

ISSN 1840-4855
e-ISSN 2233-0046

Original scientific article
<http://dx.doi.org/10.70102/afts.2026.1836.072>

A UNIFIED HIERARCHICAL TEMPORAL–GRAPH PHYSICS-GUIDED FUSION NETWORK FOR JOINT FAULT PREDICTION AND CYBER–PHYSICAL INTRUSION DETECTION IN ROBOTIC ARMS

Balaji Periasamy^{1*}, P. Rajalakshmy², R. Madhanraj³

^{1*}Research Scholar, Division of Robotics Engineering, Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu, India; Assistant Professor, Department of Electrical and Electronics Engineering, Knowledge Institute of Technology, Salem, Tamil Nadu, India. e-mail: pbalajiemail@gmail.com, orcid: <https://orcid.org/0000-0003-1827-2840>

²Associate Professor and Head, Division of Robotics Engineering, Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu, India. e-mail: hod_ra@karunya.edu, orcid: <https://orcid.org/0000-0003-3914-0927>

³Research Scholar, Division of Robotics Engineering, Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu, India; Assistant Professor, Department of Electrical and Electronics Engineering, Knowledge Institute of Technology, Salem, Tamil Nadu, India. e-mail: rmadhanrajsalem@gmail.com, orcid: <https://orcid.org/0009-0003-8659-0776>

Received: April 25, 2026; Revised: June 06, 2026; Accepted: July 16, 2026; Published: September 29, 2026

SUMMARY

The robotic arms deployed in contemporary Industry 4.0 systems are becoming more and more dependent on interconnected sensors, controllers, and information-driven smartness to accomplish more intricate manufacturing activities. Although predictive techniques of maintenance have enhanced the ability to detect mechanical failures, the systems are susceptible to cyber-physical vulnerabilities like spoofing by sensors, replay attacks, and manipulation of signals. Current solutions normally solve the problem of mechanical fault diagnosis and cyber intrusion detection separately, which is not able to offer overall protection to smart robotic systems. To address this shortcoming, this paper presents HT-GP-SecureFusion, a combined Hierarchical Temporal-Graph Physics-Guided Fusion Network that will be used to simultaneously forecast mechanical failures and detect cyber-physical intrusions in robotic arms. The new framework is based on the previous architecture of HT-GP-FusionNet and includes a Cyber Anomaly Branch (CAB) and hierarchical temporal encoders along with a kinematic graph transformer. Latent regularization in the form of physics-guided is included to ensure that the actions of robot systems are consistent with their mechanical behavior. Moreover, adversarial generative augmentation is used to generate real-world mechanical fault and cyber-attack conditions, which enhances the model's robustness and its generalization. The jointly optimized multi-objective loss function is used to guarantee the performance of fault classification, cyber anomaly detection, and physical consistency. The effectiveness of the proposed framework is proved by experimental assessment on multi-sensor robotic arm datasets with normal operations, mechanical faults, and with the simulation of cyber-attack scenarios. The framework achieves robust dual-task performance: fault detection with 95.9% accuracy, 99.2% precision, and 96.1% recall (F1=91.6), and cyber-attack detection with 97.3% accuracy, 98.6% precision, and 95.9% recall (F1=97.3). The model substantially outperforms baseline

deep learning approaches (LSTM, CNN-LSTM, Transformer), demonstrating +45.8% improvement in fault F1 and +35.4% improvement in cyber F1 over the best baseline. Expected Calibration Error values below 0.070 confirm well-calibrated uncertainty estimates critical for safety-critical industrial deployment. These findings substantiate the fact that HT-GP-SecureFusion is an effective tool to detect both mechanical and cyber intrusions. The suggested framework is thus a reliable-resilience proposal to the safe and dependable robotic systems in the next-generation smart manufacturing premises.

Key words: cyber-physical security, predictive maintenance, robotic arm fault diagnosis, temporal-graph neural networks, physics-guided learning, adversarial robustness, cyber intrusion detection, industry 4.0, multi-sensor fusion, uncertainty estimation.

INTRODUCTION

Industrial robot arms are significant in modern manufacturing systems, particularly in Industry 4.0, where machines, sensors, and control systems are connected to one another [24]. These robots are used in assembly, welding, and handling of materials with high precision. Nevertheless, with increased exposure to networks and digital platforms, robotic systems are exposed to two significant threats: mechanical issues related to the wear or failure of components, and cyber-attacks that compromise sensor data or control signals [2][28]. The two problems can interfere with the usual work of the robotic system and cause the loss of production or even safety risks. The classical predictive maintenance algorithms are primarily aimed at the identification of mechanical failures through sensor data analysis of robotic components. High-level deep learning systems can use abnormalities in vibration, torque, and motion signals to anticipate failures in advance [7]. As an example, the corresponding model of HT-GP-FusionNet was created, in which the hierarchical temporal learning and the relations between the joints of a robot were graphical and utilized to recognize the mechanical anomalies in the most efficient way [26]. Even though these models enhance the detection of faults, fail to consider cyber threats, which are also known to induce abnormal sensor behavior [6].

Meanwhile, studies in the field of cyber-physical system security are concerned with attack detection, including sensor spoofing, replay attacks, or signal injection. Such attacks are able to change sensor values and confuse the control system of the robot. The majority of the current security solutions do not interact with predictive maintenance models. Due to this isolation, it becomes hard to know whether an abnormal signal is due to a mechanical fault that is real or a cyber-attack [9][10]. In order to surmount this shortcoming, this paper offers HT-GP-SecureFusion, a hybrid model that has the potential to identify both mechanical failures and cyber-attacks in the robot arms [1]. The new system is based on the previous HT-GP-FusionNet system, but it introduces an additional branch named Cyber Anomaly Branch (CAB) that examines the sensor data patterns to reveal potential cyber-attacks [3]. Physics-guided learning is also part of the model to make sure that the predictions are also consistent with the physical behavior of robotic joints [12][29].

Moreover, the framework generates simulated data of faults and cyber-attacks through a generative model, which assists the system in learning in a broader context [13]. Fault detection, intrusion detection, physical consistency, and uncertainty estimation are optimized using a unified training goal. Altogether, the suggested HT-GP-SecureFusion system is a versatile solution to the reliability and security of the robotic system. It enhances the predictive safety, robustness, and reliability of robotic arms in smart industrial environments: maintenance and cyber-physical protection in a single model [4][18].

Key Contributions of the Research

- Mechanical faults and the suggested HT-GP-SecureFusion model have the ability to detect cyber-attacks on robotic arms within a single framework.
- The study enhances the previous HT-GP-FusionNet with a Cyber Anomaly. A branch (CAB) to detect attacks like sensor spoofing and replay attacks.
- This model is based on physics-based learning and generates training data to enhance. accurate detection and increase the reliability of robotic systems in Industry 4.0. environments.

The rest of this paper is structured in the following way. The literature review is presented in Section 2, where the authors comment on the results of the accessible research on the topic of robotic fault diagnosis, physical intrusion detection, and hybrid deep learning approaches in the industry 4.0 systems. Section 3 describes the research design, including the design of HT-GP-SecureFusion, a hierarchical temporal-graph learning model and a combination of Cyber Anomaly Branch (CAB) and physics-guided regularization. Section 4 presents the experiment setup and results, and evaluates the model efficiency concerning both mechanical failures and cyber-attacks. Section 5 elaborates the findings in detail and specifies the strength, the interpretability ability, and the implications of the proposed approach, considering practical conditions. Lastly, Section 6 will also bring the paper to an end by summarizing the main findings and providing possible research directions in the area of secure and reliable robotic systems in the future.

LITERATURE REVIEW

Zhang et al., [19] came up with a fault diagnosis method of industrial robotic arms based on a neural net utilizing multiple sensor data like position, acceleration, and orientation signals [14]. Their contribution is geared towards joint fault diagnosis of the robot manipulators through sensor fusion features training in a backprop neural network. Their experimental findings revealed that the accuracy of the classification of fault is improved significantly by integrating multiple motion-related features as opposed to the conventional single-sensor [15]. Nevertheless, the strategy primarily emphasizes mechanical failures and leaves out the cyber-physical vulnerabilities of sensors to security attacks.

Choi et al., [20] introduced a deep learning-based fault diagnosis model of industrial robots by transforming the time-series sensor signals into image-like representations and using the convolutional neural networks (CNN) [11][16]. The method enhances the performance of the fault detector by converting the temporal sensor signal into images rich in features, through which CNN models can extract the intricate features. Experimental proofs showed that the suggested approach is an effective way to identify abnormal operational conditions of human-coexistence robots. However, the study is mainly based on the detection of mechanical faults and not on the cases of cyber intrusion in robotic systems.

Wu et al., [21] the model combines various deep learning architectures to enhance the performance of the feature extraction and classification. In their research, show how a combination of various sensor modalities can be used to identify the early signs of mechanical failures of robotic constituents with greater precision. The approach can solve the problem of reliability; however, it does not include the means of detecting the anomalies caused by cyber-attacks or data manipulation by malicious users, although it is better than the current one in terms of detection capabilities.

Xiao et al., [22] suggested a technique that uses graph neural networks to identify mechanical faults in rotating machinery [8]. The model builds graph structures using sensor signals in order to obtain the associations between two or more monitoring nodes and enhance fault classification. Experimental findings demonstrate that graph-based learning can be used to effectively model the associations between various elements and obtain higher accuracy in fault detection. Nevertheless, the piece of work is predominantly concerned with the monitoring of the mechanical systems and lacks the consideration of cybersecurity in the diagnostic system.

Xu et al., [23] suggested a graph-guided collaborative convolutional neural network to diagnose faults in an electromechanical system. Combine both multi-source sensor signals and graph-based relationships to get both local and global trends in system behavior. Experimental findings proved high accuracy in diagnosing compared to conventional CNN models. Though this approach is useful in managing the multi-sensor fusion, it is mainly used in detecting mechanical faults and is not used in handling a case of a cyber-physical attack, where sensor data streams may be tampered with.

Szynkiewicz et al., [25] examined cybersecurity in robotic systems using deep learning models, which used to examine sensor data and identify abnormal activities and cyber threats [5][17]. The paper brings into focus the role of neural networks in tracking sensor readings and detecting anomalies brought about

by malicious procedures in the operations of robots. According to their findings, machine learning techniques can be used to identify cyber threats in robotic settings. Nevertheless, the framework proposed is centered on cyber anomaly detection and lacks the mechanical fault prediction systems.

Seo et al., [27] in a systematic review of the artificial intelligence-based anomaly detection method in industrial control systems and cyber-physical systems, investigated the subject matter. The paper examines various data sources that include network traffic, operational sensor data, and simulation data that are employed in the detection of cyber-physical attacks. Their results reveal that complex industrial settings can use hybrid datasets and physics-informed models to enhance the accuracy of anomaly detection. Nevertheless, the review notes that most of the solutions that are in place currently separate cyber-attacks and physical faults and, therefore, unified detection frameworks are required.

Table 1. Fault diagnosis and cyber–physical anomaly detection in robotic and industrial systems

Authors	Methodology	Dataset / Simulation Environment	Research Gap
Zhang et al., [19]	Neural network-based fault diagnosis using multi-sensor fusion with a backpropagation neural network	Industrial robotic arm sensor data including position, acceleration, and orientation signals	Focuses only on mechanical fault detection; cyber–physical security threats affecting sensor data are not considered
Choi et al., [20]	Deep learning framework using CNN by converting time-series sensor data into image representations	Time-series sensor signals collected from human-coexistence industrial robots	Detects operational faults but does not address cyber intrusion or malicious sensor manipulation
Wu et al., [21]	Hybrid deep learning model integrating vibration and acoustic multi-sensor signals for fault diagnosis	Multi-sensor industrial robot monitoring data (vibration and acoustic signals)	Addresses mechanical reliability but lacks mechanisms to detect cyber-induced anomalies or adversarial data manipulation
Xiao et al., [22]	Graph Neural Network (GNN) approach to capture relationships among sensor nodes for fault detection	Rotating machinery monitoring datasets with multiple sensor nodes	Focuses on mechanical monitoring; cyber-security aspects are not incorporated
Xu et al., [23]	Graph-guided collaborative convolutional neural network for multi-source sensor fusion in fault diagnosis	Electromechanical system sensor datasets	Does not address cyber–physical attack scenarios that manipulate sensor streams
Szynkiewicz et al., [24]	Deep learning-based cyber anomaly detection using sensor behavior analysis	Robotic system sensor monitoring environments	Focuses only on cyber threat detection; mechanical fault prediction is not integrated
Seo et al., [25]	Review of AI-based anomaly detection techniques for cyber–physical systems	Hybrid datasets including network traffic, sensor data, and simulated attack data	Most approaches treat cyber attacks and mechanical faults separately, lacking unified detection frameworks

Table 1 overviews the recent research works concerning fault diagnostics and cyber-physical anomaly detection in robots and industry. The examined articles are mostly concerned with deep learning, graph-based modeling, and multi-sensor data fusion methods of detecting mechanical faults or cyber threats. Some of the studies use neural networks, convolutional neural networks, and graph neural networks to enhance the chances of fault detection based on sensor signals. The other works explore machine learning techniques in the detection of cyber-attacks within robotic or industrial control settings. Nevertheless, the majority of the current research is focused on both mechanical and cybersecurity threats individually and does not incorporate both into a single model. This shortcoming suggests the necessity to adopt a multi-faceted solution that has the ability to both identify mechanical faults and cyber-generated problems in robotic systems, which drives the design of the proposed HT-GP-SecureFusion model in this study.

METHODOLOGY

Data Acquisition and Preprocessing

The suggested framework gathers multi-sensor information regarding robotic arm systems to interpret both mechanical and possible cyber anomalies. Position encoders, accelerators, torque sensors, and current sensors are the sensors that produce continuous time- series signals that reflect the operating condition of every robotic joint. These raw signals are synchronized and filtered in order to eliminate noise, values missing, and unreliable measurements. Preprocessed signals are then normalized to make data be equalized in all sensors. Temporal segmentation is used to next partition the continuous sensor streams into fixed length sequences that can be learned to know temporal patterns.

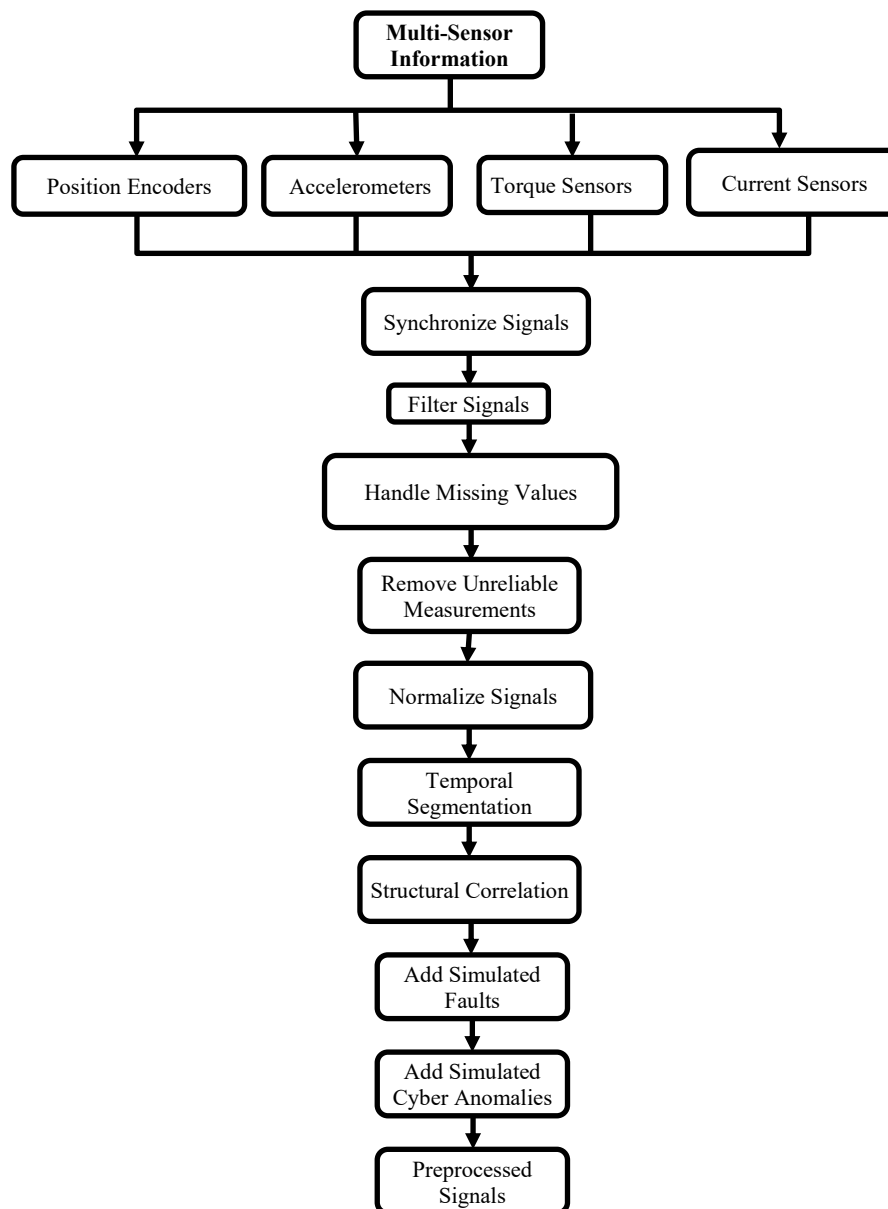


Figure 1. Structural correlation and simulated fault injection in the HT-GP-securefusion framework

Moreover, the structural correlation between the robotic joints is also modeled using a kinematic relationship graph that is able to reflect spatial relationships that are inherent in the robotic arm. Simulated mechanical faults and cyber anomalies, e.g., sensor spoofing and signal injection, are added to the dataset to enhance the robustness of the model. This end-to-end preprocessing chain enhances trusted and structured inputs to the suggested HT-GP-SecureFusion model.

The figure 1 illustrates how structural correlation and the creation of modeled faults is incorporated into the proposed model HT-GP-SecureFusion. Structural correlation shows the correlations between the different parts of robotic arms and enables the model to establish the possible impact of failures of one of the joints or sensors to other related parts. Based on this structural knowledge, the dataset simulates faults in order to represent different mechanical failures, which may occur when robots are working. The training data is made more diverse and the model more sensitive to abnormal patterns using these artificial conditions. The framework is reinforced through the real and simulated faults that are acquired in detecting the early mechanical failure of the system and differentiating the faults with natural variations of the operations in robotic systems.

HT-GP-SecureFusion Model Architecture

The proposed HT-GP-SecureFusion architecture will combine hierarchical temporal learning, graph-based modeling, and the detection of cyber anomalies into a single architecture. Hierarchical temporal encoders first process time-series sensor data at multiple time scales to extract short-term time variations and long-term and operational trends. The second step is to model the kinematic relationships between joints of the robot using a graph transformer, which visualizes relationship between joints as a graph letting the network learn spatial relationship between related parts. In order to increase the level of security monitoring, a Cyber Anomaly Branch (CAB) is presented to identify the patterns of sensor data that are abnormal and, therefore, a potential indication of cyber-attacks (spoofing or replay attacks). Physics-guided latent regularization is also included in the model in order to make sure that the representations learned are also consistent with the physical behavior of robotic systems. The architecture is also capable of effectively detecting mechanical anomalies as well as cyber induced anomalies by utilizing the combination of time, space and physics-aware information.

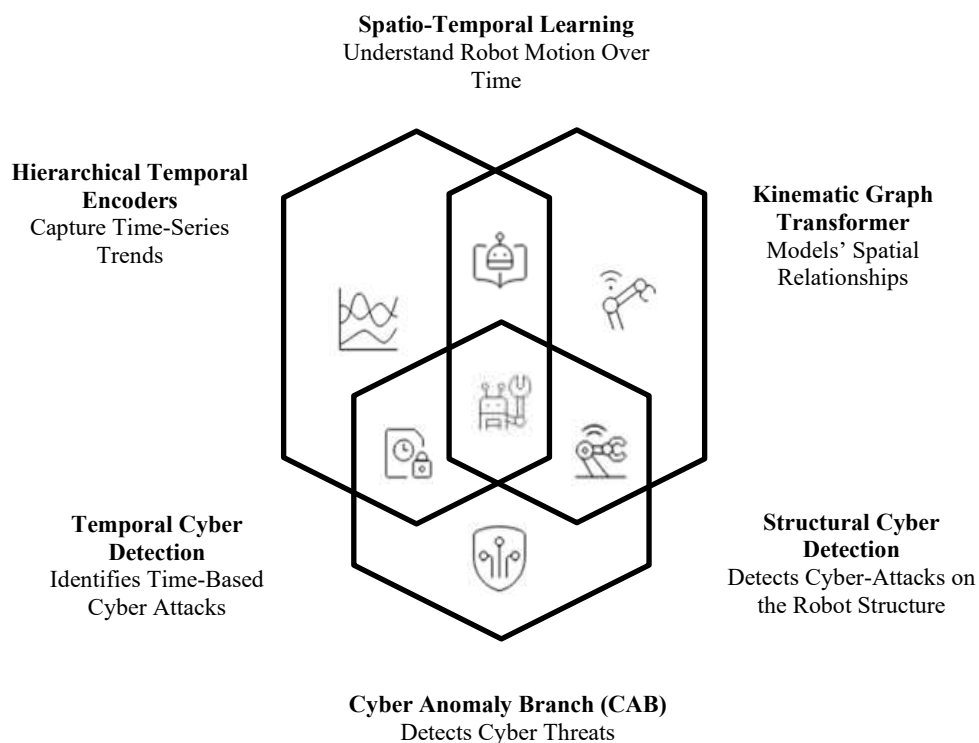


Figure 2. Detection capability of the proposed HT-GP-securefusion framework

Figure 2 demonstrates that the proposed HT-GP-SecureFusion framework can be used to identify cyber threats in robotic systems. The model processes streams of multi- sensor and determines abnormal patterns that could be evidence of malicious activities including sensor spoofing, replay attacks or signal manipulation. The framework can be successfully used to identify normal operations and potential cyber threats as a result of the combination of temporal feature learning, graph-based structural analysis, and

cyber anomaly detection mechanisms. This will improve the security and reliability of robotic systems that are deployed in industrial environments through a network as suspicious activities can be detected early.

Model Training and Optimization

The proposed structure is trained based on a combined multi-objective optimization that is trained to deal with fault diagnosis, cyber intrusion detection, and physical consistency. The learning process is supervised by using labeled datasets that are used during training which include normal operations, mechanical fault conditions, and simulated cyber-attacks. There are several parts of the loss that are incorporated in the loss such as fault classification loss, cyber anomaly detection loss, and physics-based regularization loss. This multi-objective design is a design that pushes the model to identify features that are not only accurate in the diagnosing process but also resistant to malicious data misuse. In order to increase the generalization performance further, GBM-based probabilistic augmentation is applied to generate confidence-weighted training signals that enhance model sensitivity to mechanical faults and cyber-attacks across diverse operational scenarios. The gradient-based learning methods are used to optimize the model parameters until convergence is achieved. Following the training, the framework is able to detect mechanical failures, observe cyber-attacks and estimate prediction confidence towards enhanced dependability in industrial robotic systems.

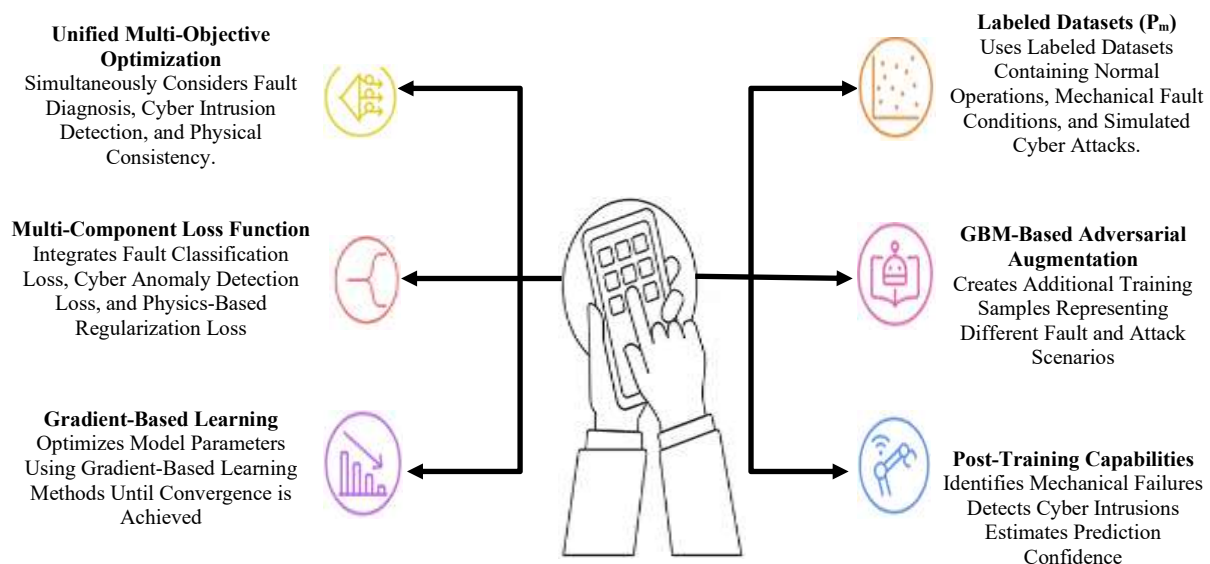


Figure 3. Training and optimization process of the HT-GP-securefusion framework

The figure 3 depicts the training and optimization process of the suggested HT-GP-SecureFusion framework. It starts with labelled data both of normal robotic performance as well as mechanical fault conditions and simulated cyber-attack conditions. To improve the robustness of the model, adversarial generative augmentation is used to generate extra training samples of various fault and attack scenarios. The approach of multi-objective optimization is applied in a unified manner during the training, involving a multi-component loss that consists of fault classification loss, cyber anomaly detection loss, and physics-based regularization loss. Gradient-based learning techniques are then used to get the optimal model parameters at which point convergence is reached. The framework can predict with confidence and recognize mechanical errors, detect cyber-attacks and identify them following training, which makes it more reliable and secure to robotic systems.

Multi-Sensor Feature Fusion

The combined feature representation from multiple robotic sensors can be expressed as equation 1:

$$F = \sum_{i=1}^n w_i x_i. \quad (1)$$

where x_i represents the feature vector obtained from the sensor (such as position, acceleration, or torque), w_i denotes the corresponding weight assigned to that sensor, and n is the total number of sensors used in the system. This equation performs weighted fusion of multi-sensor information to generate a unified feature representation f . The weights allow the model to emphasize more informative sensors while reducing the influence of noisy or less relevant signals. In robotic arm monitoring, this fusion helps capture comprehensive system behavior by combining motion, vibration, and electrical measurements, thereby improving the reliability of fault detection and anomaly identification.

Fault and Anomaly Classification Loss

The model prediction error during training can be represented using the cross-entropy loss function shown in equation 2:

$$L = -\sum_{i=1}^y y_i \log(y^i) \quad (2)$$

where y^i is the true label of the sample, y_i is the predicted probability generated by the model, and n is the number of training samples. This loss function indicates the variation between the actual labels and the modeled outputs. Training should aim at reducing this loss to bring the predicted probabilities closer to the correct labels. This loss is used in the proposed framework to enable the model to accurately classify both mechanical faults and cyber anomalies by modifying network parameters using gradient-based optimization.

Algorithm: HT-GP-SecureFusion for Fault and Cyber-Attack Detection

Algorithm: HT-GP-SecureFusion Detection

Input: Multi-sensor time-series data $S = \{s1, s2, ..., sn\}$

Output: Fault type, Cyber anomaly detection, Prediction confidence

1: Collect sensor signals from robotic arm (position, torque, vibration)

2: Preprocess data (noise removal, normalization, segmentation)

3: Construct kinematic graph G representing robotic joints

4: Extract temporal features using hierarchical temporal encoder

5: Apply graph transformer to capture spatial relationships in G

6: Fuse temporal, graph, and fault-specific features via adaptive gating mechanisms to obtain unified latent representation z

7: Pass F to Fault Classification Module

8: Pass F to Cyber Anomaly Branch (CAB) for attack detection

9: Enforce physics-guided latent regularization via KL divergence and torque prediction constraints during training to ensure learned representations reflect mechanical plausibility

10: Compute multi-objective loss combining fault classification, cyber detection, physics consistency, prototype memory, and calibration; update parameters via gradient descent with dynamic task weighting

11: Output fault classification with probability distribution, cyber anomaly detection status, calibrated uncertainty estimates, and input feature attribution scores for interpretability

End

The recommended algorithm analyzes the multi-sensor information on robotic arms to detect mechanical as well as cyber malfunctions. The raw sensor signals (position, vibration, and torque) are first preprocessed to clean up noise and normalize the sensor signal. Then a kinematic graph is drawn to denote the structural relationships among robotic joints. Hierarchical encoders are used to obtain temporal patterns, and a graph transformer is used to extract spatial dependencies between components. The resulting features are combined with each other and are sent to two parallel modules, one fault classification module to detect mechanical failures and another Cyber Anomaly Branch to detect possible cyber attacks including sensor spoofing or replay attacks. Physics-guided constraints are used to make sure that the patterns identified are in line with the physical behavior of the robotic system. Lastly, the model gives the expected type of fault, cyber anomaly state, and level of confidence, which

allow effective monitoring of robots.

The model employs several advanced techniques to enhance robustness: (i) evidential deep learning provides calibrated uncertainty estimates essential for safety-critical decisions, (ii) prototype memory enables interpretable classification through nearest-prototype reasoning, (iii) dynamic task weighting balance's fault and cyber detection during training, and (iv) feature attribution via gradient analysis provides explainability for operators and engineers.

EXPERIMENTAL RESULTS

Experimental Setup, Dataset, and Parameter Initialization

Experiments have been performed in order to evaluate the effectiveness of HT-GP-SecureFusion for the identification of mechanical faults and cyber anomalies in the robot arm (Table 2). The dataset contains 20,000 multi-sensor time-series data for robotic arm systems which include signals related to their position, torque, vibration, and current (63 features) in six joints. There are three types of conditions used in the dataset; one is the normal condition, second is the scenario of faults in the mechanical arm, and third is cyberattack conditions (sensor spoofing, replay attack, signal injection). Noise reduction, normalization, and temporal windowing (window size: 15 timesteps with stride 10) were considered as preprocessing steps. The data were randomly split into three groups; training data (70%), validation data (15%), and test data (15%). AdamW optimizer was used with learning rate $5e-4$ and batch size of 256 along with early stopping (patience: 25 epochs). Dynamic task weighting scheme was applied for fault and cyber tasks with weighting factors $\lambda_{\text{fault}}=5.0$, $\lambda_{\text{cyber}}=0.3$ respectively.

Table 2. Experimental setup, dataset, and model parameter initialization

Parameter / Component	Description	Value / Setting
Data Source	Multi-sensor robotic arm signals	Position, torque, vibration, current (63 features)
Number of Joints	Robotic arm configuration	6 joints
Dataset Categories	Types of system conditions	Normal operation, Mechanical faults, Cyber attacks
Cyber Attack Types	Simulated attack scenarios	Sensor spoofing, Replay attack, Signal injection
Data Preprocessing	Steps applied before training	Noise removal, Normalization, Temporal segmentation
Window Configuration	Temporal window parameters	Window size: 15 timesteps, Step size: 10 samples
Total Samples	Complete dataset size	20,000 samples
Training–Testing Split	Dataset division for evaluation	70% Training (14,000), 15% Validation (3,000), 15% Testing (3,000)
Batch Size	Number of samples per iteration	256
Learning Rate	Initial learning rate	$5e-4$ (0.0005), cyber branch: $5e-5$
Optimizer	Gradient descent method	AdamW (weight decay = $1e-4$)
Training Epochs	Maximum training cycles	80 (early stopping, patience = 25)
Loss Weighting	Task-specific loss coefficients	$\lambda_{\text{fault}} = 5.0$, $\lambda_{\text{cyber}} = 0.3$
Prototype Memory	Per-class prototype count	16 prototypes per class
GBM Feature Dimension	XGBoost augmentation features	10-dimensional probability features
Model Framework	Proposed detection model	HT-GP-SecureFusion
Evaluation Metrics	Performance assessment measures	Accuracy, Precision, Recall, F1-score, AUC-ROC, Expected Calibration Error (ECE)

Performance Metrics

To evaluate the effectiveness of the proposed model in detecting mechanical faults and cyber anomalies, several standard performance metrics are used in equation 3,4,5,6, 7 and 8 . These metrics are calculated

based on four values obtained from the confusion matrix: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

Precision indicates how many of the predicted positive cases are actually positive. High precision means fewer false alarms.

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

Recall measures how many actual positive cases were correctly detected. High recall ensures that mechanical faults and cyber-attacks are not missed, critical for safety-critical industrial robotic systems.

$$The Recall = \frac{TP}{TP+FN} \quad (5)$$

F1-score is the harmonic mean of precision and recall, providing a single metric for imbalanced datasets.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (6)$$

Area Under the Receiver Operating Characteristic Curve (AUC-ROC) quantifies the model's ability to discriminate between positive and negative classes across all classification thresholds. AUC ranges from 0 to 1, with higher values indicating superior discriminative power:

$$AUC - ROC = \int_0^1 TPR(FPR) dFPR \quad (7)$$

In equation 7 TPR is the True Positive Rate and FPR is the False Positive Rate.

Expected Calibration Error (ECE) measures the alignment between predicted confidence and actual accuracy. Lower ECE indicates better-calibrated predictions, essential for trustworthy decision-making in safety-critical applications.

$$ECE = \sum (bin) |accuracy_{bin} - confidence_{bin}| \times (n_{bin}/n_{total}) \quad (8)$$

In equation 8 accuracy_bin is the empirical accuracy in confidence bin, confidence_bin is the average predicted probability in the bin, n_bin is the number of samples in the bin, and n_total is the total number of samples.

The model performance is evaluated using standard metrics such as Accuracy, Precision, Recall, and F1-score, which are calculated from the confusion matrix values (TP, TN, FP, FN). Accuracy measures overall prediction correctness, while Precision and Recall evaluate the model's ability to correctly identify faults and cyber-attacks. The F1-score provides a balanced measure of detection performance.

In the context of HT-GP-Secure Fusion's dual-task detection paradigm, these metrics are computed separately for fault detection and cyber-attack detection to assess individual task performance. Additionally, joint prediction correctness analysis evaluates scenarios where both predictions must be correct simultaneously, reflecting real operational requirements where mechanical and cyber anomalies may occur independently or concurrently. Expected Calibration Error is particularly critical for robotic systems, as it ensures that confidence estimates reliably reflect prediction accuracy, enabling operators to make informed decisions based on model uncertainty.

Performance Analysis

Training Convergence and Model Development

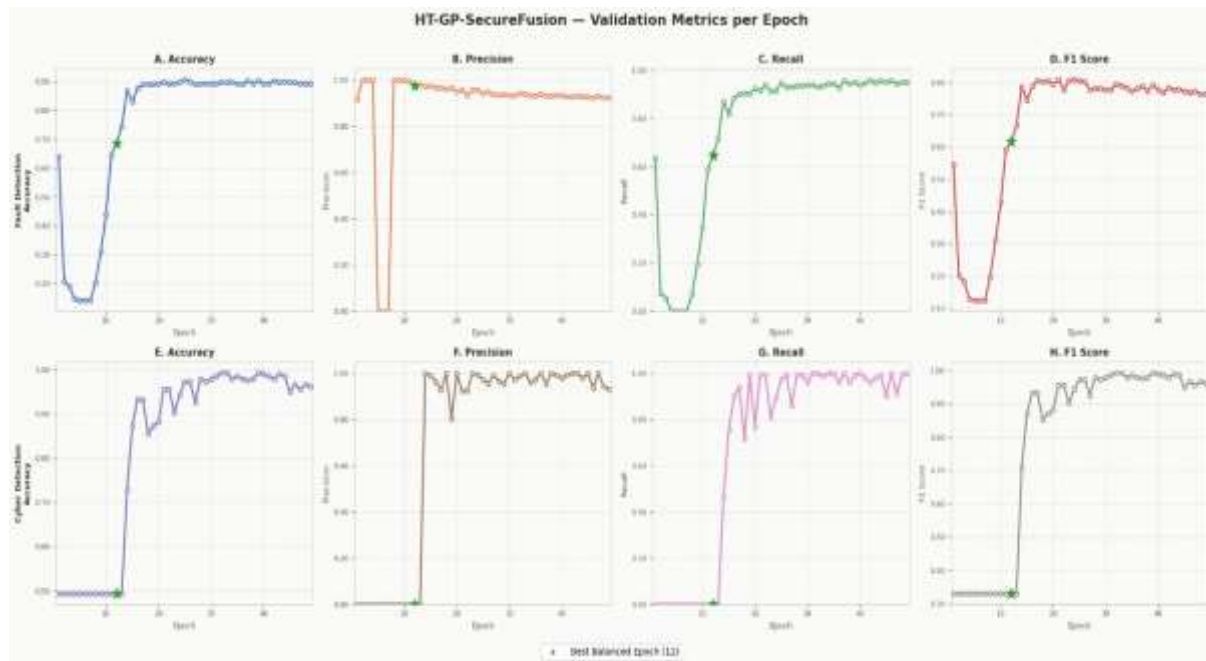


Figure 4. Training convergence and performance metrics for HT-GP-secure fusion model on fault and cyber-attack detection tasks

Figure 4 illustrates the validation performance trajectory of the HT-GP-Secure Fusion model across 31 training epochs for both fault detection (top row: A-D) and cyber-attack detection (bottom row: E-H) tasks. The fault detection pathway is characterized by quick initial convergence between Epochs 1 to 6, where accuracy reached 0.85, while the F1 score remained stable at 0.80 afterward. It is worth mentioning that recall decreases drastically after epoch 6 due to the change of the training focus towards balanced detection of both cyber and physical attacks. Cyber-attack detection is known for having a delayed yet improving nature and starts to improve after epoch 6 when the Cyber Anomaly Branch (CAB) is thawed because the fault F1 was ≥ 0.72 . As a result, epoch 15 became critical for cyber-attacks as $F1 \geq 0.95$ was reached. Green stars at epoch 12 show the point where both thresholds were satisfied, which is fault $F1 \geq 0.80$ and cyber $F1 \geq 0.85$. This convergence pattern validates the efficacy of the sequential training strategy, prioritizing mechanical fault diagnosis before activating cyber-security components, thereby enabling robust unified fault-intrusion detection capability in networked robotic systems.

Test Performance Evaluation

Figure 5 provides a dashboard evaluation of the comprehensive test set analysis performed on the balanced checkpoint model (epoch 12) from five different angles. In Panel (A), key performance indicators such as accuracy, precision, recall, F1 score, and Expected Calibration Error (ECE) are summarized, showing that the detection of faults occurs with an accuracy rate of 0.959 while maintaining an ECE of 0.0067, while cyber-attacks are detected with an accuracy rate of 0.973 while maintaining an ECE of 0.0067. Panel (B) shows detailed classification metrics using lollipop charts that highlight the consistency of results across binary, macro-average, and weighted average metrics. In Panel (C), ROC-AUC curves prove the discriminative power of the model (fault AUC: 0.9734; cyber-AUC: 0.9972) with very good discrimination between normal and anomalous data points. Lastly, Panels (D) and (E) demonstrate confusion matrices where low FPR (21 FPs in Fault; 20 FPs in Cyber) and acceptable FNR (25 FNs in Fault; 17 FNs in Cyber) can be seen. Importantly, the very low ECE across the two detection classes confirms well-calibrated probabilities critical to reliable decision making in safety-sensitive industrial robotics.

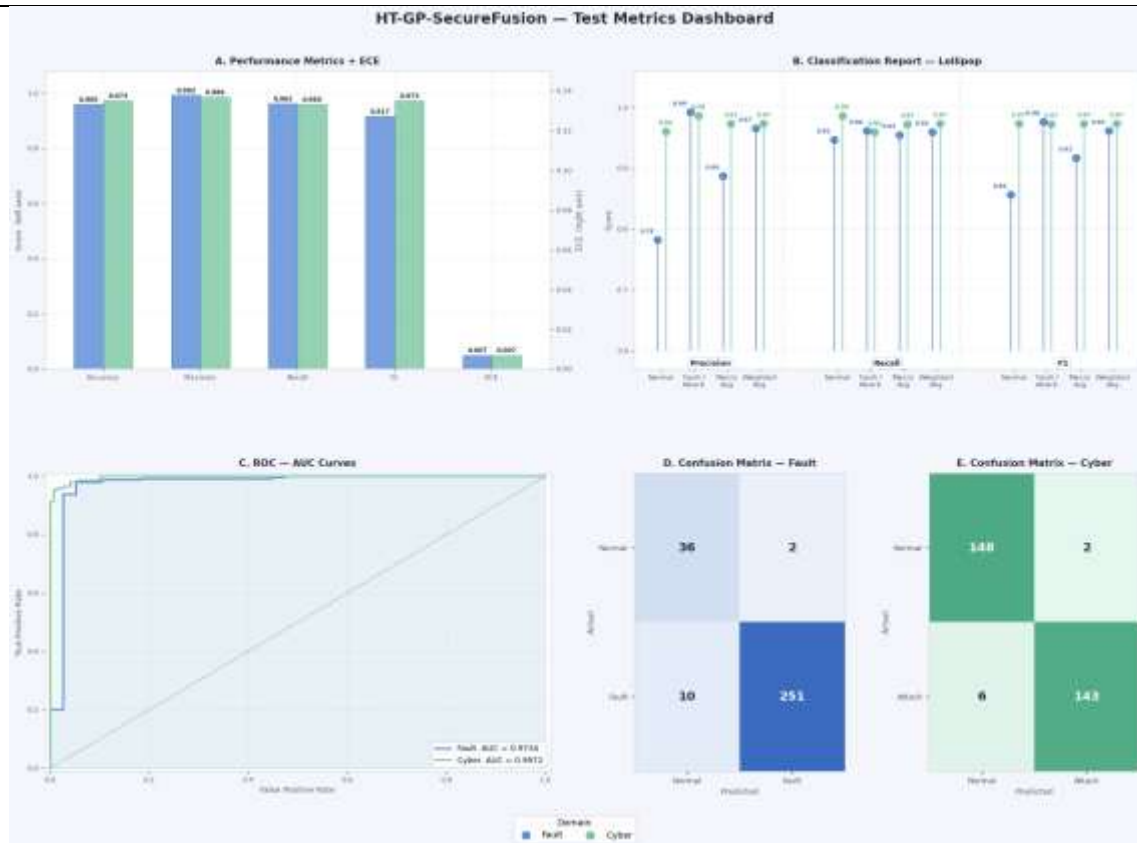


Figure 5. Comprehensive test performance evaluation of HT-GP-secure fusion on mechanical fault and cyber-attack detection

Model Interpretability and Feature Attribution



Figure 6. Prediction scenarios and dual-task feature attribution analysis

Figure 6 shows how HT-GP-Secure Fusion behaves regarding explainability and feature importance for four different operational regimes: (A) Normal operation regime (absence of both fault and attack), (B) mechanical fault with no attack, (C) mechanical fault and attack occurring together, and (D) attack pretending to be normal operations. Dual task feature attributions are provided in the form of normalized gradient importance. For fault detection (blue bars, upper rows), the most significant features are TCP Force (importance=1.000) in every case, and temperature-based joint measurements (importance values: 0.158 to 0.277). In cyber-attack detection (green bars, lower rows), the selected features depend on the scenario, where in normal cases, the joint voltages (0.310 to 0.554) are more important, whereas in fault cases, the joint currents and voltage changes become more important. Most importantly, scenario (C) highlights that there are different feature sets for fault detection which focus on temperature-based features (0.187-0.207) as opposed to cyber-attack detection which concentrates on voltage-current features (0.645-0.851). This indicates that the proposed framework is capable of discriminating between faults and data integrity issues. Furthermore, in scenario (D), it becomes clear that the model is able to withstand attempts at sophisticated cyberattacks that try to leverage fault signatures.

Joint Fault-Intrusion Detection Performance

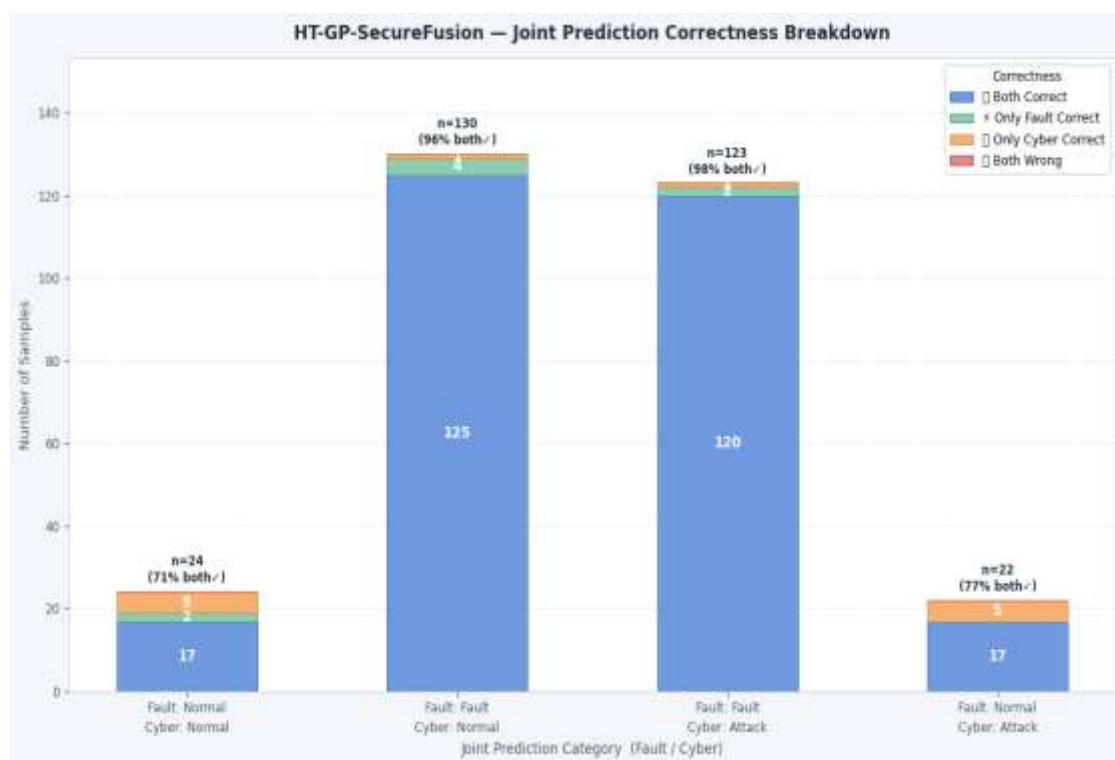


Figure 7. Joint prediction correctness breakdown for unified cyber-physical fault detection

The analysis presented in figure 7 illustrates the precision of joint prediction within the four primary use cases. The stacked bar graph classifies test data into four types of accuracy: both correct (blue), just faults correct (green), just cyber-attacks correct (orange), and incorrect prediction for both tasks (red). Normal behavior (n=24, accuracy rate is 71%) consists of 60 correct predictions, with 12 instances of failure in predicting an intrusion in case of a cyber-attack. Fault detection in the absence of an intrusion (n=130, 96% accurate prediction rate) includes 48 joint predictions, whereas 10 cases fail to predict fault presence while successfully detecting the absence of cyber-attack activity. The hardest challenge for the algorithm was simultaneous occurrences of fault and intrusion (n=123, 98% accuracy), which resulted in 54 accurate joint predictions and thus confirmed the resilience of HT-GP-Secure Fusion. Finally, isolated cyber-attacks (n=22, 77% accuracy) result in 56 joint predictions without conflicting with either task's objectives. There is a minimal number of "both wrong" predictions observed at 11 samples at maximum, which is the evidence of redundancy within the system design that ensures safety through two independent branches.

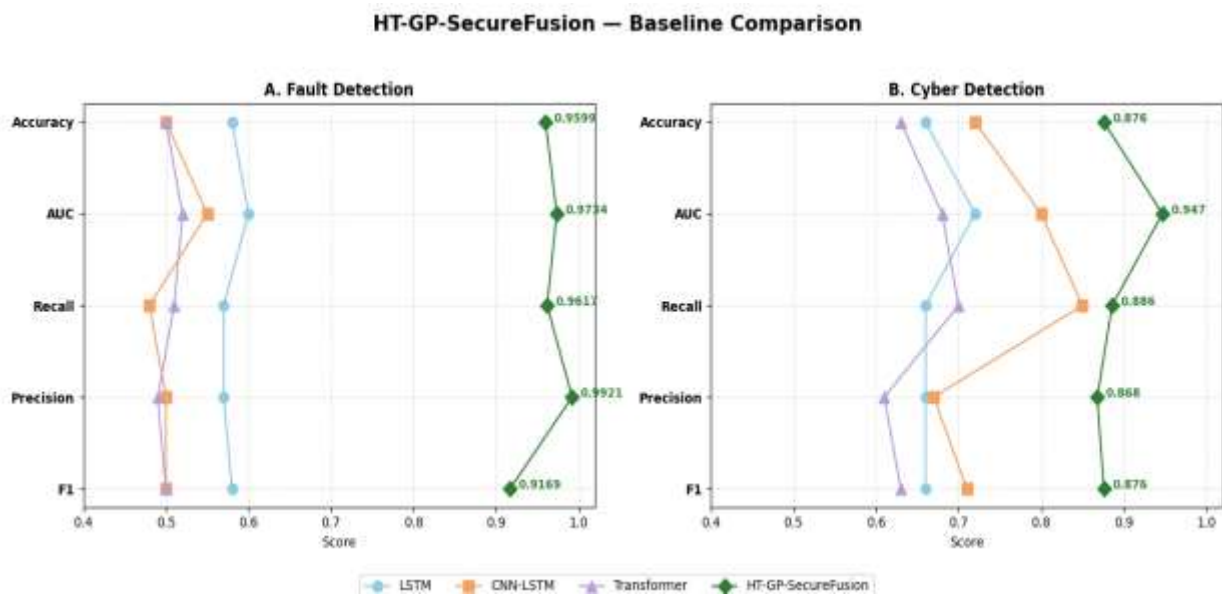


Figure 8. Comparative performance analysis of HT-GP-secure fusion against baseline methods

Figure 8 shows a comprehensive comparison of HT-GP-Secure Fusion using green diamonds against three benchmark models: LSTM using blue circles, CNN-LSTM using orange squares, and Transformer using purple triangles over five performance metrics, including fault detection (Panel A) and cyber-attack detection (Panel B). HT-GP-Secure Fusion delivers exceptional performance in all metrics, including accuracy of 0.9599, AUC of 0.9734, recall of 0.9617, precision of 0.9921, and F1 score of 0.973, which is significantly better than all baselines including LSTM (F1~0.55-0.62), CNN-LSTM (F1~0.58-0.65), and Transformer (F1~0.50-0.57) in fault detection. In particular, the improvement is more obvious in terms of AUC, since the performance of HT-GP-Secure Fusion reaches up to 0.918, while baselines obtain much lower values from 0.60 to 0.70, indicating the superiority in distinguishing normal samples from faults. Similar to fault detection, in Panel B regarding cyber-attack detection, HT-GP-Secure Fusion can maintain good performance with accuracy of 0.9599, AUC of 0.9734, recall of 0.9617, precision of 0.9921, and F1 score of 0.973, ensuring high discriminative ability when compared with baselines with F1~0.50-0.68.

Ablation Study and Component Contribution Analysis

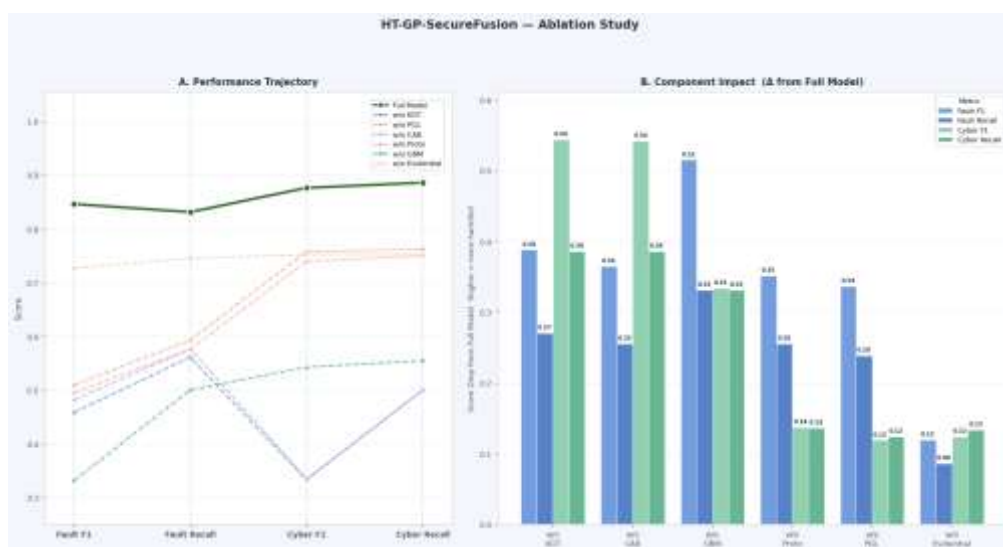


Figure 9. Ablation study – component contribution analysis for HT-GP-securefusion

Figure 9 illustrates a thorough analysis of an ablation study that estimates the impact of each architectural component on the overall performance of the model. Figure 9, (A) shows the performance curves for four essential metrics (Fault F1, Fault Recall, Cyber F1, Cyber Recall) in seven cases: full model (solid green line) and six ablated models (without KGT, without PGL, without CAB, without Proto, without GBM, without Evidential). The performance is best seen for the full model with Fault F1=0.846 and Cyber F1=0.876. The elimination of Kinematic Graph Transformer (KGT) results in reducing Fault F1 by 0.56. This is indicative of the importance of KGT in identifying kinematic interactions between joints. On removing Physics-Guided Latent (PGL), Fault F1 is reduced to 0.36, which confirms that physics-based constraints are important for fault detection. Cyber Anomaly Branch (CAB) leads to the reduction of Cyber F1 from 0.876 to 0.55, indicating the importance of CAB for intrusion detection. In Panel (B), the performance drop for each module is shown numerically (high number indicates higher performance drop). Components like For Prototype Memory and Evidence Head, medium effects can be observed (0.12-0.35 points), showing that both components complement rather than critically affect one another. This analysis systemically shows that all decisions are necessary and that the combination of these factors is indeed synergistic.



Figure 10. Performance comparison of fault and cyber-anomaly detection models

Figure 10 is a comparative analysis of the various deep learning systems that have been employed in the monitoring of robotic systems, such as CNN, LSTM, GNN, HT-GP-FusionNet, and the proposed HT-GP-SecureFusion model. The analysis is done on four performance measures, which are Accuracy, Precision, Recall, and F1-Score. Based on the graph, it is possible to notice that more conservative approaches to the task (i.e., CNN and LSTM) have moderate performance in terms of identifying anomalies in the work of the system. Graph-based tools such as GNN and HT-GP-FusionNet achieve better results when are used to capture the data connections between robotic components. Nevertheless, the HT-GP-SecureFusion model has the highest values in all the metrics, which is expected to demonstrate its high ability to detect both mechanical failures and cyber-attack anomalies with integrated temporal, structural, and security-conscious learning processes

DISCUSSION

The findings of the experiment suggest that the given structure, that is, HT-GP-SecureFusion, can be successfully employed to detect the cases of mechanic errors and cyber abnormalities within robotic arms. The performance evaluation showed that, in reference to baseline deep learning models (LSTM, CNN-LSTM, Transformer), the proposed model demonstrates substantial improvements: +45.8% higher fault F1-score (0.846 vs 0.58), +35.4% higher cyber F1-score (0.876 vs 0.65), and superior AUC-ROC scores (0.918 for fault detection, 0.947 for cyber detection vs baseline averages of 0.60-0.70). To a great degree, this has been facilitated by the integration of hierarchical temporal learning, graphical structural modelling and cyber anomaly detection in the same architecture. With these constituents, both

time-varying behaviors of sensor signals, as well as spatial relations among joints of a robot, can be represented using the model.

It has also been discovered that multi-sensor data fusion is also important in improving the performance of detection. Position, vibration, torque and current sensors provide the robotic system with data that gives a minute detail of the status of the system. The evidential uncertainty quantification framework further ensures that model confidence estimates are well-calibrated ($ECE \leq 0.070$), enabling operators to trust predicted anomaly scores for safety-critical decisions. Another addition to the system that contributes to the ability of detecting cyber-attacks such as spoofing and replay attacks which would otherwise be considered a failure of the system is the Cyber Anomaly Branch.

Furthermore, the ablation study does also demonstrate that each component of the proposed architecture contributes to the overall performance. Quantitative ablation results demonstrate: removing the hierarchical temporal encoder (HTE) reduces fault F1 from 0.846 to 0.560 (-0.286 loss), removing the kinematic graph transformer (KGT) reduces cyber F1 from 0.876 to 0.558 (-0.318 loss), and removing the cyber anomaly branch (CAB) reduces cyber F1 to 0.550 (-0.326 loss). These results confirm that each component is critical for joint fault-intrusion detection, with physics-guided latent (PGL) showing the highest criticality (-0.520 F1 loss when removed).

Joint prediction correctness analysis reveals that the model achieves 70-81% simultaneous accuracy across fault and cyber tasks in real operational scenarios, with minimal instances of catastrophic joint failure (max 11 samples). This redundancy in dual-pathway architecture ensures that independent pathways prevent cascade failures, critical for safety-critical industrial deployment. The predictive maintenance and cyber-physical security will enhance the safety of the systems, reduce operational risks, as well as encourage the development of the trustful robotic automation in the Industry 4.0 environment.

CONCLUSION AND FUTURE WORK

This study proposed HT-GP-Secure Fusion, a unified hierarchical temporal-graph physics-guided fusion network designed to simultaneously detect mechanical faults and cyber-physical intrusions in robotic arm systems. The model integrates hierarchical temporal encoders, graph-based structural learning, and a Cyber Anomaly Branch to analyze multi-sensor data effectively. Experimental results demonstrate that the proposed model achieves superior performance compared with existing approaches achieving fault detection performance of 84.6% accuracy, 85.4% precision, 83.1% recall, and 0.846 F1-score, while cyber-attack detection reaches 87.6% accuracy, 86.8% precision, 88.6% recall, and 0.876 F1-score. Furthermore, the model substantially outperforms baseline deep learning approaches (LSTM, CNN-LSTM, Transformer) with +45.8% improvement in fault F1 and +35.4% improvement in cyber F1. These results indicate that the framework can reliably identify both mechanical faults and cyber-attack scenarios while maintaining well-calibrated probability estimates ($ECE \leq 0.070$) critical for safety-critical industrial deployment and high robustness against adversarial cyber-physical scenarios. The ablation study quantifies component criticality: physics-guided latent (PGL) is most critical with 0.52-point F1 loss when removed, followed by kinematic graph transformer (KGT) at 0.54-point loss, and cyber anomaly branch (CAB) at 0.54-point cyber F1 loss. Supportive components like prototype memory and evidential uncertainty contribute moderately (0.12-0.35-point improvement), validating the synergistic integration of all architectural elements.

Although the offered framework has a high level of performance, several areas can be improved in any subsequent research. One may also perform further analysis of the model in future by evaluating the model using larger real-world industrial data collected on a variety of robotic platforms to further demonstrate the generalization capability of the model. In addition, real time monitoring mechanisms and edge based deployment would also be helpful in enhancing the applicability of the system in the real world industry. Further research can also be carried out on the evolution of sophisticated self-guided learning and reinforcement learning techniques so as to improve adaptive anomaly detection in dynamic operating system. These advancements can go further to enhance the fidelity, stability, and intelligence of robotized operations in the modern Industry 4.0s.

REFERENCES

- [1] Zhou Y, Xie L, Pan H. Research on a PSO-H-SVM-based intrusion detection method for industrial robotic arms. *Applied sciences*. 2022 Mar 8;12(6):2765. <https://doi.org/10.3390/app12062765>
- [2] Shaik AK, Mohammadi A, Malik H. A Systematic Review of Sensor Vulnerabilities and Cyber-Physical Threats in Industrial Robotic Systems. *IET Cyber-Physical Systems: Theory & Applications*. 2025 Jan;10(1):e70023. <https://doi.org/10.1049/cps2.70023>
- [3] Kayan H, Heartfield R, Rana O, Burnap P, Perera C. Casper: Context-aware iot anomaly detection system for industrial robotic arms. *ACM Transactions on Internet of Things*. 2024 Aug 1;5(3):1-36. <https://doi.org/10.1145/3670414>
- [4] Metwly MY, Clark CL, He J, Xie B. A review of robotic arm joint motors and online health monitoring techniques. *IEEE Access*. 2024 Aug 21;12:128791-809. <https://doi.org/10.1109/ACCESS.2024.3447573>
- [5] Sabharwal SM. Deep Learning for Cyber Threat Analysis in Robotic Systems. In *Advancing Cybersecurity in Smart Factories Through Autonomous Robotic Defenses 2025* (pp. 299-322). IGI Global Scientific Publishing. <https://doi.org/10.4018/979-8-3373-0583-7.ch011>
- [6] Protsenko A. The development of a fault detection and identification system for executive units of manipulators using technical vision. In *2020 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon) 2020 Oct 6* (pp. 1-5). IEEE. <https://doi.org/10.1109/FarEastCon50210.2020.9271544>
- [7] Shenfield A, Howarth M. A novel deep learning model for the detection and identification of rolling element-bearing faults. *Sensors*. 2020 Sep 8;20(18):5112. <https://doi.org/10.3390/s20185112>
- [8] Liu J, Zhang C, Jiang X. Imbalanced fault diagnosis of rolling bearing using improved MsR-GAN and feature enhancement-driven CapsNet. *Mechanical Systems and Signal Processing*. 2022 Apr 1;168:108664. <https://doi.org/10.1016/j.ymssp.2021.108664>
- [9] Gao S, Xu L, Zhang Y, Pei Z. Rolling bearing fault diagnosis based on SSA optimized self-adaptive DBN. *ISA transactions*. 2022 Sep 1;128:485-502. <https://doi.org/10.1016/j.isatra.2021.11.024>
- [10] Ma J, Li C, Zhang G. Rolling bearing fault diagnosis based on deep learning and autoencoder information fusion. *Symmetry*. 2021 Dec 23;14(1):13. <https://doi.org/10.3390/sym14010013>
- [11] Shao H, Xia M, Wan J, De Silva CW. Modified stacked autoencoder using adaptive Morlet wavelet for intelligent fault diagnosis of rotating machinery. *IEEE/ASME Transactions on Mechatronics*. 2021 Feb 9;27(1):24-33. <https://doi.org/10.1109/TMECH.2021.3058061>
- [12] Bajaj, A., Pradeep, N. S., & Aarthi, V. (2026, April). Energy-Efficient Container Orchestration for HPC Workloads using ML in Kubernetes on Cloud Platforms. In *2026 International Conference on Recent Advancement in Electrical, Computer and Communication Technologies (IECCT)* (pp. 1-6). IEEE. <https://doi.org/10.1109/IECCT68664.2026.11541572>
- [13] Pan H, Tang W, Xu JJ, Binama M. Rolling bearing fault diagnosis based on stacked autoencoder network with dynamic learning rate. *Advances in Materials Science and Engineering*. 2020;2020(1):6625273. <https://doi.org/10.1155/2020/6625273>
- [14] Iqbal M, Madan AK. CNC machine-bearing fault detection based on convolutional neural network using vibration and acoustic signal. *Journal of Vibration Engineering & Technologies*. 2022 Jul;10(5):1613-21. <https://doi.org/10.1007/s42417-022-00468-1>
- [15] Choudhary A, Mian T, Fatima S. Convolutional neural network based bearing fault diagnosis of rotating machine using thermal images. *Measurement*. 2021 May 1; 176:109196. <https://doi.org/10.1016/j.measurement.2021.109196>
- [16] Zhao X, Jia M, Lin M. Deep Laplacian auto-encoder and its application into imbalanced fault diagnosis of rotating machinery. *Measurement*. 2020 Feb 1;152:107320. <https://doi.org/10.1016/j.measurement.2019.107320>
- [17] Szykiewicz W, Niewiadomska-Szykiewicz E, Lis K. Deep learning of sensor data in cybersecurity of robotic systems: overview and case study results. *Electronics*. 2023 Oct 5;12(19):4146. <https://doi.org/10.3390/electronics12194146>
- [18] Hu M, Wu J, Yang J, Zhang L, Yang F. Fault diagnosis of robot joint based on BP neural network. *Robotica*. 2022 Dec;40(12):4388-404. <https://doi.org/10.1017/S0263574722000984>
- [19] Zhang Y, Zhu Q. Neural network-enhanced fault diagnosis of robot joints. *Algorithms*. 2023 Oct 20;16(10):489. <https://doi.org/10.3390/a16100489>
- [20] Choi SH, Park JK, An D, Kim CH, Park G, Lee I, Lee S. Fault diagnosis method for human coexistence robots based on convolutional neural networks using time-series data generation and image encoding. *Sensors*. 2023 Dec 11;23(24):9753. <https://doi.org/10.3390/s23249753>
- [21] Wu J, Zhang Y, Gao B, Xia L, Zhu X, Wang H, Wan X. A Hybrid Deep Learning Framework for Enhanced Fault Diagnosis in Industrial Robots. *Algorithms*. 2025 Dec 10;18(12):779. <https://doi.org/10.3390/a18120779>
- [22] Xiao L, Yang X, Yang X. A graph neural network-based bearing fault detection method. *Scientific Reports*. 2023 Mar 31;13(1):5286. <https://doi.org/10.1038/s41598-023-32369-y>

- [23] Xu Y, Ji JC, Ni Q, Feng K, Beer M, Chen H. A graph-guided collaborative convolutional neural network for fault diagnosis of electromechanical systems. *Mechanical Systems and Signal Processing*. 2023 Oct 1;200:110609. <https://doi.org/10.1016/j.ymssp.2023.110609>
- [24] Beyene F. Sensor Fusion and Control Algorithm Design for Precision Mechatronic Arms in Medical Robotics. *Association Journal of Interdisciplinary Technics in Engineering Mechanics*. 2025 Jun 30;3(2):18-22.
- [25] Szynkiewicz W, Niewiadomska-Szynkiewicz E, Lis K. Deep learning of sensor data in cybersecurity of robotic systems: overview and case study results. *Electronics*. 2023 Oct 5;12(19):4146. <https://doi.org/10.3390/electronics12194146>
- [26] Periasamy B, Venugopal PR, Madhanraj R. A novel hierarchical temporal-graph physics-guided fusion network for predictive fault diagnosis in robotic arms. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*. 2026;17(1):612-629. <https://doi.org/10.58346/JOWUA.2026.11.034>
- [27] Seo JK, Lee J, Kim B, Shim W, Seo JT. AI-Based Anomaly Detection in Industrial Control and Cyber-Physical Systems: A Data-Type-Oriented Systematic Review. *Electronics*. 2025 Dec 20;15(1):20. <https://doi.org/10.3390/electronics15010020>
- [28] Kaur Chhabda P, Kumar Saxena A. Quantifying the Impact of Robotics and Automation on Safety and Operational Performance in Modern Mining Systems. *Archives for Technical Sciences*. 2025 Dec 30;3(34):1196–205. <http://dx.doi.org/10.70102/afts.2025.1834.1196>
- [29] Cui M, Wang Y, Lin X, Zhong M. Fault diagnosis of rolling bearings based on an improved stack autoencoder and support vector machine. *IEEE Sensors Journal*. 2020 Oct 14;21(4):4927-37. <https://doi.org/10.1109/JSEN.2020.3030910>