Mitigating Cyber Threats in Smart Factories: A Hybrid Artificial Intelligence and Machine Learning Approach

# CHAPTER ONE

# Introduction

## 1.1 Background of the Study

The onset of the Fourth Industrial Revolution, or Industry 4.0, has redefined manufacturing by integrating cyber-physical systems, Industrial Internet of Things (IIoT), artificial intelligence (AI), and machine learning (ML) into industrial operations (de Azambuja et al., 2023). At the heart of this transformation are smart factories, where interconnected sensors, actuators, control systems, and cloud-based services communicate seamlessly to optimize processes in real-time (Goyal, Rajawat, Solanki, Zaaba, and Long, 2023). However, this hyper-connectivity also exposes critical infrastructure to an evolving range of cyber threats, which can lead to operational shutdowns, data breaches, and physical damage.

Smart factories differ from traditional manufacturing plants due to their reliance on automated decision-making, real-time analytics, and edge computing. While these advancements promote agility and efficiency, they simultaneously increase the attack surface, making these environments more vulnerable to sophisticated cyberattacks such as Advanced Persistent Threats (APTs), ransomware, Distributed Denial of Service (DDoS) attacks, and zero-day exploits (Gajiwala, 2024). The convergence of operational technology (OT) and information technology (IT) within smart factories further compounds this risk, as traditional security frameworks were not designed for such complex, integrated systems (Zhang, Yang, and Yang, 2023).

Conventional cybersecurity solutions primarily rely on rule-based systems and signature detection, which are often reactive and ineffective against novel or polymorphic threats (Wang, Kuo, and Chen, 2021). These systems struggle to cope with the velocity, volume, and variety of data generated in real-time industrial environments (Jimmy, 2023). In contrast, AI and ML offer the potential for proactive threat detection, anomaly recognition, and autonomous response, making them vital tools in modern cybersecurity strategies.

AI-driven systems can monitor vast streams of network traffic and operational data to detect behavioral anomalies and potential attacks before they manifest. Machine learning algorithms, such as supervised learning (e.g., Random Forests, Support Vector Machines), unsupervised learning (e.g., clustering, autoencoders), and deep learning models (e.g., Long Short-Term Memory networks), have shown promising results in reducing false positives and improving real-time detection of cyber threats (Chen, Shen, Wang, Ke, and Xu, 2024).

A compelling case for the role of AI in cybersecurity is found in smart grid and IIoT environments, which share many vulnerabilities with smart factories. According to (Vilkelyte et al., 2024), AI and ML are increasingly employed in grid applications to detect and neutralize threats targeting interconnected components. The paper emphasizes the need for intelligent and adaptive cybersecurity architectures capable of evolving with the threat landscape.

In smart factory settings, AI enhances incident response times, automates intrusion detection systems (IDS), and minimizes human error in threat analysis. For example, (Elbes, Hendawi, Alzu’Bi, Kanan, and Mughaid, 2023) showed that machine learning models could outperform traditional methods in threat detection across multiple attack vectors, including phishing, malware, and insider threats.

Despite its advantages, implementing AI/ML in industrial environments is not without challenges. Key issues include the need for high-quality, labeled datasets, computational overhead, explainability of model decisions, and defense against adversarial AI attacks (Asere, Nuga, and Medugu, 2025). Additionally, legacy systems in older factories may not support the computational requirements of modern AI-based security solutions, making integration and scalability difficult.

To overcome these obstacles, researchers have begun exploring hybrid AI models that combine multiple learning paradigms. These hybrid systems typically integrate supervised, unsupervised, and deep learning methods to improve detection accuracy and reduce false positives. For instance, combining Random Forest classifiers for known threats, autoencoders for anomaly detection, and LSTMs for temporal pattern recognition offers a multi-dimensional defense mechanism tailored for complex smart factory networks (Diaba, 2024).

Another pressing issue is data localization and privacy. Smart factories operating in different regulatory jurisdictions must ensure compliance with data protection laws such as the GDPR. AI-driven security systems must be designed to uphold these standards while processing massive amounts of sensitive production and operational data (Gajiwala, 2024).

As digitalization accelerates, the economic and safety risks of cybersecurity breaches also grow. The use of AI to automate detection and response is not just a technical enhancement; it is increasingly a business imperative. According to Muheidat, Mallouh, Al-Saleh, Al-Khasawneh, and Tawalbeh, 2024), AI improves cost-efficiency by reducing manual analysis and downtime and offers real-time decision-making capabilities that legacy systems cannot match.

In response to these challenges and opportunities, this study aims to develop a hybrid AI/ML model tailored to smart factory cybersecurity. The proposed model will be trained and evaluated using publicly available industrial datasets, applying ensemble methods for optimized threat detection. This approach addresses critical gaps in current research by emphasizing real-world applicability, low-latency response, and compatibility with industrial control systems.

In conclusion, smart factories represent both the pinnacle of industrial automation and a growing frontier for cyber threats. Traditional security mechanisms are insufficient in this new paradigm. Artificial intelligence and machine learning provide a robust framework for proactive, scalable, and adaptive cybersecurity. However, to realize their full potential, these technologies must be strategically integrated through hybrid architectures that are context-aware, resource-efficient, and ethically sound.

## 1.2 Problem Statement

The integration of Industry 4.0 technologies into manufacturing, particularly the convergence of cyber-physical systems (CPS), Industrial Internet of Things (IIoT), and smart automation, has significantly increased both operational efficiency and the cybersecurity attack surface in smart factories. As factories become increasingly interconnected, they are also more vulnerable to sophisticated and evolving cyber threats, including Advanced Persistent Threats (APTs), ransomware, zero-day vulnerabilities, and insider threats (Jimmy, 2023). These attacks not only compromise data confidentiality and operational integrity but can lead to physical disruptions, safety hazards, and significant financial losses.

Traditional cybersecurity approaches, largely rule-based, signature-driven, and manually configured, struggle to detect and respond to these threats in real-time. They are inherently reactive and incapable of managing the dynamic, high-velocity, and heterogeneous data environments typical of smart factory ecosystems (Gajiwala, 2024). Furthermore, these legacy systems generate high false positive rates, which contribute to alert fatigue among security personnel and delayed threat response, both of which can have catastrophic consequences in real-time production environments (Elbes et al., 2023).

While Artificial Intelligence (AI) and Machine Learning (ML) have shown immense potential in enhancing cybersecurity by enabling predictive analytics, anomaly detection, and automated incident response, current implementations remain fragmented and limited. Standalone AI models often lack the robustness, contextual adaptability, or real-time responsiveness needed for industrial control systems. In many instances, these models are either trained on irrelevant datasets, fail to adapt to evolving attack vectors, or cannot operate efficiently within the constraints of legacy manufacturing infrastructure (Diaba, 2024).

Moreover, cyber threats are becoming increasingly multi-dimensional, requiring detection mechanisms that can identify both known (signature-based) and unknown (behavioral or zero-day) attacks across various temporal and spatial layers of factory networks (Chen et al., 2024). Single-method AI approaches (e.g., only using Random Forest or LSTM) fall short of this requirement due to their inherent limitations in coverage and adaptability.

Consequently, there is a critical need for a hybrid AI/ML model that can integrate the strengths of multiple learning paradigms, supervised for known threats, unsupervised for anomalies, and deep learning for sequential patterns, to deliver comprehensive, real-time cybersecurity protection in smart factories (Vilkelytė et al., 2024).

Therefore, the core problem this research seeks to address is the absence of a scalable, context-aware, and empirically validated hybrid AI and machine learning approach specifically optimized for mitigating diverse cyber threats in smart factory environments. Without such targeted innovations, smart factories will continue to remain highly vulnerable, jeopardizing the reliability, safety, and competitive edge promised by Industry 4.0.

## 1.3 Aim and Objectives of the Study

The aim of this study is to develop and evaluate a hybrid Artificial Intelligence (AI) and Machine Learning (ML) model designed to effectively detect, prevent, and mitigate cyber threats in smart factory environments.

### 1.3.1 Objectives

1. To critically analyze existing applications of AI and ML in smart factory cybersecurity, evaluating their strengths, weaknesses, and limitations across different industrial control system contexts.
2. To design and develop a hybrid AI/ML-based cybersecurity framework, combining supervised learning (e.g., Random Forest), unsupervised learning (e.g., Autoencoders), and deep learning (e.g., Long Short-Term Memory networks) to handle both known and unknown threats across time-series factory data.
3. To implement the hybrid model using real-world smart factory datasets, such as network logs, sensor outputs, and attack simulations from Water Distribution Dataset acquired from Itrust Centre for Research in Cybersecurity.
4. To evaluate and benchmark the hybrid model’s performance against conventional or standalone AI models using predefined metrics including accuracy, precision, recall, F1-score, false positive rate, and detection time.

## 1.4 Significance of the Study

The significance of this study lies in its direct response to the growing inadequacy of traditional cybersecurity approaches in protecting smart factory environments. As smart factories adopt interconnected technologies such as industrial IoT, edge computing, and cyber-physical systems, their vulnerability to cyberattacks increases dramatically. Conventional rule-based systems are no longer capable of addressing the velocity and complexity of cyber threats targeting these real-time, data-intensive environments (Jimmy, 2023). This research is significant as it proposes a novel hybrid AI/ML model that offers adaptive, scalable, and intelligent threat detection tailored specifically to the unique architecture and needs of smart factories.

Furthermore, this study addresses the critical performance limitations observed in current standalone AI models, which often fail to generalize across industrial scenarios due to poor adaptability or lack of temporal analysis capabilities. By integrating supervised, unsupervised, and deep learning techniques, the proposed model aims to overcome these limitations and provide more robust detection of both known and novel threats (Chen et al., 2024). This makes the study particularly relevant for industries relying on continuous production and real-time data integrity, where even minor delays or false positives can lead to operational losses or safety hazards.

The research also contributes to bridging the implementation gap by focusing on deployment feasibility in real-world industrial contexts. Issues such as integration with legacy systems, computational efficiency, and adversarial robustness are explored, making the study applicable to both high-tech and resource-constrained factory settings (Asere et al., 2025). In doing so, it supports the development of cybersecurity infrastructures that are not only technically advanced but also practical and sustainable in operational environments.

In a broader context, the study’s outcomes are valuable to cybersecurity professionals, industrial engineers, and policy-makers aiming to strengthen digital resilience in manufacturing. The empirical evidence generated through performance evaluation provides a foundation for more informed decisions on deploying AI-based security frameworks in critical infrastructure. Moreover, by contributing to the growing academic discourse on hybrid AI security systems, the research lays groundwork for future innovations in autonomous, intelligent cybersecurity models (Vilkelytė et al., 2024).

## 1.5 Scope of the Study

This study is specifically scoped to the technical evaluation and implementation of a hybrid Artificial Intelligence and Machine Learning (AI/ML) model for detecting and mitigating cyber threats in smart factory environments. The research focuses on industrial settings characterized by interconnected cyber-physical systems (CPS), IoT networks, and real-time data processing architectures, which collectively form the foundation of Industry 4.0. It emphasizes the development, simulation, and performance benchmarking of a multi-layered AI model integrating supervised (e.g., Random Forest), unsupervised (e.g., Autoencoder), and deep learning (e.g., LSTM) techniques.

The scope is limited to technical threat detection mechanisms, excluding broader organizational factors such as cybersecurity policy, employee behavior, regulatory compliance, and economic cost analysis. Additionally, while AI and ML systems are explored, the ethical and legal implications of AI use in cybersecurity, such as bias in model predictions or data governance, are only discussed to the extent they affect technical deployment and model reliability (Asere et al., 2025).

The model will be trained and tested using a water distribution dataset, from <https://itrust.sutd.edu.sg/itrust-labs_datasets/dataset_info/>, an iTrust research for Cybersecurity, which simulate attack scenarios relevant to industrial control systems. As such, the study does involve direct data collection from live industrial networks collected in 2019.

Lastly, the proposed solution is scoped for simulation-based validation rather than field deployment. It is intended as a prototype for future deployment and does not extend to hardware implementation on factory equipment or edge devices. Nonetheless, scalability and resource constraints typical of industrial edge environments are considered in model selection and architecture design to enhance future applicability (Vilkelytė et al., 2024).

## 1.6 Methodology for the Study

This study adopts a quantitative, simulation-based research methodology aimed at designing, developing, and evaluating a hybrid Artificial Intelligence (AI) and Machine Learning (ML) model for cyber threat detection in smart factory environments. The choice of methodology is rooted in the need for empirical validation of model performance across multiple threat dimensions, accuracy, latency, and robustness, without disrupting real-world industrial operations.

The experimental framework involves offline analysis using the WADI (Water Distribution) dataset, sourced from the iTrust Laboratory at the Singapore University of Technology and Design (SUTD). The WADI dataset is a real-world industrial dataset derived from a scaled-down water treatment testbed that mimics industrial control system (ICS) operations, including programmable logic controllers (PLCs), SCADA systems, and networked sensors. The dataset includes both normal operational data and labeled cyberattacks, making it suitable for supervised, unsupervised, and temporal pattern learning. Access to this dataset was obtained through formal request and approval from iTrust, ensuring ethical and authorized data usage.

Data preprocessing includes cleaning, normalization, and feature selection. The model will be trained using 70% of the data and validated on 30%, with performance evaluated using metrics such as accuracy, precision, recall, F1-score, false positive rate, and detection time (Chen et al., 2024). Model implementation will be done in Python using Scikit-learn, TensorFlow/Keras, and evaluated in Jupyter Notebook. Results will be compared to standalone models to validate the benefit of hybridization.

# Chapter two

# Literature review

## 2.1 Introduction

The increasing digitalization of manufacturing processes under Industry 4.0 has elevated the importance of cybersecurity in smart factories. These environments rely heavily on interconnected systems, cyber-physical systems (CPS), industrial IoT (IIoT), and cloud-based infrastructures, which, while enhancing efficiency, significantly expand the potential attack surface for cyber threats. The convergence of operational technology (OT) and information technology (IT) introduces unique vulnerabilities that traditional security mechanisms struggle to address.

Traditional rule-based and signature-based detection systems are often reactive and unable to identify zero-day threats or subtle anomalies in real-time, dynamic factory environments (Gajiwala, 2024). In response to these limitations, researchers and practitioners are increasingly turning to Artificial Intelligence (AI) and Machine Learning (ML) to provide proactive, adaptive, and scalable cybersecurity solutions.

This chapter reviews existing literature across five critical areas: the conceptual foundation of smart factory security, theoretical models supporting AI-driven detection, empirical applications of ML techniques, analysis of related works, and identified gaps in current research. The objective is to establish a robust knowledge base that justifies the development of a hybrid AI/ML model tailored specifically to cyber threat detection in smart factories.

## 2.2 Conceptual Overview

### 2.2.1 Smart Factories

Smart factories represent the core technological innovation driving Industry 4.0, an era characterized by the fusion of cyber-physical systems (CPS), the Industrial Internet of Things (IIoT), big data analytics, and artificial intelligence (AI) into manufacturing processes. These factories function as highly digitized production environments where machines, sensors, and control systems are interconnected to enable real-time data sharing, autonomous decision-making, and self-optimization of operations (Mosleuzzaman, Arif, and Siddiki, 2024).

Unlike traditional factories, smart factories operate on the foundation of integrated information and operational technologies. CPS facilitates the connection between the physical and digital world, allowing physical assets to be monitored and controlled through computational algorithms. When integrated with IoT, these systems become capable of collecting large volumes of data from sensors and actuators across the production floor, which can then be processed in real time to optimize decision-making (Exploring the Foundation of Smart Factories in Industry 4.0: A Conceptual Review, 2024)

The functional advantages of smart factories are well established. These include improved operational efficiency, reduced downtime, predictive maintenance, real-time quality control, and increased production flexibility. For example, embedded AI systems allow for adaptive learning based on historical production patterns, enabling proactive fault detection and response without human intervention (Musaeva, Vyachina, and Aliyeva, 2024). Moreover, smart factories enable mass customization through flexible production lines that adjust based on user demand and product specifications.

However, despite these benefits, the transition to smart manufacturing introduces complex challenges, particularly in cybersecurity. Smart factories are inherently exposed to cyber risks due to their dependence on interconnected digital infrastructure. Every sensor, actuator, and communication protocol introduce a potential entry point for malicious actors. The reliance on real-time data also means that any disruption can have immediate and severe consequences on production, safety, and supply chain continuity (Masum, 2023).

Moreover, the fusion of IT (information technology) and OT (operational technology) in smart factories complicates the security landscape. Unlike IT systems, OT components are often designed with long operational lifespans and minimal consideration for cybersecurity, making them vulnerable to attacks such as ransomware, DDoS, and advanced persistent threats (APTs) (Yi and Jeong, 2022). Furthermore, OT networks typically lack built-in encryption or access control mechanisms, and updates may be limited due to system stability requirements.

The need for secure, adaptive, and intelligent cybersecurity frameworks has become urgent. As highlighted by (Jamai, Ben Azzouz, and Saidane, 2020), IIoT devices in smart factories are particularly susceptible to attacks because of their limited computing power, outdated firmware, and frequent default configurations. Given these vulnerabilities, smart factories have become high-value targets for cybercriminals and even nation-state actors seeking to exploit industrial control systems for espionage or sabotage.

In summary, smart factories are transformative in their ability to create intelligent, agile, and efficient production environments. However, this transformation comes at the cost of increased cybersecurity complexity. Addressing these vulnerabilities demands the deployment of advanced threat detection mechanisms, particularly those powered by AI and ML, capable of monitoring large-scale data flows, identifying anomalies, and responding autonomously to evolving threats in real time.

### 2.2.2 Cyber-Physical Systems (CPS)

Cyber-Physical Systems (CPS) are foundational to smart factory operations, serving as the integrative bridge between the digital and physical realms of manufacturing. These systems comprise tightly coupled components of computation, networking, and physical processes that interact in real time to monitor, analyze, and control industrial activities. In a smart factory context, CPS enables the automation, flexibility, and adaptability necessary for high-efficiency and data-driven production environments (Yao et al., 2019).

CPS is not just an enabling technology; it is the structural backbone of Industry 4.0. Its architecture typically integrates industrial IoT (IIoT), real-time sensors, edge computing, and cloud services to create an interconnected ecosystem where every physical device is mirrored by a virtual counterpart. These digital twins communicate continuously, allowing for decentralized control and automated decision-making. As noted by Renteria-Marquez, Basaldua, Aguirre, Lara-Medrano, and Tseng (2024), the use of CPS in digital twins enables predictive scheduling, remote diagnostics, and real-time control of production systems.

However, the integration of CPS into manufacturing also introduces significant cybersecurity challenges. Unlike traditional IT systems, CPS must maintain both digital security and physical system integrity. Any cyberattack on a CPS can potentially cause physical damage, production loss, or even threaten human safety. These concerns are amplified in industrial environments where uptime and deterministic response are critical. The complexity and heterogeneity of CPS components, ranging from PLCs and SCADA systems to cloud-based analytics, further increase the attack surface and make end-to-end security more difficult to enforce (Simonthomas, Subramanian, and Mathiew, 2024).

As pointed out by Contreras et al. (2023), CPS in industrial environments face a wide range of cyber threats, including denial-of-service attacks, spoofing, and ransomware. The decentralized and distributed nature of CPS makes traditional perimeter-based security models ineffective. Therefore, there is a pressing need for embedded, context-aware, and autonomous security mechanisms that operate within the CPS architecture itself.

Moreover, CPS presents a unique trade-off between performance and security. Ensuring low-latency, high-reliability communication is essential in real-time manufacturing, but adding encryption, authentication, and intrusion detection can introduce delays and overhead. As Jain (2021) emphasizes, one of the core challenges in CPS security is maintaining this balance without degrading system performance or responsiveness.

To address these risks, architectural improvements to CPS are being proposed. For instance, new frameworks like the 8C architecture add dimensions such as coalition, customer integration, and content awareness to enhance system resilience and interoperability within smart factory environments (Jiang, 2018). These enhancements allow systems to become more adaptive to dynamic operational and cybersecurity conditions.

In conclusion, CPS is both a powerful enabler and a critical vulnerability point in smart factories. While it underpins the operational intelligence of modern manufacturing, it also introduces complex cybersecurity threats that demand advanced, multi-layered protection strategies. Understanding and securing CPS is therefore essential not just for factory productivity but also for maintaining industrial safety and continuity in the face of growing cyber risks.

### 2.2.3 Cybersecurity in Industrial Environments

As industrial environments adopt advanced digital technologies under Industry 4.0, cybersecurity has emerged as one of the most critical and complex challenges. Smart factories now rely on deeply interconnected systems including cyber-physical systems (CPS), industrial Internet of Things (IIoT), cloud computing, and machine learning tools. While these technologies increase efficiency, automation, and flexibility, they also significantly expand the attack surface, exposing critical infrastructure to a wide range of cyber threats (Masum, 2023).

Unlike traditional IT systems, industrial environments operate on both digital and physical layers, making cyberattacks potentially destructive in real-time operations. Attacks targeting industrial control systems (ICS), programmable logic controllers (PLCs), or SCADA components can directly disrupt production, damage physical equipment, or endanger human safety. These risks are intensified by the inherent vulnerabilities of OT (Operational Technology), such as legacy software, weak access controls, and limited cryptographic protection (Juarez, 2019).

One of the critical gaps in securing industrial environments is the disconnect between IT and OT security priorities. IT focuses on data confidentiality and integrity, while OT emphasizes real-time availability and process continuity. This divergence makes unified cybersecurity strategies difficult to implement. Moreover, updating or patching OT systems is often delayed or avoided entirely due to fears of operational disruption (Corallo, Lazoi, Lezzi, and Luperto, 2022).

A growing body of research emphasizes the need for holistic, real-time cybersecurity frameworks that incorporate anomaly detection, predictive analytics, and automated response mechanisms. These systems must be integrated into the production environment without disrupting operations, which demands lightweight, adaptive solutions. For example, AI-driven security models can help identify abnormal behavior patterns and zero-day threats before they escalate (Goyal et al., 2023).

However, deploying AI and machine learning in industrial cybersecurity is not without its challenges. High-quality labeled datasets for training are often scarce, and real-time processing demands optimized architectures that balance performance and computational load. Additionally, adversarial attacks that manipulate AI predictions pose new risks, especially when applied to critical infrastructure control systems (da Silva Oliveira and Santos, 2022).

Regulatory frameworks like ISA/IEC 62443 are increasingly promoted to standardize industrial cybersecurity practices, offering layered defense models and certification guidelines. Still, many organizations fall short in implementation due to cost, complexity, or lack of skilled personnel. As Kulinich (2023) notes, the management of cybersecurity in smart factories must also evolve, integrating strategic planning, workforce training, and continuous threat assessment into operational protocols.

In conclusion, cybersecurity in industrial environments is a multi-faceted challenge that requires more than just technological upgrades. It demands integrated strategies combining AI-powered threat detection, cross-domain policy alignment, resilient system architectures, and continuous workforce education. Without these, the very technologies designed to enhance productivity could become the entry points for catastrophic industrial failures.

### 2.2.4 Artificial Intelligence and Machine Learning in Cybersecurity

Artificial Intelligence (AI) and Machine Learning (ML) have become transformative tools in cybersecurity, particularly in the context of smart factories where cyber-physical systems and IIoT generate vast streams of real-time data. These technologies allow for predictive analysis, anomaly detection, and autonomous decision-making in environments where manual intervention is too slow or unreliable for effective threat mitigation (Hartmann, Brock, Kühn, and Dumitrescu, 2024).

In contrast to traditional rule-based systems, which require known threat signatures, ML models learn patterns from data, enabling them to detect unknown or evolving threats. Supervised learning algorithms such as Random Forest and Support Vector Machines are often used to classify network traffic or detect intrusions, while unsupervised approaches like clustering and autoencoders identify previously unseen anomalies in complex industrial datasets (Diaba, 2024). Deep learning models, particularly LSTM networks and convolutional autoencoders, further enhance performance by capturing temporal and spatial patterns in factory network traffic (Milić, 2024).

AI-enabled cybersecurity systems in smart factories also contribute to reducing false positives and automating threat responses. According to Siam, Hassan, and Bhuiyan (2025), combining ML with anomaly detection and natural language processing can help identify subtle signs of phishing, insider threats, and malware activity, tasks where traditional systems often fail or overload operators with alerts.

Despite these benefits, challenges remain in real-world deployments. As observed by Gavrovska and Samcovic (2020), issues such as insufficient labeled data, high computational cost, and integration with legacy factory systems continue to hinder wide-scale adoption. Furthermore, AI models themselves can be targets of adversarial attacks, where inputs are subtly manipulated to deceive classification systems.

Real-world implementation also involves socio-technical barriers. A study of over 50 smart factory deployments highlighted that beyond technological readiness, factors like lack of management support, data governance concerns, and workforce resistance can undermine AI initiatives (Hartmann et al., 2024). This emphasizes the need for a systems-level approach that includes technical innovation, human training, and organizational readiness.

In conclusion, AI and ML offer powerful solutions for securing smart factory environments, particularly through real-time monitoring, predictive threat detection, and autonomous response. However, to fully realize their potential, future research must focus on improving model transparency, adversarial robustness, and deployment strategies that are sensitive to operational and organizational constraints.

### 2.2.5 Hybrid AI Models

As cyber threats targeting smart factories grow in complexity and subtlety, there is increasing recognition that no single artificial intelligence (AI) or machine learning (ML) technique is sufficient for comprehensive cybersecurity. This has led to the development and growing application of hybrid AI models, which integrate multiple AI paradigms, such as supervised, unsupervised, and deep learning methods, to improve detection accuracy, responsiveness, and resilience to evolving threats (Choudhary and Choudhary, 2024).

Hybrid models exploit the strengths and compensate for the weaknesses of individual algorithms. For example, supervised classifiers like Random Forests are effective for detecting known attacks but struggle with unknown threats. Conversely, unsupervised models such as Autoencoders can detect novel anomalies but are prone to false positives. Deep learning models like LSTMs excel in modeling temporal patterns but require large, labeled datasets and significant computational power. By combining these approaches, hybrid models can address diverse attack vectors in smart factory networks (Paul, Stanley, Kessie, and Salawudeen, 2023).

Moreover, hybrid models are better equipped to defend against adversarial machine learning (AML), where attackers manipulate inputs to deceive AI systems. Research by Paul et al. (2023) shows that hybrid architectures, incorporating techniques such as adversarial training and layered anomaly detection, demonstrate greater robustness under simulated AML attacks compared to single-method models.

In addition to boosting technical performance, hybrid AI systems also support human-in-the-loop security, where automated systems work in tandem with human analysts. This collaborative approach enhances transparency and interpretability, two major concerns in cybersecurity, by allowing humans to validate AI decisions, especially in high-stakes industrial settings (Geissler and Lukowicz, 2024). This form of hybrid intelligence not only improves model accuracy and decision trustworthiness but also contributes to energy-efficient model training, an important consideration for embedded or edge-based industrial deployments.

Additionally, the application of hybrid AI has been proven in real-world domains such as banking and autonomous driving, suggesting strong cross-domain potential for cybersecurity in smart manufacturing. For instance, the use of neural networks alongside fuzzy logic and genetic algorithms has led to more accurate, adaptable, and resilient security solutions in high-volume, high-risk digital systems (Corral de La Mata, García de Blanes Sebastián, and Carvajal Camperos, 2024).

Nonetheless, hybrid AI systems face challenges, including increased computational demand, model complexity, and difficulty in tuning multiple integrated algorithms. These concerns are particularly relevant in smart factory contexts, where real-time constraints and hardware limitations often restrict implementation. Therefore, developing lightweight, modular hybrid systems is a key direction for ongoing research (Gavrovska & Samčović, 2020).

In conclusion, hybrid AI models offer a promising path forward for smart factory cybersecurity by merging the strengths of diverse algorithms into a unified, adaptive framework. While they introduce complexity, their enhanced performance, robustness, and interpretability make them highly suitable for protecting industrial systems against sophisticated, multi-dimensional cyber threats.

## 2.3: Theoretical Framework

### 2.3.1 Anomaly Detection Theory

Anomaly detection theory serves as a core analytical framework in cybersecurity, particularly for identifying abnormal behaviors or patterns that may indicate intrusions, faults, or cyberattacks in smart factory systems. In smart industrial environments where vast amounts of sensor, control, and communication data are continuously generated, anomaly detection plays a crucial role in proactively identifying system deviations before they escalate into operational failures or security breaches.

The theory is based on the assumption that anomalies, such as sudden spikes in network traffic, unusual sensor values, or inconsistent system behavior, can be statistically distinguished from normal patterns. These deviations often act as early indicators of cyberattacks such as data injection, denial-of-service (DoS), or insider threats (Chauhan and Kumar, 2023). In smart factories, real-time anomaly detection is essential because a minor system compromise can cascade into physical disruptions, downtime, or even safety hazards.

Recent advances have extended traditional statistical anomaly detection techniques by incorporating artificial intelligence and deep learning models, which improve detection performance on high-dimensional, non-linear, and time-dependent industrial data. For instance, hybrid deep learning models can learn temporal dependencies in sensor data using autoencoders and LSTM networks, which improves the precision of detecting complex anomalies in manufacturing processes (Lee, Kim, and Kim, 2023).

A key challenge in anomaly detection is balancing sensitivity and specificity. High sensitivity ensures that most anomalies are detected, but often at the cost of increased false positives, which can overwhelm security teams. This trade-off necessitates adaptive models that prioritize context-aware detections. For example, multivariate detection methods that account for correlated variables (e.g., pressure, flow, and temperature in water systems) can improve accuracy by reducing redundant alarms and focusing on significant deviations (Feeken et al., 2022).

The latest contributions also highlight domain-specific anomaly frameworks designed for smart factory conditions. For example, the integration of MEMS vibration sensors with edge computing platforms supports real-time, on-device anomaly detection in equipment, enhancing both responsiveness and sustainability of manufacturing lines (Kim, Lee, and Park, 2024). These systems not only identify anomalies but also provide predictive maintenance capabilities, reducing unplanned downtime and improving overall operational efficiency.

Furthermore, anomaly detection theory is evolving to support semi-supervised and unsupervised learning methods due to the scarcity of labeled attack data in real-world industrial environments. These methods enable systems to model normal operations and flag deviations without requiring extensive prior knowledge of attack types (Somma, Gallien, and Stojanovic, 2025). The application of physics-inspired consistency, such as in the Temporal Differential Consistency Autoencoder (TDC-AE), also shows promise for anomaly detection in dynamical systems by aligning with system stability constraints.

In conclusion, anomaly detection theory is fundamental to the cybersecurity architecture of smart factories. As threats evolve in scale and sophistication, so too must the theoretical and practical approaches used to detect them. Continued innovation in AI-integrated, context-aware, and computationally efficient anomaly detection systems is essential to safeguarding the future of automated manufacturing.

### 2.3.2 Defense-in-Depth Model

The Defense-in-Depth (DiD) model is a layered cybersecurity strategy designed to protect complex systems, such as those found in smart factories, through multiple, redundant security mechanisms. This approach acknowledges that no single security solution is foolproof and thus distributes risk mitigation across technical, administrative, and physical layers to contain and minimize potential breaches. In industrial environments, DiD is increasingly being applied to address the convergence of IT and OT systems, each of which presents unique vulnerabilities (Rahman, 2024).

In the smart manufacturing context, DiD frameworks extend beyond firewalls and antivirus tools to include process monitoring, endpoint protection, human factor awareness, and physical access controls. This holistic approach is essential because attacks on smart factories often exploit multiple weak points, including legacy OT devices, misconfigured networks, or human error. Rahman (2024) emphasize that vulnerabilities must be identified across digital, cyber-physical, and organizational layers and that defense mechanisms should be strategically aligned to those threat vectors.

A core strength of the DiD model is its compartmentalization of risk, meaning that even if one security layer fails, such as a compromised PLC or network intrusion, other layers can prevent escalation. For example, intrusion detection systems (IDS), multi-factor authentication, network segmentation, and real-time monitoring may all work in tandem to detect, delay, and respond to an attack (Halenar, Halenarova, and Tanuska, 2023). The layered security structure also supports redundancy, improving resilience and fault tolerance within critical systems.

A modern application of the DiD model includes compliance with international standards such as IEC 62443, which provides guidelines for segmenting smart factory networks, hardening devices, and enforcing security policies. This standardized approach aligns cybersecurity implementation with operational safety and regulatory requirements (Oliveira & Santos, 2022). It also fosters a culture of continuous assessment and iterative improvement through formal certification and system maturity models.

However, DiD is not without its challenges. The effectiveness of layered defenses depends on proper configuration, coordination, and continuous monitoring. Poorly integrated or overlapping tools can create blind spots or unnecessary complexity. Moreover, human elements, such as inadequate training or inconsistent policy enforcement, can nullify technical controls. As noted by (Acton and Datta, 2024), building cybersecurity awareness and institutionalizing best practices across personnel is just as important as technical defenses, especially at the endpoint level where breaches often begin.

In conclusion, the Defense-in-Depth model is a foundational theoretical and practical approach for securing smart factories. It offers a robust structure for mitigating diverse threats through multiple, overlapping layers of defense, each reinforcing the other. To ensure its effectiveness, however, DiD must be dynamic, well-integrated, and supported by organizational commitment and continuous assessment.

## 2.4 Empirical Review

### 2.4.1 AI and ML in Real-Time Threat Detection

Real-time threat detection is a core requirement in smart factories, where uninterrupted operations depend on the constant availability, integrity, and security of interconnected cyber-physical systems. Traditional security approaches, such as static rule-based intrusion detection systems (IDS), are increasingly insufficient in addressing the dynamic and evolving nature of cyber threats. Artificial Intelligence (AI) and Machine Learning (ML) offer powerful alternatives, capable of analyzing complex, high-volume data streams in real-time to detect anomalies and intrusions as they happen (Dontha, 2023).

AI-enabled models provide adaptive learning and continuous monitoring, which are essential for identifying both known and previously unseen threats. Machine learning classifiers such as Random Forest, Support Vector Machines, and Naïve Bayes have been used effectively in critical infrastructure for threat recognition, particularly when integrated into centralized control systems (Perrone, Flammini, and Setola, 2021). These models allow security operators to receive alerts only when threat confidence exceeds certain thresholds, helping reduce false alarms and improving decision-making efficiency.

Hybrid AI approaches that combine machine learning with deep learning are showing superior results in situational awareness and early detection of security breaches. In one study, a framework integrating convolutional neural networks (CNNs) with classical ML techniques achieved over 95% accuracy in identifying malicious activities in industrial control systems (ICS), demonstrating their scalability and robustness in high-volume production environments (Lu, Wu, and Chen, 2024).

Edge-based and lightweight AI systems are also becoming more common in smart factories. These solutions allow for local processing of data, which improves response time and reduces dependency on centralized systems, an advantage in environments where latency can disrupt production. As reported by Gavrovska & Samčović (2020), migrating ML-based inference closer to industrial endpoints (e.g., sensors, PLCs) has helped address threats such as DDoS attacks and unauthorized data access in IIoT-based smart factories.

In addition, use cases involving honeypots combined with ML have proven particularly effective against botnet and Distributed Denial of Service (DDoS) attacks. A model developed by Lee, Abdullah, Jhanjhi, and Kok (2021) achieved over 96% accuracy with low false positive rates by leveraging Random Forest classifiers on botnet behaviour datasets in simulated smart factory environments. This highlights the practical feasibility of hybrid AI deployments in real-time network protection.

However, challenges persist. These include the scarcity of high-quality labelled datasets, model explainability, and vulnerability to adversarial manipulation. Many ML models are black boxes, making it difficult to trace or explain decisions in high-stakes environments. Furthermore, deploying real-time AI systems in resource-constrained industrial settings demands optimization of model size, memory usage, and inference time, areas still under active research (Kavitha and Thejas, 2024)

In conclusion, AI and ML have proven effective in enhancing real-time threat detection in smart factories by improving accuracy, speed, and adaptability. Hybrid models, edge deployments, and integrated detection frameworks are at the forefront of current innovation. However, for widespread adoption, challenges related to interpretability, resource constraints, and adversarial resilience must be carefully addressed.

### 2.4.2 Performance Metrics from Recent Studies

Evaluating the effectiveness of AI and machine learning (ML) models in cybersecurity, especially within smart factory environments, relies on well-defined performance metrics. These metrics serve not only as benchmarks for algorithm accuracy but also reflect the practicality and reliability of threat detection systems under real-world conditions. The most widely used metrics include accuracy, precision, recall, F1-score, false positive rate (FPR), and computational efficiency, each offering unique insights into model behaviour and trade-offs.

Accuracy refers to the proportion of correctly classified instances among all samples. While often used as a headline figure, accuracy alone can be misleading in imbalanced datasets where benign activity vastly outnumbers malicious events. This has prompted researchers to rely more on precision (true positives over predicted positives) and recall (true positives over actual positives) to assess a model’s reliability in identifying rare but critical threats (Alapati and Dhanasekaran, 2024).

The F1-score, a harmonic mean of precision and recall, is particularly valuable in cybersecurity because it balances false alarms and missed detections, both of which carry operational risks. In industrial systems, a low F1-score may result in undetected intrusions or excessive alerts that overload security teams, undermining response effectiveness (Chen et al., 2024).

Recent studies highlight the growing importance of false positive rate (FPR) and false negative rate (FNR). A low FPR ensures that benign activity is not flagged erroneously, which is critical in smart factories where automation and uptime are essential. A model with a high FPR can interrupt processes or initiate unnecessary security responses. In contrast, a high FNR, where real threats are missed, poses even greater risk, as it allows attackers to operate undetected. Optimal models must therefore minimize both rates, a goal that has led to the adoption of ensemble learning and hybrid AI architectures (Das and Panda, 2025).

Computational efficiency is another critical metric, especially for edge-based deployments in smart factories. Real-time detection models must operate under tight latency and resource constraints. Studies such as those by Paul et al. (2023) emphasize that even high-performing models must be optimized for runtime and memory usage to be viable for industrial use. Trade-offs between detection accuracy and computational load are often resolved using model compression or lightweight AI frameworks.

Recent evaluations confirm that deep learning models, such as LSTMs and CNNs, often outperform traditional algorithms in terms of detection accuracy, achieving scores above 95% in some cases. However, they also require more computational power and are more opaque in terms of interpretability (Rishad, 2025). Interpretability is increasingly seen as an essential metric, particularly in high-stakes industrial environments where understanding a model’s reasoning can guide response actions and enhance trust.

In summary, performance evaluation in AI-driven cybersecurity must go beyond accuracy to include a comprehensive analysis of detection quality, false alarm control, computational feasibility, and interpretability. These metrics guide not only academic validation but also the deployment readiness of security solutions in real-time industrial environments.

### 2.4.3 Practical Deployments in Industrial Settings

The deployment of AI and machine learning (ML) technologies for cybersecurity in real-world industrial settings has evolved from conceptual experimentation to operational integration. These technologies are increasingly applied in smart factories to enhance threat detection, process monitoring, and system resilience. However, real-world implementation poses both opportunities and persistent challenges in scalability, interpretability, and organizational alignment.

One of the most significant deployments involves AI-driven intrusion detection systems (IDS) embedded into smart manufacturing networks. These systems leverage anomaly detection and threat modeling to prevent unauthorized access, data breaches, and cyber-physical disruptions. As Hassan, Collins, Babatunde, Alabi, and Mustapha (2024) observed, the application of deep learning and predictive analytics in real-world factories enables adaptive and near real-time response to sophisticated cyber threats, significantly improving operational continuity and system resilience.

Yet, actual deployments remain uneven across industries. Hartmann et al. (2024), in their review of over 50 industrial AI case studies, found that technical capability alone is insufficient for success. Challenges such as unclear use-case definition, limited AI skills within operational teams, resistance from factory-floor workers, and lack of strategic alignment frequently lead to failed or underperforming AI projects. This underscores the importance of a balanced approach that incorporates technology, people, and organizational structures into deployment strategies.

Emerging frameworks such as FactoryML are being developed to bridge the gap between R&D models and production environments. This system simplifies the packaging and integration of ML models with Programmable Logic Controllers (PLCs), making AI deployment in operational environments more feasible and reproducible (Wewer, Mahapatra, Esterle, and Larsen, 2024). FactoryML demonstrates that real-world deployment is not only a question of algorithm performance but also of communication protocol compatibility and runtime stability in heterogeneous factory systems.

Several use cases also show the effectiveness of edge AI for industrial environments. AI models deployed at the edge, near sensors or controllers, offer reduced latency and lower bandwidth requirements, enabling localized response to cyber threats. This deployment model is particularly useful in time-critical environments, such as power grids or high-speed assembly lines, where centralized processing would cause unacceptable delays (Dontha, 2023).

Nonetheless, real-world deployments must contend with limitations such as adversarial robustness, dataset generalization, and maintenance of AI models over time. Industrial AI systems often deteriorate in performance if not regularly updated, and few factories have processes in place for continuous model retraining. Security audits and feedback loops are essential to maintain effectiveness in the face of changing threats (Madupati, 2025).

In summary, while AI and ML are increasingly deployed for cybersecurity in smart factories, success hinges on the integration of technical solutions with robust infrastructure, workforce readiness, and agile deployment frameworks. Tools like FactoryML and edge-based AI illustrate promising directions, but sustainable deployment requires addressing human, organizational, and computational limitations alongside technical innovation.

## 2.5 Related Works

Numerous studies have explored the integration of AI and machine learning (ML) techniques into industrial cybersecurity, particularly within smart factory contexts. These works address real-time threat detection, anomaly detection, predictive maintenance, and adaptive defence systems, reflecting the shifting landscape of cybersecurity under Industry 4.0.

Gavrovska and Samčović (2020) proposed intelligent automation using ML and deep learning for cybersecurity in IIoT environments. Their study emphasized DDoS detection and edge-level inference for protecting surveillance and communication systems. The authors highlighted how migrating detection closer to edge devices enhances speed and reduces central processing loads.

Kiangala and Wang (2020) introduced an adaptive ML-based framework for configuring SCADA systems in resource-constrained manufacturing settings. Their work demonstrated how integrating ML into human-machine interfaces can transition legacy systems into self-configurable, smart platforms, offering a practical roadmap for scalable AI adoption.

Sarker (2020) provided a comprehensive theoretical framework for AI-driven cybersecurity, detailing how ML, expert systems, and natural language processing (NLP) can collectively support intelligent threat detection, access control, and anomaly identification. The work is foundational in framing cybersecurity as an intelligent, adaptive process rather than a static rule-based system.

Pajola, Conti, and Conti (2024) explored vulnerabilities in AI systems themselves, focusing on adversarial machine learning. They argued that without robust defences, AI systems deployed in industrial environments can be manipulated to misclassify inputs, turning intelligent defences into potential attack surfaces.

Bajaj (2025) analysed how predictive maintenance and AI-enhanced monitoring reduce downtime and improve product quality. His case-based analysis of smart factories showed significant gains in operational efficiency through ML-driven scheduling, automated quality checks, and decision-support tools.

Mishra, (2025) designed a hybrid AI threat intelligence system combining supervised and unsupervised models (Random Forest and Isolation Forest) for real-time cloud-based cybersecurity. The system achieved 95% accuracy and demonstrated the effectiveness of ensemble models in adapting to evolving threats.

da Rocha, Canciglieri, Szejka, dos Santos Coelho, and Canciglieri Junior (2020) proposed a failure prediction system in smart manufacturing using ensemble ML. Their implementation demonstrated that AI could proactively detect system degradation, enabling preemptive maintenance to avoid process disruption and scrap generation.

Hartmann et al. (2024) reviewed over 50 AI deployments in smart factories and found that technical success was often limited by poor organizational alignment and lack of operator trust. This underscores the need to pair technical innovation with human-centric design and workforce training.

Dontha (2023) investigated the dual application of AI in automation and cybersecurity. Her study revealed that AI simultaneously improves efficiency and strengthens anomaly detection capabilities across industrial systems, emphasizing the importance of integrated AI deployment across IT and OT domains.

Lăzăroiu et al. (2024) presented a futuristic model of cognitive digital twins and generative AI within cyber-physical manufacturing. Their work positioned AI not just as a security layer but as a predictive, adaptive decision engine across the industrial metaverse, enabling real-time data fusion, planning, and robotic control.

In summary, these studies collectively demonstrate the evolution of cybersecurity in smart factories from static detection toward intelligent, self-learning, and proactive protection systems. The emerging trend favours hybrid and edge-deployed AI, enhanced by organizational readiness, workforce integration, and robust adversarial defence mechanisms.

## 2.6 Gap Identification

While significant progress has been made in applying artificial intelligence (AI) and machine learning (ML) to cybersecurity in industrial contexts, several critical gaps remain, particularly in real-time, adaptive, and scalable threat detection for smart factories.

A major limitation identified across the literature is the lack of generalizability and contextual adaptation in most AI models. Existing solutions are often designed and validated on narrow datasets that do not capture the heterogeneity of industrial operations, leading to reduced performance when applied in live, complex environments (Hartmann et al., 2024). Many real-world implementations face failure due to data insufficiency, model rigidity, or poor integration with legacy systems.

Additionally, current AI cybersecurity models are often single-paradigm and lack the capability to detect both known and novel threats simultaneously. Supervised models require large volumes of labeled data and struggle with zero-day threats, while unsupervised models, though better at anomaly detection, frequently generate high false positive rates (Goyal et al., 2023). Hybrid models that integrate these strengths are emerging but remain underexplored and undervalidated in dynamic factory networks.

Another critical gap is the absence of adaptive, edge-deployable solutions that can respond to threats with low latency. Although some studies highlight the potential of edge computing and embedded AI for real-time responsiveness (Dontha, 2023), few have demonstrated successful, scalable deployments. This disconnect between centralized AI design and decentralized industrial needs weakens the protective architecture in smart factories.

Furthermore, there is a growing concern over the interpretability and trust of AI decisions in cybersecurity applications. Black-box models such as deep neural networks, while accurate, often lack explainability, an essential feature in industrial settings where human operators need to understand and validate AI-triggered responses (Rathore Davis et al., 2023).

Lastly, despite widespread academic interest, limited empirical evaluation exists for AI-based cybersecurity models in full-scale, real-world smart factories. The dominance of synthetic datasets and simulation-based validation leads to questions about deployment-readiness and robustness under industrial constraints like limited compute, noisy environments, or system downtimes (Mosleuzzaman et al., 2024).

In light of these challenges, this study addresses the following key gaps:

1. The need for a context-aware hybrid AI model that combines supervised, unsupervised, and deep learning techniques for real-time, multi-vector cyber threat detection.
2. The absence of empirical validation using realistic industrial datasets such as WADI.
3. The lack of deployment frameworks that support low-latency inference, model explainability, and scalability in industrial settings.

By responding to these gaps, this research aims to develop a novel AI/ML cybersecurity framework tailored to the operational and architectural realities of smart factories.

# CHAPTER THREE

# METHODOLOGY

## 3.1 Introduction

This chapter outlines the research methodology adopted to design, develop, and evaluate a hybrid AI/ML model for cyber threat detection in smart factory environments. The study employs a quantitative, simulation-based approach using the WADI dataset, which simulates real-world industrial control scenarios. The methodology is structured to support model training, testing, and performance benchmarking using supervised, unsupervised, and deep learning techniques. This approach ensures empirical validation of the model’s effectiveness in addressing cybersecurity threats in line with the study’s objectives.

## 3.2 Research Design

The research adopts a quantitative, experimental research design focused on simulation-based evaluation of a hybrid AI/ML cybersecurity model within a smart factory context. This design was selected for its capacity to support empirical testing, structured hypothesis validation, and repeatable performance comparisons using real-world industrial datasets. The goal is to build, train, and validate a composite detection system capable of identifying diverse cyber threats in real-time, leveraging supervised, unsupervised, and deep learning components.

In alignment with best practices from industrial AI cybersecurity research, the design follows a multi-phase structure:

1. data collection and preprocessing using the WADI dataset,
2. model development involving algorithm integration (e.g., Random Forest, Autoencoders, LSTM),
3. training and testing under controlled conditions, and
4. performance benchmarking using established metrics such as accuracy, F1-score, and false positive rate.

This design supports iterative prototyping and comparative analysis, therefore evaluating the contribution of each AI/ML component both independently and within the hybrid system. By training on historical cyber-attack data and testing on unseen portions, the study applies principles of cross-validation to improve generalizability and avoid overfitting (Sarker, 2020).

Furthermore, the experimental design incorporates realistic operational constraints, including time-series dependencies and latency considerations, which are crucial in industrial environments. These factors are embedded into the model evaluation phase using sequential deep learning (LSTM) to detect temporal anomalies in control system behavior (Lăzăroiu et al., 2024).

This design also addresses the need for deployment readiness, recognizing that the practical value of the research lies not only in algorithmic accuracy but also in system robustness and adaptability. To that end, the study emphasizes interpretability and computational efficiency as part of its evaluation criteria, aligning with the challenges highlighted by recent industrial AI deployments (Lin et al., 2022).

## 3.3 Data Collection and Dataset Description

This study relies on secondary data obtained from the iTrust Centre for Research in Cyber Security at the Singapore University of Technology and Design (SUTD). The specific dataset used is the Water Distribution (WaDi) dataset, am industrial dataset designed and collected for cybersecurity research in cyber-physical systems (CPS) from real scenarios.

The dataset was formally requested through the iTrust dataset portal - https://itrust.sutd.edu.sg/itrust-labs\_datasets/dataset\_info/, and access was granted via direct communication with the lab (as documented in the email response dated 10 May 2025 in the appendix A. The dataset links were provided by the iTrust team following the submission of the required information through their online access form.

### 3.3.1 Dataset Prediction

The WaDi dataset originates from a testbed simulating a water distribution plant, which operates using industrial components such as Programmable Logic Controllers (PLCs), Supervisory Control and Data Acquisition (SCADA) systems, sensors, actuators, and networking devices. It contains time-series data reflecting both normal operation and various stages of cyberattack scenarios, including integrity attacks, DoS, and replay attacks.

The dataset includes:

1. Multivariate sensor and actuator data.
2. Logs labelled with “attack” and “normal” states.
3. Time-aligned network traces and physical system responses.

Its design reflects real-world operational dynamics, making it an excellent proxy for evaluating AI and ML cybersecurity models in smart factory-like environments (Mosleuzzaman et al., 2024).

### 3.3.2 Justification for Dataset Selection

The WADI dataset was selected because:

1. It captures realistic industrial control behaviours and complex temporal attack patterns.
2. It includes a balance of labelled normal and attack periods, ideal for both supervised and unsupervised learning.
3. It reflects CPS vulnerabilities common to smart factory systems, such as sensor spoofing and control hijacking (Chauhan and Kumar, 2023).
4. It has been widely used in cybersecurity literature, ensuring the study aligns with recognized benchmarks.

The dataset supports the training, validation, and testing phases of this research. It is particularly suitable for evaluating the hybrid model proposed in this study, composed of Random Forest (RF), Autoencoder, and Long Short-Term Memory (LSTM) algorithms, against diverse cyber threat vectors under industrial operational constraints.

In conclusion, the use of the WADI dataset strengthens the reliability and real-world relevance of the study’s findings. By applying the hybrid AI/ML model to this dataset, the research aims to demonstrate not only detection accuracy but also resilience and practical adaptability in smart factory security systems.

## 3.4 Model Architecture

The architecture of the proposed cybersecurity model follows a hybrid design, integrating three core AI/ML components: Random Forest (RF) for supervised classification, Autoencoder for unsupervised anomaly detection, and Long Short-Term Memory (LSTM) for time-series behavioural modelling. This architecture is intentionally modular, allowing the system to leverage the unique strengths of each algorithm while compensating for their individual limitations.

### 3.4.1 Random Forest: Supervised Learning

Random Forest (RF) is a powerful supervised machine learning algorithm based on ensemble learning, where multiple decision trees are trained and aggregated to produce more accurate and robust predictions. It performs both classification and regression by constructing a forest of decision trees during training and outputting the majority vote (classification) or mean prediction (regression) of the individual trees (Zheng et al., 2021).

1. Working Architecture
   1. Data Bootstrapping: The algorithm generates multiple subsets of the original dataset using bootstrapping (random sampling with replacement).
   2. Tree Construction: For each subset, a decision tree is constructed. At each node, only a random subset of features is considered for splitting, which increases tree diversity.
2. Voting Mechanism:
   1. Classification: Each tree votes on the class; the class with the most votes is chosen.
   2. Regression: The average output from all trees is taken as the final prediction.

This randomness in both data and feature selection reduces overfitting, enhances generalization, and increases accuracy across diverse datasets (Xiao et al., 2021).

1. Pseudocode (Simplified)

1. Input: Dataset D with features F and labels Y

2. 1. For i = 1 to N (number of trees):

3. a. Create bootstrap sample Di from D

4. b. Train a decision tree Ti on Di:

5. i. At each node, randomly select a subset fi ⊂ F

6. ii. Choose the best split among fi using Gini or Entropy

7. iii. Grow tree to full depth or stopping condition

8. 2. For new input x:

9. a. Predict class using each tree Ti(x)

10. b. Return majority vote (classification) or mean (regression)

11.

Random Forest is selected in this study for several compelling reasons:

1. High Accuracy & Robustness: RF consistently demonstrates superior classification accuracy in intrusion detection tasks. A comparative study of supervised algorithms found RF achieving up to 99.7% accuracy in detecting network attacks like DoS and R2L (Kaddoura, Arid, and Moukhtar, 2022).
2. Feature Importance Evaluation: RF can rank the importance of different input variables, helping to identify the most critical features influencing cyber threats (Labu and Ahammed, 2024).
3. Noise Tolerance and Overfitting Resistance: Its ensemble nature and randomness reduce model variance and help prevent overfitting, even with noisy or imbalanced industrial datasets like WADI (Chai and Zhao, 2020).
4. Computational Efficiency: Despite being an ensemble model, Random Forest is relatively fast to train and can be easily parallelized for industrial-scale implementation (Zheng, Yi, and Deng, 2021).

In the context of this study, developing a hybrid AI/ML model for cyber threat detection in smart factories, the Random Forest algorithm serves as the supervised learning backbone of the system. It is responsible for:

1. Classifying labelled events from the WADI dataset as either benign or malicious.
2. Providing a high-precision first filter for known threat types, which complements the Autoencoder (anomaly detection) and LSTM (behavioural sequence modelling) layers.
3. Supporting explainable AI goals by enabling variable importance analysis and rule-based decision interpretation, key for operator trust in industrial settings (Owezarski, 2023).

In summary, Random Forest is integrated into this study’s architecture due to its proven reliability, scalability, and interpretability in industrial cybersecurity. It offers a high-performance supervised learning solution well-suited to the labelled component of the WADI dataset and the structured nature of smart factory cyber threats.

### 3.4.2 Autoencoder: Unsupervised Anomaly Detection

An Autoencoder is a type of unsupervised artificial neural network designed to learn efficient representations (codings) of input data by compressing it into a latent space and then reconstructing the original data. The core idea is that normal data will reconstruct well, while anomalous or unfamiliar patterns will result in a higher reconstruction error, making autoencoders particularly useful for anomaly detection in industrial control systems (Po and Kim, 2023).

1. Architecture

Autoencoders consist of three primary components:

* 1. Encoder: Compresses the input X into a lower-dimensional latent representation Z.
  2. Latent Space: Encoded compressed features that capture essential characteristics.
  3. Decoder: Attempts to reconstruct the input from the latent code Z.

This process allows the model to learn the underlying structure of “normal” data during training. During inference, high reconstruction error suggests an anomaly (Bae, Jang, Kim, and Joe, 2019).

1. Mathematical Formulation
   1. Given an input :
      1. Encoder:
      2. Decoder:
      3. Reconstruction Error:

Where,

We, Wd are weight matrices,

be, bd are biases,

Ϭ is an activation function (e.g, ReLu or Sigmoid),

||.|| is the mean squared error (MSE)

Anomalies are detected when exceeds a predefined threshold

1. Step-by-Step Pseudocode

1. Input: Training data X\_normal

2. 1. Initialize encoder and decoder parameters

3. 2. for epoch in range(num\_epochs):

4. a. Encode input: Z = encoder(X\_normal)

5. b. Decode: X\_hat = decoder(Z)

6. c. Compute loss: L = MSE(X\_normal, X\_hat)

7. d. Backpropagate and update weights

8.

9. # Inference

10. Input: New data X\_new

11. 1. Compute reconstruction: X\_hat\_new = autoencoder(X\_new)

12. 2. Compute error: E = MSE(X\_new, X\_hat\_new)

13. 3. if E > threshold:

14. classify as anomaly

15. else:

16. classify as normal

17.

1. Justification for the use of Autoencoder
   1. Unsupervised Learning Advantage: Autoencoders eliminate the need for labelled attack data, making them ideal for detecting zero-day attacks and unknown threats in industrial settings (Russo, Zanasi, Marasco, and Colajanni, 2024).
   2. Suitability for Industrial Time-Series: They handle high-dimensional, temporal sensor data effectively, especially when integrated with time-dependent structures such as LSTM or CNN (Waters et al., 2022).
   3. High Sensitivity to Deviations: Autoencoders demonstrate high anomaly detection accuracy in smart factory datasets like WADI, achieving F1-scores exceeding 90% under optimal tuning (Ahmad, Kovalenko, and Makarov, 2024).
   4. Low False Positive Rate: Compared to traditional clustering (e.g., DBSCAN), autoencoders yield significantly lower false positives, a crucial benefit in industrial environments where unnecessary alerts can disrupt operations (Bae et al., 2019).

In this study, the Autoencoder module is the unsupervised component of the proposed hybrid model. It learns the “normal” behaviour of system variables from the WADI dataset and flags anomalous states based on reconstruction errors. Its integration offers:

1. Defence against novel threats missed by supervised classifiers.
2. A scalable and low-maintenance detection layer, suitable for edge deployments.
3. Enhanced model diversity and ensemble performance when combined with Random Forest and LSTM.

In conclusion, Autoencoders provide a powerful, interpretable, and adaptive approach to anomaly detection in smart factories. Their ability to detect subtle, previously unseen threats complements the hybrid cybersecurity model’s mission to safeguard critical infrastructure with both precision and breadth.

### 3.4.3 Long Short-Term Memory (LSTM): Behavioural Sequence Modelling

Long Short-Term Memory (LSTM) is a specialized form of recurrent neural network (RNN) that is designed to model long-term dependencies in sequential data. Unlike traditional RNNs, which suffer from vanishing gradient issues, LSTM cells are structured with gating mechanisms that allow the network to remember or forget temporal patterns over extended time steps. This makes LSTM particularly effective in analysing time-series data such as sensor logs in industrial control systems, where detecting gradual or delayed attack patterns is critical (Chiranjeevi, Ramya, Balaji, Shashank, and Reddy, 2024).

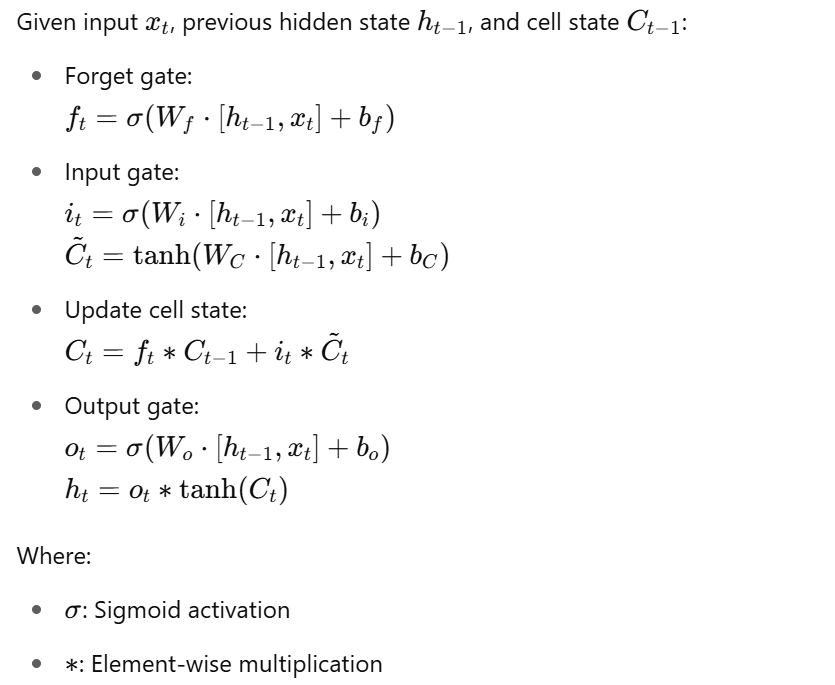
1. Architecture:

An LSTM unit consists of:

* 1. Input gate: Controls how much of the new input to incorporate into memory.
  2. Forget gate: Decides which information from the previous state to discard.
  3. Output gate: Determines the output and the influence of current memory.

Each LSTM cell maintains a cell state Ct and a hidden state ht, which are updated over time based on the input xt.

1. Mathematical formular for LSTM



These formulas allow LSTM to model both short-term and long-term dependencies within time-series data (Lee et al., 2023).

1. Pseudocode of LSTM

1. Input: Time-series input data X = {x\_1, x\_2, ..., x\_T}

3. 1. Initialize LSTM parameters: W\_f, W\_i, W\_C, W\_o, b\_f, b\_i, b\_C, b\_o

4. 2. Initialize h\_0 = 0, C\_0 = 0

5.

6. For t = 1 to T:

7. f\_t = sigmoid(W\_f \* [h\_{t-1}, x\_t] + b\_f)

8. i\_t = sigmoid(W\_i \* [h\_{t-1}, x\_t] + b\_i)

9. C̃\_t = tanh(W\_C \* [h\_{t-1}, x\_t] + b\_C)

10. C\_t = f\_t \* C\_{t-1} + i\_t \* C̃\_t

11. o\_t = sigmoid(W\_o \* [h\_{t-1}, x\_t] + b\_o)

12. h\_t = o\_t \* tanh(C\_t)

13.

14. Output: Anomaly score = function(h\_T, reconstruction error or prediction deviation)

1. Justification for the adoption of LSTM
   1. Temporal Modelling: LSTM can learn time-dependent patterns in factory process data, which is critical for identifying delayed or evolving cyber-attacks that would be missed by static classifiers (Hwang, Jang, Jang, Kim, and Ha, 2022).
   2. Multivariate Sensor Inputs: Industrial systems like the WADI dataset include dozens of interdependent signals. LSTM handles multivariate sequences better than traditional techniques (Fachrezi, Ihsan, and Astuti, 2024).
   3. High Detection Accuracy: Studies have shown LSTM-based models outperform other deep learning methods in real-time anomaly detection, achieving high F1-scores and low false positives (Amer, Al-Rimy, El-Sappagh, Almalki, and Alghamdi, 2024).
   4. Edge Deployment Potential: With optimization, LSTM models can run on lightweight embedded systems, making them viable for real-time monitoring in industrial settings (Fachrezi et al., 2024).

Within this hybrid framework, the LSTM component is responsible for temporal behaviour modelling and long-range anomaly recognition. It complements:

1. Random Forest, which handles static classification of known threats.
2. Autoencoder, which captures point-based anomalies.

LSTM enhances the model’s ability to track event progression over time, ensuring detection of stealthy, time-lagged threats that may evolve across several operational cycles in the WADI water distribution system.

In summary, LSTM is essential for dynamic cybersecurity modelling in smart factories. It offers time-aware anomaly detection, adapts to complex data streams, and strengthens the hybrid model’s capacity for robust, real-time defence.

## 3.5 Model Training and Testing

The model training and testing process for this study involves three major stages: data preprocessing, model training, and model evaluation, each tailored to handle the complexity and scale of industrial time-series data, using the WADI dataset. The goal is to evaluate the effectiveness of the hybrid model, comprising Random Forest, Autoencoder, and LSTM, in identifying cyber threats in smart factory systems.

### 3.5.1 Data Preprocessing

The WADI dataset contains high-dimensional, multivariate time-series data that include both normal and attack states. Preprocessing steps include:

1. Missing value imputation
2. Min-Max normalization
3. Label encoding for classification tasks
4. Sliding window segmentation for LSTM input preparation

This process ensures the dataset is structured for both supervised and unsupervised learning models.

### 3.5.2 Training Strategy

Each model in the hybrid architecture is trained independently, then integrated via ensemble decision logic:

1. Random Forest is trained on the labelled subset of WADI, using standard 70:30 train-test splits and 10-fold cross-validation to ensure robustness.
2. Autoencoder is trained only on the normal behaviour data to learn reconstruction patterns and detect anomalies based on deviation thresholds.
3. LSTM is trained on sequentially ordered sensor data to capture temporal dependencies and predict the next states of sensor readings.

Hyperparameters for all models are tuned using grid search and validation curves, optimizing for accuracy and F1-score, while also tracking computational cost for potential deployment at the edge.

### 3.5.3 Model Evaluation

Model performance is assessed using multiple metrics:

1. Accuracy, Precision, Recall, F1-Score
2. False Positive Rate (FPR) and False Negative Rate (FNR)
3. Latency and throughput for real-time readiness

These metrics are chosen to reflect not only classification power but also operational suitability for smart factory deployment environments. In training, Autoencoder and LSTM are especially evaluated for their ability to generalize across unseen attack vectors, a critical need in industrial settings where novel threats frequently emerge.

The hybrid model is benchmarked against baseline classifiers (e.g., standalone SVM and KNN models) to demonstrate the benefit of combining detection paradigms. The results are used to validate the hypothesis that a hybrid architecture significantly improves both detection accuracy and reliability.

### 3.6 Model Evaluation

The models will be trained on 70% of the dataset and evaluated using the remaining 30% through 5-fold cross-validation. This ensures that the evaluation is robust and minimizes the chances of overfitting or underfitting.

To assess and compare the performance of both the individual models (Random Forest, Autoencoder, LSTM) and the hybrid model, the following evaluation metrics will be applied:

1. Accuracy: The ratio of correctly predicted instances (both attacks and normal traffic) to the total number of instances.

Equation 3. 1 Accuracy formular ​

1. Precision: Measures the proportion of correctly predicted positive (attack) observations to all predicted positive observations. A high precision means fewer false alarms.

……. Equation 3. 2 Precision formular

1. Recall (Sensitivity or True Positive Rate): The proportion of actual attack instances that were correctly identified by the model.

…. Equation 3. 3 Recall Formular

1. F1-Score: The harmonic mean of Precision and Recall. It provides a balance between the two, especially in datasets where class imbalance (i.e., many more normal events than attacks) is present.

… Equation 3. 4 F1-Score Formular

1. False Positive Rate (FPR): The ratio of normal (benign) activities that were incorrectly classified as attacks.

… Equation 3. 5 False Positive Rate Formular

1. Detection Time (Latency): Measures the time taken by the model to detect and respond to a cyber threat. This is especially critical in real-time environments like smart factories, where even milliseconds can impact operations.

These metrics will be used to create a comparative analysis between models, identifying which architecture or combination offers the highest accuracy, lowest false positive rate, and fastest detection time.

## 3.7 Ethical Considerations

No personal or sensitive data is used in this research. However, the study will uphold ethical standards of AI research, as well as Ensuring model transparency and reproducibility, preventing data misuse and adhering to responsible AI principles as suggested by (Olafuyi, 2023) and (Nacheva & Azeroual, 2024)

# References

Acton, T., and Datta, P. M. (2024). Endpoint cybersecurity: When smart devices turn stupid. *Journal of Information Technology Teaching Cases*. https://doi.org/10.1177/20438869241242142

Asere, G. F., Nuga, K. A., and Medugu, M. (2025). The Role of Artificial Intelligence in Cybersecurity: Understanding the Dynamics, Impacts, and Remediations. *Journal of Computer, Software, and Program*, *2*(1), 1–9. https://doi.org/10.69739/JCSP.V2I1.120

Chauhan, N., and Kumar, K. S. (2023). Cyber Attacks Detection and Prevention using Cryptography Algorithms for Industrial Automation. *2023 2nd International Conference for Innovation in Technology, INOCON 2023*. https://doi.org/10.1109/INOCON57975.2023.10101087

Chen, H., Shen, Z., Wang, Y., Ke, H., and Xu, J. (2024). Threat Detection Driven by Artificial Intelligence: Enhancing Cybersecurity with Machine Learning Algorithms. *World Journal of Innovation and Modern Technology*, *7*(6), 58–70. https://doi.org/10.53469/WJIMT.2024.07(06).09

Choudhary, L., and Choudhary, J. S. (2024). Deep Learning Meets Machine Learning: A Synergistic Approach towards Artificial Intelligence. *Journal of Scientific Research and Reports*, *30*(11), 865–875. https://doi.org/10.9734/JSRR/2024/V30I112614

Contreras, J., Salkuti, S., Fernando, L., Monteiro, R., Rodrigues, Y. R., and Souza, A. C. Z. D. (2023). Cybersecurity in Cyber–Physical Power Systems. *Energies*, *16*(12). https://doi.org/10.3390/EN16124556

Corallo, A., Lazoi, M., Lezzi, M., and Luperto, A. (2022). Cybersecurity awareness in the context of the Industrial Internet of Things: A systematic literature review. *Comput. Ind.*, *137*. https://doi.org/10.1016/J.COMPIND.2022.103614

Corral de La Mata, D., García de Blanes Sebastián, M., and Carvajal Camperos, M. (2024). Hybrid artificial intelligence: Application in the banking sector. *Revista  de  Ciencias  Sociales*, *30*(3), 22–36. https://doi.org/10.31876/RCS.V30I3.42674

da Silva Oliveira, A., and Santos, H. (2022). Continuous Industrial Sector Cybersecurity Assessment Paradigm\* : Proposed Model of Cybersecurity Certification. *2022 18th International Conference on the Design of Reliable Communication Networks (DRCN)*, 1–6. https://doi.org/10.1109/DRCN53993.2022.9758022

de Azambuja, A. J. G., Plesker, C., Schützer, K., Anderl, R., Schleich, B., and Almeida, V. R. (2023). Artificial Intelligence-Based Cyber Security in the Context of Industry 4.0—A Survey. *Electronics (Switzerland)*, *12*(8). https://doi.org/10.3390/ELECTRONICS12081920

Diaba, S. Y. (2024). The Role of Machine Learning Algorithms in Smart Grid Cybersecurity. *2024 International Workshop on Artificial Intelligence and Machine Learning for Energy Transformation (AIE)*, 1–7. https://doi.org/10.1109/AIE61866.2024.10561386

Elbes, M., Hendawi, S., Alzu’Bi, S., Kanan, T., and Mughaid, A. (2023). Unleashing the Full Potential of Artificial Intelligence and Machine Learning in Cybersecurity Vulnerability Management. *2023 International Conference on Information Technology (ICIT)*, 276–283. https://doi.org/10.1109/ICIT58056.2023.10225910

Exploring the Foundation of Smart Factories in Industry 4.0: A Conceptual Review. (2024). Exploring the Foundation of Smart Factories in Industry 4.0: A Conceptual Review. *CONFERENCE PROCEEDING*, 874–879. https://doi.org/10.62919/MRTY9098

Feeken, L., Kern, E., Szanto, A., Winnicki, A., Kao, C. Y., Wudka, B., … Burghardt, C. (2022). Detecting and Processing Anomalies in a Factory of the Future. *Applied Sciences*, *12*(16). https://doi.org/10.3390/APP12168181

Gajiwala, C. (2024). Artificial Intelligence in Cybersecurity : Advancing Threat Modeling and Vulnerability Assessment. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, *10*(5), 778–788. https://doi.org/10.32628/CSEIT241051066

Gavrovska, A., and Andreja Samcovic. (2020). Intelligent Automation Using Machine and Deep Learning in Cybersecurity of Industrial IoT. *Cyber Security of Industrial Control Systems in the Future Internet Environment*, 156–174. https://doi.org/10.4018/978-1-7998-2910-2.CH008

Geissler, D., and Lukowicz, P. (2024). Leveraging Hybrid Intelligence Towards Sustainable and Energy-Efficient Machine Learning. *ArXiv*, *abs/2407.10580*. https://doi.org/10.48550/ARXIV.2407.10580

Goyal, S. B., Rajawat, A. S., Solanki, R. K., Zaaba, M. A. M., and Long, Z. A. (2023). Integrating AI With Cyber Security for Smart Industry 4.0 Application. *2023 International Conference on Inventive Computation Technologies (ICICT)*, 1223–1232. https://doi.org/10.1109/ICICT57646.2023.10134374

Halenar, I., Halenarova, L., and Tanuska, P. (2023). Communication Safety of Cybernetic Systems in a Smart Factory Environment. *Machines*, *11*(3). https://doi.org/10.3390/MACHINES11030379

Hartmann, S., Brock, J., Kühn, A., and Dumitrescu, R. (2024). Applying Artificial Intelligence in the Smart Factory: Lessons Learned from real-world use cases. *Procedia CIRP*, *58*(27), 100–105. https://doi.org/10.1016/J.PROCIR.2024.10.062

Jain, R. (2021). Challenges and Applications of Cyber Physical Systems. *Advances in Systems Analysis, Software Engineering, and High Performance Computing*, 1–17. https://doi.org/10.4018/978-1-7998-6721-0.CH001

Jamai, I., Ben Azzouz, L., and Saidane, L. A. (2020). Security issues in Industry 4.0. *2020 International Wireless Communications and Mobile Computing (IWCMC)*, 481–488. https://doi.org/10.1109/IWCMC48107.2020.9148447

Jiang, J. R. (2018). An improved cyber-physical systems architecture for Industry 4.0 smart factories. *Advances in Mechanical Engineering*, *10*(6). https://doi.org/10.1177/1687814018784192

Jimmy, F. (2023). The Role of Artificial Intelligence in Predicting Cyber Threats. *International Journal of Scientific Research and Management (IJSRM)*, *11*(08), 935–953. https://doi.org/10.18535/IJSRM/V11I08.EC04

Juarez, F. A. B. (2019). Cybersecurity in an Industrial Internet of Things Environment (IIoT) Challenges for Standards Systems and Evaluation Models. *2019 8th International Conference On Software Process Improvement (CIMPS)*, 1–6. https://doi.org/10.1109/CIMPS49236.2019.9082437

Kim, H. J., Lee, J. H., and Park, J. G. (2024). SITRAN: Self-Supervised IDS With Transferable Techniques for 5G Industrial Environments. *IEEE Internet of Things Journal*. https://doi.org/10.1109/JIOT.2024.3437448

Kulinich, T. (2023). SPECIAL FEATURES OF INDUSTRIAL MANAGEMENT IN THE CONTEXT OF APPLYING THE “SMART FACTORY” SYSTEM. *Наукові Праці Міжрегіональної Академії Управління Персоналом. Економічні Науки*, (4 (71)), 53–61. https://doi.org/10.32689/2523-4536/71-7

Lee, K. S., Kim, S. B., and Kim, H. W. (2023). Enhanced Anomaly Detection in Manufacturing Processes Through Hybrid Deep Learning Techniques. *IEEE Access*, *11*, 93368–93380. https://doi.org/10.1109/ACCESS.2023.3308698

Masum, R. (2023). Cyber Security in Smart Manufacturing (Threats, Landscapes Challenges). *ArXiv*, *abs/2304.10180*. https://doi.org/10.48550/ARXIV.2304.10180

Milić, S. (2024). Artificial intelligence in the smart environments. *Veštacka Inteligencija - Prakticna Primena i Izazovi - Zbornik Radova*, 9–18. https://doi.org/10.5937/VI24009M

Mosleuzzaman, M., Arif, I., and Siddiki, A. (2024). DESIGN AND DEVELOPMENT OF A SMART FACTORY USING INDUSTRY 4.0 TECHNOLOGIES. *ACADEMIC JOURNAL ON BUSINESS ADMINISTRATION, INNOVATION &amp; SUSTAINABILITY*, *4*(04), 89–108. https://doi.org/10.69593/AJBAIS.V4I04.131

Muheidat, F., Mallouh, M. A., Al-Saleh, O., Al-Khasawneh, O., and Tawalbeh, L. A. (2024). Applying AI and Machine Learning to Enhance Automated Cybersecurity and Network Threat Identification. *Procedia Computer Science*, *251*, 287–294. https://doi.org/10.1016/J.PROCS.2024.11.112

Musaeva, K., Vyachina, I., and Aliyeva, M. (2024). Smart factories and their impact on modern manufacturing enterprises: Prospects and challenges in the era of the digital economy. *E3S Web of Conferences*, *537*. https://doi.org/10.1051/E3SCONF/202453707010

Paul, E. M., Stanley, U. M., Kessie, J. D., and Salawudeen, M. D. (2023). Adversarial machine learning in cybersecurity: Mitigating evolving threats in AI-powered defense systems. *World Journal of Advanced Engineering Technology and Sciences*, *10*(2), 309–325. https://doi.org/10.30574/WJAETS.2023.10.2.0294

Rahman, M. H. (2024). *Cyber-Physical Security Vulnerabilities Identification and Classification in Smart Manufacturing -- A Defense-in-Depth Driven Framework and Taxonomy* (Mohammed Shafae University of Massachusetts Dartmouth, The University of Arizona). Mohammed Shafae University of Massachusetts Dartmouth, The University of Arizona. Retrieved from https://doi.org/

Renteria-Marquez, I. A., Basaldua, J. C. G. M., Aguirre, O., Lara-Medrano, B. E., and Tseng, T. L. (2024). Cyber-Physical Production System Framework for Production Scheduling in Smart Factories. *2024 Winter Simulation Conference (WSC)*, 1740–1748. https://doi.org/10.1109/WSC63780.2024.10838901

Siam, A. Al, Hassan, M. M., and Bhuiyan, T. (2025). Artificial Intelligence for Cybersecurity: A State of the Art. *2025 IEEE 4th International Conference on AI in Cybersecurity (ICAIC)*, 1–7. https://doi.org/10.1109/ICAIC63015.2025.10848980

Simonthomas, S., Subramanian, R., and Mathiew, S. A. (2024). A Survey of Enhancing Cyber Physical System Security in Smart grid. *2024 International Conference on Communication, Computing and Internet of Things (IC3IoT)*, 1–6. https://doi.org/10.1109/IC3IOT60841.2024.10550307

Somma, M., Gallien, T., and Stojanovic, B. (2025). Anomaly Detection in Complex Dynamical Systems: A Systematic Framework Using Embedding Theory and Physics-Inspired Consistency. *ArXiv.Org*. https://doi.org/10.48550/ARXIV.2502.19307

Vilkelyte, Z., Wojciechowski, J., Bojarczak, P., Fallah, S. El, Kharbach, J., Jamil, M. O., and Vanagas, J. (2024). A Review on Improvement in Detection of Cyberattacks Using Artificial Intelligence for the Grid Applications. *2024 IEEE Open Conference of Electrical, Electronic and Information Sciences (EStream)*, 1–5. https://doi.org/10.1109/ESTREAM61684.2024.10542612

Wang, H. W., Kuo, S. Y., and Chen, L. B. (2021). Exploring the relationship between internal information security, response cost, and security intention in container shipping. *Applied Sciences (Switzerland)*, *11*(6). https://doi.org/10.3390/APP11062609

Yao, X., Zhou, J., Lin, Y., Li, Y., Yu, H., and Liu, Y. (2019). Smart manufacturing based on cyber-physical systems and beyond. *Journal of Intelligent Manufacturing*, *30*(8), 2805–2817. https://doi.org/10.1007/S10845-017-1384-5

Yi, K. J., and Jeong, Y. S. (2022). Smart factory: security issues, challenges, and solutions. *Journal of Ambient Intelligence and Humanized Computing*, *13*(10), 4625–4638. https://doi.org/10.1007/S12652-021-03457-6

Zhang, G., Yang, Y., and Yang, G. (2023). Smart supply chain management in Industry 4.0: the review, research agenda and strategies in North America. *Annals of Operations Research*, *322*(2), 1075–1117. https://doi.org/10.1007/S10479-022-04689-1