Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026
DOI : https://doi.org/10.33395/sinkron.v10i3.16211
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Intelligent Fault Diagnosis in Multi-Setpoint Water Level Systems Using LSTM-Autoencoder

Muhammad Giriarda Abrari[1)*], Fitria Suryatini[2)], Hasbi Fajrul Hakim[3)]
[1,2,3)]Departement of Automation Engineering, Politeknik Manufaktur Bandung, Indonesia
[1)]giriarda.abrari@polman-bandung.ac.id, [2)]fitria@ae.polman-bandung.ac.id, [3)]hasbifajrul03@gmail.com

Submitted :May 13, 2026 | **Accepted** : May 18, 2026 | **Published** : July 5, 2026

Abstract: Fault detection in multi-setpoint industrial process control systems is complicated by the fact that normal sensor behavior shifts substantially across operating points, making conventional threshold-based alarms and single-condition models unreliable when setpoints change frequently. Recent studies on LSTM-Autoencoder for industrial anomaly detection have demonstrated promising results, yet most are evaluated under fixed operating conditions and do not examine how feature engineering choices affect detection performance across diverse setpoints. This study aims to determine whether physics-informed derived features improve LSTM-AE fault detection performance in a real-time multi-setpoint water level control system, and whether the improvement holds under practical deployment conditions. The proposed framework augments seven raw PLC sensor readings with three derived variables: delta flow, level error, and frequency-per-flow and applies a per-setpoint windowing strategy to prevent cross-setpoint data contamination during training. An ablation study compares the eleven-feature model against a seven-feature baseline under three labeling scenarios reflecting varying preprocessing quality. The eleven-feature model achieves an AUC of 1.0000 and F1-score of 0.9993 under onset-cut evaluation, and reduces the false positive rate from 18.48% to 15.21% under corrected labeling while maintaining perfect recall. Real-time validation across thirty fault injection experiments confirms a 100% detection rate with a mean latency of 6.37 ± 2.04 seconds, 38.2% faster than the baseline. These results confirm that derived features meaningfully improve both classification quality and temporal detection performance, though adaptive thresholding at high-variability setpoints remains an open challenge for future work.

Keywords: LSTM-Autoencoder; Fault Detection; Multi-Setpoint Control; Feature Engineering; Anomaly Detection

INTRODUCTION

Accurate liquid level and flow control are essential in industrial processes to ensure efficiency and product quality, as precise regulation directly determines process stability and operational safety (Nugraha et al., 2025). Within these architectures, Programmable Logic Controllers (PLC) are implemented to provide precise regulation of actuators based on real-time sensor feedback, enabling the system to maintain stability and performance across multiple operating setpoints required by dynamic production demands (Nugraha et al., 2025). As industrial operations move toward continuous and uninterrupted processes, early detection of faults particularly leakage through bypass valves or pipeline breaches has become a high-priority requirement to prevent setpoint deviation, pump overload, and cascading system failures (Adhitya & Ihsan, 2025; Essamlali et al., 2024).

Existing approaches to fault detection in process control systems range from static threshold alarms to supervised classification models, yet both categories share a common limitation when applied to multi-setpoint operation (Cui et al., 2018; Karadimos & Anthopoulos, 2023). Static thresholds defined at one operating point produce excessive false alarms or missed detections when the system transitions to a different setpoint, as normal sensor behavior shifts considerably with the target level (Nizam et al., 2022). Supervised classifiers, while more adaptive, require labeled fault samples from each operating condition, which are rarely available in operational settings where fault events are infrequent by design (Kim & Heo, 2022). Unsupervised learning approaches that train only on normal data offer a more practical alternative, but their ability to maintain consistent detection

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

performance across multiple operating conditions without cross-contamination between setpoints during training remains an open challenge (Belay et al., 2023; Goetz & Humm, 2023).

Machine learning-based approaches domain-specific derived features particularly the Long Short-Term Memory Autoencoder (LSTM-AE), have demonstrated strong capabilities for unsupervised anomaly detection in multivariate time-series data (Abdennebi et al., 2023; Belay et al., 2023; Wei et al., 2023). LSTM-AE operates by training exclusively on normal operating data, learning to reconstruct expected sensor patterns, and identifying fault conditions through elevated reconstruction error eliminating the need for labeled fault samples during training (Aslam et al., 2024). This approach has been validated across multiple domains including water leak detection (Kammoun et al., 2023; Lee et al., 2023), indoor air quality monitoring (Wei et al., 2023), smart water metering systems (Nelago Kanyama et al., 2024), industrial pipeline operations , and fault diagnosis in industrial rotating machinery (Khalid Fahmi et al., 2024; Qu et al., 2025; Sunal et al., 2022). However, most existing studies evaluate models under fixed single setpoint or stable repetitive operating conditions and do not systematically compare the contribution of domain specific derived features against raw sensor inputs, nor do they examine the sensitivity of performance metrics to labeling quality at the fault onset boundary (Soni et al., 2022; Zhao et al., 2024).

The research gap addressed in this study is threefold. First, no existing work has proposed a per-setpoint windowing strategy that prevents cross-setpoint data contamination during LSTM-AE training for multi-setpoint process control systems. Second, the role of physics-informed derived features such as differential flow, level error, and frequency to flow ratio in improving model separability has not been systematically isolated from raw sensor features through a dedicated ablation study (Kim & Heo, 2022). Third, a unified evaluation framework that assesses LSTM-AE performance under varying labeling scenarios, particularly the distinction between the pre fault steady state phase and genuine fault onset, is absent from the current literature (Belay et al., 2023; Soni et al., 2022).

This study proposes a real-time fault detection framework based on LSTM-AE for a multi-setpoint water level control system operating at six setpoints from 10 to 60 cm. The main contributions are: (1) a per-setpoint windowing methodology that enables a single LSTM-AE model to generalize across all operating setpoints without cross-contamination; (2) a physics-informed feature engineering approach introducing three derived variables delta flow, level error, and frequency-per-flow ratio to enhance fault discriminability; (3) a systematic ablation study comparing LSTM-AE trained on eight baseline features (comprising seven raw sensor inputs and one operating setpoint) against LSTM-AE trained on eleven features that incorporate the three additional derived variables, evaluated under three labeling scenarios that reflect varying data preprocessing quality; and (4) real-time deployment validation through thirty online fault injection experiments measuring detection latency and system recovery time across all setpoints.

LITERATURE REVIEW

The adoption of unsupervised deep learning for anomaly detection in industrial systems has grown substantially in recent years, driven by the practical constraint that labeled fault data is rarely available in real operational environments (Aslam et al., 2024; Belay et al., 2023). Among the architectures proposed for this purpose, the Long Short-Term Memory Autoencoder (LSTM-AE) has emerged as a widely adopted approach due to its ability to model complex temporal dependencies in multivariate sensor data and detect deviations through reconstruction error (Khalid Fahmi et al., 2024; Wei et al., 2023). Applications of LSTM-AE have been reported across diverse industrial domains, ranging from water distribution networks and environmental monitoring to pipeline operations and process control systems (Adhitya & Ihsan, 2025; Kammoun et al., 2023; Nelago Kanyama et al., 2024). Table 1 summarizes representative recent studies related to the proposed method and identifies the gaps that motivate the present work.

Table 1
State of the Art

Study	Data / Context	Method	Contribution	Limitation
(Wei et al., 2023)	Indoor air quality, multivariate time-series sensors	LSTM-AE with reconstruction loss-based threshold	Demonstrated LSTM-AE effectiveness for multivariate time-series anomaly detection; threshold determined from reconstruction loss rate achieving 99.50% accuracy	Evaluated under single fixed environmental condition; no multi-setpoint generalization
(Kammoun et al., 2023)	Water distribution network, hydraulic sensor data	Unsupervised LSTM-AE with multithresholdi ng per zone	Applied LSTM-AE for leak detection without labeled fault data; threshold defined per measurement point in water network	Single operating condition per zone; no cross-condition evaluation or

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026
DOI : https://doi.org/10.33395/sinkron.v10i3.16211
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Study	Data / Context	Method	Contribution	Limitation
				derived feature analysis
(Nelago Kanyama et al., 2024)	Smart water metering, monthly consumption data	LSTM-AE with data augmentation	Addressed small dataset limitation in water anomaly detection; improved model performance through synthetic data generation	Monthly temporal resolution; not applicable to real-time PLC-based control systems
(Aslam et al., 2024)	Industrial control systems (ICS), multivariate sensor	Improved Autoencoder with z-score threshold	Proposed statistical threshold (mean ± std) from training reconstruction error; reduced false positive rate in ICS anomaly detection	Evaluated on benchmark ICS dataset under fixed operating conditions; no multi-setpoint or real-time deployment testing
(Khalid Fahmi et al., 2024)	Gas turbine power plant, operational sensor data	DLSTM-Autoencoder for industrial fault detection	Deployed LSTM-AE in real industrial environment for turbine fault detection with real-time inference	Domain specific architecture; not evaluated across multiple operating setpoints or with physics-derived features
(Adhitya & Ihsan, 2025)	Oil and gas pipeline, multivariate operational data	Single-layer vs multi-layer LSTM-AE with sliding window	Compared architecture depth effect on anomaly detection; single-layer LSTM-AE achieved better stability and lower anomaly rate in repetitive pipeline operations	No comparison of raw vs derived sensor features; single stable operating condition without multi-setpoint validation

The synthesis from Table 1 reveals three consistent gaps across existing work. First, all reviewed studies evaluate their models under fixed or single operating conditions, leaving the challenge of multi-setpoint generalization unaddressed. Second, none of the studies systematically isolates the contribution of physics-informed derived features against raw sensor inputs through an ablation study within the same experimental framework. Third, the effect of data labeling quality at the fault onset boundary on model evaluation metrics has not been examined, despite its practical significance in real operational deployments where the pre-fault steady-state phase is commonly mislabeled. This study directly addresses these three gaps through per-setpoint windowing, feature ablation comparison, and multi-scenario evaluation on a PLC-based water level control plant.

METHOD

This study proposes a real-time fault detection framework based on LSTM-Autoencoder for a multi-setpoint water level control system. The overall system integrates five functional layers as illustrated in Fig. 1: sensor data acquisition, PLC-based process control, middleware data routing, machine learning inference, and fault status output. Sensor readings are acquired by the PLC Omron CP1H, transmitted to a PostgreSQL database via KEPServerEX using the FINS and OPC UA protocols, processed by the Python-based ML engine, and the resulting fault status is written back to KEPServerEX for monitoring purposes.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v10i3.16211 p-ISSN : 2541-044X

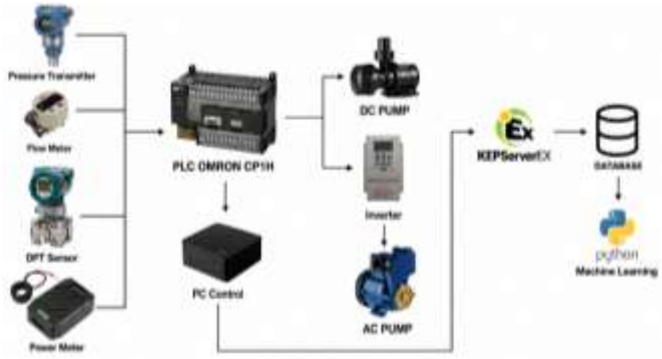

Fig. 1 System Architecture

1. Research Design

This study applies an experimental design combining offline model training and online deployment validation. The offline phase covers data collection, preprocessing, feature engineering, model training, threshold determination, and comparative evaluation. The online phase involves deploying the trained model in a real-time environment and measuring detection latency through thirty fault injection experiments across all six setpoints. The overall research workflow is presented in Fig. 2.

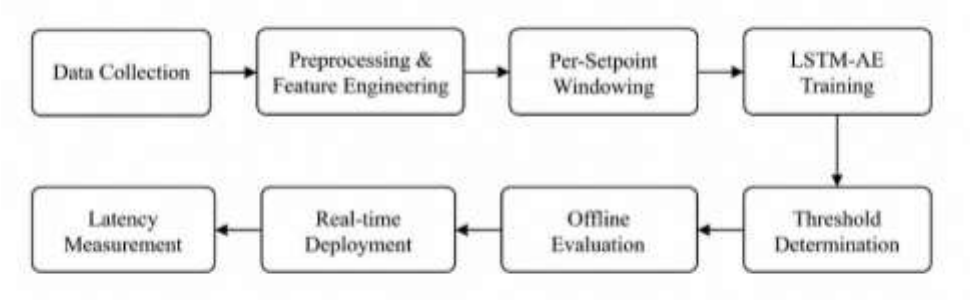

Fig. 2 Research workflow

2. Data Acquisition

Data were collected from a laboratory-scale water level control plant equipped with four sensor types: a pressure transmitter, two flow meters (inlet and outlet), a differential pressure transmitter for level measurement, and a power meter. The PLC Omron CP1H samples all sensor signals at a rate of 0.5 seconds per reading and stores integer-encoded values in a PostgreSQL database table via KEPServerEX. Data were collected across six operating setpoints 10, 20, 30, 40, 50, and 60 cm under two conditions normal operation and fault condition simulated by opening a bypass valve after the system reached steady state. Table 2 summarizes the dataset composition.

Table 2
Dataset Summary

Setpoint (cm)	Normal Samples	Fault Samples
10	7.048	6.576
20	7.052	5.542
30	10.542	13.325
40	12.921	12.952
50	7.893	6.032
60	6.999	6.512
Total	52.455	50.939

3. Data Preprocessing and Feature Engineering

Raw sensor readings consist of seven variables flow in, flow out, frequency, level, pressure, current, and power. Three additional derived features are computed from the raw measurements to capture physical relationships within the system:

$$\text{delta\_flow} = \text{flowin} - \text{flowout} \tag{1}$$

$$\text{level\_error} = \text{level} - \text{setpoint} \tag{2}$$

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026
DOI : https://doi.org/10.33395/sinkron.v10i3.16211
e-ISSN : 2541-2019
p-ISSN : 2541-044X

$$\text{freq\_per\_flow} = \frac{\text{frekuensi}}{\text{flowout} + \varepsilon}, \quad \varepsilon = 1.0 \tag{3}$$

The variable ε is set to 1.0 to prevent division by zero when flow out reads zero at low setpoints a condition observed at SP 10 cm due to sensor resolution limitations. A setpoint variable is also appended, yielding eleven features in total for the proposed model. The seven-feature variant retains only the raw sensor readings without derived variables, serving as the ablation baseline.

All features are normalized using StandardScaler fitted exclusively on the training partition of normal data. To prevent cross-setpoint contamination where a sliding window spans data from two different operating conditions windowing is performed per setpoint prior to concatenation, as illustrated in Fig. 3. Each setpoint's data is split 80/20 for training and validation before merging into the final training and validation sets.

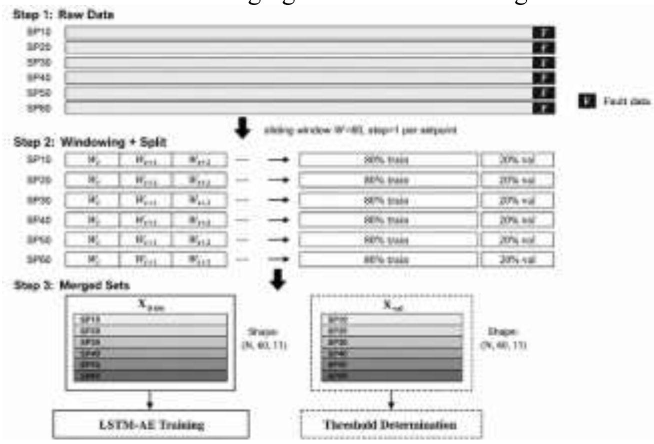

Fig. 3 Per-setpoint windowing strategy

4. Sliding Window Configuration

Time-series segmentation employs a sliding window of length W = 60 timesteps with a step size of 1, corresponding to a 30-second observation window at 0.5-second sampling intervals. This window length was selected to encompass sufficient temporal context for the LSTM layers to capture the dynamic behavior of the PID controller response while remaining computationally feasible for real-time inference. The resulting input tensor shape for model training is (samples, 60, 11) for the eleven-feature model and (samples, 60, 7) for the seven-feature baseline.

5. LSTM-AE Model Architecture

The proposed model follows an encoder-decoder architecture where both components are implemented using stacked LSTM layers, as depicted in Fig. 4. The encoder compresses the input window into a latent representation through two LSTM layers with 64 and 32 units respectively. The bottleneck representation is then replicated across the time dimension using a Repeat Vector layer and decoded by two symmetric LSTM layers with 32 and 64 units. A Time Distributed Dense layer reconstructs the output to match the input dimensionality.

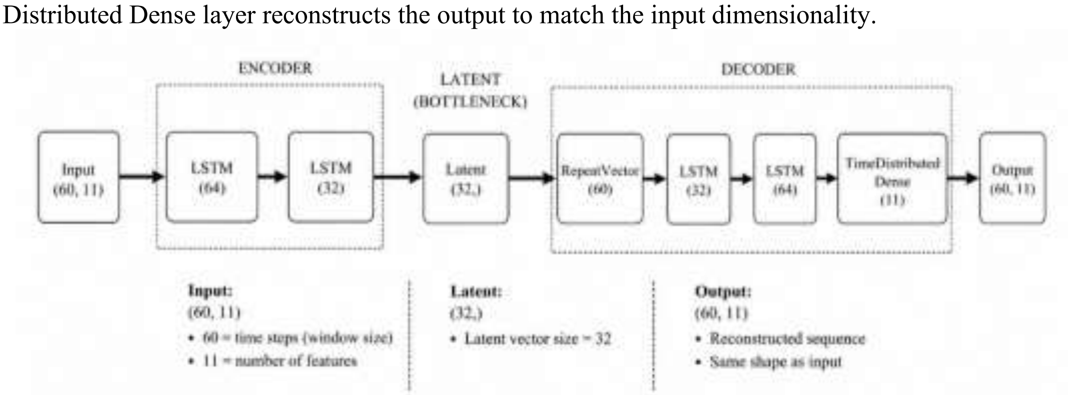

Fig. 4 Structure of LSTM-Autoencoder

The model is trained using the Adam optimizer with mean squared error (MSE) as the reconstruction loss function. Early stopping with patience of 10 epochs monitors validation loss to prevent overfitting, and ReduceLROnPlateau with a factor of 0.5 and patience of 5 epochs is applied to adaptively reduce the learning rate.

*name of corresponding author

DOI : https://doi.org/10.33395/sinkron.v10i3.16211
e-ISSN : 2541-2019
p-ISSN : 2541-044X

6. Threshold Determination

The anomaly threshold is determined from the distribution of reconstruction errors on the validation set using the statistical 3-sigma rule (Aslam et al., 2024):

$$\text{Threshold} = \mu_{\text{val}} + 3\sigma_{\text{val}} \tag{4}$$

where μ_val and σ_val represent the mean and standard deviation of per-window MSE values computed from the validation normal data. This formulation ensures that approximately 99.73% of normal reconstruction errors remain below the threshold under the assumption of a Gaussian error distribution, limiting the theoretical false alarm rate to 0.27% (Aslam et al., 2024; Belay et al., 2023). The threshold is computed independently for each model variant and stored for deployment.

7. Evaluation Framework

Model performance is evaluated under three labeling scenarios designed to assess the sensitivity of detection metrics to data preprocessing quality at the fault onset boundary, as illustrated in Fig. 5. Each scenario reflects a different assumption about how the pre-fault steady-state phase is handled prior to evaluation.

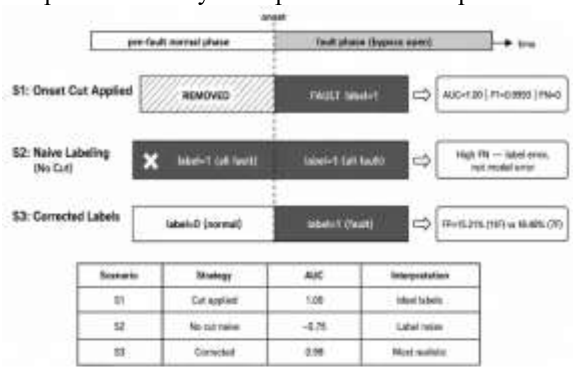

Fig. 5 Three labeling scenarios applied to fault recordings

Scenario 1 (S1) applies automatic onset detection to the fault recording, where a fixed number of rows preceding the detected onset point are removed before window construction. This ensures that all windows labeled as anomalous contain only confirmed fault behavior, producing the cleanest possible label assignment for evaluation.

Scenario 2 (S2) uses the unprocessed fault recording without any onset cutting, with every row assigned an anomaly label regardless of whether it originates from the pre-fault or post-fault phase. This represents a naive labeling approach and serves as a stress test to expose the consequence of ignoring the transient boundary between normal and fault states.

Scenario 3 (S3) applies corrected labeling to the uncut fault recording: rows preceding the detected onset receive a normal label, while rows following the onset receive an anomaly label. This scenario represents the most operationally realistic condition, as the full recording is preserved but labels are assigned according to the actual system state at each timestep.

Performance metrics computed consistently across all three scenarios are defined as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{5}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{6}$$

$$F_1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{7}$$

$$\text{AUC-ROC} = \int_0^1 \text{TPR}(FPR)\, d(FPR) \tag{8}$$

$$\text{Separability Ratio} = \frac{\mu_{\text{fault}}}{\mu_{\text{normal}}} \tag{9}$$

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v10i3.16211 p-ISSN : 2541-044X

where TP, FP, FN, and TN denote true positive, false positive, false negative, and true negative counts respectively. The separability ratio quantifies the degree of discrimination between normal and fault reconstruction error distributions, with higher values indicating greater model sensitivity to fault-induced deviations. Evaluation across all three scenarios enables a complete characterization of model robustness S1 reveals peak detection capability under ideal labeling, S2 exposes vulnerability to label noise, and S3 provides the most meaningful estimate of real-world deployment performance.

8. Real-time Deployment Validation

The trained model is deployed in a Python runtime environment that reads sensor data from PostgreSQL at 0.5-second intervals, computes derived features, maintains a sliding buffer of 60 timesteps, and writes the fault status and MSE value back to KEPServerEX via OPC UA. Thirty online fault injection experiments are conducted five per setpoint where a bypass valve is opened after the system reaches steady state. Detection latency is measured as the elapsed time between valve opening and the first fault status output, and recovery time is measured from valve closing to the first normal status. All experiments are logged to CSV files for subsequent analysis.

RESULT

The results are organized into three subsections threshold determination and offline model evaluation, ablation comparison between the proposed eleven-feature model and the seven-feature baseline, and real-time deployment validation using thirty online fault injection experiments.

1. Threshold Determination

The anomaly detection threshold for each model was computed from the validation set reconstruction error distribution using the three-sigma statistical rule (Aslam et al., 2024). For the eleven-feature model, the threshold was obtained as:

$$\tau_{11F} = \mu_{val} + 3\sigma_{val} = 0.2110 + 3 \times 0.2067 = 0.8312 \quad (10)$$

For the seven-feature baseline:

$$\tau_{7F} = \mu_{val} + 3\sigma_{val} = 0.1155 + 3 \times 0.0950 = 0.4014 \quad (11)$$

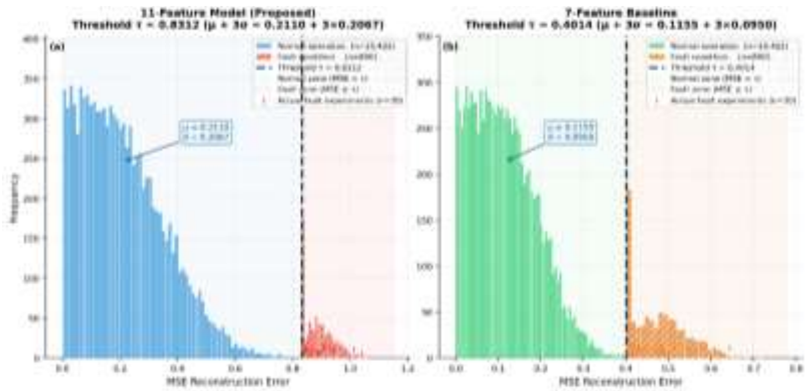

Fig. 6 MSE reconstruction error distribution

The eleven-feature model produces a separability ratio of 66.91× between fault and normal mean MSE, compared to 32.18× for the seven-feature baseline, confirming that the derived features substantially enhance fault discriminability.

2. Offline Evaluation

Both model variants were evaluated under three labeling scenarios. Table 3 summarizes all metrics, and Fig. 7 presents the confusion matrices. Fig. 8 shows the ROC curves, and Fig. 9 compares all metrics visually.

Table 3
Offline evaluation results

Scenario	Model	AUC-ROC	Precision	Recall	F1-Score	FP	FN	FPR
S1 (Cut)	11-Feature	1.0000	0.9985	1.0000	0.9993	74	0	0.73%
S1 (Cut)	7-Feature	1.0000	0.9970	1.0000	0.9985	150	0	1.44%
S2 (Naïve)	11-Feature	0.9799	0.9306	0.9309	0.9635	74	3.883	0.73%
S2 (Naïve)	7-Feature	0.9851	0.9972	0.9402	0.9678	150	3.366	1.44%
S3 (Corrected)	11-Feature	0.8943	0.9519	1.0000	0.9754	2.500	0	15.21%
S3 (Corrected)	7-Feature	0.9082	0.9411	1.0000	0.9697	3.092	0	18.48%

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026
DOI : https://doi.org/10.33395/sinkron.v10i3.16211
e-ISSN : 2541-2019
p-ISSN : 2541-044X

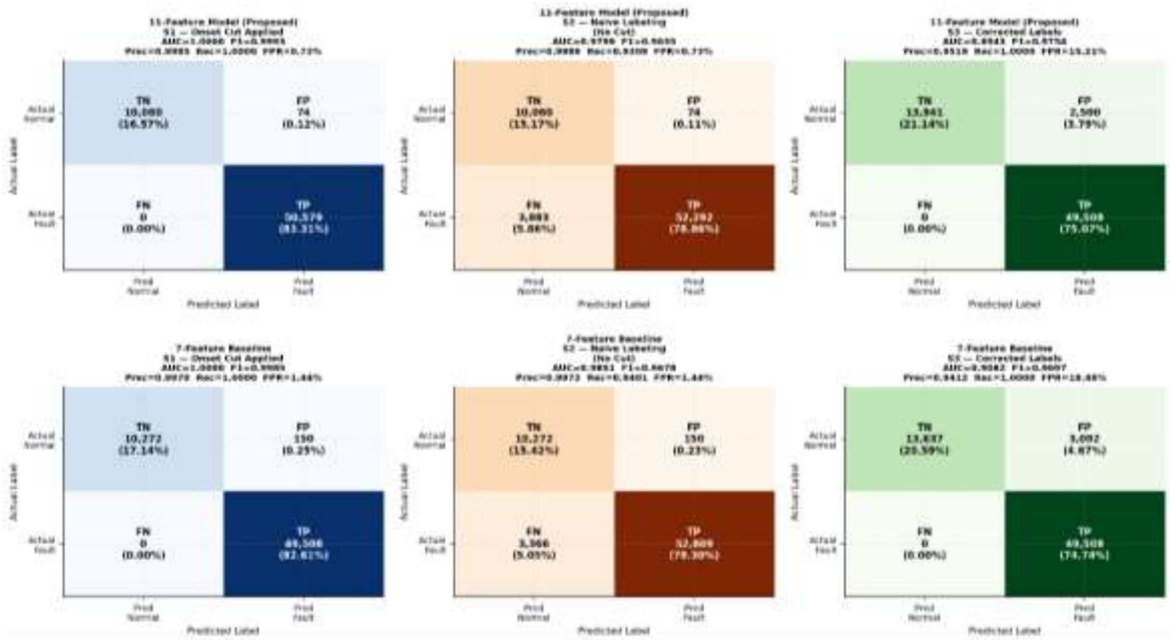

Fig. 7 Confusion matrices of the eleven-feature model and seven-feature baseline

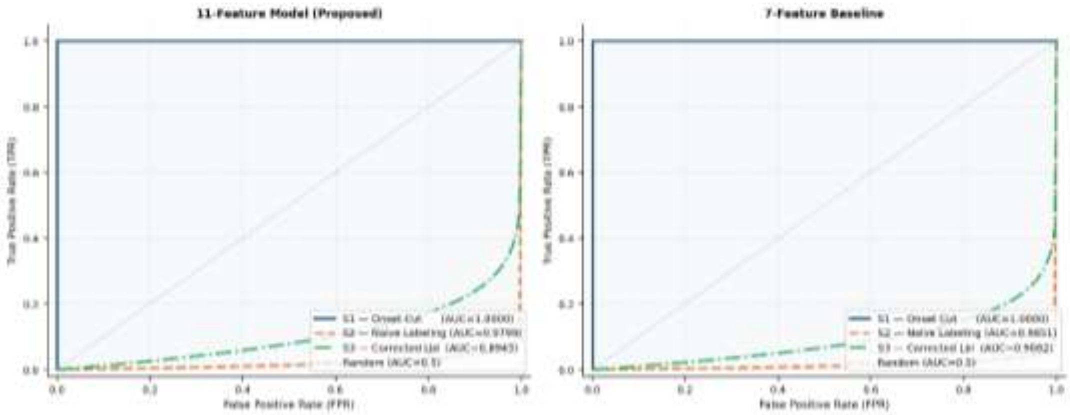

Fig. 8 ROC curves for the eleven-feature model and seven-feature baseline

Under S1, both models achieve perfect recall with zero false negatives. The eleven-feature model produces fewer false positives (74 vs 150), indicating superior specificity under clean label conditions. The AUC of 1.0000 under S1 is not a product of data leakage but reflects genuine separation between populations after onset removal, the test set contains only windows where bypass flow imbalance is fully established a condition never encountered during training on stable normal data. The separability ratios of 66.91× and 32.18× for the eleven-feature and seven-feature models respectively confirm that the two distributions do not overlap under this condition. The S3 results therefore serve as the primary deployment readiness benchmark.

Under S2, apparent metric degradation occurs due to incorrect labeling of pre-fault normal windows as anomalous. Since the LSTM-AE correctly reconstructs these windows which are genuinely normal in behavior they are classified as normal despite carrying a fault label, producing high false negatives against a flawed ground truth rather than reflecting genuine model failure. Under S3, the most realistic evaluation condition, both models maintain recall at 1.000 with the eleven-feature model achieving a lower false positive rate of 15.21% compared to 18.48% for the seven-feature baseline.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v10i3.16211 p-ISSN : 2541-044X

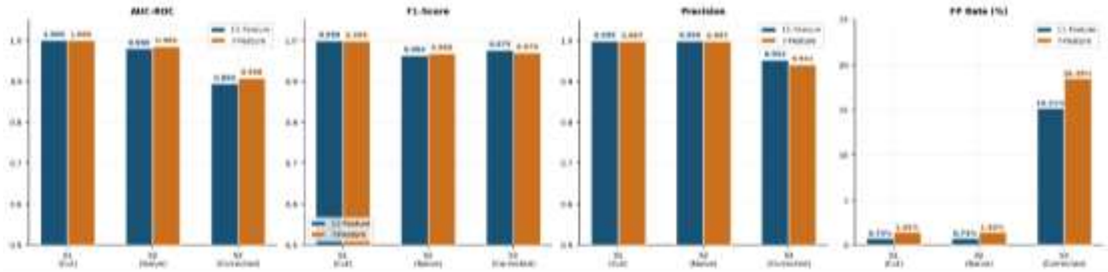

Fig. 9 Comparison of AUC-ROC, F1-score, Precision, and False Positive Rate

3. Ablation Study Derived Features vs Raw Sensor Features

A direct comparison between both models across all scenarios confirms that the three derived features delta flow, level error, and frequency-per-flow consistently reduce false positives without compromising recall. Under S1, false positives decrease from 150 to 74 (50.7% reduction). Under S3, the false positive rate is reduced by 3.27 percentage points (15.21% vs 18.48%). The delta flow feature provides an early fault signal by detecting flow imbalance at the moment the bypass valve opens, before the level error becomes measurable by the differential pressure transmitter, which explains the consistently lower detection latency of the eleven-feature model observed in real-time testing.

4. Real-Time Deployment Validation

Both models were deployed in a real-time environment and validated through 30 online fault injection experiments (5 per setpoint). All 60 experiments resulted in successful detection, yielding a detection rate of 100% (30/30) for both models. Table 5 summarizes per-setpoint deployment results, and Fig. 10 presents the comparison.

Table 4
Real-time deployment performance per operating setpoint

Setpoint (cm)	11F Latency (s)	11F Recovery (s)	7F Latency (s)	7F Recovery (s)
10	4.6 ± 2.1	43.6 ± 14.7	14.0 ± 3.7	53.0 ± 9.2
20	7.6 ± 2.6	38.8 ± 2.2	13.4 ± 2.9	36.4 ± 13.5
30	7.4 ± 1.1	39.2 ± 2.0	11.4 ± 2.5	55.4 ± 16.5
40	7.2 ± 1.3	41.4 ± 1.7	10.8 ± 4.6	59.0 ± 32.7
50	6.0 ± 1.6	43.4 ± 3.0	4.4 ± 3.8	61.2 ± 14.1
60	5.4 ± 1.7	40.8 ± 1.3	7.8 ± 5.3	48.2 ± 1.9
Overall	6.37 ± 2.04	41.2 ± 6.3	10.30 ± 5.05	52.2 ± 18.7

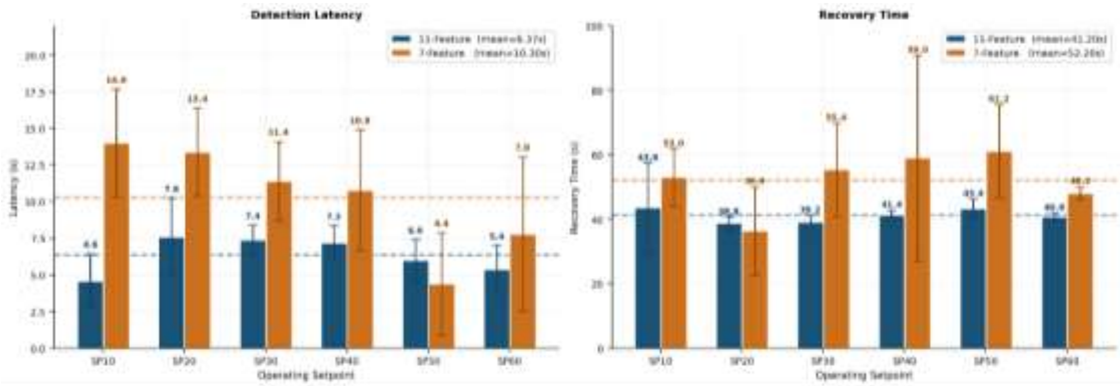

Fig. 10 Mean detection latency and recovery time per operating setpoint

The eleven-feature model achieves a mean detection latency of 6.37 ± 2.04 seconds, which is 3.93 seconds faster (38.2% improvement) than the seven-feature baseline at 10.30 ± 5.05 seconds. The lower standard deviation (2.04 vs 5.05 s) further demonstrates greater consistency across all setpoints and experiments. Recovery time similarly improves by 11.0 seconds on average (41.2 s vs 52.2 s).

5. Normal Operation Stability Test

To evaluate false alarm behavior under sustained normal operation, both models were tested continuously per setpoint without any fault injection. The results are presented in Fig. 11.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026
DOI : https://doi.org/10.33395/sinkron.v10i3.16211

e-ISSN : 2541-2019
p-ISSN : 2541-044X

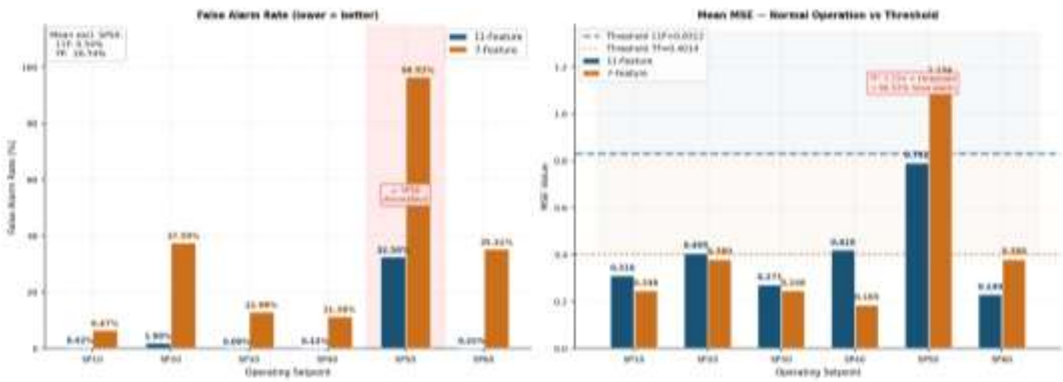

Fig. 11 Normal operation stability test results

For five of the six setpoints, the eleven-feature model maintains a false alarm rate below 2%, confirming reliable threshold behavior during normal operation. The seven-feature baseline exhibits significantly higher false alarm rates, reaching 37.59% at SP20 and 35.31% at SP60. At SP50, both models show elevated false alarm rates of 32.56% (eleven-feature) and 96.52% (seven-feature). The mean normal-operation MSE at SP50 reaches 0.792 for the eleven-feature model approaching the threshold of 0.8312 and 1.154 for the seven-feature baseline, which substantially exceeds the threshold of 0.4014. This is attributed to the PID tuning at SP50 (Kp=1, Ki=68, Kd=6) generating higher amplitude oscillations than other setpoints, elevating normal reconstruction error toward the threshold boundary.

DISCUSSIONS

The most practically significant observation is the consistent advantage of the eleven-feature model in detection latency not just in mean value, but in standard deviation. A mean latency of 6.37 seconds means little in industrial deployment if individual experiments swing between 3 and 12 seconds unpredictably. The seven-feature baseline shows exactly this problem a standard deviation of 5.05 seconds against a mean of 10.30 seconds, meaning that in the worst observed case, the system took 19 seconds to trigger an alarm. For a water level plant where bypass leakage can accelerate level drop at several centimeters per minute, a 19-second delay carries real operational consequences. The eleven-feature model narrows this window substantially, and the improvement is traceable to the delta flow feature because flow imbalance registers immediately when the bypass valve opens before the differential pressure transmitter detects any level change the reconstruction error begins rising earlier, compressing time-to-detection across experiments.

The three-scenario evaluation raises a methodological point that extends beyond this study. In reconstruction-based anomaly detection, the boundary between normal and fault behavior is rarely a sharp step there is always a transition region where the system has deviated from steady state but the deviation has not yet become unambiguous. How a researcher decides to label that region has outsized consequences on reported metrics. Scenario 2 makes this visible both models appear to perform poorly with F1-scores of 0.9635 and 0.9678, yet both are actually behaving correctly by classifying pre-fault windows as normal. The fault belongs to the labeling procedure, not the model. This dynamic is a systematic risk in any unsupervised anomaly detector evaluated on field data where fault onset is recorded with limited temporal precision. Reporting all three scenarios rather than selecting only the most favorable one provides a more honest characterization of where performance numbers actually come from (Belay et al., 2023; Soni et al., 2022).

The SP50 false alarm rate of 32.56% for the eleven-feature model during normal operation requires concrete examination rather than brief acknowledgment. Of 7,834 normal windows evaluated at SP50, approximately 2,551 exceeded the threshold equivalent to roughly one false alarm window every three observations during stable operation. The root cause is the PID tuning at this setpoint (Kp=1, Ki=68, Kd=6), which produces control oscillations with amplitudes larger than other setpoints, pushing normal reconstruction error to a mean of 0.792 only 4.8% below the threshold of 0.8312. For the seven-feature baseline, the situation is more severe a mean normal MSE of 1.154 against a threshold of 0.4014 renders the model entirely within the fault zone at SP50 during normal conditions, making it operationally unusable at this point. The eleven-feature model, while imperfect, maintains a detectable boundary that setpoint-specific threshold recalibration could correct without retraining. A targeted fix recomputing the threshold using only SP50 validation data rather than the pooled six-setpoint distribution represents the most immediate practical improvement for deployment at this operating point (Aslam et al., 2024).

One finding that runs counter to common intuition is that a deeper LSTM-AE architecture did not necessarily produce faster detection. The latency advantage of the eleven-feature model comes from richer input information, not from architectural complexity. The encoder-decoder structure LSTM(64) to LSTM(32) as bottleneck,

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v10i3.16211 p-ISSN : 2541-044X

symmetrically decoded is deliberately moderate. This design choice to keep the architecture fixed across both models and vary only the input feature set isolates the contribution of feature engineering from architectural effects, making the ablation study genuinely interpretable. This aligns with findings from pipeline anomaly detection studies showing that simpler architectures generalize more stably in repetitive industrial systems (Adhitya & Ihsan, 2025).

CONCLUSION

This study set out to answer a specific question does adding physics-informed derived features to an LSTM-Autoencoder meaningfully improve fault detection performance in a multi-setpoint water level control system, and does the improvement hold under real-time deployment conditions? The results demonstrate that physics-informed derived features improve fault detection performance under both offline and real-time deployment conditions, although several limitations remain.

The eleven-feature model reduced the false positive rate from 18.48% to 15.21% under realistic labeling conditions, detected bypass faults 3.93 seconds faster on average, and did so with more than twice the consistency of the seven-feature baseline as measured by standard deviation of latency. Both models achieved 100% detection across thirty experiments and all six operating setpoints, confirming that the per-setpoint windowing strategy enabled effective generalization without setpoint-specific retraining. The three-scenario evaluation framework contributed a finding of broader methodological relevance apparent performance degradation in Scenario 2 reflects a labeling artifact rather than model failure, and studies reporting reconstruction-based anomaly detector performance on field data should account for onset boundary uncertainty in their evaluation design.

The framework exhibits a known limitation at SP50, where the PID tuning parameters (Kp=1, Ki=68, Kd=6) generate control oscillations with substantially higher amplitudes than other operating setpoints, causing the normal reconstruction error to reach an average of 0.792, only 4.8% below the detection threshold of 0.8312. Consequently, the eleven-feature model produces a false alarm rate of 32.56% during normal operation at SP50, incorrectly triggering fault alarms in approximately one out of every three windows despite the absence of actual bypass leakage. This behavior reflects a common limitation of threshold-based anomaly detection in systems with heterogeneous process dynamics, where a single threshold calibrated from pooled multi-setpoint statistics becomes overly restrictive for operating points with inherently higher variance. To mitigate this issue, future implementations should adopt a setpoint-conditioned adaptive threshold in which the detection boundary τ is computed independently for each operating setpoint using its own validation MSE distribution, thereby reducing false alarms without requiring model retraining or sacrificing detection sensitivity at stable setpoints. Furthermore, for systems with continuously changing PID tuning or time-varying plant dynamics, a rolling adaptive threshold can be applied by periodically recalculating τ from a sliding window of recent normal-state reconstruction errors, allowing the threshold to adapt automatically to evolving process variability. Together, these strategies provide a practical and implementable solution for addressing the SP50 limitation without modifying the underlying LSTM-AE architecture.

REFERENCES

Abdennebi, A., Tuncay, A., Yilmaz, C., Koyuncu, A., & Gungor, O. (2023). LSTM-AE for Anomaly Detection on Multivariate Telemetry Data. *Proceedings - 2023 IEEE/ACIS 21st International Conference on Software Engineering Research, Management and Applications, SERA 2023*, 90–97. https://doi.org/10.1109/SERA57763.2023.10197673

Adhitya, M. H., & Ihsan, A. F. (2025). Anomaly Detection of Oil and Gas Pipeline Operational Data Using Multi-Layer LSTM Autoencoder Method. *2025 5th International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA)*, 155–160. https://doi.org/10.1109/ICICyTA68677.2025.11362820

Aslam, M. M., Tufail, A., De Silva, L. C., Haji Mohd Apong, R. A. A., & Namoun, A. (2024). An improved autoencoder-based approach for anomaly detection in industrial control systems. *Systems Science and Control Engineering*, *12*(1). https://doi.org/10.1080/21642583.2024.2334303

Belay, M. A., Blakseth, S. S., Rasheed, A., & Salvo Rossi, P. (2023). Unsupervised Anomaly Detection for IoT-Based Multivariate Time Series: Existing Solutions, Performance Analysis and Future Directions. *Sensors*, *23*(5). https://doi.org/10.3390/s23052844

Cui, Y., Member, S., Ab, G., & Tjernberg, L. B. (2018). An Anomaly Detection Approach Based on Machine Learning and SCADA Data for Condition Monitoring of Wind Turbines. *2018 IEEE International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, 1–6.

Essamlali, I., Nhaila, H., & Khaili, M. El. (2024). Heliyon Advances in machine learning and IoT for water quality monitoring : A comprehensive review. *Heliyon*, *10*(6), e27920. https://doi.org/10.1016/j.heliyon.2024.e27920

Goetz, C., & Humm, B. (2023). Decentralized Real-Time Anomaly Detection in Cyber-Physical Production Systems under Industry Constraints. *Sensors*, *23*(9). https://doi.org/10.3390/s23094207

Kammoun, M., Kammoun, A., & Abid, M. (2023). LSTM-AE-WLDL: Unsupervised LSTM Auto-Encoders for Leak Detection and Location in Water Distribution Networks. *Water Resources Management*, *37*(2), 731–

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026
DOI : https://doi.org/10.33395/sinkron.v10i3.16211
e-ISSN : 2541-2019
p-ISSN : 2541-044X

746. https://doi.org/10.1007/s11269-022-03397-6

Karadimos, P., & Anthopoulos, L. (2023). Machine Learning-Based Energy Consumption Estimation of Wastewater Treatment Plants in Greece. *Energies*, *16*(21), 7408. https://doi.org/10.3390/en16217408

Khalid Fahmi, A. T. W., Reza Kashyzadeh, K., & Ghorbani, S. (2024). Fault detection in the gas turbine of the Kirkuk power plant: An anomaly detection approach using DLSTM-Autoencoder. *Engineering Failure Analysis*, *160*(October 2023), 108213. https://doi.org/10.1016/j.engfailanal.2024.108213

Kim, D., & Heo, T. Y. (2022). Anomaly Detection with Feature Extraction Based on Machine Learning Using Hydraulic System IoT Sensor Data. *Sensors*, *22*(7). https://doi.org/10.3390/s22072479

Lee, D., Choo, H., & Jeong, J. (2023). Leak Detection and Classification of Water Pipeline Data Using LSTM Auto-Encoder with Xavier Initialization. *2023 IEEE/ACIS 8th International Conference on Big Data, Cloud Computing, and Data Science, BCD 2023*, 69–74. https://doi.org/10.1109/BCD57833.2023.10466341

Nelago Kanyama, M., Bhunu Shava, F., Gamundani, A. M., & Hartmann, A. (2024). Enhancing Anomaly Detection in Smart Water Metering Networks with LSTM-Autoencoder and Data Augmentation Techniques. *Proceedings of 2024 4th International Multidisciplinary Information Technology and Engineering Conference, IMITEC 2024*, 20–28. https://doi.org/10.1109/IMITEC60221.2024.10851168

Nizam, H., Zafar, S., Lv, Z., Wang, F., & Hu, X. (2022). Real-Time Deep Anomaly Detection Framework for Multivariate Time-Series Data in Industrial IoT. *IEEE Sensors Journal*, *22*(23), 22836–22849. https://doi.org/10.1109/JSEN.2022.3211874

Nugraha, N. W., Suryatini, F., Lilansa, N., & Madani, F. A. (2025). *Implementation of Industrial IoT Integration Using Node-RED and PLC on Cascade Control Level and Flow Plant*. *11*(6), 191–205. https://doi.org/10.29303/jppipa.v11i6.11623

Qu, Y., Fu, S., Yong, M., Tian, J., Lv, Z., & Li, R. (2025). Health Indicator Construction and Remaining Useful Life Prediction Based on MSC-LSTM-AE Model for Working Bearings. *IEEE Sensors Journal*, *25*(9), 15525–15535. https://doi.org/10.1109/JSEN.2025.3548675

Soni, M., Khan, I. R., Basir, S., Chadha, R., Alguno, A. C., & Bhowmik, T. (2022). Light-Weighted Deep Learning Model to Detect Fault in IoT-Based Industrial Equipment. *Computational Intelligence and Neuroscience*, *2022*. https://doi.org/10.1155/2022/2455259

Sunal, C. E., Dyo, V., & Velisavljevic, V. (2022). Review of Machine Learning Based Fault Detection for Centrifugal Pump Induction Motors. *IEEE Access*, *10*(June), 71344–71355. https://doi.org/10.1109/ACCESS.2022.3187718

Wei, Y., Jang-Jaccard, J., Xu, W., Sabrina, F., Camtepe, S., & Boulic, M. (2023). LSTM-Autoencoder-Based Anomaly Detection for Indoor Air Quality Time-Series Data. *IEEE Sensors Journal*, *23*(4), 3787–3800. https://doi.org/10.1109/JSEN.2022.3230361

Zhao, Z., Xiao, Z., & Tao, J. (2024). MSDG: Multi-Scale Dynamic Graph Neural Network for Industrial Time Series Anomaly Detection. *Sensors*, *24*(22), 1–18. https://doi.org/10.3390/s24227218

*name of corresponding author