abnormal. Stone et al. [17] propose a health monitoring and fault identification method for thermal degradation effects in the insulation systems of form-wound stator coils. Through periodic off-line inspections of these coils, the researchers analyzed thermal aging data of the insulation materials up to the point of failure. Attallah et al. [18] develop an online non-invasive diagnostic approach for three-phase induction machines (IMs) based on infrared thermography. This methodology integrates three convolutional neural network (CNN)-based deep learning architectures (Inception, Xception, and MobileNet) to simultaneously identify the (1) IM operational status, (2) fault types, and (3) the location and severity of inter-turn faults (ITFs).

Despite significant progress, existing methods still face challenges in achieving high-precision detection and early warning under complex operating conditions. To address these limitations, this paper proposes a stator thermal fault detection model for steam turbine generators, which integrates an improved Transformer architecture with a Gaussian mixture model to enhance feature representation and anomaly discrimination.

3 Materials and Methods

3.1 Locally Physics-Guided Normalization

A Locally Physics-Guided Normalization (LPGN) method integrating physical priors with local trend enhancement is proposed for preprocessing thermal fault detection data in turbine generator stators, effectively improving subsequent Transformer model learning. While conventional normalization methods (e.g., Min-Max or Z-score) can standardize data distributions, they fail to account for physical characteristics of temperature fields, such as: (1) varying thermal inertia across stator regions, (2) thermal diffusion coupling effects, and (3) potential masking of anomalies by global averaging. The proposed LPGN method synergistically combines local trend enhancement with thermal-physical structural guidance, enabling heightened model sensitivity to subtle yet critical temperature rise anomalies.

Given the original temperature sequence T = $\{T_1, T_2,..., T_n\}$ with corresponding spatial positions X = $\{x_1, x_2,..., x_n\}$, the normalization method is defined as follows:

$$T_i^{(norm)} = \frac{T_i - \mu_i^{(local)}}{\sigma_i^{(local)} + \alpha \cdot D_i + \beta} \tag{1}$$

$$\mu_i^{(local)} = \frac{1}{|N(i)|} \sum j \in N(i) \tag{2}$$

$$\sigma_i^{(local)} = \sqrt{\frac{1}{|N(i)|} \sum_{j \in N(i)} (T_j - \mu_i^{(local)})^2} \tag{3}$$

Where:

T_i…The local sliding window mean centered at point i;

σ…The corresponding local standard deviation;

D_i…The estimated thermal diffusion coefficient at that location (reflecting thermal stability);

α, β…Tunable hyperparameters (α=0.1, β=1e-5);

N(i)…The neighborhood set of point i (spatially topologically connected).

Compared to global means, local means exhibit greater sensitivity to abrupt anomalies, the incorporation of thermal diffusion coefficient D_i directs network attention to thermally unstable, failure-prone regions, the β parameter controls amplification effects under minimal standard deviation to maintain numerical stability. This approach is well-suited for spatio-temporal temperature field inputs and enhances dynamic feature extraction in Transformer-based models.

3.2 Improved Transformer model

In recent years, large models with Transformer networks as their core architecture have garnered significant attention due to their outstanding creativity and logical reasoning capabilities. The Transformer network introduces a self-attention mechanism that enables more effective and efficient learning of interrelationships within input sequences [19]. Additionally, the parallel processing of positional information in Transformer networks substantially improves training speed and efficiency. To enhance the prediction accuracy of braking torque in the Transformer encoder, we have improved the encoder component by designing a parallel module based on a gated mechanism that combines self-attention and convolutional neural networks. This module separately extracts global features through the self-attention mechanism and local features via convolutional neural networks. Figure 1 illustrates the improved Transformer structure with a feature fusion gate.

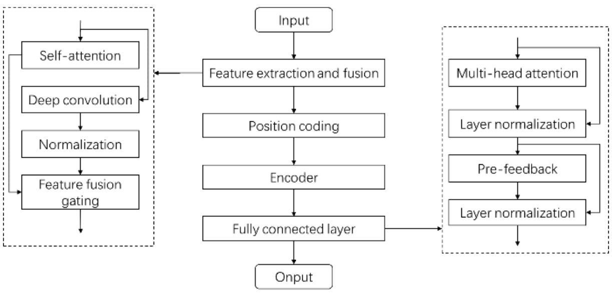

Fig. 1 *Improve the Transformer structure diagram*

Simply incorporating global and local features proves insufficient to enhance the Transformer's prediction accuracy. To address this limitation, we introduce a gating mechanism and propose an adaptive feature fusion approach tailored for multi-dimensional characteristics, whose architecture is depicted in Figure 2.

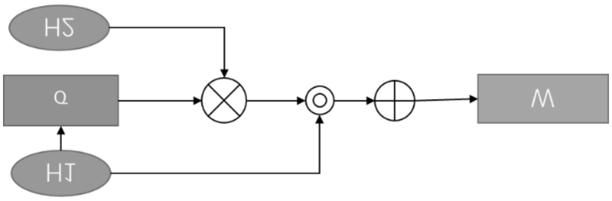

Fig. 2 *Feature fusion gating structure diagram*

The fusion process consists of three key stages: First, during the weight computation phase, gate control weights are generated by applying Sigmoid activation to input feature h_1, which quantitatively evaluate the importance of another feature h_2. Subsequently, in the adaptive fusion phase, these weights undergo element-wise multiplication with h_2 before being added to h_1, achieving dynamic feature integration. Finally, the fused features are transformed through a linear mapping layer to produce the final output [20]. This architecture effectively regulates the relative influence of different features during fusion, thereby enhancing the model's capability to extract critical information.

3.3 Gaussian Mixture Model

By modeling the temporal features extracted by the Transformer, the Gaussian Mixture Model (GMM) can represent them as a mixed probability model composed of multiple Gaussian distributions, thereby effectively distinguishing between normal operating states and potential thermal fault characteristics. This approach requires minimal prior labeling, making it suitable for unsupervised or semi-supervised scenarios, and can autonomously learn the distribution structure within the feature space, demonstrating high sensitivity to subtle and concealed thermal anomalies. By combining the powerful feature extraction capabilities of the Transformer with the probabilistic modeling advantages of GMM, the method achieves high-precision identification and early warning of stator thermal faults. The process of the Gaussian mixture model is as follows:

$$P(X_t) = \sum_{k=1}^{K} \omega_{k,t}\eta\left(X_t, \mu_{k,t}, C_{k,t}\right) \tag{4}$$

Where:

K…The number of Gaussian distributions, which corresponds to the quantity of sub-models in the mixture model;

$P(X_t)$…The probability of observing value X_t at time t;

$\omega_{k,t}$…The distribution weight at time t, signifying the likelihood of a sample belonging to that particular sub-model;

$\eta(X_t, \mu_{k,t}, C_{k,t})$…The probability density function;

$\mu_{k,t}$…The expectation of the Gaussian distribution;

$C_{k,t}$…The covariance at time t.

The number of Gaussian distributions used as the background model is defined as follows:

$$B = \arg_b min\left(\sum_{k=1}^{b} \omega_{k,t} > T_{\mathrm{th}}\right) \tag{5}$$

Where:
T_{th}…The set distribution weight threshold.

The way the new observations update the Gaussian distribution is as follows:

$$\omega_{k,t+1} = (1-\alpha)\cdot\omega_{k,t} + \alpha\cdot M_{k,t+1} \tag{6}$$

$$\mu_{k,t+1} = (1-\rho)\cdot\mu_{k,t} + \rho\cdot X_{t+1} \tag{7}$$

$$\sigma_{k,t+1}^2 = (1-\rho)\sigma_{k,t}^2 + \rho\left(X_{t+1}-\mu_{k,t+1}\right)\times\left(X_{t+1}-\mu_{k,t+1}\right) \tag{8}$$

$$M_{k,t+1} = \begin{cases} 1, & \text{matched, Gaussian, distribution} \\ 0, & \text{others} \end{cases} \tag{9}$$

$$\rho = \frac{\alpha}{\omega_{k,t}} \tag{10}$$

Where:
α…The learning rate;
ϱ…The parameter update speed;
$M_{k,\,t+1}$…The adaptation of the observed values and the sub-model, where 1 indicates a match and 0 indicates a mismatch.

4 Results and Analysis

4.1 Data Sets and Experiments

This study utilizes operational data from a steam turbine generator in northern China (2023) as training data, comprising temperature measurements from 48 monitoring points with a 1-minute sampling interval. The raw data undergoes sliding window segmentation with a window length of 120 minutes and a step size of 30 minutes, resulting in time-series samples of 120 data points per temperature sensor [21]. This process completes data preprocessing and feature extraction. Prior to fault diagnosis, the collected dataset (totaling 800 initial samples) requires partitioning. To facilitate subsequent data augmentation, 80% of samples from each operational state of the guide vanes are allocated for augmented learning, while the remaining 20% serve as the test set.

The deep learning experiments in this study were conducted on a computing platform equipped with an NVIDIA RTX 4080 GPU, utilizing Python 3.10 and PyTorch 2.1 as the primary development environment. The training process employed the Adam optimizer with an initial learning rate of 1e-4, a batch size of 16, and a maximum of 100 training epochs, incorporating early stopping to prevent overfitting. The Transformer model was configured with a hidden layer dimension of 256, 4 attention heads, and 3 layers. All experiments were executed on the Ubuntu 22.04 operating system. The variations of the loss values during the training process are shown in Figure 3.

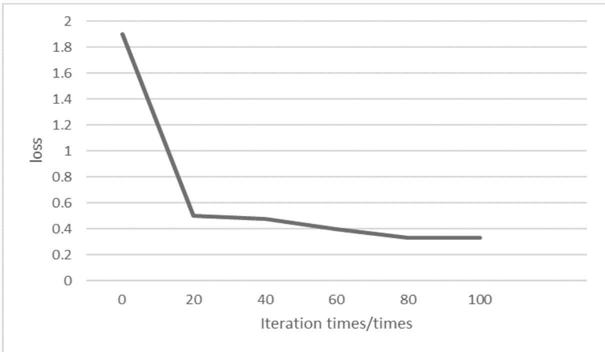

Fig. 3 *The change of the loss value*

4.2 Analysis of Experimental Results

The results of the ablation experiment are shown in Table 1. The ablation study results demonstrate that each introduced module significantly enhances model performance. The baseline Transformer model (without optimizations) achieved Precision, Recall, and Accuracy scores of 86.2%, 84.2%, and 87.4% respectively. After incorporating the physics-guided normalization method (LPGN), these metrics improved to 88.1%, 86.9%, and 89.6%, confirming LPGN's effectiveness in enhancing thermal anomaly detection. Subsequent integration of the enhanced Transformer architecture yielded more substantial gains - Precision sharply increased to 93.5% while Accuracy reached 93.3%, indicating superior temporal feature extraction capabilities. The final configuration, which added Gaussian Mixture Model (GMM) for feature distribution modeling and post-processing, achieved 97.6% Precision and 96.5% Accuracy, significantly outperforming other combinations and demonstrating GMM's exceptional discriminative power for abnormal patterns. Collectively, the synergistic effects of these modules substantially improved both the accuracy and robustness of stator thermal fault detection.

Tab. 1 *Ablation experiment results*

Model	Precision	Recall	Accuracy
Base	86.2	84.2	87.4
Base+ LPGN	88.1	86.9	89.6
Base+LPGN+Improved Transformer	93.5	88.1	93.3
Base+LPGN+Improved Transformer+GMM	97.6	90.3	96.5

For comparative analysis, we concurrently trained both ARIMA (AutoRegressive Integrated Moving Average) and LSTM (Long Short-Term Memory) models. The predictive performance was evaluated by calculating the average Root Mean Square Error (RMSE) across test datasets, where lower RMSE values indicate stronger algorithmic representation capabilities. Table 2 presents the comparative RMSE results of these different prediction algorithms.

Tab. 2 *Comparative test results*

Algorithm	RMSE
ARIMA	0.0625
LSTM	0.0762
Ours	0.0438

The RMSE (Root Mean Square Error) results reveal significant performance differences among prediction algorithms for stator thermal fault detection. The ARIMA model achieved an RMSE of 0.0625, outperforming LSTM's 0.0762 but remaining higher than our proposed method's 0.0438. As a conventional time-series analysis approach, ARIMA suits linear and weakly nonlinear sequence modeling but struggles to capture the complex nonlinear dynamics and long-term dependencies in stator temperature data, particularly showing limitations in multivariate interactions and abrupt change prediction. While LSTM, as a classical recurrent neural network architecture, possesses nonlinear processing and long-sequence dependency capabilities, its suboptimal performance in this task stems from insufficient sensitivity to local anomalies and inadequate extraction of multi-scale spatial thermal features. Additionally, LSTM's complex training process with parameter redundancy often leads to overfitting or gradient vanishing issues. In contrast, our method integrates physics-guided normalization (LPGN), enhanced Transformer architecture, and GMM probabilistic modeling, demonstrating superior advantages in critical temporal feature extraction, anomaly response, and stability - ultimately achieving the lowest RMSE and exhibiting stronger predictive accuracy and generalization capability.

5 Conclusion

This study proposes an integrated detection method combining Physics-Guided Normalization (LPGN), enhanced Transformer architecture, and Gaussian Mixture Model (GMM) for thermal fault detection in the turbine generator stators. The incorporation of localized physical information for feature normalization significantly improves the model's sensitivity to abnormal temperature rise patterns. The optimized Transformer structure further enhances the extraction efficiency of critical temporal features, while the integration of GMM endows the output layer with superior distribution modeling and anomaly discrimination capabilities. Ablation experiments demonstrate that the progressive integration of these modules substantially improves key metrics including Precision, Recall, and Accuracy. Compared with conventional ARIMA and LSTM approaches, the proposed method achieves optimal RMSE performance, validating its superior robustness in complex nonlinear scenarios. Overall, this framework not only advances the accuracy and reliability of thermal fault detection, but also establishes a robust technical pathway for intelligent diagnostic systems in practical power equipment health monitoring.

Funding

Research on the detection of nutrient elements in saline alkali soil by combining machine learning algorithms with LIBS technology (JJKH20241490KJ).

References

[1] ŽIVKOVIĆ, V. B., GRKOVIĆ, V. R., & KLJAJIĆ, M. V. (2024). The instigating factors behind the occurrence of vibration in steam turbines: A review analysis. *Thermal Science*, Vol. 28, No. 6 Part A, pp. 4451-4471. https://doi.org/10.2298/TSCI240105098Z

[2] WOLNIK, T., & JAREK, T. (2022). Solid rotor core vs. lamination rotor core in fractional-slot pmsm motor with high power density. *Energies*, Vol. 15, No. 15, pp. 5729. https://doi.org/10.3390/en15155729

[3] CHAUDHURI, A., DATTA, R., KUMAR, M. P., DAVIM, J. P., & PRAMANIK, S. (2022).

Energy conversion strategies for wind energy system: Electrical, mechanical and material aspects. *Materials*, Vol. 15, No. 3, pp. 1232. https://doi.org/10.3390/ma15031232

[4] AMIRYAR, M. E., & PULLEN, K. R. (2017). A review of flywheel energy storage system technologies and their applications. *Applied Sciences*, Vol. 7, No. 3, pp. 286. https://doi.org/10.3390/app7030286

[5] LIU, H., WANG, Y., ZENG, T., WANG, H., CHAN, S. C., & RAN, L. (2024). Wind turbine generator failure analysis and fault diagnosis: A review. *IET Renewable Power Generation*, Vol. 18, No. 15, pp. 3127-3148. https://doi.org/10.1049/rpg2.13104

[6] YANG, Q., SALENBIEN, R., SMITH, K. M., & TUNZI, M. (2024). Identifying untraced faults associated with high return temperatures from heating systems in buildings connected to district heating networks. *Energy*, Vol. 309, pp. 133097. https://doi.org/10.1016/j.energy.2024.133097

[7] LI, T., ZHAO, Y., ZHANG, C., ZHOU, K., & ZHANG, X. (2022). A semantic model-based fault detection approach for building energy systems. *Building and Environment*, Vol. 207, No. Part B, pp. 108548. https://doi.org/10.1016/j.buildenv.2021.108548

[8] LI, X.B., TIAN, F.Q., GUO, Y.K. (2023). Key Technologies for Health Management and Intelligent Operation and Maintenance of Power Equipment in New Power Systems. *Power System Technology*, Vol. 47, No. 9, pp. 3710-3726. DOI: 10.13335/j.1000-3673.pst.2022.2451

[9] CHEN, J., ZHANG, L., LI, Y., SHI, Y., GAO, X., & HU, Y. (2022). A review of computing-based automated fault detection and diagnosis of heating, ventilation and air conditioning systems. *Renewable and Sustainable Energy Reviews*, Vol. 161, pp. 112395. https://doi.org/10.1016/j.rser.2022.112395

[10] YAN, W., WANG, J., LU, S., ZHOU, M., & PENG, X. (2023). A review of real-time fault diagnosis methods for industrial smart manufacturing. *Processes*, Vol. 11, No. 2, pp. 369. https://doi.org/10.3390/pr11020369

[11] CHEN, Z., O'NEILL, Z., WEN, J., PRADHAN, O., YANG, T., LU, X., HERR, T. (2023). A review of data-driven fault detection and diagnostics for building HVAC systems. *Applied Energy*, Vol. 339, pp. 121030. https://doi.org/10.1016/j.apenergy.2023.121030

[12] CHEN, Y., MO, S., LI, W., HUANG, L., WEN, S., & HE, Z. (2023). Applications of distributed fiber Bragg gratings to the measurements of in-tube fluid temperature distribution. *Applied Thermal Engineering*, Vol. 220, pp. 119724. https://doi.org/10.1016/j.applthermaleng.2022.119724

[13] LIU, C., QIU, S., WANG, G., CHENG, C., TANG, J., & YANG, M. (2023). Simultaneous distributed acoustic and temperature sensing system based on ultra-weak chirped fiber Bragg grating array. *Optics Express*, Vol. 31, No. 11, pp. 18516-18524. https://doi.org/10.1364/OE.487205

[14] EHYA, H., NYSVEEN, A., & ANTONINO-DAVIU, J. A. (2021). Advanced fault detection of synchronous generators using stray magnetic field. *IEEE Transactions on Industrial Electronics*, Vol. 69, No. 11, pp. 11675-11685. DOI: 10.1109/TIE.2021.3118363

[15] MICHALSKI, M. A., MELANI, A. H., DA SILVA, R. F., DE SOUZA, G. F., & HAMAJI, F. H. (2022). Fault detection and diagnosis based on unsupervised machine learning methods: A kaplan turbine case study. *Energies*, Vol. 15, No. 1, pp. 80. https://doi.org/10.3390/en15010080

[16] EL IDRISSI, A., DEROUICH, A., MAHFOUD, S., EL OUANJLI, N., BYOU, A., BANAKHR, F. A., & MOSAAD, M. I. (2024). Stator imbalance defects diagnosis of induction machine using thermography and machine learning algorithms. *IEEE Access*, Vol. 12, pp. 51606-51618. DOI: 10.1109/ACCESS.2024.3382118

[17] STONE, E. J., PANAGIOTOU, P. A., MÜHLTHALER, J., MILLS, A. R., LAMBOURNE, A., & JEWELL, G. W. (2024). Identification of Thermal Failure Profiles in Stator Winding Insulation by Impedance Spectroscopy. *2024 International Conference on Electrical Machines (ICEM)*, Torino, Italy, IEEE, pp. 1-7. DOI: 10.1109/ICEM60801.2024.10700564

[18] ATTALLAH, O., IBRAHIM, R. A., & ZAKZOUK, N. E. (2022). Fault diagnosis for induction generator-based wind turbine using ensemble deep learning techniques. *Energy Reports*, Vol. 8, pp. 12787-12798. https://doi.org/10.1016/j.egyr.2022.09.139

[19] CHALUPA, M., ŠVÁSTA, A., KROBOT, Z., VEVERKA, J., KNOBLOCH, R., SVOBODA, A., ET AL. (2025). Creating a 3D Model of a Hovercraft for Research into Structural Shape Optimization and Material Design of Structural Parts. *Manufacturing Technology*, Vol. 25, No. 1, pp. 2-13. DOI: 10.21062/mft.2025.007

[20] WANG, Z., YIN, H. (2025). Titanium Alloy Turning Machining Model and Quality Analysis Based on Finite Element Analysis. *Manufacturing Technology*, Vol. 25, No. 3, pp. 413-423. DOI: 10.21062/mft.2025.036

[21] MRÁZEK, J., PEXA, M., ČEDÍK, J. (2024). The Effect of a Mixture of Methanol and Gasoline on the Operation of an Engine in an Electric Power Generator System. *Manufacturing Technology*, Vol. 24, No. 4, pp. 636-644. DOI: 10.21062/mft.2024.070