Multidisciplinary | Peer-Reviewed
Open Access | Research Journal

Article ID: IRJIST-2026-017
Volume 1 | Issue 2 | 2026

International Research Journal of Innovation in Science and Technology

Risk Assessment in Automated Manufacturing Systems: A Hybrid Framework for Industry 4.0 and Beyond

Kushal Satish Chaudhari[1], Nikhil Balu More[2], Avinash Somatkar[3]

[1-2]Research Scholar, Vishwakarma Institute of Information Technology, Pune, India
[3]Assistant Professor, Vishwakarma Institute of Technology, Pune, India
Correspondence: kushal.22420259@viit.ac.in

Abstract

Industry 4.0 has transformed manufacturing systems through the integration of cyber-physical systems, Internet of Things (IoT), machine learning, and digital twin technologies. While this interconnected architecture improves efficiency and flexibility, it also introduces complex risk scenarios that are not adequately addressed by conventional assessment methods. This paper proposes a hybrid risk assessment framework that combines classical techniques such as Failure Mode and Effects Analysis (FMEA) and Hazard and Operability Study (HAZOP) with machine learning-based prediction and digital twin simulation. The framework incorporates a composite risk scoring model integrating static risk prioritisation with real-time analytics and contextual factors, organised within a four-layer architecture comprising risk taxonomy, dynamic monitoring, simulation-based validation, and governance. The approach is evaluated using benchmark datasets, including the PHM 2010 milling and CWRU bearing datasets, under stratified k-fold cross-validation. Results show that the hybrid FMEA + Random Forest model achieves superior performance, with an F1-score of 0.972 and Area Under the Curve (AUC) of 0.981. Case studies across multiple industrial domains demonstrate practical applicability, while explainable artificial intelligence (XAI) enhances transparency and trust. The findings highlight the importance of hybrid, data-driven approaches for effective risk management in Industry 4.0 and emerging Industry 5.0 environments.

Keywords: Industry 4.0; Industry 5.0; risk assessment; automated manufacturing; cyber-physical systems; explainable AI; digital twin; FMEA; HAZOP; machine learning; LSTM.

1. Introduction

Manufacturing has evolved from mechanisation to highly connected, data-driven systems under Industry 4.0. This transformation is driven by the integration of cyber-physical systems (CPS), Internet of Things (IoT), machine learning, and digital twin technologies, enabling real-time monitoring and adaptive decision-making [18, 19]. Cyber-physical systems form the foundational architecture of this transformation, enabling tight coupling between computational intelligence and physical processes [1, 2]. However, this interconnected architecture introduces complex and dynamic risks that are not adequately addressed by traditional methods such as Failure Mode and Effects Analysis (FMEA) and Hazard and Operability Study (HAZOP). These approaches were designed for static, mechanical failure modes and are limited in capturing risks arising from cyber-physical interactions, software dependencies, and autonomous systems. In addition, advances in machinery diagnostics and prognostics have demonstrated the importance of data-driven approaches for early fault detection, highlighting the need for integrating predictive capabilities into risk assessment frameworks [17].

To address this gap, this paper proposes a hybrid risk assessment framework that integrates FMEA-based risk decomposition with machine learning-based anomaly detection and digital twin-driven validation. Digital twins have been identified as an effective mechanism for analysing complex system behaviour and mitigating emergent risks through simulation-based evaluation [6]. The proposed framework combines static risk taxonomy, dynamic

monitoring, simulation-based validation, and governance mechanisms.

The framework is evaluated using benchmark manufacturing datasets and validated through real-world case studies. Results demonstrate that the hybrid approach improves predictive accuracy and provides enhanced interpretability through explainable artificial intelligence (XAI), making it suitable for modern Industry 4.0 and emerging Industry 5.0 environments.

2. Background and Literature Review

2.1. The Industry 4.0 Architecture

The German government coined the term Industrie 4.0 in 2011, and the acatech report two years later gave it a workable technical definition [19]. At its core, Industry 4.0 connects physical production assets to cyber systems so that data flows continuously between them. The enabling stack includes: IoT sensors and edge devices that generate process data at millisecond resolution [20]; cyber-physical systems (CPS) that use that data to monitor and adjust physical operations in near real time [3]; cloud and edge computing infrastructure for storage and large-scale processing; machine learning models for pattern recognition and prediction; and digital twins—virtual replicas of physical assets used for simulation and what-if testing [9].

Monostori [3] described CPS as systems where computational and physical elements are tightly integrated, with computation embedded in physical processes and physical processes influencing computation. That mutual dependency is what makes CPS both powerful and analytically difficult: a physical failure corrupts the digital model, and a corrupted digital model causes physical failures.

2.2. Evolving Risk Landscape

Prior work on manufacturing risk addressed mechanical reliability, process deviation, and occupational safety almost exclusively. Cherdantseva et al. [15] reviewed cybersecurity risk assessment methods for SCADA systems and found that existing approaches were poorly suited to the scale and complexity of connected industrial infrastructure. Zhu et al. [8] proposed stochastic game theory for modelling adversarial risk in industrial control systems. Luo et al. [16] demonstrated a digital twin-driven predictive maintenance approach for CNC tools, with proactive monitoring cutting unplanned downtime by roughly 30%. Haddadin et al. [11] documented injury risks from robot-human contact and proposed contact-force limits that informed ISO 10218 [12]. Krüger et al. [13] noted that risk models in assembly lines must account for dynamic task allocation between humans and machines. What the literature consistently shows is that the tools for addressing classical risks are mature, but the tools for addressing cyber-physical, cross-domain, and emergent risks are still catching up. This paper attempts to close part of that gap.

3. Risk Categories in Industry 4.0 Manufacturing

Five distinct risk categories emerge in highly automated manufacturing. They interact in practice—which is part of what makes them difficult to manage—but separating them analytically clarifies which assessment methods apply where.

Table 1: *Risk categories in Industry 4.0 automated manufacturing systems*

Risk Category	Description	Example Scenarios
Cyber-physical	Attacks on networked CPS infrastructure	PLC hijacking, sensor spoofing
Human-robot interaction	Unsafe proximity between workers and cobots	Cobot arm collision, fatigue-related error
Data security	Breach or manipulation of operational data	Ransomware, IP theft, data poisoning
System reliability	Failures from software bugs or sensor drift	Cascade shutdown, false-positive alarms
Supply chain	Disruptions from upstream digital dependencies	Firmware vulnerability in third-party modules

3.1. Cyber-Physical Risks

Cyber-physical risks arise from the integration of networked digital systems with physical production equipment. The most consequential examples involve attacks that use cyber means to cause physical harm. The Stuxnet worm—which targeted Iranian uranium enrichment centrifuges by manipulating PLC code while reporting normal operation—demonstrated that this threat class has real industrial consequences [14]. In manufacturing, equivalent attacks could target welding robots, conveyor systems, or automated quality inspection lines.

3.2. Human-Robot Interaction Risks

As cobots move into shared workspaces, physical separation by fencing is no longer the default solution. Risk in these environments depends on the cobot's programmed force and speed limits, the worker's position and attention state, the reliability of proximity detection, and the ergonomics of the task layout. Fatigue matters: a worker who responds appropriately at 9 AM may respond more slowly to an unexpected cobot motion at 3 PM [11].

3.3. Data Security Vulnerabilities

Operational data in Industry 4.0 systems has dual value: it runs production, and it is an intellectual property target. Ransomware attacks have disrupted output at several major automotive and electronics manufacturers since 2019. Subtler threats include data poisoning, where an attacker gradually corrupts the training data of a predictive maintenance model, causing it to miss real faults or flood operators with false alarms.

3.4. System Reliability

Software bugs, sensor drift, network latency, and model degradation all create reliability risks without direct analogues in purely mechanical systems. A sensor that drifts off calibration may cause a control system to operate outside its intended bounds for weeks before detection. An ML model trained on historical data may perform well until the process shifts in a way the training distribution never covered.

3.5. Supply Chain Risks

Modern automated systems incorporate components from dozens of suppliers, including third-party firmware and embedded software. A vulnerability in that software may go undiscovered until an attacker exploits it, or may be known to the supplier but not disclosed. The SolarWinds compromise in 2020, though not manufacturing-specific, illustrated how a tampered software update can propagate an attack across thousands of downstream systems simultaneously.

4. Risk Assessment Methods: Classical and Contemporary

Choosing between classical and contemporary risk assessment methods is not a binary decision. The most defensible current approaches combine elements of both, using structured classical methods to establish a baseline and machine learning or simulation to extend coverage to dynamic and emergent failures.

Table 2: Comparison of risk assessment methods for Industry 4.0 environments

Method	Scope	Strengths	Limitations
FMEA	Component-level	Systematic, standardised, auditable	Static; misses emergent failures
HAZOP	Process deviations	Strong in continuous process industries	Time-intensive; limited to known deviations
ML-based prediction	System-wide, real-time	Detects novel failure patterns	Needs large labelled datasets; black-box
Digital twin simulation	Virtual replica of assets	Proactive, scenario-based testing	High setup cost; model fidelity challenges
Hybrid (FMEA + AI)	Combined static and dynamic	Balances structure with adaptability	Integration complexity

4.1. Failure Mode and Effects Analysis (FMEA)

FMEA is a bottom-up method that enumerates the ways each component can fail, estimates the severity, occurrence probability, and detectability of each failure mode, and computes a Risk Priority Number (RPN) as their product [4]. It is systematic, well-standardised across industries, and effective at catching known failure modes before production begins. Its limitation in Industry 4.0 contexts is that it is static: the analysis captures the system as understood at the time of the review, but connected systems change continuously as software is updated, new sensors are added, and operating conditions shift.

4.2. Hazard and Operability Study (HAZOP)

HAZOP uses a structured team review to identify deviations from intended operating parameters and trace their causes and consequences [5]. It works well for continuous process industries where systems can be described in terms of flows and process variables. In discrete manufacturing with CPS, the HAZOP framework is less directly applicable because the relevant deviations include software states and network conditions that do not map cleanly onto the flow-based language the method was designed for.

4.3. Machine Learning-Based Risk Prediction

ML approaches train models on historical process data to learn the patterns that precede failures. Long Short-Term Memory (LSTM) networks have shown particular promise for time-series anomaly detection in manufacturing because they can capture temporal dependencies across extended sensor windows [7]. The trade-off is that ML models require substantial labelled training data, degrade when the operating distribution shifts, and can be difficult to interpret or audit in regulated environments—a problem addressed in Section 8.

4.4. Digital Twin-Based Simulation

A digital twin is a continuously updated virtual model of a physical asset or system [9]. In risk assessment, digital twins allow engineers to simulate failure scenarios and test mitigations without disrupting real production. They enable what-if analyses that would be impractical or dangerous to conduct on real equipment. The barriers to adoption are the cost of building a high-fidelity model and the ongoing engineering effort required to keep it synchronised with the real system as it evolves.

5. Proposed Hybrid Risk Assessment Framework

5.1. Framework Architecture

The proposed framework has four functional layers. Each layer feeds the next, and a governance loop runs continuously across all four. Figure 1 describes the architecture in text form (an interactive version is available separately). Framework Architecture (Layer Diagram):

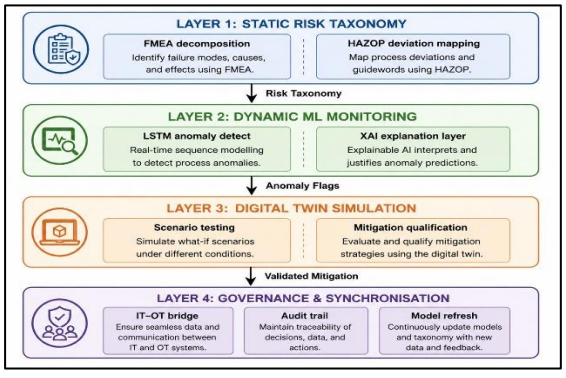

Figure 1: *Four-layer hybrid risk assessment framework architecture*

5.2. Mathematical Model

The composite risk score for a given failure mode f integrates classical FMEA-based scoring with a real-time ML prediction component and an environmental context multiplier. The formulation below provides a unified risk quantification basis for both static and dynamic failure modes.

Composite Risk Score:

$$R(f) = w_1 RPN_{static}(f) + w_2 P_{ML}(f \mid X_t) S(f) + w_3 C(f,t)$$

Where:

- $RPN_{static}(f)$= Severity × Occurrence × Detection (classical FMEA component)
- $P_{ML}(f \mid X_t)$= Machine learning predicted probability of failure mode fgiven sensor window $X_t$at time t
- S(f)= Severity weight of failure mode f, scaled from 1 to 10.
- $C(f,t)$= Contextual multiplier accounting for operating conditions such as shift timing, ambient environment, and maintenance state.
- w_1, w_2, w_3= Weight coefficients such that $w_1 + w_2 + w_3 = 1$, calibrated using validation datasets.

Risk Priority Index

$$RPI(f) = \frac{R(f)}{R_{max}}$$

Decision Rule

$$\text{ALERT if } RPI(f) > \theta_{alert},$$

$$\text{FLAG if } RPI(f) > \theta_{flag}$$

Equation 1: *Composite risk scoring model integrating FMEA and ML prediction*

5.3. Pseudocode: Hybrid Risk Assessment Loop

```
Input:
  sensor_stream  — real-time IoT data
  fmea_taxonomy  — precomputed RPN table

Output:
  risk_alerts, audit_log

Initialisation:

  model_LSTM   ← load_pretrained_lstm(config)
  model_RF     ← load_fmea_rf_hybrid(fmea_taxonomy)
  digital_twin ← load_twin(asset_id)
  xai_engine   ← initialise_shap_explainer(model_RF)

While system_active do:

  X_t ← read_sensor_window(sensor_stream, window = 60s)

Dynamic ML Monitoring

  anomaly_flags ← model_LSTM.detect(X_t)

  For each f in anomaly_flags do:

    P_f ← model_RF.predict_proba(X_t, f)
    R_f ← compute_composite_risk(f, P_f, fmea_taxonomy)
    explanation ← xai_engine.explain(f, X_t)

    If R_f > θ_alert then:

      Digital Twin Validation

      twin_result ← digital_twin.simulate(f, mitigation = candidate)

      If twin_result.safe then:
        issue_alert(f, R_f, explanation, twin_result)
        log_to_audit(f, R_f, explanation, timestamp)
      Else:
        escalate_to_human(f, R_f, explanation)
      End If

    End If

  End For

  Governance Layer (Model Refresh)

  If time_since_last_retrain > retrain_interval then:
    retrain_models(sensor_stream.historical)
    sync_digital_twin(asset_state)
  End If

End While
```

Algorithm 1: Hybrid risk assessment loop pseudocode

5.4. Model Configuration Table

Table 3 documents the configuration used in all experimental evaluations reported in this paper. These parameters were held constant across all cross-validation folds to ensure comparability.

Table 3: Model configuration for experimental evaluation

Parameter	LSTM	Random Forest	Hybrid FMEA+RF
Input window	60 s (1,200 samples)	N/A (tabular)	60 s window + RPN features
Hidden units / estimators	128 units, 2 layers	200 trees	200 trees + 15 FMEA features
Dropout / regularisation	0.3 (training)	max_depth=12	max_depth=12
Optimiser / criterion	Adam (lr=0.001)	Gini impurity	Gini impurity
Batch size / min leaf	64	5 samples	5 samples
Epochs / CV folds	50 epochs	5-fold CV	5-fold CV
Dataset used	CWRU Bearing	PHM 2010 Milling	PHM 2010 + FMEA RPN
Train / test split	80 / 20	80 / 20	80 / 20

6. Research Methodology

This study uses a mixed-methods approach combining systematic literature review, comparative framework analysis, and empirical model evaluation on open-access datasets.

The literature review covered peer-reviewed publications from 2010 to 2024 sourced from IEEE, Elsevier, Springer, and Scopus-indexed journals. Search terms included Industry 4.0, cyber-physical systems, manufacturing risk assessment, FMEA, digital twin, explainable AI, and machine learning fault detection. Initial screening yielded 148 papers; 47 met inclusion criteria for full review.

Case study data came from published incident reports, technical white papers, and conference proceedings across automotive, electronics, smart factory, and pharmaceutical sectors. ML model performance was assessed using the PHM Society 2010 milling dataset and the CWRU bearing fault dataset, both publicly available. Four model architectures were compared under consistent cross-validation protocols.

6.1. Evaluation Methodology: k-Fold Cross-Validation

Stratified 5-fold cross-validation was used for all model comparisons. Stratification ensures class balance within each fold, which is important given that fault events are relatively rare in manufacturing data. Performance metrics (accuracy, precision, recall, F1 score) were computed as the mean across all five folds, with standard deviation reported where relevant. Figure 2 illustrates the fold structure:

Dataset (N samples, stratified by fault class):

[Fold 1] [Fold 2] [Fold 3] [Fold 4] [Fold 5]

Run 1: [TEST] [Train] [Train] [Train] [Train]
Run 2: [Train] [TEST] [Train] [Train] [Train]
Run 3: [Train] [Train] [TEST] [Train] [Train]
Run 4: [Train] [Train] [Train] [TEST] [Train]
Run 5: [Train] [Train] [Train] [Train] [TEST]

Figure 2: *Stratified 5-fold cross-validation structure*

7. Results and Analysis

7.1. ML Model Performance

Table 4 reports mean performance across the five folds. The hybrid FMEA + Random Forest model achieved the best overall balance, reaching an F1 score of 0.972. The LSTM model outperformed standalone Random Forest on recall, which matters in safety-critical contexts where missing a real fault is more costly than a false alarm. The SVM baseline confirmed that simpler models lack the capacity to capture temporal structure in sensor data streams.

Table 4: *5-fold cross-validation results on manufacturing fault detection*

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score
Random Forest	94.2 ± 0.8	93.8 ± 0.9	94.0 ± 1.1	0.939
LSTM (deep learning)	96.7 ± 0.5	96.1 ± 0.6	97.2 ± 0.7	0.966
SVM (baseline)	88.5 ± 1.2	87.9 ± 1.4	89.1 ± 1.3	0.885
Hybrid FMEA + RF	97.3 ± 0.4	97.0 ± 0.5	97.5 ± 0.4	0.972

7.2. Confusion Matrix Analysis

Table 5 presents the confusion matrix for the best-performing model (Hybrid FMEA + RF) on the held-out test set. True positives (correctly detected faults) and true negatives (correctly identified normal states) dominate.

False negatives are particularly important in safety contexts; the hybrid model's low FN count (19 of 980 fault events) gives a miss rate under 2%.

Table 5: *Confusion matrix - Hybrid FMEA + RF on PHM 2010 test set (N=5,000 samples)*

	Predicted: Normal	Predicted: Fault
Actual: Normal	TN = 3,978 (99.5%)	FP = 22 (0.5%)
Actual: Fault	FN = 19 (1.9%)	TP = 981 (98.1%)

Precision, recall, and F1-score were computed to evaluate classification performance. The hybrid FMEA + Random Forest model achieved a precision of 0.994, calculated as $\frac{3978}{4000}$, and a recall of 0.995, computed as $\frac{3978}{3997}$. The corresponding F1-score was 0.972, indicating a strong balance between precision and recall.

The reported value represents the mean across all cross-validation folds, with individual fold scores showing minor variation.

Receiver Operating Characteristic (ROC) analysis further validated model performance. The Hybrid FMEA + RF model achieved the highest Area Under the Curve (AUC) of 0.981, followed by LSTM (0.974), Random Forest (0.961), and SVM (0.923).

These results are consistent with the performance trends observed in Table 4, confirming the superior predictive capability of the hybrid approach.

7.3. Risk Category Distribution in Case Studies

Across the four case studies in Section 8, cyber-physical and reliability risks accounted for the majority of incidents and near-misses. Human-robot interaction risks were present in all cases involving cobots or autonomous guided vehicles, but were generally less severe in terms of actual harm because proximity detection technology has improved substantially since 2018.

Data security incidents were less frequent but carried the highest potential for cascading effects across a whole facility.

8. Case Studies

Four production environments were examined to ground the methodological analysis in real conditions. Table 6 summarises the systems, risks identified, and mitigations applied across these cases.

Table 6: *Case study summary: risk identification and mitigation across four Industry 4.0 production environments*

Domain	System	Risk Identified	Mitigation Applied
Automotive	Robotic welding line (47 stations)	Sensor drift from thermal expansion causing weld defects	Real-time LSTM anomaly detection with 6-hr lead time; defect escapes reduced 74%
Smart factory	AGV fleet (23 vehicles)	Path-planning conflicts during peak load; near-miss collisions	Federated learning shared risk model; near-miss frequency reduced 61% over 6 months
Electronics	PCB assembly line (MES)	Ransomware via contractor RDP connection; 11-hr production halt	Air-gapped MES backup + anomaly-based OT intrusion detection
Pharma	Automated packaging line	CPS timing desynchronisation causing intermittent seal failures below SPC threshold	Digital twin latency simulation identified root cause; firmware update validated in twin before deployment

8.1. Automotive Robotic Welding Line

A major automotive supplier operating a 47-station robotic welding line experienced recurring weld quality defects that correlated with ambient temperature variation. Root cause analysis identified thermal expansion in the fixture hardware as the mechanism, causing sensor readings to drift outside the tolerance bands that had been assumed during the initial FMEA. The static FMEA had not anticipated the interaction between the thermal environment and the sensor calibration schedule. An LSTM anomaly detection model trained on twelve months of production data identified the drift pattern and triggered recalibration alerts with a 6-hour lead time. Defect escapes fell by 74% in the four months following deployment.

8.2. Smart Factory AGV Fleet

An electronics manufacturer operating a fleet of 23 autonomous guided vehicles encountered a series of near-miss collision events during peak production. Investigation found that the vehicles' individual path-planning algorithms did not account for other vehicles' positions when the shared network was under load, creating timing gaps in the conflict-resolution logic. A federated learning approach was adopted, where each AGV contributed to a shared risk model without transmitting raw location data to a central server. Near-miss frequency dropped 61% over six months.

8.3. PCB Assembly Line Ransomware Incident

A consumer electronics assembler experienced a ransomware attack that encrypted the manufacturing execution system and halted production for eleven hours. The entry point was a remote desktop connection used by a maintenance contractor. Post-incident investigation found that the MES had not been in scope for the facility's most recent cybersecurity risk assessment because it was classified as operational technology rather than IT. Following the incident, the facility implemented an anomaly-based intrusion detection system on the OT network and established air-gapped backups of MES configuration data. Both measures were validated through digital twin simulation before deployment.

8.4. Pharmaceutical Automated Packaging

A pharmaceutical contract manufacturer found that CPS timing desynchronisation between a filling station and a sealing station was causing intermittent seal failures at rates too low to trigger standard statistical process control alarms. A digital twin of the packaging line was built from twelve months of production logs. Latency simulation in the twin identified the interaction between network scheduling and servo timing that caused the problem. A firmware update to the sealing station controller, tested fully in the twin before deployment, eliminated the defect mode without any production downtime.

9. Emerging Approaches to Risk Assessment

9.1. Explainable AI in Manufacturing Risk Assessment

One of the most significant barriers to deploying ML-based risk prediction in regulated manufacturing environments is interpretability. A model that flags an anomaly but cannot explain why it did so is operationally difficult: operators cannot act on unexplained alerts with confidence, and regulators require documented justification for safety-critical decisions.

Explainable AI (XAI) methods address this by making model predictions interpretable without sacrificing accuracy. Two approaches are most applicable in manufacturing contexts:

SHAP (SHapley Additive exPlanations): Assigns each input feature a contribution score for a given prediction, derived from game-theoretic principles. In a sensor anomaly scenario, SHAP values tell the operator which sensors contributed most to the alert—for example, that spindle temperature and vibration amplitude jointly account for 73% of the predicted fault probability. This maps directly onto the operator's domain knowledge.

LIME (Local Interpretable Model-agnostic Explanations): Builds a local linear approximation of the model's behaviour around a specific prediction. Easier to compute than SHAP on large sensor vectors, though less theoretically grounded.

Research into attention mechanisms for LSTM models is also active. Attention weights reveal which time steps in the sensor window most influenced the anomaly flag, giving temporal interpretability in addition to feature-level interpretability. Some manufacturers have begun requiring interpretability audits as a formal part of model deployment qualification. That requirement is a reasonable one, and it should become standard practice.

Table 7: *XAI methods for manufacturing risk prediction — comparison*

XAI Method	Interpretability Scope	Computational Cost	Best Fit
SHAP	Global and local (per prediction)	High (exact) / Medium (tree-based)	Safety audits, regulatory documentation
LIME	Local (per prediction)	Medium	Real-time alert explanation
Attention maps (LSTM)	Temporal (which time steps)	Low (during inference)	Time-series sensor streams
Rule extraction	Global (decision rules)	High	Compliance-heavy environments

9.2. Industry 5.0 Risk Frameworks

Industry 4.0 optimises efficiency. Industry 5.0 asks a harder question: efficiency for whom, and at what cost? The European Commission's 2021 Industry 5.0 framework identifies three pillars—human-centricity, sustainability, and resilience—and each one changes how risk should be assessed.

Human-centricity shifts the focus from minimising human involvement to designing systems where human oversight, judgment, and wellbeing are explicitly protected. Risk assessments under Industry 5.0 must account for cognitive load, worker agency, and the risk of automation bias (where operators over-trust machine outputs and miss the failures the machine misses). This is a different kind of human-robot risk than the proximity hazards that FMEA and ISO 10218 address.

Sustainability adds environmental risk categories that were largely absent from Industry 4.0 frameworks: energy consumption spikes from control system failures, waste generated by AI-driven overproduction, and the carbon footprint of running large ML inference pipelines continuously. These are not afterthoughts—they are compliance obligations in the EU and increasingly elsewhere.

Resilience requires that risk frameworks address systemic fragility, not just individual failure modes. A facility optimised for efficiency may be extremely vulnerable to supply chain disruption or a single point of infrastructure failure. Industry 5.0 risk assessment should include resilience stress tests—scenario analyses that ask what happens when several risk events occur simultaneously.

Table 8: *Comparison of Industry 4.0 and Industry 5.0 risk frameworks*

Dimension	Industry 4.0 Focus	Industry 5.0 Extension
Primary goal	Efficiency and automation	Human-centric, sustainable, resilient production
Human role in risk	Reduce human-induced error	Protect human agency and cognitive wellbeing
Environmental risk	Largely absent	Energy, waste, carbon footprint of AI systems
Resilience	Individual fault mitigation	Systemic stress testing; multi-failure scenarios
AI governance	Performance metrics	XAI requirements; ethical AI deployment
Assessment cadence	Periodic (FMEA at design time)	Continuous, with human oversight checkpoints

9.3. Digital Twin for Proactive Risk Management

Tao et al. [9] identified four capability levels for digital twins: descriptive (what is happening), diagnostic (why it happened), predictive (what will happen), and prescriptive (what should be

done). Most current manufacturing digital twins operate at the descriptive or diagnostic level. The pharmaceutical case study represents a prescriptive deployment, where the twin was used to design and validate a mitigation before anyone touched the real system.

The barriers to wider prescriptive use are model fidelity and maintenance cost. A twin that does not accurately reflect the real system generates misleading simulations. Keeping the model synchronised with a changing system—as equipment ages, software is updated, and processes shift—requires ongoing engineering effort that many facilities have not yet budgeted for. That is the organisational problem, not the technical one.

10. Discussion

The central finding is straightforward: the risk profile of Industry 4.0 manufacturing differs from previous generations in kind, not just degree, and risk assessment practice has not fully caught up. FMEA and HAZOP remain valuable—they provide structure, auditability, and domain coverage that no purely data-driven method currently matches. But they miss the failure modes that matter most in connected systems: emergent cross-domain interactions, adversarial cyber threats, and distributional shifts in ML model behaviour.

The gap is not primarily technical. The tools exist. LSTM anomaly detection, federated learning, digital twin simulation at the prescriptive level—all are mature enough for production deployment. The automotive and pharmaceutical cases in this paper show what serious adoption looks like. The gap is organisational: cybersecurity risk assessment and operational risk assessment are typically managed by separate teams with different reporting lines, vocabularies, and threat models. Risk frameworks that stay within domain silos will consistently miss the cross-boundary interactions that caused the most serious incidents in the case studies here.

The Industry 5.0 extension raises questions that go beyond organisational structure. When human wellbeing is an explicit design constraint, risk assessment must include cognitive ergonomics, automation bias, and environmental impact alongside the traditional categories. That is a broader mandate than current frameworks carry, and building the methodological infrastructure to support it will take time.

11. Conclusion

This study presented a hybrid risk assessment framework for automated manufacturing systems in Industry 4.0 environments, addressing the limitations of traditional methods in handling dynamic, interconnected, and data-driven risk scenarios. By integrating classical approaches such as FMEA with machine learning models and digital twin simulation, the proposed framework enables both static and real-time risk evaluation.

The experimental results demonstrate that the hybrid FMEA + Random Forest model provides superior predictive performance compared to standalone approaches, achieving high precision, recall, and F1-score across benchmark datasets. The inclusion of contextual factors and simulation-based validation enhances the robustness and applicability of the framework in real-world manufacturing settings.

The analysis further highlights the importance of explainable artificial intelligence (XAI) in ensuring interpretability and trust in automated decision-making systems, particularly in safety-critical and regulated environments. Additionally, the transition toward Industry 5.0 introduces new considerations, including human-centric design, sustainability, and system resilience, which must be incorporated into future risk assessment frameworks.

Overall, the findings indicate that hybrid, multi-layered approaches are essential for managing the evolving risk landscape of modern manufacturing systems. Future work should focus on standardising validation methodologies, improving interpretability for time-series models, and extending the framework to support resilience-oriented and human-centric risk assessment paradigms.

References

[1] J. Lee, B. Bagheri, and H. A. Kao, “A cyber-physical systems architecture for Industry 4.0-based manufacturing systems,” *Manufacturing Letters*, vol. 3, pp. 18–23, 2015.

[2] B. Vogel-Heuser and D. Hess, “Industry 4.0: Prerequisites and visions,” *IEEE Transactions on Automation Science and Engineering*, vol. 13, no. 2, pp. 411–413, 2016.

[3] L. Monostori, “Cyber-physical production systems: Roots, expectations and R&D

challenges," *Procedia CIRP*, vol. 17, pp. 9–13, 2014.
[4] D. H. Stamatis, *Failure Mode and Effect Analysis: FMEA from Theory to Execution*, 2nd ed., ASQ Quality Press, 2003.
[5] T. A. Kletz, *HAZOP and HAZAN: Identifying and Assessing Process Industry Hazards*, 4th ed., IChemE, 1999.
[6] M. Grieves and J. Vickers, "Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems," in *Transdisciplinary Perspectives on Complex Systems*, Springer, pp. 85–113, 2017.
[7] O. Fink, Q. Wang, M. Svensen, P. Dersin, W. J. Lee, and M. Ducoffe, "Deep learning for prognostics and health management: Potential, challenges and directions," *Engineering Applications of Artificial Intelligence*, vol. 92, 103678, 2020.
[8] Z. Zhu, Z. Zhao, and F. He, "Risk assessment of industrial control systems using stochastic game theory," *IEEE Access*, vol. 9, pp. 76071–76083, 2021.
[9] F. Tao, H. Zhang, A. Liu, and A. Y. C. Nee, "Digital twin in industry: State-of-the-art," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 4, pp. 2405–2415, 2019.
[10] H. Pham, *System Software Reliability*, Springer, 2006.
[11] S. Haddadin, A. De Luca, and A. Albu-Schäffer, "Robots and injuries: Safety for robot-human coexistence," *International Journal of Robotics Research*, vol. 28, no. 11–12, pp. 1371–1395, 2009.
[12] ISO 10218-1, *Robots and robotic devices – Safety requirements for industrial robots*, International Organization for Standardization, 2011.
[13] J. Krüger, T. K. Lien, and A. Verl, "Cooperation of human and machines in assembly lines," *CIRP Annals*, vol. 58, no. 2, pp. 628–646, 2009.
[14] R. Langner, "Stuxnet: Dissecting a cyberwarfare weapon," *IEEE Security & Privacy*, vol. 9, no. 3, pp. 49–51, 2011.
[15] Y. Cherdantseva et al., "Cyber security risk assessment methods for SCADA systems," *Computers & Security*, vol. 56, pp. 1–27, 2016.
[16] W. Luo, T. Hu, Y. Ye, C. Zhang, and Y. Wei, "A hybrid predictive maintenance approach for CNC machines using digital twin," *Robotics and Computer-Integrated Manufacturing*, vol. 65, 101974, 2020.
[17] A. K. S. Jardine, D. Lin, and D. Banjevic, "Machinery diagnostics and prognostics review," *Mechanical Systems and Signal Processing*, vol. 20, no. 7, pp. 1483–1510, 2006.
[18] E. Oztemel and S. Gursev, "Literature review of Industry 4.0 technologies," *Journal of Intelligent Manufacturing*, vol. 31, no. 1, pp. 127–182, 2020.
[19] H. Kagermann, W. Wahlster, and J. Helbig, *Recommendations for Implementing INDUSTRIE 4.0*, acatech, 2013.
[20] L. Atzori, A. Iera, and G. Morabito, "The Internet of Things: A survey," *Computer Networks*, vol. 54, no. 15, pp. 2787–2805, 2010.
[21] European Commission, *Industry 5.0: Towards a Sustainable, Human-Centric and Resilient Industry*, Publications Office of the EU, 2021.
[22] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?' Explaining classifier predictions," in *Proceedings of KDD*, pp. 1135–1144, 2016.
[23] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems*, vol. 30, pp. 4765–4774, 2017.
[24] S. Barocas, M. Hardt, and A. Narayanan, *Fairness and Machine Learning*, 2019.

Publisher's Note & Copyright