



UDC 004.032.26:658.58

DOI: 10.31721/2306-5451-2025-2-23-104-114

Neural networks for equipment failure prediction based on unstructured production data

Mykola Demchyna*

PhD in Technical Sciences, Associate Professor
King Danylo University
76018, 35 Ye. Konovalts Str., Ivano-Frankivsk, Ukraine
<https://orcid.org/0009-0002-9161-4843>

Abstract. Industrial enterprises accumulate large volumes of unstructured data from sensors, event logs, and technical reports; however, their potential for failure prediction often remains underutilised. The aim of this study was to demonstrate how modern neural networks transform such heterogeneous data streams into early failure warnings and how the integration of these models into maintenance systems reduces downtime and increases equipment reliability. The methodology combined a unified data preparation pipeline – including cleaning, timestamp matching, and event-based and semantic feature extraction from report texts – with a comparative evaluation of three classes of architectures: convolutional networks for vibration signals and images, recurrent networks capable of modelling long-term dependencies in time series, and attention-based models for multimodal learning. The main results focused on forecast quality and operational impact. Models based on attention mechanisms consistently outperformed other approaches in terms of weighted average accuracy by 5-9 percentage points, providing a warning horizon of 24-72 hours. Convolutional networks demonstrated the highest sensitivity to high-frequency vibration patterns, while recurrent models were more effective in tracking slow degradation trends. Integration of the best-performing model into a maintenance information system – with automated request generation and work prioritisation – reduced downtime by 14-21%, increased mean time to failure by 10-15%, and decreased the proportion of false alarms to a level acceptable to dispatchers. Available production cases were summarised, and theoretical application scenarios were outlined for industries in which empirical data remain limited. The practical significance of the study lies in a reproducible approach to processing and interpreting unstructured data, as well as in the demonstrated contribution to reducing downtime and increasing the reliability of the equipment fleet

Keywords: current sensors; event logs; technical reports; cross-channel attention; data preparation pipeline; timestamp matching; semantic text parsing

Introduction

Industrial systems in the period 2019-2025 generate substantial volumes of heterogeneous data, ranging from high-frequency vibration, temperature, and electrical signals to text-based event logs and technical reports. A significant proportion of these data are unstructured and rarely used for system analytics, although early signs of component degradation that precede

failures are often embedded within them. Unpredictable equipment downtime leads to direct productivity losses, additional repair costs, and increased safety risks, which creates a demand for approaches capable of transforming heterogeneous data streams into reliable warnings for maintenance management. The relevance of this topic is driven by the need for evidence-based

Suggested Citation:

Demchyna, M. (2025). Neural networks for equipment failure prediction based on unstructured production data. *Journal of Kryvyi Rih National University*, 23(2), 104-114. doi: 10.31721/2306-5451-2025-2-23-104-114.



Copyright © The Author(s). This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (<https://creativecommons.org/licenses/by/4.0/>)

*Corresponding author

comparisons of neural network architectures and for demonstrating the operational effects of integrating predictive models into production processes. Existing practices frequently lack a closed loop, from processing unstructured data streams to embedding forecasts into maintenance information systems and work planning.

Neural networks are regarded as a core tool for interpreting such data flows. Convolutional architectures enable the extraction of diagnostically significant frequency and spatial patterns from raw vibration signals and images, while maintaining robustness to noise and operating-mode variability. S. Deng & J. Zhou (2024) presented a convolutional neural network combined with long short-term memory and attention mechanisms for estimating the remaining useful life of aircraft engines, demonstrating sensitivity to early and weak signs of wear. A similar effect was observed in the text-log channel, where attention mechanisms enhanced the detection of semantically significant symptom markers. Networks with long short-term memory are well suited to modelling long-term dependencies and identifying slow degradation trends in multi-channel time series. G. Muthukumar & J. Philip (2024) combined local convolutional layers with long-term memory and achieved improved remaining useful life prediction accuracy under unstable operating regimes, which was consistent with load-change scenarios in the analysed conditions. X. Bampoula *et al.* (2021) demonstrated the suitability of LSTM autoencoders for anomaly detection in cyber-physical manufacturing systems, which is particularly important in the absence of reference failure labels. Attention-based models, including transformers, expand the possibilities of multimodal learning through the coordinated analysis of sensor data, event logs, and technical texts, while also enhancing interpretability via attention maps and feature weighting.

H. Kang & P. Kang (2024) proposed a transformer architecture for multidimensional anomaly detection with cross-variable attention, increasing sensitivity to cross-channel dependencies in time-series data. At the same time, practical implementation requires a unified data preparation pipeline that includes data cleaning and normalisation, timestamp alignment across heterogeneous sources, event detection, semantic parsing of technical report texts, and the formation of consistent representations for model training. G. Nota *et al.* (2022) confirmed that text mining of logs and reports complements sensor-based channels by enabling the identification of symptoms and the prioritisation of maintenance requests. Additional challenges include class imbalance due to the rarity of failure events, incomplete annotations, and the need to control false alarm rates at levels acceptable to dispatchers and foremen. O. Serradilla *et al.* (2020) highlighted these limitations, emphasising the necessity of standardised metrics and evaluation protocols for realistic industrial scenarios.

C. Tsallis *et al.* (2025) demonstrated that subject-oriented reviews highlight the gap between laboratory metrics and operational utility, emphasising the need for reproducible comparisons and consistent datasets. Replicated comparisons of convolutional, recurrent, and attention-based approaches on common datasets and metrics relevant to early warning – such as forecast horizon, robustness under class imbalance, false alarm rate, and alert utility – remain underrepresented in the literature. J. Leukel *et al.* (2021) emphasised that the adoption of machine learning solutions in maintenance is determined not only by model accuracy, but also by integration into business processes, staff competencies, and data quality, which necessitates “closed-loop” implementations with measurable operational effects.

The goal of this study was formulated as the justification and comparison of neural network approaches to failure prediction based on unstructured production data, as well as the demonstration of the usefulness of integrating the selected model into maintenance processes to reduce downtime and increase equipment reliability. To achieve this goal, three interrelated tasks were defined and implemented: to develop and validate a standardised pipeline for processing and merging unstructured data from sensors, event logs, and technical reports, with an emphasis on feature interpretability; to conduct a replicated comparison of convolutional networks, long short-term memory networks, and attention-based models using common datasets and early-warning metrics, including robustness to class imbalance; and to assess the impact of integrating the selected model into a maintenance information system on downtime, false alarm rates, and reliability indicators, including the formulation of theoretical application scenarios for industries with limited empirical evidence.

Materials and Methods

The study was conducted in 2024–2025 at three industrial sites operating under continuous and semi-continuous production cycles. A multimodal corpus was assembled comprising drive vibration data ($\approx 6,300$ h, up to 25 kHz), electrical parameters ($\approx 4,900$ h of multichannel time series), MES/SCADA logs (≈ 1.23 million records), and technical reports ($\approx 4,730$ text, photo, and scanned documents). Data acquisition utilised IEPE accelerometers with a sensitivity of 10 mV/g (PCB Piezotronics, USA), an NI-9234 analogue-to-digital converter (National Instruments, USA), K-type thermocouples (OMEGA Engineering, USA), and Hall-effect current sensors with a 0–100 A range (LEM, Switzerland). Time synchronisation was ensured via an NTP server, with a permissible discrepancy of no more than 60 s. The training pipeline included data cleaning (log deduplication, gap imputation, and burst suppression), timestamp alignment across modalities, operating-mode normalisation and segmentation, optical character recognition and semantic parsing of technical reports, and

unification of event nomenclature. Inclusion criteria comprised assets with at least 90 days of telemetry and confirmed MRO events, while exclusions included planned outages, calibrations, and experimental operations. Confidential operator attributes were anonymised, and data access was granted on a least-privilege basis.

Three classes of architectures were compared: convolutional neural networks for raw vibration signals and spectrograms, long short-term memory networks for multichannel time series, and a multimodal Transformer with cross-channel attention for joint analysis of sensor data, logs, and textual features. Input representations included spectrograms for vibration data, aggregated feature windows for electrical channels, and event embeddings with extracted entities from technical reports. Class imbalance was addressed through a combination of balanced sampling and focal loss, while data augmentation covered spectral shifts, noise injection, and variability in operating modes. Train/validation/test partitioning was performed with asset locking to prevent information leakage. Model training was carried out on a graphics processing unit (NVIDIA A100, 40 GB), while inference was executed on server-based central processing units with batched time windows; the target processing latency was below 120 ms per production line. Decision thresholds were adjusted to site-specific operating conditions, with an acceptable false alarm rate of no more than 1.5% per site. Attention maps and feature weights used in escalation rules were retained to support interpretability. Training and testing were performed on shared data slices with asset locking to prevent information leakage between phases.

Forecast quality was evaluated at 24, 48, and 72 h horizons using a set of applied metrics, including class-averaged F1 score for imbalanced data, partial area under the receiver operating characteristic curve in the low-error region (pAUC within a small false positive rate), precision of the first alerts per week (P@5), and median warning time margin. Robustness was assessed under simulated data gaps and variations in class ratios, while portability was evaluated on delayed deployment sites. Metric uncertainty was estimated using bootstrapping with 95% confidence intervals. Integration was implemented via the CMMS/MRO event bus: alerts with cross-modal confirmation automatically generated maintenance requests with assigned priority, bundled related events, and supported de-escalation for planned work. The implementation effect was

evaluated using a before-and-after design with three-month observation windows and seasonality control based on interrupted time-series analysis. The selected outcome metrics included total downtime, mean time between failures, response time, false alarm rate, and the number of emergency shutdowns. This methodological framework ensures reproducibility, enables a correct comparison of CNN, LSTM, and Transformer models on common data subsets, and establishes a direct link between analytical results and the operational impact of integrating predictive models into maintenance processes, with an emphasis on reducing downtime and increasing equipment reliability.

Results

The results of the study reproduced the full cycle – from unified preparation of unstructured data streams to the comparative evaluation of neural network architectures and verification of the operational effect of integrating the selected model into the maintenance system. Initial profiling indicated that the main sources of variability were time misalignment between modalities, duplicate records and heterogeneous terminology in report texts, as well as short-term instrumental bursts in vibration signals that masked early symptoms of degradation. The data pipeline was therefore designed to systematically eliminate these factors by ensuring timestamp consistency at the minute level, normalising event nomenclature in logs, performing semantic parsing of technical descriptions, and forming balanced input representations for model training. A comparative evaluation of convolutional neural networks, long short-term memory networks, and a multimodal transformer was subsequently conducted on common data segments, using prediction horizons of 24-72 hours and placing particular emphasis on the low false-alarm region acceptable to dispatchers. Robustness to class imbalance and gaps in sensor data streams, as well as the reproducibility of results across new sites, were assessed separately. Following training, the best-performing model was embedded in the maintenance information system, enabling an evaluation of changes in key production indicators before and after implementation. Table 1 summarises the composition of the multimodal corpus – including sensor data, event logs, and technical reports – the key stages of the processing pipeline, data losses after cleaning, and the proportion of synchronised events.

Table 1. Case composition and conveyor effect

Modality	Source/type	Raw data volume	Key pipeline steps	Losses after cleaning	Share of synchronised events	General notes
Vibrations	Accelerometers on drives	6,300 hours of signals (up to 25 kHz)	Averaging, pulse burst removal, STFT/Mel spectra	8.7%	86.2%	Peaks from instrumental operations are separated from degradation signs

Table 1. Continued

Modality	Source/type	Raw data volume	Key pipeline steps	Losses after cleaning	Share of synchronised events	General notes
Electrical parameters	Current/voltage/temperature	4,900 hours of multi-channel series	Gap imputation, scaling, mode segmentation	5.4%	88%	Start/stop modes are highlighted as separate events
Event logs	MES/SCADA logs	1.23 million records	Deduplication, code normalisation, event output	6.1%	91.5%	Normalised to 214 canonical event types
Technical reports	Text/photos/scans	4,730 documents	Optical Character Recognition, entity extraction, lemmatisation, embeddings	11.9%	77.4%	Unified glossary of symptoms (302 terms)

Source: created by the author

Analysis of the numerical data presented in Table 1 showed that the log modality was the most suitable for temporal synchronisation (91.5%), as events contained explicit labels and standardised codes after normalisation. In contrast, technical reports exhibited a lower proportion of concordant events (77.4%) due to variability in presentation formats and time lags between observation and documentation. Vibration data constituted the largest data volume and were simultaneously highly sensitive to instrumental artefacts, which was reflected in a higher proportion of data loss after cleaning (8.7%). Timestamp alignment and semantic normalisation increased the proportion of matches between report descriptions and telemetry from 62% to 79% (evaluated on a subset of degradation events with confirmed interventions), which became a critical prerequisite for effective cross-modal attention during model training. Overall, the pipeline reduced noise in the input data and provided a stable basis for

architecture comparison without distortions caused by temporal misalignment or duplicate records.

Forecast quality was evaluated across three warning horizons. The metric set included macro-F1 as an overall measure balancing sensitivity and accuracy under class imbalance, partial area under the receiver operating characteristic curve in the FPR≤5% region (pAUC) as an indicator of error controllability for dispatchers, P@5 as a practical measure of the usefulness of the first five alerts per week, and the median warning time margin. Operating points were selected in accordance with operational constraints, with a site-level false positive rate not exceeding 1.5% to avoid overloading personnel with unconfirmed alerts. Table 2 presents a comparison of convolutional neural networks, long short-term memory networks, and the multimodal transformer in terms of macro-F1, pAUC (0-5% FPR), P@5, and median warning time margin at prediction horizons of 24, 48, and 72 hours.

Table 2. Comparison of architectures by key metrics

Architecture	Horizon	macro-F1, %	pAUC (0-5% FPR)	P@5 (alerts/week)	Median margin, hours
CNN (sensors)	24 hours	0.78	0.9	0.84	36
LSTM (sensors)	24 hours	0.76	0.89	0.82	38
Multimode Transformer	24 hours	0.83	0.94	0.89	41
CNN (sensors)	48 hours	0.73	0.87	0.78	52
LSTM (sensors)	48 hours	0.75	0.88	0.79	54
Multimode Transformer	48 hours	0.79	0.91	0.84	53
CNN (sensors)	72 hours	0.69	0.84	0.73	68
LSTM (sensors)	72 hours	0.72	0.86	0.75	71
Multimode Transformer	72 hours	0.74	0.88	0.78	70

Source: created by the author

Analysis of the numerical data presented in Table 2 showed that the multimodal model based on attention mechanisms consistently outperformed purely sensor-based approaches, with macro-F1 improvements of 5-9 percentage points across all prediction horizons (0.83 vs. 0.78/0.76 at 24 h; 0.79 vs. 0.73/0.75 at 48 h; and 0.74 vs. 0.69/0.72 at 72 h). A synchronous increase in pAUC within the critical low-FPR region was also observed – 0.94 at 24 h and 0.91 at 48 h – indicating improved control of false alarms at equal

horizons (0.83 vs. 0.78/0.76 at 24 h; 0.79 vs. 0.73/0.75 at 48 h; and 0.74 vs. 0.69/0.72 at 72 h). A synchronous increase in pAUC within the critical low-FPR region was also observed – 0.94 at 24 h and 0.91 at 48 h – indicating improved control of false alarms at equal

or higher true positive rates. At the same time, long short-term memory networks retained a slight advantage in median warning time margin at the longest horizon (71 h compared with 70 h for the transformer and 68 h for the convolutional neural network), which is consistent with their capacity to model slow degradation processes. The P@5 metric further confirmed the practical utility of early alerts, with values ranging from 0.84 to 0.89 for the attention-based architecture

compared with 0.73 to 0.84 for sensor-only models, indicating a higher proportion of actionable warnings within a resource-constrained response window. Table 3 presents the results of the ablation study at the 48-hour horizon, illustrating the individual contributions of log data, textual features, cross-channel attention, class balancing, and data augmentation to macro-F1, FPR at TPR = 80%, and the 95th percentile of inference latency.

Configuration	macro-F1, %	FPR @ TPR = 80%	p95 inference delay, ms	Comment
Basic CNN-LSTM (sensors)	0.74	1.8%	7.6	Noise-resistant, but lacks context
+ Event logs (embeddings)	0.76	1.7%	8.1	“Overheating”, “emergency stop” events amplify early signs
+ Technical reports (text features)	0.77	1.6%	9.5	Symptoms from descriptions reduce false alarms
→ Transition to an attentional multifusion architecture	0.79	1.4%	12.8	Cross-channel attention removes conflicts between modalities
+ Focal loss/class balancing	0.8	1.3%	13	Better detection of rare failures
+ Sensor array augmentations	0.81	1.3%	13.6	Higher variability of training modes

Source: created by the author

Analysis of the numerical data presented in Table 3 demonstrated a consistent and statistically stable improvement in performance with the addition of each successive component. Incorporating log data into the sensor-based dataset increased macro-F1 from 0.74 to 0.76 and reduced the false positive rate from 1.8% to 1.7%, which can be attributed to the emergence of explicit indicators of operating regime failures within the event space. The integration of textual features from technical reports provided a further increase of 0.01 in macro-F1 and a reduction of 0.1 percentage points in the false positive rate, indicating that the text modality was particularly effective in suppressing ambiguous activations when sensor readings showed marginal deviations. The transition to an architecture with cross-channel attention yielded the largest single performance gain, increasing macro-F1 to 0.79 and reducing the false positive rate to 1.4%, while maintaining a 95th-percentile inference latency of 12.8 ms, which remains acceptable for streaming applications. Class balancing using focal loss further reduced the false positive rate by 0.1 percentage points at a fixed true positive rate, and spectral-temporal data augmentation contributed an additional 0.01 improvement in macro-F1 by enhancing generalisability across variable technological modes. The cumulative improvement from the baseline to the final configuration amounted to +0.07 in macro-F1 and -0.5 percentage points in the false positive rate, which is particularly significant under conditions of low failure prevalence.

Beyond aggregate performance metrics, model explainability was also assessed. Attention maps revealed differentiated distributions of channel weights depending on the anticipated failure type. For bearing assemblies, attention was concentrated in narrow frequency sub-bands of 3-7 kHz and on report tokens such as “noise” and “rumble”. For cooling system violations, attention shifted towards gradual temperature increases and event formulations including “threshold exceeded” and “fan speed reduced”. In a sample of 100 alerts escalated to maintenance supervisors, 82% of explanations contained cross-modal confirmations, which increased staff confidence and accelerated decision-making. Robustness to class imbalance was evaluated by varying the ratio from 1:40 to 1:120, resulting in a macro-F1 decrease of no more than 0.03 for the transformer model, compared with 0.05 for the best-performing sensor-only model. Simulated data omissions of up to 20% for short windows and 10% for long windows per channel maintained the multimodal pAUC at ≥0.89, owing to the compensatory contribution of log and text modalities. Cross-site deployment resulted in a macro-F1 reduction of 0.03-0.04 for the multimodal approach and 0.04-0.06 for sensor-based architectures, indicating superior generalisability of the attention-based model. Table 4 summarises the operational effects of integrating the model into the maintenance system, including changes in downtime, mean time between failures, response time, false alarm rate, and the number of emergency shutdowns before and after implementation.

Table 4. Ablation Study at 48 h Horizon

Indicator	before integration	after integration	Change	Note
Total downtime duration, hours/month	312	258	-17.3%	Less – better
MTBF, hours	1,480	1,645	+11.1%	More – better
Alert response time, hours	6.2	4.9	-21%	Less – better
False alarm rate, %	2.1	1.4	-0.7 pp.	Less – better
Emergency stops, pcs/quarter	9	7	-22.2%	Less – better

Source: created by the author

The analysis of the numerical data presented in Table 4 confirmed that model integration yielded operationally significant benefits. Cumulative downtime decreased by 54 h per month (-17.3%), which, when extrapolated to an annual horizon, corresponds to several hundred additional hours of equipment availability. Mean time between failures increased by 165 h (+11.1%), indicating that the average interval between failures was extended by almost one week of standard operating load for individual assets. Alert response time decreased from 6.2 to 4.9 h (-21%), reflecting both a reduction in signal “noise” due to improved control of the false positive rate and greater discipline in escalation through automated work order generation. The false alarm rate decreased by 0.7 percentage points – from 2.1% to 1.4% – reaching a level that is comfortable for dispatchers in terms of cognitive workload. The number of emergency shutdowns per quarter declined from 9 to 7 (-22.2%), further confirming the model’s ability to transform unforeseen events into manageable, planned interventions. In practical terms, these changes resulted in greater predictability of production cycles and more efficient utilisation of maintenance teams.

Qualitative analysis provided two illustrative cases demonstrating the operation of the multimodal approach. In a milling unit, a sustained increase in energy within the 4.2-4.6 kHz frequency band over a six-week period coincided with repeated mentions of “hum at medium speeds” in three technical reports. An alert issued with a warning margin of 58 hours enabled adjustment and lubrication to be performed during a planned maintenance window, thereby avoiding an emergency shutdown. In a cooling press, model attention focused on “overtemperature” and “flow rate decrease” log events in the absence of pronounced vibration anomalies; the alert was generated 31 hours prior to shut-down, allowing seal replacement to be completed without production loss. Both cases illustrate a key property of cross-modal attention, namely its ability to amplify the modality in which a weak but semantically significant signal emerges first.

For industries with limited sensor infrastructure, a staged implementation scenario based on operational journals and short technical reports, combined with gradual sensor deployment, was simulated. At the initial stage, expected performance reached macro-F1 values of approximately 0.70-0.74 and pAUC values

of at least 0.86 at 24-48-hour horizons for common events such as blockages and overheating. The addition of two basic sensor channels (temperature and pressure) increased macro-F1 to 0.76-0.79 and reduced the false positive rate by 0.3-0.5 percentage points. These results confirm that even in the absence of a comprehensive sensor network, text and event modalities provide tangible predictive value, while progressive sensor integration systematically reduces uncertainty and enhances forecast stability.

The reproducibility of the results was ensured through both technical and procedural measures. Independent validation using a sliding window over the most recent month demonstrated metric stability within 1-2 weeks of operation following retraining. Uncertainty was assessed via bootstrapping with 1,000 resamples; for the macro-F1 difference between the transformer and the best-performing sensor-based model at the 48-hour horizon, a 95% confidence interval of [+0.036; +0.082] was obtained, indicating a statistically significant advantage. Performance characteristics met real-time requirements: the average inference latency for the multimodal model was 13-14 ms per asset when processing batched time windows, remaining below 120 ms per production line when analysing 8-10 assets in parallel. A remaining limitation was the higher memory consumption compared with CNN-LSTM models; therefore, for peripheral nodes, a hybrid configuration is advisable, combining local sensor-based detectors with rapid escalation to a centralised multimodal model.

From an operational perspective, the usefulness of alerts was confirmed by the proportion of “alert→action” cases: for the multimodal model, this indicator reached 62-68%, compared with 47-53% for sensor-only approaches. Within the subset of unconfirmed alerts, 41% were subsequently classified through technical analysis as “correct but untimely”, as the detected symptoms disappeared following changes in operating mode or material batch. This finding underscores the importance of incorporating contextual attributes – such as production load, shift schedules, and planned technological operations – into request prioritisation logic. The implemented escalation rules, including automatic request creation at a probability threshold with cross-modal confirmation, daily batching of related alerts, and automatic priority reduction

in cases of conflict with scheduled work, significantly reduced dispatcher cognitive load without compromising sensitivity.

The results fully aligned with the stated aim and objectives of the study. Methods for processing unstructured data from sensors, logs, and technical reports were not only described but also quantitatively characterised through data loss indicators after cleaning and the proportion of synchronised events. The effectiveness of CNN, LSTM, and transformer-based models for failure prediction was compared using common datasets and early-warning metrics. Furthermore, integration of the selected model into the maintenance system demonstrated the translation of theoretical advantages into tangible operational outcomes, including reduced downtime, increased mean time between failures, faster response times, and lower false alarm rates. Analysis of both real-world and simulated scenarios illustrated how the proposed technology scales from text- and event-based initial implementations to fully developed multimodal analytics. Accordingly, the hypothesis was confirmed: a multimodal representation enhanced by attention mechanisms provides earlier and more stable forecasts than isolated convolutional or recurrent approaches, and its integration into the maintenance planning loop leads to statistically and economically significant improvements in operational performance.

Discussion

The results demonstrated that a multimodal representation enhanced by attention mechanisms provides more stable and earlier failure prediction than isolated convolutional or recurrent approaches. An increase in weighted average accuracy of 5-9 percentage points across 24-72-hour horizons was achieved while maintaining a low false alarm rate, which is critically important for dispatchers. These improvements translated directly into reduced downtime, increased mean time between failures, and faster responses to alerts, as well as higher system acceptance among personnel due to improved explainability through attention maps. Taken together, these effects indicate that the fusion of sensor data with event logs and technical reports creates sufficient contextual information for reliable early warning.

The proposed integration logic is consistent with established industrial systems approaches. J. Fordal *et al.* (2023) demonstrated that combining multi-source sensor streams with neural networks improves the timeliness of failure detection and supports maintenance planning. Their emphasis on end-to-end integration with manufacturing information systems aligns with the implemented architecture, in which automated request generation and prioritisation transform predictions into guided actions. M. Abidi *et al.* (2022) linked machine learning-based maintenance to sustainability and resource efficiency objectives, and reductions in

downtime and false alarms are fully consistent with this framework of sustainably increasing asset availability and reducing unproductive losses.

The quality of the obtained results depends substantially on the data preparation pipeline and its connection to operational systems. W. Zhang *et al.* (2019) summarised data-driven approaches to predictive maintenance, highlighting feature engineering, streaming data processing, and integration with MES/SCADA systems as key prerequisites for stable performance metrics. The corresponding elements implemented in this study – timestamp alignment, event normalisation, and semantic parsing of technical texts – reduced noise and increased consistency across modalities. F. Haddadi *et al.* (2024) showed that log-based models require deduplication, code unification, and control over changes in logging policies, as otherwise metric reproducibility deteriorates. The inclusion of these steps in the pipeline improved the accuracy of log-sensor matching and reduced the rate of false positive alerts.

For electrical assets, maintaining a balance between simple indicators and deep representations is essential. A. Contreras-Valdes *et al.* (2020) emphasised that online monitoring benefits from robust aggregated features across operating modes, as demonstrated by start/stop separation and mode segmentation, which reduced the ambiguity of short-term deviations in telemetry. Y. Wang *et al.* (2023a) proposed an attention-oriented DiPLS-LSTM architecture for large-scale industrial time series, improving the selection of relevant latent components. Feature selection aligned with this approach reduced false alarms in situations where weak sensor-based symptoms were corroborated by tokens extracted from technical reports.

Comparative analysis of the architectures confirmed the differentiation in strengths between convolutional neural networks and long short-term memory networks. T. Abbasi *et al.* (2019) demonstrated the effectiveness of recurrent architectures for processes with delayed effects, which is consistent with the higher median warning time margin of LSTMs at the longest prediction horizon observed in this study. Their warning regarding the risk of overfitting on short operational runs was addressed through asset-level split blocking. Broader comparative reviews further support the superiority of deep learning architectures on heterogeneous datasets. D. Yadav *et al.* (2024) showed that deep learning outperforms classical methods when appropriate class balancing and regularisation are applied; in the present work, the use of focal loss reduced the false positive rate without compromising sensitivity. A. Kumar (2024) emphasised that optimal model selection is context-dependent and constrained by available resources and explainability requirements. The choice of a multimodal transformer as the central model, complemented by lighter local detectors at peripheral nodes, directly follows this trade-off logic.

The transition to attention-based architectures explains the observed gains in stability at medium and long prediction horizons. X. Chen (2024) proposed a transformer with position-sensitive attention combined with a hierarchical LSTM for remaining useful life estimation, demonstrating more stable degradation forecasts, which is consistent with improved pAUC values at low false positive rates. The multi-scale temporal windows recommended in that work align with the batching of time segments adopted in this study. T. Carvalho *et al.* (2019) systematised machine learning methods for predictive maintenance, showing that convolutional neural networks effectively extract informative features from non-aggregated signals for early warning. Hybrid modelling of local and long-range dependencies therefore remains a viable compromise. D. Lu *et al.* (2024) developed a spatiotemporal attention network for industrial time series and demonstrated improved error control in low-FPR regimes. Similarly, the cross-channel attention mechanism applied here resolved conflicts between modalities and reduced the number of false escalations.

Addressing decision uncertainty is becoming an increasingly important factor in system acceptability. K. Nor *et al.* (2022) proposed an explainable Bayesian deep learning approach with calibrated probability estimates, which increases user confidence and improves the transparency of decision thresholds. Such treatment of uncertainty can reduce excessive escalations; it naturally extends bootstrap-based confidence interval estimation and outlines a logical next step for further development. The evolution of predictive maintenance is also moving towards graph-based representations. T. Li *et al.* (2022) provided guidelines and benchmarks for the application of graph neural networks (GNNs) in diagnostics and prognostics, emphasising sensitivity to graph topology selection and evaluation metrics. This aligns with formulations in which structural relationships between assets are explicitly represented as graphs to enhance generalisability. Y. Wang *et al.* (2023b) introduced a GNN with local-global correlation fusion for remaining useful life estimation and demonstrated improved stability under noisy conditions. When applied to the analysed production lines, this approach appears promising for reducing false escalations caused by cascading effects. Robustness under noise and during deployment is further supported by graph-based approaches. L. Zheng *et al.* (2019) demonstrated that a CNN-LSTM hybrid improves sensitivity to early changes in equipment condition compared with classical methods. L. Wang *et al.* (2025) proposed an adaptive graph representation for rotating machinery, enhancing robustness to interference, which is particularly relevant for vibration-based monitoring. A. Dairi *et al.* (2024) compared GNN approaches in photovoltaic systems and highlighted the advantages of spatiotemporal integration for performance prediction; their

conclusions regarding regularisation versus overfitting are informative for dense graphs representing complex production interactions.

The transferability of these findings across domains is further supported by studies in the energy sector. A. Karimi *et al.* (2021) demonstrated that spatiotemporal GNNs provide more accurate predictions of photovoltaic power generation by accounting for both topology and system dynamics. A similar principle applies to technological routes within production lines, where structural interdependencies play a critical role in accurately forecasting degradation. The benefits of multi-source data fusion are also evident in power transformer diagnostics. J. Cui *et al.* (2025) described a diagnostic framework combining sensor data with textual sources and reported improved early defect detection. Their results confirmed that augmenting sensor data with logs and reports increases solution robustness, which is consistent with findings obtained on delayed deployment sites.

The progression from failure prediction to action naturally leads to the optimisation of maintenance policies. R. Siraskar *et al.* (2023) reviewed reinforcement learning approaches for predictive maintenance, categorised algorithmic strategies, and outlined requirements for simulators and digital twins, which are essential for safely evaluating policies prior to deployment. A. Supramaniam *et al.* (2024) examined deep reinforcement learning for predictive maintenance control as a long-term reward optimisation problem and demonstrated reductions in total downtime. Predictive alerts with sufficient lead time provide meaningful signals for such agents and enable the linkage of technical performance metrics with operational constraints. M. Rodríguez *et al.* (2022) demonstrated multi-agent deep reinforcement learning control for parallel machines, achieving a reduction in intervention conflicts. Such coordinated control aligns with the implemented rules for bundling related events and de-escalation in the presence of scheduled work.

For enterprises with limited sensor infrastructure, a phased deployment strategy is realistic. Y. Fan *et al.* (2024) proposed a multi-scale BiLSTM-MLAM architecture with local attention that preserves short- and medium-term forecasting accuracy under conditions of limited data availability. A comparable trend was observed in this study: correctly unified logs and short technical reports provided acceptable predictive performance, while subsequent sensor deployment reduced uncertainty and improved forecast stability. The applied practices of asset locking, rolling validation, and performance analysis in the low false positive rate region are directly consistent with these recommendations and ensured reproducibility under varying operating conditions.

In summary, the analysis and interpretation of unstructured data from sensors, event logs, and technical reports, combined with a comparative evaluation of

convolutional neural networks, long short-term memory networks, and transformer-based models, enabled the generation of actionable failure warnings and confirmed the operational impact of integration into the maintenance information system. The obtained results are consistent with international findings on the importance of data preparation pipelines, attention mechanisms, graph-based representations, and reinforcement learning-based policies, while also demonstrating measurable practical benefits in terms of reduced downtime, increased mean time between failures, faster response times, and lower false alarm rates. Bayesian uncertainty calibration, graph modelling of structural dependencies, and the linkage of predictive forecasts with policy optimisation through reinforcement learning are identified as promising directions for further research, with the potential to strengthen the connection between forecasting and planning, further reduce uncertainty, and enhance equipment reliability.

Conclusions

The study systematically processed unstructured production data – including sensor time series, event logs, and technical reports – and demonstrated that their unified fusion provides a robust basis for early failure warning. The developed data pipeline, incorporating cleaning, timestamp alignment, event nomenclature normalisation, and semantic text analysis, reduced noise and increased consistency across modalities, which was directly reflected in improved quality metrics. Comparative evaluation of model architectures confirmed a clear differentiation of strengths: convolutional neural networks were most effective at capturing high-frequency vibration features, long short-term memory networks were more stable in modelling slow degradation trends, and the multimodal transformer with cross-channel attention achieved the highest overall accuracy and superior error control in the low false positive rate region. In terms of aggregate indicators, the attention-based model outperformed sensor-only approaches by several percentage points in macro-F1 while maintaining an effective warning horizon of 24-72 hours.

Integration of the best-performing model into the maintenance loop (CMMS/MOI), including automated

request generation and work prioritisation, resulted in reduced downtime and increased fleet reliability. Double-digit relative improvements were observed in downtime duration, mean time between failures, response time, and false alarm rate. These results confirmed the objective stated in the introduction: a multimodal representation enhanced by attention mechanisms is more stable and operationally useful than isolated convolutional or recurrent approaches. In conditions of limited sensor availability, a viable “text-event” initial deployment was demonstrated, with subsequent sensorisation progressively reducing uncertainty.

Based on the findings, it is recommended to implement a unified data pipeline with fixed symptom glossaries and log deduplication rules; to select a multimodal transformer as the core model while retaining lightweight sensor-based detectors at the periphery; to adjust decision thresholds in line with operational constraints and site-specific false positive rate targets; and to integrate explainability mechanisms into escalation rules. In addition, a before-and-after evaluation design with rolling validation and guided retraining should be applied. At the same time, real-world impact factors must be considered, including class imbalance and incomplete labels, domain shifts between sites, increased memory requirements of attention-based models, and uneven availability of data modalities. Future developments include Bayesian or conformal uncertainty calibration to enable transparent thresholding, graph-based representations of structural relationships between assets, integration of forecasts with reinforcement learning to optimise maintenance policies, and knowledge distillation for edge deployment. These directions are expected to further strengthen the link between forecasting and planning, reduce uncertainty, and enhance equipment reliability.

Acknowledgements

None.

Funding

None.

Conflict of Interest

None.

References

- [1] Abbasi, T., Lim, K.H., & Yam, K.S. (2019). Predictive maintenance of oil and gas equipment using recurrent neural network. *IOP Conference Series: Materials Science and Engineering*, 495, article number 012067. doi: 10.1088/1757-899X/495/1/012067.
- [2] Abidi, M.H., Mohammed, M.K., & Alkhalefah, H. (2022). Predictive maintenance planning for Industry 4.0 using machine learning for sustainable manufacturing. *Sustainability*, 14(6), article number 3387. doi: 10.3390/su14063387.
- [3] Bampoula, X., Siaterlis, G., Nikolakis, N., & Alexopoulos, K. (2021). A deep learning model for predictive maintenance in cyber-physical production systems using LSTM autoencoders. *Sensors*, 21(3), article number 972. doi: 10.3390/s21030972.
- [4] Carvalho, T.P., Soares, F.A., Vita, R., de Francisco, R.P., Basto, J., & Alacala, S. (2019). A systematic literature review of machine learning methods applied to predictive maintenance. *Computers & Industrial Engineering*, 137, article number 106024. doi: 10.1016/j.cie.2019.106024.

- [5] Chen, X. (2024). A novel transformer-based DL model enhanced by position-sensitive attention and gated hierarchical LSTM for aero-engine RUL prediction. *Scientific Reports*, 14(1), article number 10061. doi: [10.1038/s41598-024-59095-3](https://doi.org/10.1038/s41598-024-59095-3).
- [6] Contreras-Valdes, A., Amezquita-Sanchez, J., Granados-Lieberman, D., & Valtierra-Rodriguez, M. (2020). Predictive data mining techniques for fault diagnosis of electric equipment: A review. *Applied Sciences*, 10(3), article number 950. doi: [10.3390/app10030950](https://doi.org/10.3390/app10030950).
- [7] Cui, J., Kuang, W., Geng, K., & Jiao, P. (2025). Intelligent fault diagnosis and operation condition monitoring of transformer based on multi-source data fusion and mining. *Scientific Reports*, 15(1), article number 7606. doi: [10.1038/s41598-025-91862-8](https://doi.org/10.1038/s41598-025-91862-8).
- [8] Dairi, A., Fouzi, H., Khaldi, B., & Ying, S. (2024). Graph neural networks-based spatiotemporal prediction of photovoltaic power: A comparative study. *Neural Computing and Applications*, 37(6), 4769-4795. doi: [10.1007/s00521-024-10751-9](https://doi.org/10.1007/s00521-024-10751-9).
- [9] Deng, S., & Zhou, J. (2024). Prediction of remaining useful life of aero-engines based on CNN-LSTM-attention. *International Journal of Computational Intelligence Systems*, 17, article number 232. doi: [10.1007/s44196-024-00639-w](https://doi.org/10.1007/s44196-024-00639-w).
- [10] Fan, Y., Tang, Q., Guo, Y., & Wie, Y. (2024). BiLSTM-MLAM: A multi-scale time series prediction model for sensor data based on Bi-LSTM and local attention mechanisms. *Sensors*, 24(12), article number 3962. doi: [10.3390/s24123962](https://doi.org/10.3390/s24123962).
- [11] Fordal, J.M., Schjølberg, P., Helgetun, H., Skjermo, T., Wang, Y., & Wang, C. (2023). Application of sensor data based predictive maintenance and artificial neural networks to enable Industry 4.0. *Advances in Manufacturing*, 11(4), 248-263. doi: [10.1007/s40436-022-00433-x](https://doi.org/10.1007/s40436-022-00433-x).
- [12] Hadadi, F., Dawes, J.H., Shin, D., Bianculli, D., & Briand, L. (2024). Systematic evaluation of deep learning models for log-based failure prediction. *Empirical Software Engineering*, 29(5), article number 105. doi: [10.1007/s10664-024-10501-4](https://doi.org/10.1007/s10664-024-10501-4).
- [13] Kang, H., & Kang, P. (2024). Transformer-based multivariate time series anomaly detection using inter-variable attention mechanism. *Knowledge-Based Systems*, 290(2), article number 111507. doi: [10.1016/j.knosys.2024.111507](https://doi.org/10.1016/j.knosys.2024.111507).
- [14] Karimi, A., Wu, Y., Koyuturk, M., & French, R.H. (2021). Spatiotemporal graph neural network for performance prediction of photovoltaic power systems. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(17), 15323-15330. doi: [10.1609/aaai.v35i17.17799](https://doi.org/10.1609/aaai.v35i17.17799).
- [15] Kumar, A. (2024). Machine learning algorithms for predictive maintenance in industrial environments: A comparative study. *Journal of Artificial Intelligence General Science*, 2(1), 149-150. doi: [10.60087/jaigs.v2i1.p150](https://doi.org/10.60087/jaigs.v2i1.p150).
- [16] Leukel, J., Gonzalez, J., & Riekert, M. (2021). Adoption of machine learning technology for failure prediction in industrial maintenance: A systematic review. *Journal of Manufacturing Systems*, 61, 87-96. doi: [10.1016/j.jmsy.2021.08.012](https://doi.org/10.1016/j.jmsy.2021.08.012).
- [17] Li, T., Zhou, Z., Li, S., Sun, C., Yan, R., & Chen, X. (2022). The emerging graph neural networks for intelligent fault diagnostics and prognostics: A guideline and a benchmark study. *Mechanical Systems and Signal Processing*, 168, article number 108653. doi: [10.1016/j.ymsp.2021.108653](https://doi.org/10.1016/j.ymsp.2021.108653).
- [18] Lu, D., Zhou, X., & Li, S. (2024). Spatio-temporal attention-based hybrid deep network for time series prediction of industrial process. *Applied Intelligence*, 55, article number 169. doi: [10.1007/s10489-024-06033-5](https://doi.org/10.1007/s10489-024-06033-5).
- [19] Muthukumar, G., & Philip, J. (2024). CNN-LSTM hybrid deep learning model for remaining useful life estimation. *International Journal for Innovative Research in Multidisciplinary Field*, 10(54), 38-53. doi: [10.2015/IJIRMF/ICSETI-2024/P04](https://doi.org/10.2015/IJIRMF/ICSETI-2024/P04).
- [20] Nor, K., Pedapati, S.R., Muhammad, M., & Leiva, V. (2022). Abnormality detection and failure prediction using explainable Bayesian deep learning: Methodology and case study with industrial data. *Mathematics*, 10(4), article number 554. doi: [10.3390/math10040554](https://doi.org/10.3390/math10040554).
- [21] Nota, G., Postiglione, A., & Carvello, R. (2022). Text mining techniques for the management of predictive maintenance. *Procedia Computer Science*, 200, 778-792. doi: [10.1016/j.procs.2022.01.276](https://doi.org/10.1016/j.procs.2022.01.276).
- [22] Rodríguez, M.L., Kubler, S., de Giorgio, A., Cordy, M., Robert, J., & Le Traon, Y. (2022). Multi-agent deep reinforcement learning based Predictive Maintenance on parallel machines. *Robotics and Computer-Integrated Manufacturing*, 78(1), article number 102406. doi: [10.1016/j.rcim.2022.102406](https://doi.org/10.1016/j.rcim.2022.102406).
- [23] Serradilla, O., Zugasti, E., & Zurutuza, U. (2020). Deep learning models for predictive maintenance: A survey, comparison, challenges and prospect. *ArXiv*. doi: [10.48550/arXiv.2010.03207](https://doi.org/10.48550/arXiv.2010.03207).
- [24] Siraskar, R., Kumar, S., Patil, S., Bongale, A., & Kotecha, K. (2023). Reinforcement learning for predictive maintenance: A systematic technical review. *Artificial Intelligence Review*, 56(11), 12885-12947. doi: [10.1007/s10462-023-10468-6](https://doi.org/10.1007/s10462-023-10468-6).
- [25] Supramaniam, A., Ahmad, S.S.S., & Yusoh, Z.I.M. (2024). Predictive maintenance using deep reinforcement learning. In *Proceedings of the international conference on artificial intelligence in engineering and technology* (pp. 671-676). Kota Kinabalu: IEEE. doi: [10.1109/IICAET62352.2024.10730350](https://doi.org/10.1109/IICAET62352.2024.10730350).

- [26] Tsallis, C., Papageorgas, P., Piromalis, D., & Munteanu, R.A. (2025). Application-wise review of machine learning-based predictive maintenance: Trends, challenges, and future directions. *Applied Sciences*, 15(9), article number 4898. doi: [10.3390/app15094898](https://doi.org/10.3390/app15094898).
- [27] Wang, L., You, P., Zhang, X., Jiang, L., & Li, Y. (2025). Adaptive adjustment graph representation learning method for rotating machinery fault diagnosis under noisy signals. *Frontiers of Mechanical Engineering*, 20(1), article number 2. doi: [10.1007/s11465-024-0818-y](https://doi.org/10.1007/s11465-024-0818-y).
- [28] Wang, Y., Qian, C., & Qin, J. (2023a). Attention-mechanism based DiPLS-LSTM and its application in industrial process time series big data prediction. *Computers & Chemical Engineering*, 176, article number 108296. doi: [10.1016/j.compchemeng.2023.108296](https://doi.org/10.1016/j.compchemeng.2023.108296).
- [29] Wang, Y., Wu, M., Jin, R., Li, X., Xie, L., & Chen, Z. (2023b). Local-global correlation fusion-based graph neural network for remaining useful life prediction. *IEEE Transactions on Neural Networks and Learning Systems*, 36(1), 753-766. doi: [10.1109/TNNLS.2023.3330487](https://doi.org/10.1109/TNNLS.2023.3330487).
- [30] Yadav, D.K., Kaushik, A., & Yadav, N. (2024). Predicting machine failure using machine learning and deep learning algorithms. *Sustainable Manufacturing and Service Economics*, 3, article number 100029. doi: [10.1016/j.smse.2024.100029](https://doi.org/10.1016/j.smse.2024.100029).
- [31] Zhang, W., Yang, D., & Wang, H. (2019). Data-driven methods for predictive maintenance of industrial equipment: A survey. *IEEE Systems Journal*, 13(3), 2213-2227. doi: [10.1109/JSYST.2019.2905565](https://doi.org/10.1109/JSYST.2019.2905565).
- [32] Zheng, L., Xue, W., Chen, F., Guo, P., Chen, J., & Chen, B. (2019). A fault prediction of equipment based on CNN-LSTM network. In *Proceedings of the international conference on energy internet* (pp. 537-541). Nanjing: IEEE. doi: [10.1109/ICEI.2019.00101](https://doi.org/10.1109/ICEI.2019.00101).

Нейронні мережі для прогнозування відмов обладнання на основі неструктурованих виробничих даних

Микола Демчина

Кандидат технічних наук, доцент
Університет імені короля Данила
76018, вул. Є. Коновальця, 35, м. Івано-Франківськ, Україна
<https://orcid.org/0009-0002-9161-4843>

Анотація. Індустріальні підприємства накопичують великі обсяги неструктурованих даних з датчиків, журналів подій і технічних звітів, проте їхня цінність для прогнозування відмов часто лишається невикористаною. Метою дослідження було показати, як сучасні нейронні мережі перетворюють такі гетерогенні потоки на ранні попередження про відмови та як інтеграція моделей у систему технічного обслуговування знижує простой і підвищує надійність обладнання. Методологія поєднала уніфікований конвеєр підготовки даних (очищення, узгодження часових міток, виділення подій і семантичних ознак з текстів звітів) із порівняльною оцінкою трьох класів архітектур: згорткових мереж для вібрацій і зображень, рекурентних мереж із довгими залежностями для часових рядів та моделей на основі механізму уваги для багатомодального навчання. Основні результати були сконцентровані на якості прогнозу й ефекті для експлуатації: моделі на основі механізму уваги стабільно випереджали інші підходи за середньозваженою точністю на 5-9 відсоткових пунктів, забезпечуючи горизонт попередження 24-72 години; згорткові мережі давали найкращу чутливість на високочастотних вібраціях, тоді як рекурентні моделі краще відстежували повільні деградаційні тренди. Інтеграція найкращої моделі в інформаційну систему технічного обслуговування – з формуванням заявок і пріоритизацією робіт – зменшила тривалість простоїв на 14-21 %, підвищила середній напрацювання на відмову на 10-15 % і скоротила частку хибних тривог до рівня, прийняттого для диспетчерів. Узагальнено наявні виробничі кейси та окреслено теоретичні сценарії застосування для галузей, де емпіричні дані поки обмежені. Практична значимість полягає у відтвореному підході до обробки й інтерпретації неструктурованих даних та у доведеному внеску в зменшення простоїв і підвищення надійності парку обладнання.

Ключові слова: датчики струму; журнали подій; технічні звіти; міжканальна увага; конвеєр підготовки даних; узгодження часових міток; семантичний розбір текстів