Transformer neural networks in Industry 4.0*

Denys Olianin[1,*,†], Halyna Tsupryk[1,†], Tetiana Hovorushchenko[2,†], Oksana Bahrii-Zaiats[3,†] and Igor Andrushchak[4,†]

[1] Ternopil Ivan Pului National Technical University, Ruska St, 56, 46051, Ternopil, Ukraine
[2] Khmelnytskyi National University, Institutska str., 11, Khmelnytskyi, 29016, Ukraine
[3] I. Horbachevsky Ternopil National Medical University, Maidan Voli St., 1, Ternopil, 46002, Ukraine
[4] Lutsk National Technical University, Lvivska St., 75, Lutsk, 43018, Ukraine

Abstract

This paper investigates the emerging role of transformer-based neural networks in advancing Industry 4.0, where the rapid generation of unstructured data—from maintenance logs to sensor readings—necessitates robust analytical tools. We begin by examining the core architecture of transformers, highlighting self-attention and parallel processing as key innovations that facilitate long-range dependency modeling. We then explore how these capabilities address critical Industry 4.0 needs, such as predictive maintenance, anomaly detection, and defect localization. Case studies in text-based log analysis demonstrate significant accuracy gains—often exceeding 90%—when compared to conventional methods like regex-based parsing or RNNs. Similarly, Vision Transformers (ViTs) excel in quality control tasks, accurately identifying subtle defects under challenging conditions. In the realm of time-series sensor data, transformer architectures capture intricate temporal correlations, enabling early fault detection and forecasting of remaining useful life. Despite these successes, the paper also notes practical challenges, including computational demands, the need for interpretability, and deployment constraints on industrial edge devices. Concluding remarks emphasize that transformer-based AI stands as a cornerstone for Industry 4.0, offering a flexible and powerful framework to tackle the complexity of modern manufacturing environments while suggesting avenues for future research in optimization and domain-specific customization.

Keywords

Transformer neural networks, predictive maintenance, anomaly detection, computer vision, IIoT

1. Introduction

Modern digital technologies have fundamentally transformed both everyday life and industrial operations. On the consumer side, innovations such as smart home systems and widespread internet connectivity have enhanced daily living. In parallel, Industry 4.0 harnesses these technological advancements to optimize manufacturing processes, increase production efficiency, and facilitate a more effective creation of goods. Neural network integration to industrial processes become popular in recent years, highlighting its versatility and usability [10-13]. These studies illustrate the successful use of neural networks to model various physical and mechanical properties of composite materials, such as epoxy composites, under different treatments and testing conditions, underscoring their practical utility in engineering and materials science domains. In this framework, the adoption of cutting-edge technologies—particularly transformer-based artificial intelligence—plays a pivotal role in the fourth industrial revolution. This paper explores how transformer-based AI can drive industrial innovation, improve data analysis, and streamline decision-making processes, ultimately contributing to the advancement of modern manufacturing practices.

* *CITI'2025: 3rd International Workshop on Computer Information Technologies in Industry 4.0, June 11–12, 2025, Ternopil, Ukraine*

[1*] Corresponding author.
[†] These authors contributed equally.
denys.olianin@gmail.com (D. Olianin); tsupryk_h@tntu.edu.ua (H. Tsupryk); tat_yana@ukr.net (T. Hovorushchenko); bagrijzayats@tdmu.edu.ua (O. Bahrii-Zaiats); 9000@lntu.edu.ua (I. Andrushchak)
0009-0002-5266-3941 (D. Olianin); 0000-0002-8106-5628 (H. Tsupryk); 0000-0002-7942-1857 (T. Hovorushchenko); 0000-0002-5533-3561 (O. Bahrii-Zaiats); 0000-0002-8751-4420 (I. Andrushchak)

2. Transformer neural network

First introduced in work [1], transformer neural network (TNN) was used for natural language processing (NLP) problems as translating text to different language. At their core lies the self-attention mechanism, which enables the model to evaluate the relevance of each element in the input sequence relative to every other element. This mechanism works by assigning attention scores, allowing the model to weigh the influence of each component, regardless of its position in the sequence.

Unlike traditional recurrent neural networks (RNN) that process data sequentially, transformers handle entire sequences simultaneously through parallel processing. This not only accelerates training but also enhances the model's ability to capture long-range dependencies—connections between distant elements in the data—that are often critical for understanding complex patterns. Additionally, transformers incorporate positional encodings to maintain the order of input elements, ensuring that the temporal or sequential relationships are preserved.

This combination of self-attention, parallel processing, and positional encoding makes transformers exceptionally powerful for tasks ranging from natural language processing to real-time analysis of industrial data, where understanding intricate and distant relationships within large datasets is essential.

Since introduction of TNN, a lot of modification were developed, for example encoder only and decoder only models were created for different tasks. Some modifications turned transformer-based networks into image recognition models, called Vision Transformers (ViT), and other modifications can create a tool for a prediction.

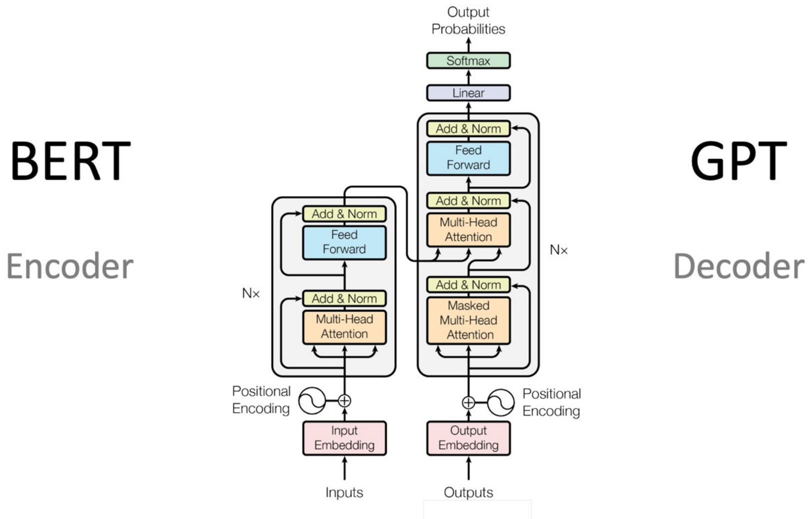

Figure 1: Traditional architecture of transformer neural network.

3. Relevance to Industry 4.0

Main benefit of Industry 4.0 approach to the production is producing higher volume of products with higher pace than regular plants. Key component to this is the Internet of things, which connects every automized part of the plant in manufacturing. These parts are covered by various sensors that monitor every step in the manufacturing process as well as state of the machine itself. If some part of the manufacturing line failed it can hold process for some valuable time.

3.1. Transformers for unstructured text data in Industry 4.0

Recent studies showed that transformer-based neural networks that are designed for forecasting and predicting failures can solve this problem with high accuracy. In paper [2] Erik Giovani proposed modified transformer-network without encoder and decoder for forecasting failure in rotational machinery. Proposed changes were made due to preprocessed data from vibrational sensors and absence in need of text outputs. After evaluation model showed 98% accuracy in machinery failure prediction.

Industrial operations also produce extensive free-text data—maintenance reports, error logs, incident descriptions, etc.—which contain valuable knowledge for predictive maintenance and process improvement. Transformers originally developed for NLP are now used to parse and analyze this text. For example, BERT and its variants have been applied to maintenance logs to perform tasks like anomaly detection and knowledge extraction.

In work [3] used a BERT-based model for system log anomaly detection. By capturing the contextual semantics of log messages with self-attention, their approach achieved high precision, recall, and F_1-scores (all >0.90) on benchmark datasets, outperforming traditional log-parsing methods. In their framework, a large language model (LLM) first structures raw logs, and then LogBERT (a Transformer network) encodes the log sequences to detect abnormal patterns, confirming the effectiveness of transformers in filtering noise and improving downstream fault detection. This demonstrates that transformers can automatically learn patterns of "normal" vs. "faulty" sequences from unlabeled log data, which is crucial for proactive monitoring in Industry 4.0.

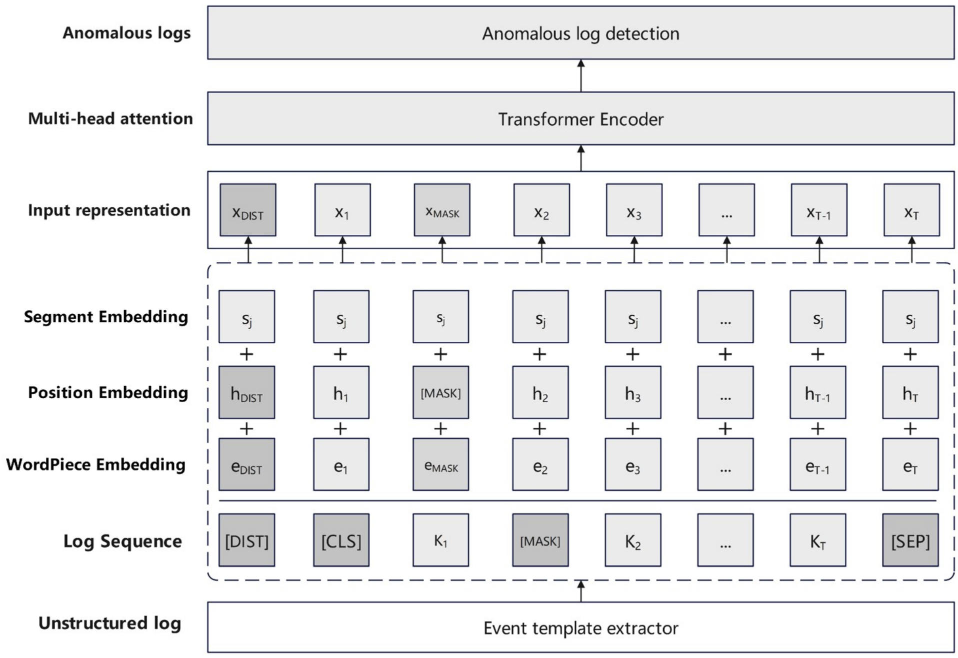

Figure 2: The logBERT framework from work [4].

In another work [4] developed a transformer-based pipeline to build a fault knowledge graph from equipment maintenance text. They fine-tuned RoBERTa (a BERT variant) to perform Named Entity Recognition (NER) on Chinese industrial fault records, extracting key entities like equipment names, fault causes, and remedies. The transformer model, combined with a BiLSTM-CRF tagger, yielded very high accuracy (average F_1 ≈ 0.95) in identifying these entities. This was a significant improvement over RNN-based baselines; for instance, adding the pre-trained BERT encoder boosted NER F_1 from ~0.816 to ~0.941 in their experiments. The authors report that the transformer's contextual embeddings enabled the system to effectively extract fault-related information from unstructured text that would be difficult to parse with manual templates. Such extracted knowledge can feed into predictive maintenance systems or structured knowledge graphs for decision support.

Another work [5]: NVIDIA's engineering team demonstrated a real-world proof-of-concept using transformers for predictive maintenance on server hardware. They applied NLP to kernel log data from NVIDIA DGX systems to automatically pinpoint log lines indicating root causes of failures. Using a fine-tuned BERT sequence classifier, the pilot system was able to highlight important log messages and classify whether a given log segment was a root-cause indicator. This approach aimed to drastically reduce manual log analysis by maintenance engineers. In the POC, the BERT-based model achieved near-perfect accuracy (over 99.9% on test data) in distinguishing root-cause log entries from normal logs, illustrating transformers' capacity to learn subtle patterns in unstructured log text that correlate with equipment faults. This real-world example underscores that transformer models can be successfully deployed to improve industrial troubleshooting and downtime reduction.

3.2. Transformers in Computer Vision for Quality Control

Another rich source of unstructured data in Industry 4.0 is images and video from production lines (e.g. for product inspection, defect detection, and surveillance). Vision transformers, which adapt the transformer architecture to image patches, have lately been applied to these tasks with great success.

Smith in his work [6] proposed using a modern ViT architecture for automated surface inspection in leather manufacturing. In this setting, images of leather sheets must be analyzed for anomalies (scratches, wrinkles, etc.) that affect quality. The researchers trained ViT models on a small dataset of low-resolution images (including an open-source leather defect dataset and the MVTec anomaly detection dataset) and compared them against lightweight methods based on convolution neural network (CNN). Experiments showed that the transformer-based model outperformed state-of-the-art lightweight methods in classification accuracy for defect vs. normal samples. Notably, the ViT achieved higher accuracy despite the challenges of low-res images and limited training data, thanks to its strong ability to capture both local and global patterns via self-attention. In addition to superior accuracy, the ViT approach proved advantageous in its low computational load and robustness to smaller image sizes, which is beneficial for real-time inspection on the factory floor. This study validates that transformers can surpass traditional vision models for industrial quality control, even under practical constraints.

Beyond classifying items as defective or not, transformers can localize subtle faults in images. The self-attention mechanism enables the model to consider relationships across an entire image, which helps in detecting small, context-dependent anomalies. For instance, the ViT model in work [6] could not only classify leather samples but also highlight the regions where defects were present (despite the resolution being low). Same research has found ViTs effective for tasks like surface crack detection and weld x-ray analysis, often outperforming CNNs in detecting anomalies under varying conditions. These advances suggest that in Industry 4.0 settings (smart factories, automated inspection stations), vision transformers offer a new level of accuracy and reliability for ensuring product quality.

A practical consideration for Industry 4.0 vision is how easily models can be adapted to new parts or visual conditions. Transformers trained on large datasets (e.g. ImageNet or proprietary multi-product image data) can be fine-tuned on specific defect detection tasks with relatively small sample sizes. This transfer learning approach has been shown to work well with ViTs, which can leverage pretrained attention weights to quickly learn new defect patterns. Researchers have reported success using pretrained ViT backbones for quick deployment in scenarios like PCB inspection and assembly line monitoring, reducing the need for collecting massive task-specific image datasets. This flexibility aligns with Industry 4.0's requirement for adaptable AI systems that can keep up with changing products and materials.

3.3. Transformers for Time-Series Sensor Data and Anomaly detection

Industrial IoT sensors continuously generate time-series data (vibrations, temperatures, pressures, etc.). Although such data is numerical, it can be considered unstructured in the sense of being unlabeled, high-dimensional streams requiring advanced pattern recognition. Transformer models have started to play a key role in analyzing these sequences for anomaly detection and forecasting.

A major advantage of transformers is the ability to capture long-range dependencies in sequences was presented in work[7]. Unlike traditional RNN/LSTM models that can struggle with long sequences, transformers use self-attention to learn relationships between events far apart in time. Recent work by Zia et al[7]. applies a transformer-based architecture to multivariate IoT sensor data for anomaly detection, specifically introducing a Transformer-based adversarial perturbation model for detecting abnormal patterns in complex sensor streams. They note that the transformer's self-attention is remarkably effective at extracting intricate temporal patterns and correlations from sequential sensor data. This allows the model to recognize subtle precursors to failures that span long durations or involve multiple interdependent parameters—something critical for predictive maintenance of industrial equipment.

Because of these properties, transformer-based models have been shown to outperform classical approaches in IoT anomaly detection tasks. Zia et al. report that their transformer not only captures long-term dependencies but also integrates with techniques like adversarial training to enhance robustness, yielding overall better performance in dynamic IoT environments. In practical terms, this means fewer false alarms and missed detections when monitoring complex machinery. Another study by Alexopoulos et al. (2024) combined LSTM autoencoders with a Transformer encoder to forecast equipment failures (predictive maintenance) on a real manufacturing dataset.

In work [8] The transformer encoder helped model the interactions of multiple sensor signals over time, improving the accuracy of remaining useful life (RUL) predictions for components. These studies validate that transformers can learn the multi-scale temporal features of industrial processes (e.g. periodic cycles, gradual drifts, sudden spikes) more effectively than earlier methods, leading to more reliable anomaly detection and forecasting in Industry 4.0 applications.

Another architecture based on Transformer-based neural network is introduced in work[9], as novel data driven framework for real-time data processing and machine failure prediction. Implemented proposed neural network system has proven system's accuracy rate in 70.84% and improved product yield from 78.38% to 89.62%.

Case Example where transformer neural networks can be applied is Equipment Monitoring: In power generation and energy distribution, transformers (the electrical kind) and other assets are monitored via sensors. Research in the past few years has begun using transformer neural networks to analyze such sensor data for early warning of faults. For example, an IoT-based monitoring system might deploy a tiny transformer model at the edge to flag anomalies in real time. Initial case studies have shown that even lightweight transformer models can run on industrial edge devices and detect anomalies like motor vibration irregularities or temperature excursions with high accuracy, thanks to the efficiency of the attention mechanism in focusing on crucial signal features. This suggests a broad potential for transformer networks to be embedded in Industry 4.0 sensor platforms for on-the-fly analytics and condition monitoring.

Conclusions

Over the past four to five years, transformer neural networks have proven highly effective for mining unstructured industrial data, ranging from textual maintenance logs to visual inspections and sensor signals. Their ability to learn rich representations and capture long-range interactions has advanced automation and intelligence within smart factories. In numerous tasks—such as log anomaly detection, fault information extraction, and visual defect classification—transformers consistently outperform traditional methods like regex-based parsing, CNNs, and RNNs.

A key factor behind these strong results is the self-attention mechanism, which enables models to incorporate the global context of data (e.g., entire log sequences or full image patches). This leads to a deeper understanding of subtle patterns that might otherwise remain hidden—particularly crucial in Industry 4.0, where minor anomalies can signal potential system failures. Additionally, transformers' architecture is versatile, facilitating the development of hybrid and domain-specific models that combine representation learning with specialized components (e.g., CRF layers for sequence labeling or autoencoders for anomaly detection).

Despite these strengths, several challenges warrant further exploration. Transformers can be computationally intensive, particularly when dealing with high-dimensional data in real-time scenarios. Interpretability also remains a pressing issue; in industrial environments, decision-making often requires transparent insights into how models arrive at specific conclusions. Future research may focus on optimizing transformer architectures for edge devices, improving explainability, and incorporating domain knowledge more seamlessly into model design.

Overall, transformer-based AI has emerged as a vital driver of Industry 4.0, offering a scalable and powerful framework for analyzing diverse, unstructured data streams. Continued advancements in model efficiency, explainability, and domain-specific customization are poised to further enhance their impact across modern manufacturing and beyond.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] Wang, Xin, Tapani Ahonen, and Jari Nurmi. "Applying CDMA technique to network-on-chip." IEEE transactions on very large scale integration (VLSI) systems 15.10 (2007): 1091-1100.

[2] Vaswani, Ashish; Shazeer, Noam; Parmar, Niki; Uszkoreit, Jakob; Jones, Llion; Gomez, Aidan N; Kaiser, Łukasz; Polosukhin, Illia (December 2017). "Attention is All you Need". In I. Guyon and U. Von Luxburg and S. Bengio and H. Wallach and R. Fergus and S. Vishwanathan and R. Garnett (ed.). 31st Conference on Neural Information Processing Systems (NIPS). Advances in Neural Information Processing Systems. Vol. 30. Curran Associates, Inc. ArXiv:1706.03762

[3] Sperandio Nascimento, Erick Giovani, Julian Santana Liang, Ilan Sousa Figueiredo, and Lilian Lefol Nani Guarieiro. "T4PdM: a Deep Neural Network based on the Transformer Architecture for Fault Diagnosis of Rotating Machinery". arXiv preprint arXiv:2204.03725 (2022).

[4] Zhou, Y.; Chen, Y.; Rao, X.; Zhou, Y.; Li, Y.; Hu, C. Leveraging Large Language Models and BERT for Log Parsing and Anomaly Detection. Mathematics 2024, 12, 2758.

[5] Gao, F.; Zhang, L.; Wang, W.; Zhang, B.; Liu, W.; Zhang, J.; Xie, L. Named Entity Recognition for Equipment Fault Diagnosis Based on RoBERTa-wwm-ext and Deep Learning Integration. Electronics 2024, 13, 3935. https://doi.org/10.3390/electronics13193935

[6] NVIDIA. "Enabling Predictive Maintenance Using Root Cause Analysis, NLP, and NVIDIA Morpheus." NVIDIA Developer Blog (n.d.). URL: https://developer.nvidia.com/blog/enabling-predictive-maintenance-using-root-cause-analysis-nlp-and-nvidia-morpheus/.

[7] Smith, A.D.; Du, S.; Kurien, A. Vision Transformers for Anomaly Detection and Localisation in Leather Surface Defect Classification Based on Low-Resolution Images and a Small Dataset. Appl. Sci. 2023, 13, 8716. https://doi.org/10.3390/app13158716

[8] Zia, S.; Bibi, N.; Alhazmi, S.; Muhammad, N.; Alhazmi, A. Enhanced Anomaly Detection in IoT Through Transformer-Based Adversarial Perturbations Model. Electronics 2025, 14, 1094.

[9] Bampoula X, Nikolakis N, Alexopoulos K. Condition Monitoring and Predictive Maintenance of Assets in Manufacturing Using LSTM-Autoencoders and Transformer Encoders. Sensors (Basel). 2024;24(10):3215. Published 2024 May 18. doi:10.3390/s24103215

[10] Poland, David & Puglisi, Lemuel & Ravì, Daniele. (2024). Industrial Machines Health Prognosis Using a Transformer-Based Framework. 776-781. 10.1109/MetroXRAINE62247.2024.10796172.

[11] Stukhliak, P., Martsenyuk, V., Totosko, O., Stukhlyak, D., & Didych, I. (2024). The use of neural networks for modeling the thermophusical characteristics of epoxy composites treated with electric spark water hammer. CEUR Workshop Proceedings, 3742, 13-24.

[12] Stukhliak, P., Totosko, O., Vynokurova, O., & Stukhlyyak, D. (2024). Investigation of tribotechnical characteristics of epoxy composites using neural networks. CEUR Workshop Proceedings, 3842, 157-170.

[13] Stukhliak, P., Totosko O., Stukhliak, D., Vynokurova, O., & Lytvynenko, I. (2024). Use of neural networks for modelling the mechanical characteristics of epoxy composites treated with electric spark water hammer. CEUR Workshop Proceedings, 3896, 405-418.
[14] Analytical stress analysis of the furan epoxy composite coatings subjected to tensile test. Dolgov, N., Stukhlyak, P., Totosko, O., Melnychenko, O., Stukhlyak, D., Chykhira, I. Mechanics of Advanced Materials and Structure. (2023), https://doi.org/10.1080/15376494.2023.2239811