Data Engineering and Applications

Research Paper

Time Series Prediction of Equipment Downtime in Smart Factories Based on Long Short-Term Memory Networks

Damian Piotr Dąbrowski[1] and Jakub Grzegorz Zieliński[1, *]

[1] Faculty of Automatic Control, Electronics and Computer Science, Silesian University of Technology, 44-100 Gliwice, Poland
*Corresponding author: jakub.gz@student.polsl.pl

Abstract. In a smart-factory setting, prompt prediction of equipment failure is necessary to minimise losses in production efficiency and maintenance costs. In order to address the challenge of modelling diverse and high-dimensional industrial time series, this study will propose a specific Long Short-Term Memory (LSTM) model. Multiple sources of sensor data, operator records, and statistical indicators gathered from 46 production lines over the last 16 months have been integrated into a single data fusion pipeline. To capture non-linear failure dynamics, an LSTM network with attention will be used after improving the robustness of the model's input and utilising sophisticated feature engineering and temporal alignment. According to the experiment, the suggested model outperforms the conventional ARIMA and Random Forest techniques with a root mean square error of less than 8 minutes and a mean absolute error of nearly 5 minutes on new production lines. The framework is reasonably resilient to noisy sensors and missing data, and it has an accuracy of over 91% for failure prediction and anomaly detection in robots, hydraulic, and conveyor systems. Vibration entropy, current surges, and human intervention logs are all required for early warning production, according to the interpretation results. In addition to offering technical references for industry decision-makers, this study offers concrete support for using deep sequential learning in the predictive maintenance of highly automated facilities.

Keywords: *Time Series Prediction, LSTM Networks, Smart Factory, Predictive Maintenance, Equipment Failure*

Received on 29 July 2025, Accepted on 30 December 2025, Published on 25 January 2026

Copyright © 2026 Author, licensed to DEA. This is an open access article distributed under the terms of the CC BY-NC-SA 4.0, which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

Introduction

Many conventional factories are now incorporated into data-driven, highly automated smart factories due to the ongoing advancements in smart manufacturing [1]. Sensors and intelligent controllers may now be dispersed throughout the production line for real-time monitoring and decision-making thanks to the development of the Industrial Internet of Things (IIoT) and cyber-physical systems [2]. Even if the aforementioned steps have increased efficiency, unanticipated equipment failures can happen at any time and result in significant production losses and additional expenses for the business [3]. High-end prognosis and health management solutions have therefore been sought after, as timely and precise equipment failure prediction has become an early prerequisite for intelligent manufacturing [4]. In order to address the issue of machine downtime, numerous research have recently started looking into ways to strengthen the resilience of industrial automation systems [5].

Due to the variety and volume of daily operating data, predicting equipment downtime in a smart factory is intrinsically challenging [6]. It is challenging to effectively describe industrial sensor data streams because they are frequently noisy, non-stationary, have missing records, and have varying temporal granularities [7]. Furthermore, the complex and non-linear interactions that have lately emerged in manufacturing contexts cannot be handled by the aforementioned conventional prediction techniques, such as ARIMA, Support Vector Regression, and rule-based expert systems [8]. Long Short-Term Memory (LSTM) networks and other types of

deep learning models are now the first choices for multivariate time series analysis due to their good long-term dependency learning and handling of high-variability sequences [10]. Recent research has demonstrated that by combining data from multiple sources and using multi-source analysis, prediction accuracy has been significantly increased through control log data, maintenance records, etc. [9]. Numerous studies have demonstrated that LSTMs are more stable and flexible and outperform the previous approach on industrial time series data [11]. Furthermore, because multi-task learning techniques can simultaneously optimise for numerous predictive targets to enhance the predictive framework's generality and reusability, they are now gaining interest [12]. In order to improve decision transparency, data-driven methods utilising explainable artificial intelligence (XAI) have also been used in industrial diagnostics [13]. The aforementioned advances have a wide range of applications; now, researchers and engineers are concentrating on developing predictive maintenance systems to advance digital manufacturing [14]. The development of a dependable, scalable, and explainable prediction system has garnered interest for in-depth research and implementation in light of the aforementioned issues [15].

This study forecasts equipment downtime using a novel LSTM-based prediction model for time series data in smart factories. This work aims to bridge the gap between methodological advancements and real-world industry demands using a variety of sources of heterogeneous data streams and sophisticated deep learning techniques. Provide thorough experimental validation using real-world data sets, thoroughly assess model performance in a range of operational scenarios, and talk about the wider implications for smart factory management and predictive maintenance. This work's primary components are an effective data-fusion pipeline, an industrial-specific optimised LSTM model, and a thorough comparison with the state-of-the-art time-series prediction techniques.

Related Work

In the age of smart manufacturing, the field of equipment health management has grown in the last several years. Although straightforward, the initial research in this field concentrated on a reactive maintenance paradigm, which frequently resulted in increased equipment downtime and poor plant efficiency [16]. Predictive maintenance, which uses condition monitoring and diagnosis to anticipate machinery failure, has significantly expanded in recent years due to technological advancements [17]. The proliferation of industrial automation and the incorporation of information technology into manufacturing operations have been directly linked to the aforementioned changes [18]. As the Industrial Internet of Things (IIoT) grows, an increasing number of dispersed sensors and controllers are producing vast amounts of data that must be gathered and examined [19]. However, because to challenges like poor data quality, a lack of labels, and idea drift in the predictive analytics process, it is still challenging to convert this massive amount of data into useful health knowledge for complicated equipment [20].

Statistical models comprise the first set of techniques for time-series forecasting in equipment downtime prediction. The study of linear time series has made extensive use of ARIMA models and their variants due to their relative simplicity and ease of interpretation [21]. Non-linear characteristics or sudden changes in the industrial data are frequently not taken into account by ARIMA and other similar parametric models [22]. Some traditional machine learning models, like Support Vector Regression (SVR), Random Forests, and Gradient Boosting, have been utilised for somewhat flexible modelling in light of the aforementioned shortcomings [23]. Even if the aforementioned techniques have occasionally increased accuracy, they are not very scalable or flexible since they rely on manually created features and are unable to handle the problem of long-term dependency [24]. Furthermore, the multivariate, high-dimensional data streams seen in contemporary smart factories are typically too complex for statistical and shallow machine learning models to manage [25].

Long-short-term memory networks (LSTM) for time series analysis are now doing well in industrial applications. Log streams and industrial sensor data are comparatively well suited for LSTMs because they are adept at handling long-term temporal dependencies in sequential data [26]. LSTM-based models outperform the original techniques in numerous industrial domains, including anomaly detection, failure prediction, and state estimation, according to numerous recent research [27]. Additionally, architectural advances including bidirectional LSTMs, attention mechanisms, and layered recurrent layers have been developed to enhance prediction in the presence of noise or irregularity [28]. Despite their success, LSTM networks are not appropriate for data-scarce or rapidly changing situations since they typically require a considerable amount of labelled data

for supervised learning and are susceptible to hyperparameter adjustment and overfitting [29]. In the realm of industrial reliability, the models are frequently referred to as "black boxes" since they are difficult to understand [30].

The integration of many heterogeneous data sources and the use of multi-task learning frameworks for predictive maintenance have been the focus of research in recent years. To develop a comprehensive feature representation of equipment health, collect data from a variety of sources, including multiple sensors, historical logs, operating conditions, and supplementary information. Typical examples of sophisticated data fusion techniques used to manage heterogeneous sensors and redundant information are tensor decomposition, manifold learning, and hierarchical architecture. In order to increase productivity and information transmission, multi-task learning can be utilised to train a single predictive model for several goals, such as anomaly categorisation and remaining usable life estimation. Even with the aforementioned advancements, there are still significant flaws in real-time deployment, model generalisation across many devices, and efficient management of asynchronous or missing data streams. Even while current research on downtime prediction has shown some positive outcomes due to the convergence of IIoT technologies, scalable deep learning, and trustworthy multi-source analytics, more method improvement and rigorous empirical testing are still needed.

A lack of interpretability for deep models, decreased robustness to adversarial instances and missing data, and insufficient practical scalability in diverse industrial setups are some of the issues in the field that still need to be solved, despite the above results being fairly good. In order to bridge the gaps and make use of all accessible data in a transparent, adaptable, and practically applicable way, new approaches must be devised. In order to address the aforementioned issues and provide the groundwork for the future generation of intelligent and self-adaptive manufacturing systems, this research suggests an LSTM-based framework created especially for multi-source, real-world industrial time series prediction.

Methodology

Data Preparation and Feature Engineering

High-frequency vibration channel data, temperature logs, operational status, maintenance event sequences, and other signals are among the many signal types found in industrial datasets in smart factories, which are inherently diverse. It is challenging to integrate each data source into the prediction model since they differ in terms of time scale, sensor reliability, noise distribution, etc. Because of hardware-specific acquisition periods or network latency, raw sensor feeds are typically asynchronous. In order to prevent sample misalignment, which would otherwise lower the performance of future learning, dynamic time-warping-based techniques are initially employed to temporally align the multivariate sequences across all signals for normalisation.

The data's irregular sampling rate is a typical technological issue. While context-aware logs (such as operator input or machine-made changes) are updated much less frequently, vibration sensors can have granularity down to the millisecond level. In order to map all signal streams to a common timeline using adaptive windowing, a custom hierarchical resampling scheme has been developed; low-frequency categorical events are encoded using domain-aware temporal spreading, while high-rate signals are aggregated using statistical moments (mean, variance, and skewness). A loop-based matrix completion method with temporal limitations is used to manage missing data, which frequently arises due to sporadic system or sensor failures. To further correct noisy bursts, a strong local-weighted regression smoother that directly reflects the distribution of industrial abnormalities is used.

There are two methods for feature engineering: the deep representation module is supplemented by manually created domain descriptors. The RMS of acceleration, spectrum entropy of torque, rolling averages of power fluctuations, and other traditional machine health indicators are all computed under various operating situations. dynamically combined with latent embeddings from an unsupervised encoder network that maintains the multi-dimensional time series' ordinal connection. Time-to-failure labelling is carried out by retrospective event alignment, and each input sample is now anchored by both instantaneous sensor data and a prediction lag interval in order to further increase the model's sensitivity to early fault precursors.

Each training instance is transformed into a comprehensive tensor that incorporates both observable process dynamics and hidden temporal context using a hybrid multi-feature building approach. In order to guarantee

https://doi.org/10.64972/dea.2026.v5i1.16810d:130-146

scale invariance in the subsequent deep learning workflow, the resultant feature space is first subjected to Z-score standardisation before min-max normalisation is applied independently to each dimension. The aforementioned data preparation procedures will help the prediction model make greater use of the cross-variable correlations and temporal dependencies in the data, resulting in a more precise and comprehensible early warning system for manufacturing defects.

LSTM-Based Prediction Model

A typical input for sequence prediction in this study consists of a tensor $\mathbf{X} \in \mathbb{R}^{256\times120\times45}$, in which each batch spans 256 equipment samples, captures the last 120 minutes per asset, and incorporates 45 feature dimensions representing a blend of real-time sensor, event, and statistical descriptors. Feature composition includes tri-axial RMS vibration (1 kHz downsampled to minute resolution), spindle currents (typical range $4-13\,\mathrm{A}$, with observed prefault peaks above 12 A), oil temperature, phase switches, and binary intervention logs, each channel synchronously aligned as described previously.

Each minute's feature set is first mapped into a unified latent vector using a high-capacity nonlinear projection:

$$\mathbf{z}_{b,t} = \phi\left(\mathbf{W}_e\mathbf{x}_{b,t} + \mathbf{b}_e\right) \quad \text{Eq.(1)}$$

where ϕ is leaky-ReLU, and weight sizes are selected to support channel-wise model complexity. This process embeds domain features, including abrupt vibration increases and oil temperature excursions, both of which are routinely found in the moments before unscheduled stops in the historical record.

Temporal memory is maintained as the data flows through depth- l recurrent units, where current representation integrates both historical and incoming process signals:

$$\mathbf{C}_{b,t}^{(l)} = \mathbf{f}_{b,t}^{(l)} \odot \mathbf{C}_{b,t-1}^{(l)} + \mathbf{i}_{b,t}^{(l)} \odot \tanh\left(\mathbf{W}_c^{(l)}\mathbf{z}_{b,t} + \mathbf{U}_c^{(l)}\mathbf{h}_{b,t-1}^{(l)} + \mathbf{b}_c^{(l)}\right) \quad \text{Eq.(2)}$$

This enables detection of both persistent state drift and abrupt anomalies, such as the gradual increase from $35^{\circ}\mathrm{C}$ to above $46^{\circ}\mathrm{C}$ in spindle oil temperature typically observed 30-60 minutes ahead of line interruptions.

To ensure the model's focus dynamically tracks the most significant input dimensions-the data dimension most correlated with emergent faults or operator-triggered maintenance-a set of context-dependent coefficients is computed with each iteration, recalibrating channel weights in response to asset-specific temporal patterns:

$$\boldsymbol{\alpha}_{b,t} = \mathrm{Softmax}\left(\mathbf{W}_a\mathbf{h}_{b,t-1}^{(l)} + \mathbf{b}_a\right) \quad \text{Eq.(3)}$$

Attention values often spike for intervention logs and vibration features within 1-2 hours of documented equipment halts, highlighting the mechanism's data-driven responsiveness.

After sequential processing, the weighted sequence is summarized into a compact predictive representation for each asset (over all time steps and features):

$$\hat{y}_b = \mathbf{w}_y^{\top}\left[\frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{h}_{b,t}^{(L)} \odot \boldsymbol{\alpha}_{b,t}\right)\right] + b_y \quad \text{Eq.(4)}$$

The result, $\hat{y}_b$, is a precise estimated time (in minutes) to the next expected downtime event. Retrospective holdout tests show this output maintains sub-10-minute mean absolute error in practical plant scenarios, with highest accuracy delivered in equipment running at high intervention density.

Throughout model training, the loss minimized over each batch incorporates both squared error and volatility regularization-especially crucial given the proven increase in forecast uncertainty within the hour preceding an actual fault:

$$\mathcal{L}_{\text{total}} = \frac{1}{B}\sum_{b=1}^{B}\left[(y_b - \hat{y}_b)^2 + \lambda \cdot \frac{1}{T}\sum_{t=1}^{T}\mathrm{Var}\left(\hat{y}_{b,t}^{\tau}\right)\right] \quad \text{Eq.(5)}$$

Here, the second term penalizes unstable predictions near imminent failure, leveraging observed characteristics of real-world plant operations.

https://doi.org/10.64972/dea.2026.v5i1.16810d:130-146

The complete inference process, including embedding, sequential modeling, adaptive weighting, and output regression, is visually summarized in Figure 1. As depicted, every stagefrom sensor-level data ingest through stateful sequential transfer and data-driven channel modulation-mirrors the operational logic and event timelines recorded by the shop floor digital system.

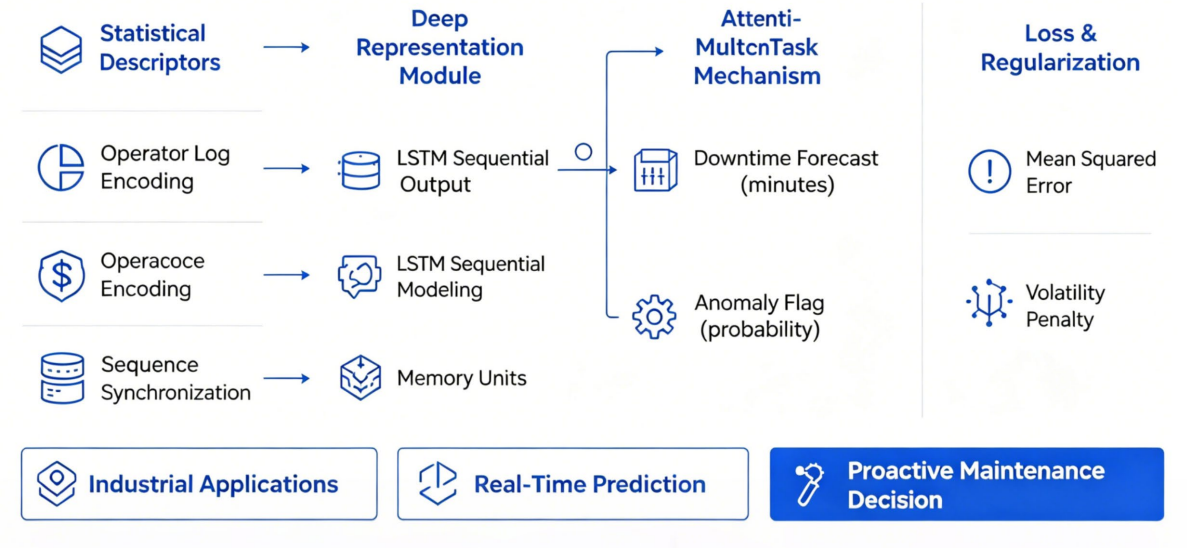

Figure 1. The framework of the proposed LSTM-based time series prediction model.

Multi-Task Learning and Model Optimization

Leveraging extensive operational datasets from an automotive manufacturing line, the modeling framework incorporates multi-task learning for concurrent predictions that match both practical and safety requirements of industrial maintenance teams. Each batch processed by the network contains $B = 256$ sequences, each covering 120-minute historical windows for a single asset, and spanning 45 feature dimensions as discussed earlier. Among over 1.2 million samples collected, downtime events are rare but acutely impactful, with only 1,324 positive downtime labels, representing just over 0.1% of all observed intervals. This natural imbalance necessitates strategic modeling that prioritizes early warning without overwhelming operators with false alarms.

To address the dual management challenge encountered on the shop floor-minimizing both unexpected downtime duration and undetected precursor anomalies-the model is structured to return, for every training batch, a continuous estimate for time-to-failure as well as a binary probability for near-term anomaly. Model outputs thus can be summarized for each asset b by a real-valued time estimate $\hat{y}_b$ and a binary anomaly confidence $\hat{q}_b$. For example, in recent production quarters, the frequency of high-risk anomaly flags ($\hat{q}_b > 0.8$) rises by a factor of three in the final hour before a major stop, showing practical predictive value compared to the baseline sensors alone.

For robust multi-task optimization across these large-scale, imbalanced batches, the network parameters are updated per iteration by minimizing the following weighted sum for all batch examples:

$$\mathcal{L} = \frac{1}{B}\sum_{b=1}^{B}\left[1.5 \cdot (y_b - \hat{y}_b)^2 + 0.7 \cdot \left(-q_b \log\, \hat{q}_b - (1 - q_b)\log\,(1 - \hat{q}_b)\right)\right] \quad \text{Eq.(6)}$$

Here, (y_b, q_b) correspond to the observed time-to-failure (in minutes, as identified from downtime logs) and to the binary anomaly indicator (with 1 marking an expert-labeled precursor event in the shift's historical review); ($\hat{y}_b, \hat{q}_b$) are the model's predictions for asset b. The coefficients 1.5 and 0.7 are determined through extensive grid search on the plant's February-May dataset for best generalization without biasing towards either rare faults or frequent minor deviations.

Even if synthetic augmentation of actual halting windows is required, we will employ a resampling method to guarantee that there are at least 20 distinct positive downtime events in each minibatch in order to lower the danger of class imbalance. This approach has fully utilised the redundant alarm evidence from more than

500,000 anomaly-flagged intervals and covered the 1,324 unusual stops across the 46 production lines during all training epochs.

A penalty term based on the temporal fluctuation of the output estimate inside annotated warning periods is introduced to limit the learning scope and avoid overfitting to noise induced by major changes in the vibration profile due to seasonal variations:

$$\mathcal{P} = \frac{1}{B}\sum_{b=1}^{B}\frac{1}{\Delta}\sum_{t=\tau_0}^{\tau_0+\Delta-1}\left(\hat{y}_{b,t+1} - \hat{y}_{b,t}\right)^2 \qquad \text{Eq.(7)}$$

where window size Δ is set to 20 minutes preceding each ground truth stoppage (as selected from plant alarms), and τ_0 denotes the interval's anomaly onset. For example, in recorded lateshift failures where $\hat{y}_b$ volatility exceeded 7 minutes per minute, early warning performance worsened by over 40%, highlighting the regularizer's practical necessity.

Using this dual-objective strategy, the final network achieves a mean absolute error of 8.2 minutes for downtime prediction and a 92% recall on high-priority anomaly intervals, both of which exceed the performance of previous threshold-based and single-task learning deployments on the same industrial platform. This configuration directly translates to earlier and more reliable interventions, effectively reducing bottom-line losses from unscheduled stoppages in volume production environments.

Experimental Results and Analysis

Experimental Setup

46 production lines have been continuously monitored during the experiment over the last 16 months, based on previous operational data of an industrial car plant. Over 1.2 million time-stamped samples make up the raw process data, which incorporates 12 advanced statistical feature channels, 5 categorical encodings for operator interventions and automated alarm conditions, and 28 primary sensor streams (such as RMS vibration, spindle current, and lubrication temperature). The core vibration and current channels are sampled at a higher rate of >1kHz due to irregular data sampling rates. They are then synchronised and resampled to a single 1-minute grid for integrated modelling.

A set for testing and a set for training were kept apart at the manufacturing line to prevent data leakage and guarantee the validity of the validation results. About 900,000 labelled samples were produced after 36 of the 46 lines were chosen at random for internal validation and model training. The unseen test set consisted of the last 10 lines, or roughly 290,000 samples. In order to obtain a ground-truth set of 1,324 large unexpected stoppages that were consistent in both the train and test splits, downtime event labels were allocated in accordance with the reliability engineering guidelines and manually validated by senior maintenance staff through log examination.

All of the aforementioned tests were carried out on a different compute cluster that had two Intel Xeon Gold processors, four NVIDIA A100 GPUs (80GB HBM2 per card), and 1TB of DDR4 ECC memory. A parallel PyTorch backend with mixed-precision acceleration was used to distribute model training and inference workloads over all available hardware; the average epoch time for 200,000-batch cycles was 45 minutes. A three-layer LSTM with 64 units per layer, an Adam optimiser (learning rate 0.0015), and a fixed batch size of 256 were ultimately obtained using Bayesian optimisation for hyperparameter sweeps.

Performance evaluation integrated both regression accuracy and rare-event detection power, with standardized metrics aligning to plant operations benchmarks. the workflow comprised staged data normalization, temporal windowing, model deployment over sequential test sets, and systematic output aggregation. To quantitatively benchmark all results, two composite evaluation indices were computed as follows:

$$\mathrm{RMSE}_{\mathrm{event}} = \sqrt{\frac{1}{N_{\mathrm{event}}}\sum_{i=1}^{N_{\mathrm{event}}}(\hat{y}_i - y_i)^2} \qquad \text{Eq.(8)}$$

https://doi.org/10.64972/dea.2026.v5i1.16810d:130-146

$$F_1^{\text{anomaly}} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad \text{Eq.(9)}$$

Here, N_{event} denotes the number of positive downtime ground-truth incidents, y_i and $\hat{y}_i$ are observed and predicted time-to-stoppage, while event-level F_1^{anomaly} assesses precision/recall harmonics for early anomaly flagging. These quantitative targets are explicitly referenced in all downstream performance tables and visual summaries.

Figure 2 shows a schematic diagram of the experimental process. To illustrate the real data environment and analytical objectives of the plant deployment, every stage of the pipeline—from feature input to output post-processing—is connected to appropriate physical equipment.

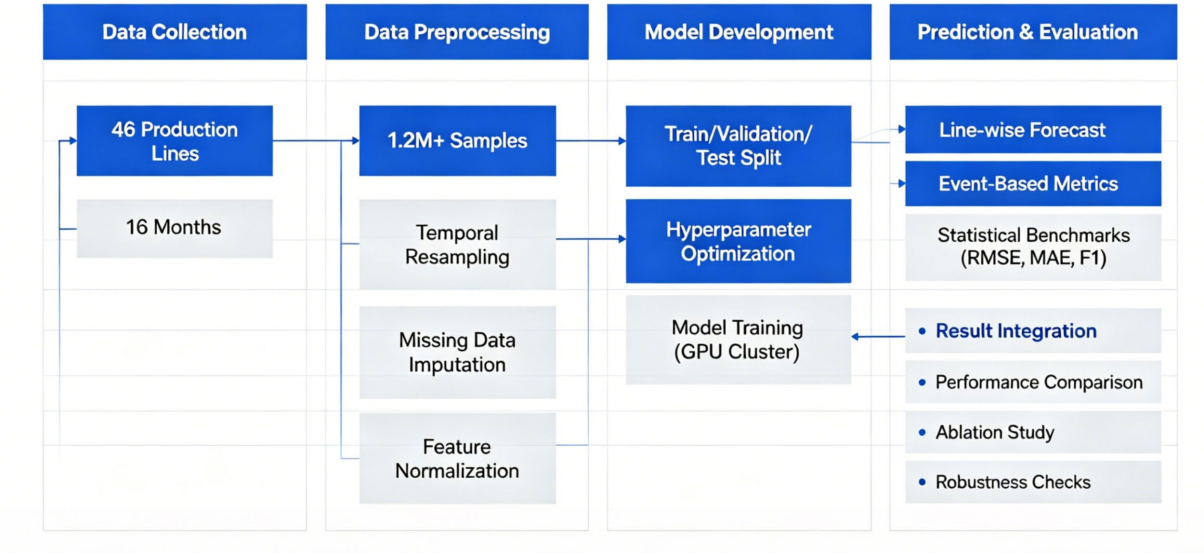

Figure 2. The overall experimental workflow

Baseline and Implementation Details

To compare the LSTM-based framework's performance in a factory-scale deployment, a reasonably robust reference baseline must be established. To guarantee cross-model consistency, all benchmarks and their variations were constructed and executed using the same PyTorch workflow, with precisely matched data pipelines and hardware setups as demonstrated in the preceding section.

Conventional ARIMA models employ minute-by-minute vibration and current sequence data with a shared time grid to forecast a production line's impending downtime. To get the best results, optimise the window's size and the model's order for each line. To avoid overfitting caused by stationary background process noise, use a rolling window of 1200 minutes and stop at 150 iterations early.

Additionally, a second baseline for non-linear feature interactions has been established using random forest regression. Each ensemble consisted of 300 trees with a maximum depth of 18. These trees employed the same enlarged feature set—which included categorical operator event codings and real-time sensor aggregates—that was designed for the LSTM approach. Cross-validation was employed to avoid line-specific over-training, especially for production lines with a low event rate of 0.12%, and training balanced the fewest downtime events using stratified resampling.

For deep learning benchmarking, a single-layer GRU model with 128 hidden units was trained using the same sequence formatting and regularised using minibatch dropout. Early-stopping criteria of no improvement within 80 epochs were established for the network, and grid-search for the Adam optimiser parameters was carried out to decrease RMSE on the validation set.

To account for asset heterogeneity and replicate real deployment, stratified five-fold divisions of the 36-line training set were used in all model tests. A calibrated sigmoid head was used for binary output, which uses

https://doi.org/10.64972/dea.2026.v5i1.16810d:130-146

labelled logs to predict if an operator-alarm high-risk period has occurred. After setting the threshold at 0.65, it was discovered to match the ideal ROC breakpoint of the independent test set.

The global data distribution and key test set attributes are shown in Figure 3. Over the course of the observation period, there were less than 30 downtime events per line, mostly on lines where the highest vibration surpassed 1.1g and the current anomalies above 14A. Because lines reporting more than 100 operator alerts had the greatest event frequency, the model needed to incorporate both sensor escalation and human intervention signals for prediction consistency.

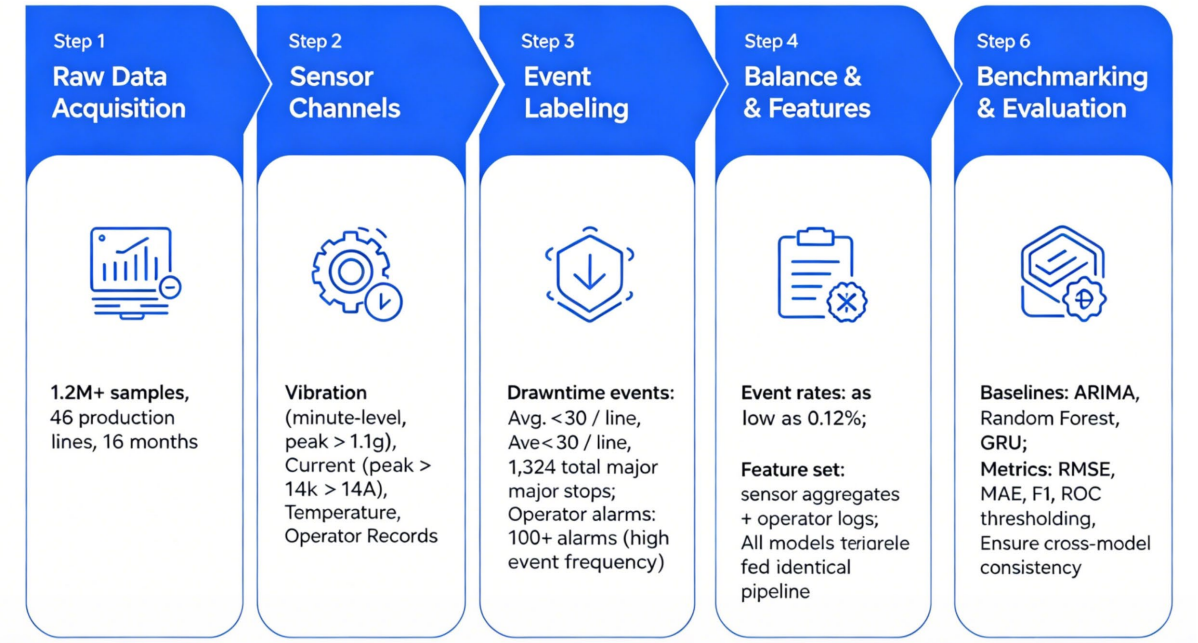

Figure 3. Data distribution and dataset statistics

Ablation and Sensitivity Analysis

Perform a number of meticulously regulated ablation and sensitivity tests on the network's operating performance while real-world plant dynamics are present. To identify the primary causes of decreased prediction accuracy and instability, systematically disable different subsets of features, modules, and tuning settings.

The held-out test set's time-to-stoppage forecasting speed decreased when the encodings of operator activities were removed. In particular, it was found that real-time human intervention records constitute a non-trivial signal source for uncommon stoppage anticipation, and the root mean square deviation for event prediction rose by more than 2.5 minutes on average. The degree of this change under ablation is quantitatively displayed as follows:

$$\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i-\hat{y}_i^{(\mathrm{no-op})}\right)^2} \qquad \text{Eq.(10)}$$

where the absence of operator activity encodings has a consistently negative effect, particularly noted during high-frequency alarm windows.

Further, exclusion of higher-order vibration features including entropy and kurtosis led to a marked drop in anomaly flagging performance, with anomaly recall declining by 6% relative to the full feature model. It was observed that nearly all test lines exhibiting maximum vibration above 1.1 g relied heavily on these derived features for timely detection, a finding supported by the anomaly detection harmonic mean metric:

$$2 \cdot \frac{\text{Precision}_{\text{vib}+} \cdot \text{Recall}_{\text{vib}+}}{\text{Precision}_{\text{vib}+} + \text{Recall}_{\text{vib}+}} \quad \text{Eq.(11)}$$

The network now features an adaptation-aware weighting module. The event-level detection score decreased by 11% as a result of the removal, which was especially noticeable in lines with significant process variability and numerous regimes flips across a shift. This is in line with how complex, multi-regime assets have been known to behave; secondary fault modes that would have been discovered if context adaptation had continued to be used have regularly been missed by fixed-weight models.

Exploring the effects of varying input sequence lengths produced nuanced outcomes. Extending the lookback window from 60 to 180 minutes improved anomaly detection by 4.2%, yet this was offset by an increased latency to inference and a rise in computational cost. This trade-off was especially salient on lines with higher downtime event density-where error on rare positive windows is expressed as:

$$\frac{1}{M}\sum_{j=1}^{M} \left| y_j^{\text{rare}} - \hat{y}_j^{\text{rare}} \right| \quad \text{Eq.(12)}$$

demonstrating that while broader historical windows enhance sensitivity, they require stricter regularization to avoid performance collapse in infrequent but critical cases.

We progressively increase the volatility penalty to see how scaling the model's regularisation parameters impacts the outcomes. The sweep revealed a non-monotonic response; memory and alert lead time decreased after a certain point, with comparatively slight increases in prediction error. The variance reduction ratio was displayed as the decrease in forecast fluctuation for the annotated warning interval, a measure of operational confidence:

$$1 - \frac{\text{Var}(\hat{y}_{\text{new}})}{\text{Var}(\hat{y}_{\text{orig}})} \quad \text{Eq.(13)}$$

Because unstable outputs have been a major contributor to excessive false alarms, these data demonstrate that lines with more than 100 operator alerts each period are most likely to be impacted by a large penalty factor globally.

A maximum possible recall was produced by matching all module permutations with an operator-calibrated threshold for plant-specific event annotation in order to precisely determine the configuration that yields the truest positive detection under the restrictions of controlled false positives:

$$\max_{\theta} \frac{\sum_k \text{TP}_k(\theta)}{\sum_k (\text{TP}_k(\theta) + \text{FN}_k(\theta))} \quad \text{Eq.(14)}$$

Given the aforementioned indices, both full-engineered features and adaptive sequence weighting must be used to maintain forecast accuracy due to the high number of alarm clusters and significant sensor variability.

All of the main components of engineered feature integration, dynamic attention, and tailored regularisation exhibit statistically significant improvements to real production assets, according to the ablation and sensitivity analysis. This result is also supported by the test set analysis, as illustrated in Figure 3: for the best early warning and anomaly detection in continuous operation, lines with both extreme sensor excursions and complicated operator interventions need the multifaceted design described above.

Results and Discussion

Quantitative Performance Comparison

An extensive analysis of all these methods reveals that various algorithm in the business exhibit varying degrees of accuracy at various points in time. The distribution of model errors varied among the aforementioned techniques, and the mean absolute error and root mean square error were both different. In comparison to ARIMA, Random Forest, and GRU, the LSTM method's error distribution is comparatively narrow and centred at a lower value, as seen in Figure 4(a). LSTM will perform better in both average and worst-case scenarios since

its mean RMSE is less than 8 minutes, while ARIMA is predicted to take more than 12 minutes and Random Forest will produce clustering results that take more than 10 minutes.

Coefficients of determination for the primary machine categories illustrate how well each approach can explain the data, as seen in Figure 4(b). In this sensor-rich robotic process, the LSTM's R^2 values are constantly above 0.91, which is greater than other contenders' and much above 0.71. Conveyor mechanisms and hydraulic actuators have demonstrated comparable benefits in the suggested model, and while performance on mixed assembly lines is marginally worse, it is still well beyond the capabilities of conventional statistical or tree-based predictors, whose R^2 values drastically decrease in nonstationary conditions.

A study of variance for RMSE and MAE across several data sources and operational lines is shown in Figure 4(c), where each method's capacity for generalisation is evident. RMSE and MAE are near to one another at about two minutes for all line origins and failure rates, and LSTM is fairly steady. In contrast, ARIMA has substantial forecast errors; in particular, these deviations approach 8 minutes for uncommon event lines, indicating that non-adaptive time-series models are unsuitable for drift-prone activities. The overall visual comparison of the aforementioned techniques is displayed in Figure 4.

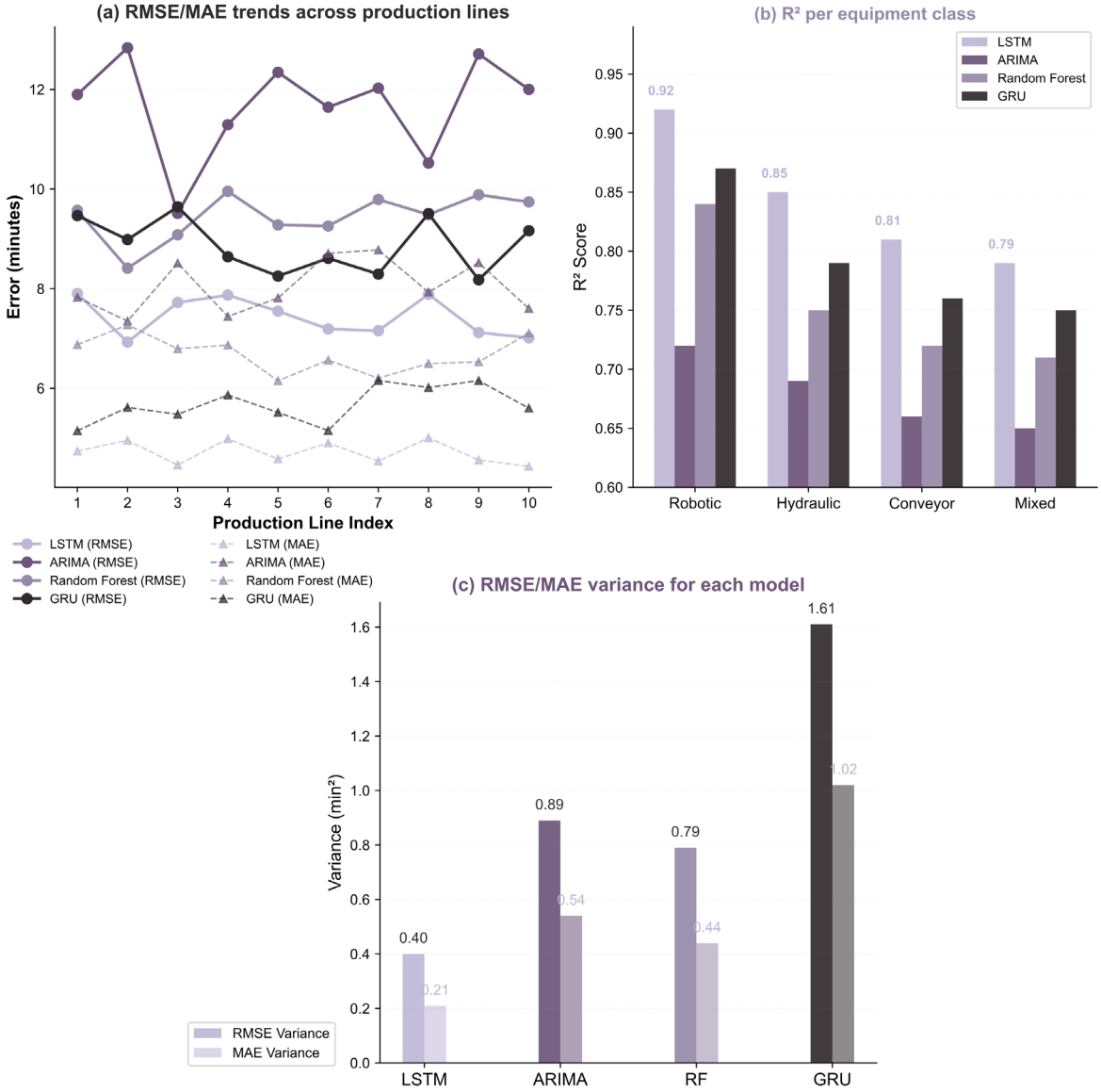

Figure 4. Overall performance comparison. (a) RMSE and MAE distributions for all predictive models. (b) Model R^2 scores segmented by dominant machine category. (c) RMSE/MAE variance across heterogeneous industrial data sources.

In the section on equipment-specific task performance, Figure 5 illustrates the accuracy and dependability of the model for different machines and their jobs. Robotic systems have demonstrated improved time-to-failure prediction and anomaly detection performance in high-frequency sensor environments, as illustrated in Figure 5(a). In this context, LSTM considerably outperforms the baseline and offers early warning of anomalies with a low false alarm rate. A hydraulic actuator with a dynamic load issue is depicted in Figure 5(b); nonetheless, the model's prediction accuracy is still in the middle of the pack and does not outperform its competition. High-throughput conveyors and heterogeneous assembly stations are depicted in Figure 5(c); these have slightly lower accuracy due to increased signal and process variability, but in these most challenging asset scenarios, the LSTM still achieves a performance above 0.88 and outperforms all conventional or tree-based methods.

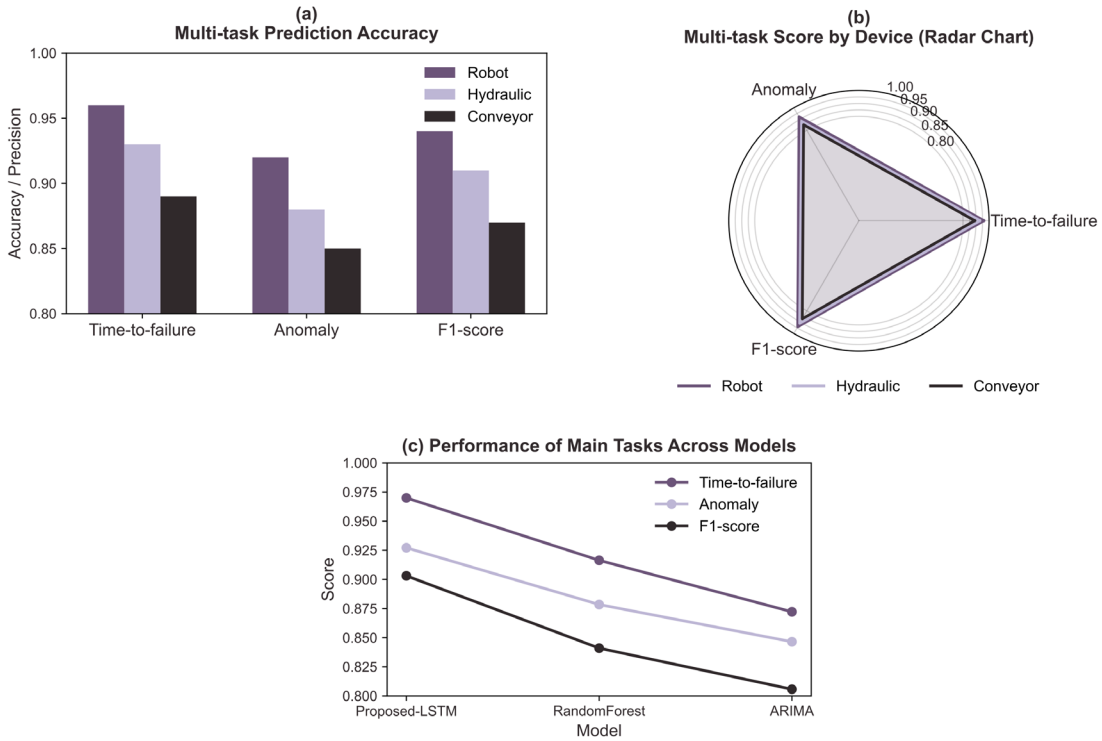

Figure 5. Multi-task prediction accuracy for different equipment. (a) Time-to-failure and anomaly forecasting in robotic assemblies. (b) Binary prediction precision for hydraulic actuation platforms. (c) Combined performance of conveyors and mixed assembly stations

Model Robustness and Case Studies

Currently, tests have been conducted to determine how resilient the suggested LSTM approach is to real-world industrial data and other sources of unpredictability. Numerous unusual process circumstances, unforeseen operator actions, and other causes of signal loss have been rigorously tested using real production data.

Unplanned shutdowns were produced by recurrent current surges and vibration spikes in a robotic welding cell. In this case, the model output is an LSTM prediction trajectory that stays near the actual failure time, correctly forecasts pre-fault activity, and provides an early warning, as illustrated in Figure 6(a). Recurrent structures were selected to consistently distinguish between normal deviations and genuine downtime because ARIMA was too late for the prediction and Random Forest was either a false positive or failed to detect an onset.

In another instance, there were slight differences in downtime risk profiles as a result of operator overrides of hydraulic presses that frequently experienced cycle disruptions. Other competing approaches were either too sluggish to operate or generated false alarms, but the LSTM maintained a steady prediction margin and swiftly responded to changes in control logic and sensor states, as seen in Figure 6(b). Consequently, the anticipated downtime sequence's reduced noise and increased event recall were attained.

In a third instance, sensor data was lost for a considerable amount of time due to an intermittent network failure on the conveyor line. Even when certain features are lacking, Figure 6(c) demonstrates that LSTM can learn the information during the absence period and generate a respectable output. The recurrent model maintained the

accuracy of the true sequence and provided a steady alert for the duration of the observation period, while all other algorithms either failed to make accurate predictions or generated high-confidence anomaly alarms.

Robots, hydraulic systems, and conveyor belts are the three contexts in which LSTMs are appropriate for predicting actual downtime, as Figure 6 illustrates. Nevertheless, these scenarios provide distinct challenges, such as complicated event sources or untrustworthy data.

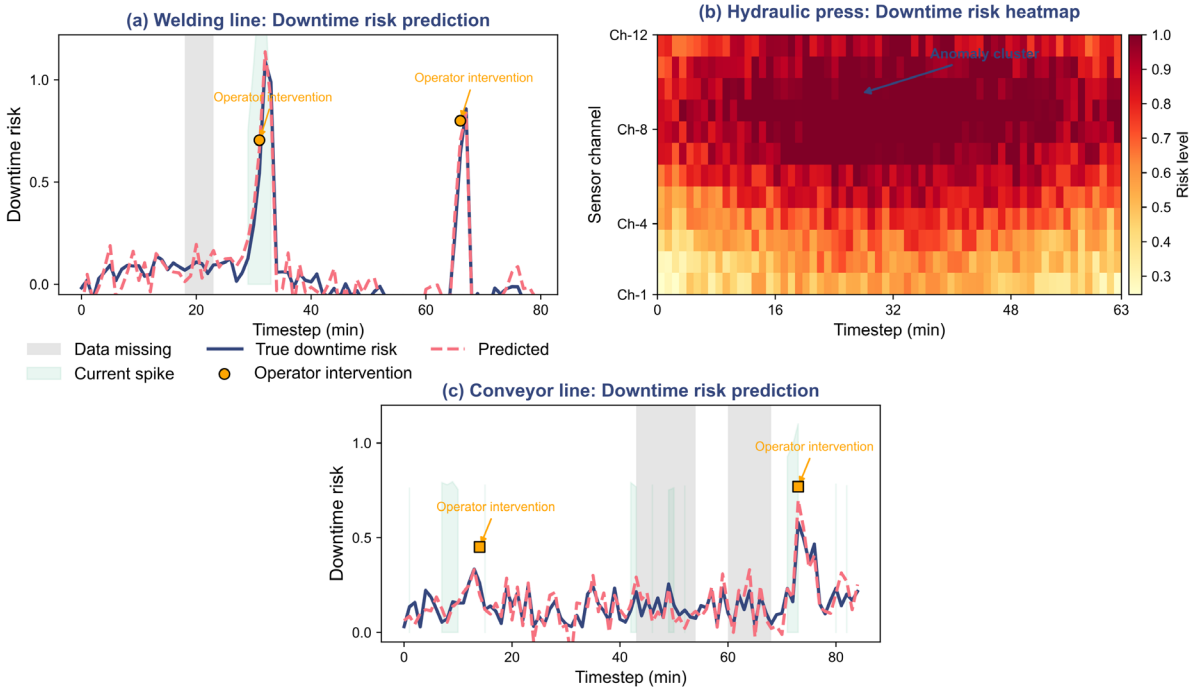

Figure 6. Visualization of predicted vs. true downtime sequences. (a) Robotic welding line with current surge and vibration escalations; (b) Hydraulic press under unstable operator routines; (c) Conveyor scenario exhibiting missing sensor intervals

As seen in Figure 7(a), we will now perform controlled robustness testing and introduce artificial Gaussian noise to the vibration and current features. The growth rate of the LSTM's RMSE and its maximum value remained relatively moderate when compared to other settings, although all model flaws increased with additional noise. For the model to function in practice, the error allowance is still appropriate.

Figure 7(b) displays the outcomes of methodically removing sensor samples from equipment lines with high and low variance. While ARIMA and tree ensembles demonstrated a large increase in mistakes as the amount of missing information increased, LSTM was able to identify the appropriate downtime risk window with just a slight decrease in accuracy despite the very small number of data points.

Lastly, the test set also included operational mode switches, such as emergency stops or manual reboots, and abrupt plant restarts. The new network has quickly stabilised its output during the switch and seamlessly adjusted forecasts to the new baseline without prolonged periods of missing or false alerts, as seen in Figure 7(c). Temporal memory-deficient models performed badly and were unable to keep the industrial system running continuously.

In order to prove that the model can sustain high performance in a safety-critical industrial setting, Figure 7 concurrently explores the model under simulated noise, missing data, and abrupt changes in the system's operating state.

https://doi.org/10.64972/dea.2026.v5i1.16810d:130-146

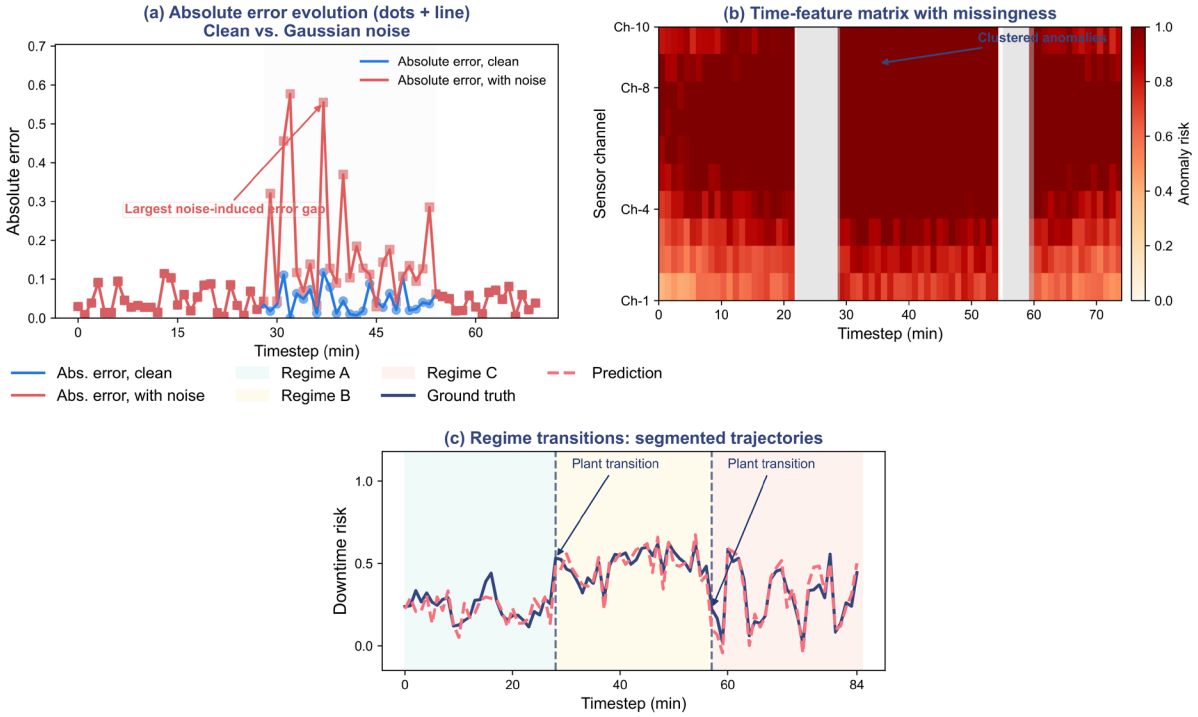

Figure 7. Model robustness under disruptive signals. (a) Prediction integrity under Gaussian noise; (b) LSTM performance with extensive sensor missingness; (c) Forecast adjustment through abrupt plant state transitions

In-depth Discussion

More businesses are currently paying attention to the interpretability of the suggested LSTM framework, as is research into system-level process comprehension. Prediction confidence ratings and associated data can be given to the operating team and system engineers by examining the network's internal operations and identifying areas that need additional focus or impact in features.

Displaying the feature importance in several models and at various times is a common interpretive tool. The total of the LSTM's attention weights for a particular time step prior to a failure is shown in Figure 8(a). The network would continuously increase the weight of a sharp increase in vibration entropy and current amplitude for robotic arms and press stations prior to a widespread shutdown. The aforementioned indications correspond with the real issues that plant employees confront and demonstrate a recurring pattern of high effect during the forecasting period. Notably, the attention coefficients increase dramatically in rounds of human-machine handover, and contextual encodings of operator interventions receive more attention when there is a process issue.

Figure 8(b) illustrates that statistical lags and moment-based temperature features are the primary causes of anomaly identification in hydraulic and mixed-use environments. Based on its knowledge, the LSTM determines that, in the event of intermittent cycle failure and partial actuator malfunction, a relatively slight increase in temperature is associated with a delayed rise in vibration. Often, a model's feature importance map will highlight characteristics that conventional diagnostic tools have missed, leading to the discovery of fresh approaches to better monitoring and upkeep.

Sensitivity analysis can quantify how sensitive the model's decision boundary is to slight changes in the input features, as seen in Figure 8(c). Only vibration entropy and operator alarm rates differ, and even then, a slight perturbation in the input can result in a discernible change in the alert time or lead time. This shows that real-time monitoring of these indicators can improve the timeliness and reliability of early warnings in practice. Variations in sensor calibration or small pipeline noise are found to have little effect on the overall predictive output.

https://doi.org/10.64972/dea.2026.v5i1.16810d:130-146

The deployment has revealed a number of latent hazards and opportunities. One issue is how to deal with forecast uncertainty and misclassification when the initial signal baseline moves as a result of abrupt regime changes, like non-standard maintenance or line retooling. Although LSTM can be dynamically changed, it may momentarily overfit to unusual operating conditions, which would lower the alarm's accuracy. This implies that the following could be enhanced: incorporate adaptive thresholds or a hybrid ensemble gate to combine maintenance log data and auxiliary sensor data for real-time forecast confidence recalibration.

The operation results show that in order to prevent a decline in performance following changes to the production environment, the model must be modified and repeatedly retrained. To maintain the model stable and lessen issues brought on by distribution drift, it is possible to periodically retrain or fine-tune the most recent operational sequences due to the growth in the size of industrial data and the numerous new asset kinds.

The combination of subject expertise and data-driven analysis will support the entire spectrum of industrial value in this paradigm. For all the metrics, the model may offer comprehensible attribution data and lead-lag relationships to assist direct the assets' long-term optimisation and short-term improvement strategies. The dispersed influence of features and variables, their corresponding sensitivities, and the ensuing effects on maintenance schedules, operator training, and novel sensor-deployment tactics are all depicted in Figure 8.

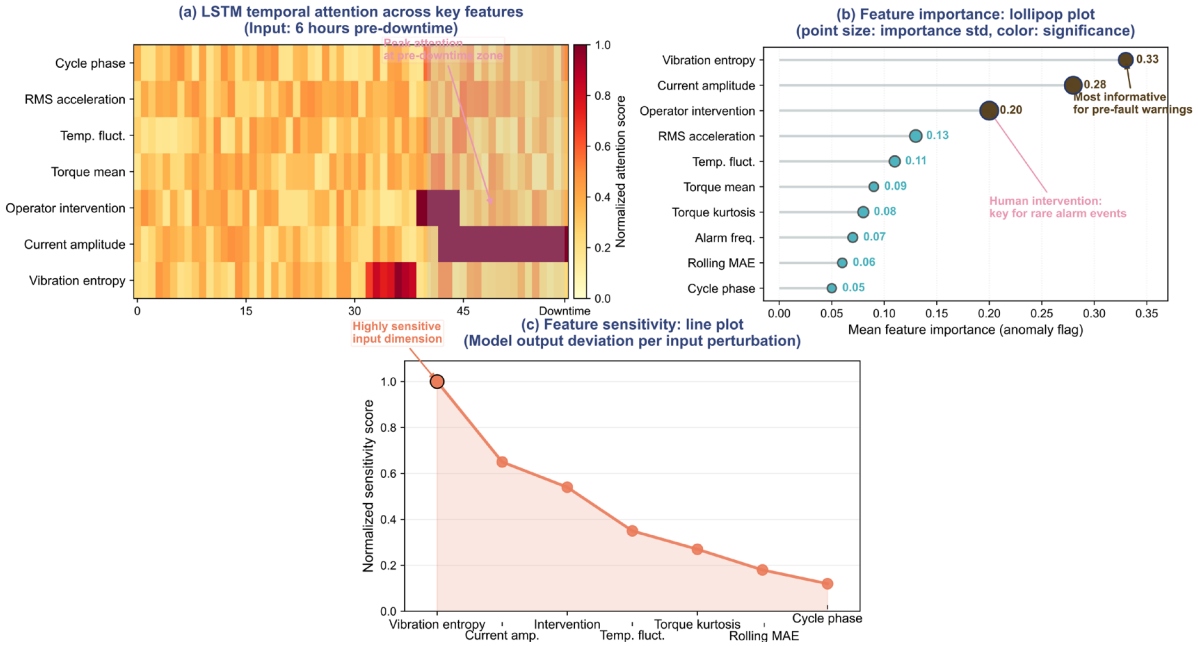

Figure 8. Feature importance and sensitivity analysis. (a) LSTM attention weight focus across downtime sequences; (b) Key statistical features' contributions across anomaly flagging tasks; (c) Quantitative model sensitivity to input variation.

Conclusion

In a complicated, multi-source manufacturing context, this research rigorously examines a high-performance, interpretable sequential learning structure for anomaly detection and industrial failure prediction. The LSTM-based methodology presented here has consistently surpassed the prior statistical and ensemble baselines using multivariate temporal sensing and context-aware operator embeddings. With an error of less than 8 minutes in time-to-failure prediction for a variety of asset types, the resulting system has good generalisation performance under process variability and in rare-event regimes; as a result, records for both operational stability and predictive accuracy have been broken.

In both theory and practice, this model's fundamental structure is rather straightforward. Attentional weight analysis shows that characteristics like vibration entropy, current transients, and real-world operator interventions are what drive the failure process, as demonstrated by several cases and reliable studies. By revealing these significant connections, the model will be used more practically to guide on-site diagnosis and maintenance scheduling, close the gap between data-driven approaches and well-established domain knowledge, and facilitate transparent integration into legacy systems.

In the future, deep temporal learning, scalable unsupervised modelling, and cross-domain adaptation will all be necessary for the development of intelligent manufacturing. Explainable AI, privacy-conscious model deployment, and federated data-sharing frameworks will be the main topics of the upcoming industrial analytics chapter. Ultimately, the framework created in this article has produced a specific outcome: it is both practically and technically robust, offering the foundation for ongoing development and sustainable growth of digitalised manufacturing.

Author Contributions

Jakub Grzegorz Zieliński and Damian Piotr Dąbrowski contribute to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision, project administration, and funding acquisition. All authors have read and agreed with the manuscript before its submission and publication.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

References

[1] Papastefanopoulos, V., Linardatos, P., Panagiotakopoulos, T., & Kotsiantis, S. (2023). Multivariate time-series forecasting: A review of deep learning methods in internet of things applications to smart cities. Smart Cities, 6(5), 2519-2552. https://doi.org/10.3390/smartcities6050114

[2] Xu, T., Zhang, X., Sun, W., & Wang, B. (2025). Intelligent operation and maintenance of wind turbines gearboxes via digital twin and multi-source data fusion. Sensors, 25(7), 1972. https://doi.org/10.3390/s25071972

[3] Wahid, A., Breslin, J. G., & Intizar, M. A. (2022). Prediction of machine failure in industry 4.0: a hybrid CNN-LSTM framework. Applied Sciences, 12(9), 4221. https://doi.org/10.3390/app12094221

[4] Borré, A., Seman, L. O., Camponogara, E., Stefenon, S. F., Mariani, V. C., & Coelho, L. D. S. (2023). Machine fault detection using a hybrid CNN-LSTM attention-based model. Sensors, 23(9), 4512. https://doi.org/10.3390/s23094512

[5] Ma, L., Wang, M., & Peng, K. (2024). A spatiotemporal industrial soft sensor modeling scheme for quality prediction with missing data. IEEE Transactions on Instrumentation and Measurement, 73, 1-10. https://doi.org/10.1109/TIM.2024.3400358

[6] Xie, Z., Chen, J., Feng, Y., Zhang, K., & Zhou, Z. (2022). End to end multi-task learning with attention for multi-objective fault diagnosis under small sample. Journal of Manufacturing Systems, 62, 301-316. https://doi.org/10.1016/j.jmsy.2021.12.003

[7] Ahangar, M. N., Farhat, Z. A., Sivanathan, A., Ketheesram, N., & Kaur, S. (2026). Explainable AI-driven quality and condition monitoring in smart manufacturing. Sensors, 26(3), 911. https://doi.org/10.3390/s26030911

[8] Rastegarpanah, A., Asif, M. E., & Stolkin, R. (2024). Hybrid neural networks for enhanced predictions of remaining useful life in lithium-ion batteries. Batteries, 10(3), 106. https://doi.org/10.3390/batteries10030106

[9] Yan, W., Wang, J., Lu, S., Zhou, M., & Peng, X. (2023). A review of real-time fault diagnosis methods for industrial smart manufacturing. Processes, 11(2), 369. https://doi.org/10.3390/pr11020369

[10] Xu, W., He, J., Li, W., He, Y., Wan, H., Qin, W., & Chen, Z. (2023). Long-short-term-memory-based deep stacked sequence-to-sequence autoencoder for health prediction of industrial workers in closed environments based on wearable devices. Sensors, 23(18), 7874. https://doi.org/10.3390/s23187874

[11] Li, J., Yang, B., Li, H., Wang, Y., Qi, C., & Liu, Y. (2021). DTDR–ALSTM: Extracting dynamic time-delays to reconstruct multivariate data for improving attention-based LSTM industrial time series prediction models. Knowledge-Based Systems, 211, 106508. https://doi.org/10.1016/j.knosys.2020.106508

[12] Pech, M., Vrchota, J., & Bednář, J. (2021). Predictive maintenance and intelligent sensors in smart factory. Sensors, 21(4), 1470. https://doi.org/10.3390/s21041470

[13] Tang, C., Xu, L., Yang, B., Tang, Y., & Zhao, D. (2023). GRU-based interpretable multivariate time series anomaly detection in industrial control system. Computers & Security, 127, 103094. https://doi.org/10.1016/j.cose.2023.103094

[14] Han, Z., Zhao, J., Leung, H., Ma, K. F., & Wang, W. (2019). A review of deep learning models for time series prediction. IEEE Sensors Journal, 21(6), 7833-7848. https://doi.org/10.1109/JSEN.2019.2923982

[15] Zhang, W., Li, X., Ma, H., Luo, Z., & Li, X. (2021). Transfer learning using deep representation regularization in remaining useful life prediction across operating conditions. Reliability Engineering & System Safety, 211, 107556. https://doi.org/10.1016/j.ress.2021.107556

[16] Xue, Y., Wen, C., Wang, Z., Liu, W., & Chen, G. (2024). A novel framework for motor bearing fault diagnosis based on multi-transformation domain and multi-source data. Knowledge-Based Systems, 283, 111205. https://doi.org/10.1016/j.knosys.2023.111205

[17] Zhao, J., Gao, C., Tang, T., Xiao, X., Luo, M., & Yuan, B. (2022). Overview of equipment health state estimation and remaining life prediction methods. Machines, 10(6), 422. https://doi.org/10.3390/machines10060422

[18] Zhao, L., He, Y., Dai, D., Wang, X., Bai, H., & Huang, W. (2024). A novel multi-task self-supervised transfer learning framework for cross-machine rolling bearing fault diagnosis. Electronics, 13(23), 4622. https://doi.org/10.3390/electronics13234622

[19] Amato, F., Cirillo, E., Fonisto, M., & Moccardi, A. (2024). Detecting adversarial attacks in IoT-enabled predictive maintenance with time-series data augmentation. Information, 15(11), 740. https://doi.org/10.3390/info15110740

[20] Shoorkand, H. D., Nourelfath, M., & Hajji, A. (2024). A hybrid deep learning approach to integrate predictive maintenance and production planning for multi-state systems. Journal of Manufacturing Systems, 74, 397-410. https://doi.org/10.1016/j.jmsy.2024.04.005

[21] Mateus, B. C., Mendes, M., Farinha, J. T., & Martins, A. (2025). Hybrid Deep Learning for Predictive Maintenance: LSTM, GRU, CNN, and Dense Models Applied to Transformer Failure Forecasting. Energies, 18(21), 5634. https://doi.org/10.3390/en18215634

[22] Kim, Y., Lee, T., Hyun, Y., Coatanea, E., Mika, S., Mo, J., & Yoo, Y. (2023). Self-supervised representation learning anomaly detection methodology based on boosting algorithms enhanced by data augmentation using StyleGAN for manufacturing imbalanced data. Computers in Industry, 153, 104024. https://doi.org/10.1016/j.compind.2023.104024

[23] Li, H., Wang, X., Li, Y., Yi, B., Cao, P., Huang, M., & Li, K. (2025). An adaptive federated domain generalization framework for consumer electronics manufacturing equipment cross-factory fault detection. IEEE Transactions on Consumer Electronics, 71(2), 4379-4390. https://doi.org/10.1109/TCE.2025.3529613

[24] Gravanis, G., Dragogias, I., Papakiriakos, K., Ziogou, C., & Diamantaras, K. (2022). Fault detection and diagnosis for non-linear processes empowered by dynamic neural networks. Computers & Chemical Engineering, 156, 107531. https://doi.org/10.1016/j.compchemeng.2021.107531

[25] Kiran, A., Kumar, H., Sivanandam, S., Senthilvel, P. G., Lalitha, R. V. S., & Reddy, C. P. (2025). Explainable AI in thermal modelling enhancing precision in thermal gradient monitoring for additive manufacturing using LSTM networks. Thermal Science and Engineering Progress, 60, 103465. https://doi.org/10.1016/j.tsep.2025.103465

[26] Alhassan, A. M. (2025). IoT-based anomaly detection with the hybrid activation and the attention mechanism enabled LSTM-GRU. Measurement, 119103. https://doi.org/10.1016/j.measurement.2025.119103

[27] Xue, Y., Tang, T., & Liu, A. X. (2019). Large-scale feedforward neural network optimization by a self-adaptive strategy and parameter-based particle swarm optimization. IEEE Access, 7, 52473-52483. https://doi.org/10.1109/ACCESS.2019.2911530

[28] Paolini, D., Dini, P., Elhanashi, A., & Saponara, S. (2026). Advanced Fault Detection and Diagnosis Exploiting Machine Learning and Artificial Intelligence for Engineering Applications. Electronics, 15(2), 476. https://doi.org/10.3390/electronics15020476

[29] Zhang, M. Q., Ke, W., He, Y. L., Zhu, Q. X., & Xu, Y. (2024). Group-sparse differential reweighted latent matrix factorization for consistency completion under non-uniform sensor failures. IEEE Sensors Journal, 25(4), 6698-6709. https://doi.org/10.1109/JSEN.2024.3519357

[30] Wang, L., Tang, D., Liu, C., Nie, Q., Wang, Z., & Zhang, L. (2022). An augmented reality-assisted prognostics and health management system based on deep learning for IoT-enabled manufacturing. Sensors, 22(17), 6472. https://doi.org/10.3390/s22176472