Literature Review on Anomaly Detection and Fault Prediction in Cyber-Physical Systems Using Advanced Machine Learning Techniques

Santoshkumar Vaman Chobe[1], *Weiwei Jiang*[2]

[1]Lincoln University College, Malaysia; Department of Information Technology, Pimpri Chinchwad College of Engineering and Research, Pune;

[2]School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 10086, China;

[1]pdf.Santoshkumar@lincoln.edu.my, sanchobe@gmail.com, [2]jww@bupt.edu.cn

Abstract

Cyber-Physical Systems (CPSs) are increasingly forming the backbone of critical infrastructures by tightly integrating computational algorithms with physical components through sensing, control, and actuation mechanisms. These systems enable intelligent, real-time decision-making across sectors such as manufacturing, energy, healthcare, transportation, and defense. However, the interconnected and complex nature of CPS introduces numerous challenges, including susceptibility to faults, cyber-attacks, system malfunctions, and unpredictable environmental conditions. Ensuring the resilience and reliability of CPS thus becomes a priority.

Anomaly detection and fault prediction are pivotal in achieving these goals. Traditional rule-based and statistical methods often fail to generalize across the diverse and high-dimensional data typically observed in CPS. Machine learning (ML) techniques, particularly advanced models like deep learning (DL), reinforcement learning (RL), and hybrid approaches, are showing promising results in learning hidden patterns, capturing temporal-spatial dependencies, and enabling adaptive decision-making.

This literature review focuses on the advancements made in applying ML for efficient anomaly detection and fault prediction in CPS. It systematically analyzes contributions in areas such as DL, RL, explainable artificial intelligence (XAI), federated learning (FL), hybrid models, and adaptive learning systems. The review highlights their comparative strengths, limitations, practical applications, and the role of explainability and real-time capability. Finally, it discusses open challenges, future research directions, and provides a comprehensive reference framework for

researchers and practitioners seeking to build trustworthy and intelligent CPS monitoring solutions.

Keywords: Cyber-Physical Systems, Deep Learning, Explainable AI, Federated Learning, Reinforcement Learning.

Introduction

The evolution of Cyber-Physical Systems (CPS) has transformed industries by integrating computational intelligence with physical processes. However, their complexity makes them vulnerable to anomalies and faults, which can lead to catastrophic failures and security breaches [1] [2] [3] [4]. Traditional anomaly detection techniques often fall short in handling the dynamic and heterogeneous nature of CPS data.

Traditional monitoring systems based on rule sets or heuristics lack the flexibility to adapt to the ever-evolving operational patterns and fail to capture subtle or nonlinear relationships among variables. This has spurred interest in data-driven approaches, particularly machine learning, which leverages historical and streaming sensor data to model behavior, detect anomalies, and predict failures. Deep learning models can uncover intricate spatiotemporal correlations, while reinforcement learning enables adaptive policy learning in dynamic environments.

At the same time, the adoption of ML in CPS is hindered by concerns over the interpretability of predictions, especially in regulated industries. Explainable AI addresses this gap by offering transparency into model decision-making processes. Together, ML and XAI are setting the stage for the next generation of CPS monitoring frameworks that are not only accurate and timely but also trustworthy and context-aware.

Background on CPS

CPS are complex, real-time systems composed of interconnected physical components (sensors, actuators, machines) and computational units (controllers, analytics engines) that communicate over wired or wireless networks. They are prevalent in smart manufacturing (Industry 4.0), healthcare (e.g., robotic surgeries), smart grids, transportation systems (e.g., autonomous vehicles), and more.

A typical CPS consists of three tightly integrated layers: the physical layer (real-world interactions), the network layer (data transmission and communication), and the cyber layer (data

processing, analytics, and control decisions). These layers must work synchronously, often under strict timing and safety constraints.

CPSs are exposed to uncertainties such as component aging, sensor drift, communication delays, and cyber threats. Failures in any part can cascade and disrupt the entire system. Therefore, robust monitoring mechanisms that can learn from data, generalize across unseen conditions, and adapt to evolving behavior are critical. This context forms the foundation for the growing interest in ML-based solutions tailored to CPS environments.

Anomaly Detection in CPS

Anomaly detection plays a pivotal role in safeguarding the integrity of CPS by identifying deviations from expected operational behavior. Such anomalies may result from sensor malfunctions, cyber-attacks, software bugs, or unexpected environmental interactions. Traditional anomaly detection techniques—like Principal Component Analysis (PCA), k-means clustering, and Support Vector Machines (SVMs)—offer reasonable performance for low-dimensional data but often fall short in handling the complex, high-dimensional, and temporally correlated data prevalent in CPS environments. These classical approaches are typically static, unable to learn from evolving system dynamics, and are sensitive to noise and missing data.

In contrast, recent advancements in deep learning have introduced models capable of capturing nonlinear patterns and temporal-spatial dependencies in multivariate sensor data. Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, have demonstrated the ability to detect sequence-based anomalies by modeling long-term dependencies. Autoencoders, both vanilla and variational, are effective in learning compact representations of normal behavior and flagging reconstruction errors as anomalies. Graph Neural Networks (GNNs) are especially useful in modeling complex topologies and dependencies among interconnected CPS components.

Recent Studies:

- Susumu Naito, et al. [5] proposed a Two-Stage AutoEncoder (TSAE) as an anomaly detection method for various fluid handling systems with dynamic components, such as power generation, water treatment, and chemical plants. They conducted an empirical study using publicly available datasets of water treatment systems. In the Water Distribution (WADI) Dataset [6], they found performance differences between methods;

TSAE obtained a higher F1-score than other state-of-the-art methods, that confirmed the performance superiority of TSAE.

- Eiteneuer & Niggemann [7] implemented LSTM networks for sequence-level anomaly detection in power distribution systems. Their model exhibited robustness against noisy inputs.
- Zhilei Zhao. et al. [8] applied GNNs for spatial-temporal anomaly detection in interconnected sensor networks of smart factories, demonstrating scalability and precision.

These works emphasize a shift from rigid threshold-based models to intelligent, data-driven frameworks capable of detecting subtle and context-aware anomalies in real time.

Fault Prediction in CPS

Fault prediction in CPS aims to anticipate failures before they occur, enabling proactive maintenance, reduced downtime, and enhanced safety. Unlike anomaly detection, which focuses on identifying abnormal events in real-time, fault prediction requires learning from historical data patterns to forecast future failures. This involves regression models, supervised classification, and sequence modeling techniques.

Early approaches utilized linear regression and support vector regression (SVR) to estimate failure probabilities, but these methods struggle with nonlinear relationships and delayed fault signatures. Recent work focuses on deep learning and reinforcement learning, enabling models to capture temporal dynamics and adapt to system changes.

Recent Studies:

- Hang Ruan et al. [9] developed a neural fault prediction framework tailored for industrial CPS, using multi-sensor fusion and time-series analysis to improve prediction lead time.
- Ahsan, Muhammad [10] proposed a CNN-LSTM hybrid model for fault prediction in rotating machinery. The proposed model adapts varying load conditions while having the necessary levels of accuracy.
- Limin Wang et al. [11] proposed a CNN-LSTM hybrid model for fault prediction in rotating machinery. The proposed model adapts varying load conditions while having the necessary levels of accuracy.

Deep Learning Approaches

- Deep learning has emerged as a powerful tool for anomaly detection and fault prediction in CPS due to its ability to model complex patterns in temporal and spatial data. Unlike traditional ML models that require handcrafted features, DL architectures such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory networks (LSTMs), and Transformers automatically learn hierarchical representations from raw data. CNNs are effective in capturing local spatial dependencies, whereas LSTMs excel at modeling long-term temporal dependencies. Transformers, originally designed for natural language processing, are increasingly used in CPS due to their superior performance in sequence modeling and parallelization.
- Recent research has showcased various applications of DL in CPS. For example, Orabi et al. [12] proposed a Transformer-based model for anomaly detection in smart factories, demonstrating improved accuracy and scalability. Gong et al. [13] combined CNN and LSTM to capture both spatial and temporal features for time-series prediction. Capogrosso et al. [14] explored an edge-cloud hybrid architecture using DL models for CPS monitoring, reducing latency while maintaining high prediction accuracy.

Reinforcement Learning Approaches

- RL is gaining prominence in CPS as it enables systems to make autonomous decisions in dynamic and uncertain environments. In RL, an agent learns optimal actions through trial and error by maximizing cumulative rewards. This approach is well-suited for CPS tasks like fault recovery, adaptive control, and energy-efficient operation.
- Q-learning and Deep Q-Networks (DQNs) have been widely adopted for CPS fault prediction and mitigation. Modirrousta et al. [15] applied Q-learning to predict and preempt system faults in a dynamic industrial setup. Schoepp et al. [16] proposed a RL model that ensures fault tolerance in robotic CPS, enhancing system resilience to anomalies RL's ability to operate in real-time and adapt to novel states makes it a key technology for future CPS deployments.

Explainable AI in CPS:

- As CPS applications extend to safety-critical domains like healthcare and transportation, the need for transparency and accountability in AI models becomes paramount.

Explainable AI provides mechanisms to interpret, visualize, and understand the decisions made by complex ML models. Techniques like SHAP (SHapley Additive exPlanations), LIME (Local Interpretable Model-agnostic Explanations), and Grad-CAM (Gradient-weighted Class Activation Mapping) have been employed to offer local and global model explanations.

- In the context of CPS, Asutkar et al. [17], Brusa et al. [18] utilized SHAP to enhance fault diagnosis interpretation in industrial systems. Zhou et al. [19] proposed XAI techniques to identify critical decision factors and diagnosing misclassification errors. These methods facilitate trust and compliance with industry standards by making AI decisions interpretable to engineers and stakeholders.

Federated and Distributed Learning:

- Federated learning and distributed learning frameworks are essential for CPS, where data is often generated at the edge (e.g., sensors, actuators) and may be sensitive or private. FL allows collaborative model training without centralizing the data, thereby preserving data privacy and reducing communication overhead.
- Marfo et al. [20], Hao et al. [21] implemented FL for anomaly detection in IoT-based CPS, achieving privacy-preserving accuracy improvements. Such frameworks are particularly useful in healthcare and critical infrastructure sectors where data sensitivity is high.

Hybrid AI Models:

- Hybrid AI approaches combine the strengths of multiple paradigms—such as DL, RL, and symbolic AI—to build robust CPS solutions. These models leverage DL's representational power, RL's adaptability, and symbolic AI's reasoning capabilities to enhance system performance and interpretability.
- Barros et al. [22] introduced a hybrid DL-RL model for smart grid monitoring, which improved fault detection accuracy and reduced false positives. Similarly, Vincent et al. [23] presented an AI pipeline integrating GNNs with RL to predict cyber-attacks and physical faults in Cyber-physical Systems. Hybrid models represent a promising direction for achieving high performance in heterogeneous CPS settings.

Adaptive Learning Techniques:

- Adaptive learning allows CPS models to evolve based on new patterns, changing environments, or concept drift. These techniques are crucial in dynamic systems where fixed models may become obsolete over time. Online learning, transfer learning, and ensemble updates are commonly used strategies.
- Chang et al. [24] proposed a self-updating anomaly detection system using ensemble models, which dynamically adjusted to new fault patterns. Jouini et al. [25] integrated concept drift detection with online learning algorithms to maintain model relevance over time. Adaptive learning ensures that CPS solutions remain resilient and accurate in non-stationary environments.

Real-Time and Edge AI in CPS:

- Real-time analytics is a critical requirement for CPS to support timely decision-making and control. Edge AI refers to deploying ML models directly on edge devices, reducing latency, network load, and improving responsiveness. This is particularly important for applications like autonomous vehicles, manufacturing control systems, and smart healthcare.
- Park et al. [26] deployed lightweight DL models on edge devices for real-time fault detection in smart grids. Arnika et al. [27] proposed a real-time monitoring and fault detection system for smart grids via IoT sensors along with deep learning algorithms. Carter Happer [28] introduced optimized CNN architectures that retain high classification accuracy while remaining resource-efficient. Wang et al. [29] emphasized the importance of trust and reliability in Edge-AI systems, proposing mechanisms for secure and explainable edge deployments in CPS. Real-time and edge AI are essential for meeting the performance, autonomy, and scalability demands of modern CPS.

Summary of Comparative Studies:

Table 1 summarizes the comparative evaluation of key machine learning techniques used in anomaly detection and fault prediction within Cyber-Physical Systems (CPS). The metrics compared include accuracy, real-time capabilities, explainability, and adaptability.

Table 1. Comparative Evaluation of key Machine Learning Techniques

Technique	Accuracy	Real-Time Capable	Explainability	Adaptive
LSTM	High	Yes	Moderate	Yes
CNN	Medium	Yes	Moderate	No
RL	High	Yes	Low	High
FL	Medium	Limited	Moderate	High
XAI	N/A	Dependent	High	N/A

- **LSTM (Long Short-Term Memory):** Widely used for capturing temporal dependencies in time-series CPS data. High accuracy and adaptability make LSTMs suitable for dynamic environments. However, they require significant training data and are moderately interpretable.
- **CNN (Convolutional Neural Networks):** Effective in identifying spatial features and anomalies in image-like sensor data. While capable of real-time execution, CNNs struggle with temporal dependencies and lack adaptability to concept drift without retraining.
- **Reinforcement Learning (RL):** Excels in learning optimal control strategies and adapting to changes via reward feedback. Though powerful and adaptive, RL models are often computationally heavy and lack transparency in their decision-making process.
- **Federated Learning (FL):** Supports collaborative learning across distributed devices without sharing raw data, enhancing privacy. However, real-time deployment remains limited due to communication overhead and synchronization issues.
- **Explainable AI (XAI):** Techniques like SHAP, LIME, and Grad-CAM are critical for model transparency and trust. XAI is often integrated with other models to provide post-hoc explanations, and while it enhances explainability, it is not inherently a predictive model.

This comparative analysis highlights that no single technique excels across all metrics. Hybrid approaches are increasingly being explored to balance accuracy, adaptability, real-time performance, and explainability.

Research Gaps and Future Directions:

Despite substantial progress, several research challenges persist in implementing effective ML-based anomaly detection and fault prediction in CPS:

- **Limited Interpretability of Deep Learning Models:** Most DL models operate as black boxes, offering limited insight into the rationale behind predictions. This hinders their adoption in safety-critical CPS applications.
- **Computational Demands of Reinforcement Learning:** RL-based models require extensive training and exploration, which can be computationally expensive and infeasible for real-time or edge-based deployments.
- **Privacy Concerns in Federated Learning:** Although FL enhances data privacy, ensuring end-to-end security and handling Non-Independent and Identically Distributed (non-i.i.d.) data distributions remain major challenges.
- **Real-Time Constraints in Edge Environments:** Many ML techniques fail to meet stringent real-time requirements under resource-constrained edge computing conditions. Optimizing model size and inference latency is crucial.
- **Generalizability Across CPS Domains:** Many studies are domain-specific, and solutions that generalize across industries (e.g., smart grid, healthcare, manufacturing) are still lacking.

Future work should emphasize developing interpretable, resource-efficient, and domain-adaptive models. In addition, incorporating causality, uncertainty quantification, and continual learning will be vital for next-generation CPS anomaly detection frameworks.

Conclusion:

This literature review underscores the growing role of machine learning—especially deep learning, reinforcement learning, federated learning, and explainable AI—in enhancing the reliability and intelligence of Cyber-Physical Systems. With increasing system complexity, the emphasis is shifting toward hybrid models that can learn, adapt, and explain their decisions in real-time.

While notable progress has been made, key limitations remain in interpretability, scalability, and real-time readiness. The path forward involves not only advancing algorithmic capabilities but also integrating interdisciplinary solutions that combine AI with domain-specific knowledge,

human oversight, and system-level validation. This will pave the way for deploying truly trustworthy, efficient, and robust CPS solutions across diverse sectors.

References:

[1] Kumar, S., & Patel, R. (2023). "Explainable AI for CPS Monitoring." *Proceedings of the ACM Conference on Cybersecurity.*

[2] Lee, D., & Kim, H. (2023). "Adaptive Learning for CPS Anomaly Detection." *IEEE Access.*

[3] Miller, D., & Stevens, P. (2024). "Distributed Learning for Cyber-Physical Security." *IEEE Access.*

[4] Zhao, L., & Chen, P. (2024). "Federated Learning for Anomaly Detection in CPS." *IEEE Transactions on Neural Networks.*

[5] Naito, Susumu, Yasunori Taguchi, Kouta Nakata, and Yuichi Kato. "Anomaly detection for multivariate time series on large-scale fluid handling plant using two-stage autoencoder." In *2021 International Conference on Data Mining Workshops (ICDMW)*, pp. 542-551. IEEE, 2021.

[6] Mathur, Aditya P., and Nils Ole Tippenhauer. "SWaT: A water treatment testbed for research and training on ICS security." In *2016 international workshop on cyber-physical systems for smart water networks (CySWater)*, pp. 31-36. IEEE, 2016.

[7] Eiteneuer, Benedikt, and Oliver Niggemann. "Lstm for model-based anomaly detection in cyber-physical systems." *arXiv preprint arXiv:2010.15680* (2020).

[8] Zhao, Zhilei, Zhao Xiao, and Jie Tao. "MSDG: Multi-scale dynamic graph neural network for industrial time series anomaly detection." *Sensors* 24, no. 22 (2024): 7218.

[9] Ruan, Hang, Bogdan Dorneanu, Harvey Arellano-Garcia, Pei Xiao, and Li Zhang. "Deep learning-based fault prediction in wireless sensor network embedded cyber-physical systems for industrial processes." *Ieee Access* 10 (2022): 10867-10879.

[10] Ahsan, Muhammad, Muhammad Waqar Hassan, Jose Rodriguez, and Mohamed Abdelrahem. "Enhanced fault diagnosis in rotating machinery using a hybrid CWT-LeNet-5-LSTM model: Performance across various load conditions." *IEEE Access* (2024).

[11] Wang, Limin, Xueyu Li, Ridong Zhang, and Furong Gao. "Reinforcement learning-based optimal fault-tolerant tracking control of industrial processes." *Industrial & Engineering*

Chemistry Research 62, no. 39 (2023): 16014-16024.

[12] Orabi, Moussab, Kim Phuc Tran, Philipp Egger, and Sébastien Thomassey. "Anomaly detection in smart manufacturing: An Adaptive Adversarial Transformer-based model." *Journal of Manufacturing Systems* 77 (2024): 591-611.

[13] Gong, Yuhao, Yuchen Zhang, Fei Wang, and Chi-Han Lee. "Deep learning for weather forecasting: A cnn-lstm hybrid model for predicting historical temperature data." *arXiv preprint arXiv:2410.14963* (2024).

[14] Capogrosso, Luigi, Shengjie Xu, Enrico Fraccaroli, Marco Cristani, Franco Fummi, and Samarjit Chakraborty. "Learning-enabled CPS for edge-cloud computing." In *2024 IEEE 14th International Symposium on Industrial Embedded Systems (SIES)*, pp. 132-139. IEEE, 2024.

[15] Modirrousta, M. H., M. Aliyari Shoorehdeli, Mostafa Yari, and Arash Ghahremani. "DQLAP: Deep Q-Learning Recommender Algorithm with Update Policy for a Real Steam Turbine System." *arXiv preprint arXiv:2210.06399* (2022).

[16] Schoepp, Sheila, Mehran Taghian, Shotaro Miwa, Yoshihiro Mitsuka, Shadan Golestan, and Osmar ZaÃĐane. "Enhancing hardware fault tolerance in machines with reinforcement learning policy gradient algorithms." *arXiv preprint arXiv:2407.15283* (2024).

[17] Asutkar, Supriya, and Siddharth Tallur. "An explainable unsupervised learning framework for scalable machine fault detection in Industry 4.0." *Measurement Science and Technology* 34, no. 10 (2023): 105123.

[18] Brusa, Eugenio, Luca Cibrario, Cristiana Delprete, and Luigi Gianpio Di Maggio. "Explainable AI for machine fault diagnosis: understanding features' contribution in machine learning models for industrial condition monitoring." *Applied Sciences* 13, no. 4 (2023): 2038.

[19] Zhou, Yujing, Marc L. Jacquet, Robel Dawit, Skyler Fabre, Dev Sarawat, Faheem Khan, Madison Newell et al. "Explainable Machine Learning for Cyberattack Identification from Traffic Flows." In *2025 IEEE Security and Privacy Workshops (SPW)*, pp. 329-334. IEEE, 2025.

[20] Marfo, William, Deepak K. Tosh, and Shirley V. Moore. "Federated learning for efficient condition monitoring and anomaly detection in industrial cyber-physical systems." In *2025 International Conference on Computing, Networking and Communications (ICNC)*, pp. 740-746. IEEE, 2025.

[21] Hao, Junfeng, Juan Chen, Peng Chen, Yang Wang, Xianhua Niu, Lei Xu, and Yunni Xia. "Efficiently detecting anomalies in IoT: A novel multi-task federated learning method." In *International Conference on Collaborative Computing: Networking, Applications and Worksharing*, pp. 100-117. Cham: Springer Nature Switzerland, 2023.

[22] Barros, E. B. C., W. O. Souza, D. G. Costa, G. P. Rocha Filho, G. B. Figueiredo, and M. L. M. Peixoto. "Energy management in smart grids: An Edge-Cloud Continuum approach with Deep Q-learning." *Future Generation Computer Systems* 165 (2025): 107599.

[23] Vincent, Edeh, Mehdi Korki, Mehdi Seyedmahmoudian, Alex Stojcevski, and Saad Mekhilef. "Reinforcement learning-empowered graph convolutional network framework for data integrity attack detection in cyber-physical systems." *CSEE Journal of Power and Energy Systems* 10, no. 2 (2024): 797-806.

[24] Chang, Bao Rong, Hsiu-Fen Tsai, and Guan-Ru Chen. "Self-adaptive server anomaly detection using ensemble meta-reinforcement learning." *Electronics* 13, no. 12 (2024): 2348.

[25] Jemili, Farah, Khaled Jouini, and Ouajdi Korbaa. "Intrusion detection based on concept drift detection and online incremental learning." *International Journal of Pervasive Computing and Communications* 21, no. 1 (2025): 81-115.

[26] Park, Donghyun, Seulgi Kim, Yelin An, and Jae-Yoon Jung. "LiReD: A light-weight real-time fault detection system for edge computing using LSTM recurrent neural networks." *Sensors* 18, no. 7 (2018): 2110.

[27] Tyagi, Reshu, Pawan Kumar, Pramod Kumar Sagar, Manish Saraswat, and Abhinav Bansal. "Real-time monitoring and fault detection in smart grids using IoT sensors and deep learning algorithm." In *2024 4th International Conference on Advancement in Electronics & Communication Engineering (AECE)*, pp. 1128-1132. IEEE, 2024.

[28] Happer, Carter. "Design and Deployment of Lightweight CNN Models for Real-Time Fault Detection on IIoT Edge Gateways." (2022).

[29] Wang, Xiaojie, Beibei Wang, Yu Wu, Zhaolong Ning, Song Guo, and Fei Richard Yu. "A survey on trustworthy edge intelligence: From security and reliability to transparency and sustainability." *IEEE Communications Surveys & Tutorials* (2024).