Proceedings of International Design Engineering Technical Conferences & Computers and Information in Engineering Conference
IDETC-CIE 2026
August 23–26, 2026, Houston, TX, USA

IDETC/CIE2026-194390

LIGHTWEIGHT MULTIMODAL FUSION FOR REAL-TIME CYBERSECURITY OF OPTICAL AND THERMAL IN-SITU MONITORING SIGNALS IN METAL ADDITIVE MANUFACTURING

Nishant Chinnasami
Department of Computer Science and Engineering
University of South Carolina
Columbia, South Carolina 29208
Email: NISHANTC@email.sc.edu

Rye Stahle-Smith
Department of Computer Science and Engineering
University of South Carolina
Columbia, South Carolina 29208
Email: RYES@email.sc.edu

Zaki Bushiri
Department of Computer Science and Engineering
University of South Carolina
Columbia, South Carolina 29208
Email: ZBUSHIRI@email.sc.edu

Mumin Adhami
Department of Mechanical Engineering
University of South Carolina
Columbia, South Carolina 29208
Email: MADHAMI@email.sc.edu

Austin R.J. Downey
Department of Mechanical Engineering
University of South Carolina
Email: austindowney@sc.edu

Lang Yuan
Department of Mechanical Engineering
University of South Carolina
Columbia, South Carolina 29208
Email: LANGYUAN@cec.sc.edu

Rasha Karakchi*
Department of Computer Science and Engineering
University of South Carolina
Columbia, South Carolina 29208
KARAKCHI@cec.sc.edu

*Corresponding author: Rasha Karakchi, KARAKCHI@cec.sc.edu

ABSTRACT

Industrial manufacturing systems are increasingly vulnerable to coordinated physical–cyber attacks that manipulate both computational execution and process-level sensing signals. In metal additive manufacturing environments, in-situ optical and thermal monitoring systems generate high-bandwidth sensing streams that are critical for process validation, quality assurance, and operational reliability. Ensuring the integrity and trustworthiness of these monitoring signals is therefore essential for secure manufacturing operations. This paper proposes a lightweight multimodal fused hybrid framework for real-time secure monitoring of manufacturing sensing data. The proposed model integrates statistical thresholding with machine learning to jointly analyze execution timing, optical emissions, and thermal behavior. Unlike software-only detection schemes, the framework is designed for edge deployment and executed entirely on a PYNQ-based FPGA–ARM platform. All training and evaluation experiments were performed directly on the edge device to validate practical feasibility under real deployment constraints. Experimental results across multiple industrial welding datasets demonstrate that multimodal fusion significantly reduces false negatives while maintaining low false alarm rates. By correlating computational execution characteristics with physical sensing behavior, the proposed framework enables context-aware anomaly detection that distinguishes malicious perturbations from normal manufacturing process variability. The hardware implementation achieves deterministic low-latency inference with minimal resource and power overhead. These results demonstrate that lightweight hybrid multimodal monitoring provides scalable and energy-efficient protection against physical–cyber attacks in manufacturing systems.

INTRODUCTION

Modern manufacturing systems increasingly depend on edge-class computing platforms that acquire, process, and secure high-bandwidth optical and thermal sensor streams for in-situ monitoring and quality assurance [1–4]. In processes such as additive manufacturing and robotic welding, optical and infrared cameras are integrated within controlled-atmosphere environments to monitor melt pool dynamics, surface morphology, and thermal evolution during fabrication. Performing encryption and anomaly detection directly at the edge reduces latency, preserves data privacy, and enables rapid mitigation. However, embedding security functions alongside process monitoring introduces new challenges: physical process variability and adversarial manipulation may both manifest as perturbations in computational and sensing signals [5, 6].

Although cryptographic algorithms such as AES remain mathematically secure, their implementations on resource-constrained embedded platforms may exhibit measurable variations under timing interference, fault injection, or environmental disturbance [7–11]. In manufacturing environments, this issue is amplified because benign process phenomena—such as spatter, illumination shifts, or transient thermal drift—naturally produce correlated effects across optical, thermal, and execution timing channels [12, 13]. As a result, purely timing-based detectors struggle to differentiate between malicious implementation-level perturbations and legitimate physical process dynamics.

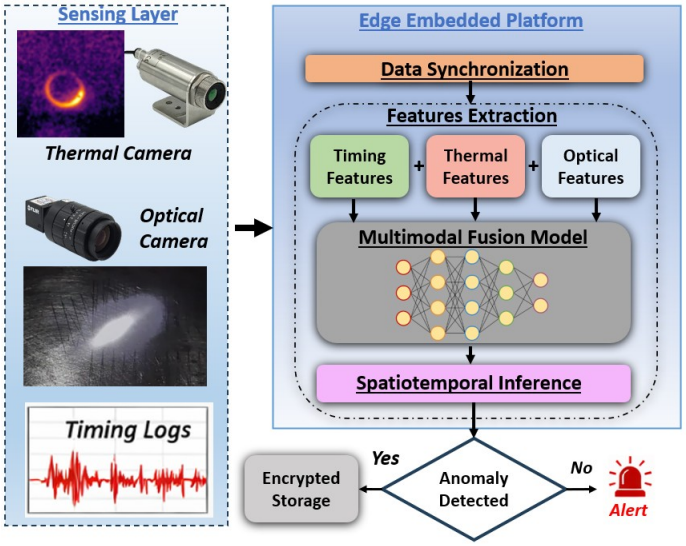

FIGURE 1. Proposed multimodal monitoring architecture. Optical and thermal sensors capture process dynamics while the embedded platform performs AES encryption. Extracted timing, optical, and thermal features are fused within a hybrid statistical–machine learning detector to produce real-time anomaly decisions.

Consider, for example, an embedded controller inside a metal additive manufacturing system performing AES encryption for secure firmware updates or authenticated communication. During fabrication, minor thermal fluctuations or spatter events may influence power stability or clock behavior, producing slight timing variations. While the cryptographic primitive remains uncompromised, these benign disturbances can resemble timing-based or fault-injection attack signatures when analyzed without physical context. This ambiguity underscores the need for detection mechanisms that incorporate process-aware information.

Existing countermeasures such as constant-time software implementations, shielding, or hardware redundancy can mitigate some implementation-level attacks but are often impractical on edge devices due to area, cost, and power constraints. Runtime monitoring is an attractive alternative. However, practical monitors must (1) discriminate malicious perturbations from benign process variability and (2) operate with deterministic, low-latency execution suitable for edge deployment.

Figure 1 illustrates the proposed lightweight multimodal fused hybrid monitoring architecture. Optical and thermal sensors capture process-level dynamics, while the embedded plat-

form performs cryptographic operations. Rather than relying exclusively on execution timing, the system constructs a multimodal feature vector combining timing, optical, and thermal signals. These features are processed through a two-stage hybrid detector consisting of a fast statistical thresholding layer and a compact machine-learning classifier. The thresholding stage provides ultra-low-latency screening of large deviations, while the learning stage models cross-modal correlations to identify subtle or coordinated disturbances.

Prior hybrid approaches combined statistical timing thresholds with compact machine-learning classifiers trained on timing and ciphertext-derived features, enabling real-time anomaly detection on edge-class System on Chips (SoCs). However, these timing-centric models assume that execution latency deviations alone are sufficient indicators of abnormal behavior and do not incorporate sensor-derived physical context [7, 14]. Consequently, they lack a mechanism to resolve ambiguity between process-induced timing jitter and genuine implementation-level faults.

To address this limitation, we propose a lightweight multimodal fused hybrid framework that jointly models cryptographic timing, optical emissions, and thermal behavior acquired from a sealed Laser Powder Bed Fusion (LPBF) imaging platform [15]. Optical images are captured between layer depositions under controlled illumination, while thermal intensity variations are recorded during active weld events within an argon-controlled chamber. By evaluating cross-modal consistency over short time windows, the detector suppresses false alarms caused by benign process variability while remaining sensitive to coordinated physical–cyber disturbances.

This paper makes four focused contributions:

1. A lightweight multimodal fusion architecture combining statistical thresholding and compact machine learning for real-time physical–cyber anomaly detection.
2. Integration of optical and thermal sensing data from a sealed LPBF platform to enable realistic multimodal evaluation.
3. A comparative analysis demonstrating that cross-modal modeling improves robustness relative to timing-only baselines.
4. Edge deployment and validation on a PYNQ-based platform, demonstrating deterministic real-time monitoring without reliance on heavyweight deep-learning models.

Threat Model

We consider coordinated cyber-physical attacks that affect both computational execution and physical sensing signals. Specifically, we evaluate:

Timing perturbations: Variations in execution latency caused by clock manipulation, workload interference, or injected faults.

Physical disturbances: Changes in thermal or optical signals due to environmental interference, sensor manipulation, or process disruption.

Cross-modal inconsistencies: Coordinated or uncoordinated deviations across modalities designed to evade unimodal detection.

Benign variations include natural process phenomena such as thermal drift, spatter, and illumination changes. Malicious samples are constructed by introducing controlled perturbations that mimic known side-channel and fault-injection attack patterns while preserving realistic signal characteristics.

PROPOSED MONITORING FRAMEWORK

Figure 1 presents the proposed lightweight multimodal fused hybrid monitoring framework for real-time physical–cyber anomaly detection in manufacturing systems. The architecture is designed explicitly for edge deployment, where sensing, encryption, and anomaly detection operate concurrently on a resource-constrained platform. Unlike timing-only detectors, the proposed framework jointly models execution-layer behavior and physical sensing data to resolve ambiguity between benign process variability and malicious perturbations.

System Overview

As illustrated in Figure 1, the monitoring pipeline consists of four tightly integrated stages: (1) multimodal feature acquisition, (2) statistical threshold screening, (3) machine-learning-based multimodal inference, and (4) hybrid decision fusion. The system operates at the block level, ensuring fine-grained analysis of both computational execution and manufacturing process dynamics. All feature extraction, inference, and performance evaluation were executed directly on an edge device (PYNQ platform), validating practical deployment feasibility.

Multimodal Feature Acquisition

For each processing block t, synchronized measurements are collected and assembled into a compact feature vector:

$$\mathbf{x}_t = [f_t^{\text{time}}, f_t^{\text{therm}}, f_t^{\text{opt}}] \tag{1}$$

where f_t^{time} denotes execution latency during encryption, f_t^{therm} captures thermal process signatures, and f_t^{opt} represents optical intensity behavior. As shown in the leftmost block of Figure 1, this stage forms the interface between the physical manufacturing environment and the embedded monitoring system.

By combining execution-layer and process-layer signals, the framework enables detection of both cyber-induced perturbations and process-level disturbances. This multimodal representation is critical for distinguishing malicious anomalies from legitimate manufacturing variability.

Stage 1: Lightweight Statistical Screening

The first detection layer performs ultra-low-latency statistical screening using the three-sigma rule:

$$T_m = \mu_m + 3\sigma_m \tag{2}$$

for each modality $m \in \{\text{time}, \text{therm}, \text{opt}\}$. If any feature exceeds its corresponding threshold, an anomaly flag is immediately generated. This stage provides deterministic and constant-time detection of large deviations, making it well-suited for hardware acceleration on edge platforms.

The statistical layer acts as a fast filter, ensuring immediate response to abrupt perturbations while maintaining minimal computational overhead.

Stage 2: Multimodal Machine Learning

While statistical thresholding captures pronounced anomalies, subtle coordinated disturbances may remain within individual bounds. The second stage therefore employs a compact Random Forest classifier to model nonlinear cross-modal relationships:

$$\hat{y}_t = \mathscr{F}(\mathbf{x}_t) \tag{3}$$

This learning layer identifies inconsistencies across timing, thermal, and optical channels that would be ambiguous under unimodal detection. By modeling cross-modal coherence, the classifier enhances sensitivity to stealthy or coordinated physical–cyber attacks.

Hybrid Fusion Strategy

The final decision is obtained through a conservative logical OR fusion rule:

$$y_t = y_t^{\text{threshold}} \vee \hat{y}_t \tag{4}$$

This hybrid strategy preserves rapid response to large deviations while improving recall through multimodal contextual analysis. The fused architecture therefore balances responsiveness, robustness, and computational efficiency.

Edge-Oriented Design and Deployment

The proposed framework is intentionally lightweight to enable execution on edge devices. Feature dimensionality is limited to three scalar modalities, and the machine learning component employs a shallow ensemble model to minimize memory and computation cost. The statistical screening layer is hardware-friendly and supports deterministic execution.

All experimental datasets were processed directly on a PYNQ-based edge platform to evaluate inference latency, throughput, and memory footprint under real deployment constraints. This ensures that reported performance results reflect practical edge execution rather than offline simulation.

By integrating thresholding and machine learning within a multimodal fused architecture, the proposed monitoring framework achieves real-time physical–cyber anomaly detection while satisfying the latency and resource constraints inherent to industrial edge environments.

RESULTS AND PERFORMANCE EVALUATION

Training and Evaluation Protocol

Data were partitioned into training and testing subsets. The training phase used a balanced distribution of benign and malicious samples to calibrate the statistical timing threshold and train the Random Forest classifier.

Performance was quantified using confusion matrix metrics (TP, FP, TN, FN), Precision, Recall, and F1-score. Average inference latency per block was measured to assess edge feasibility. Comparative experiments were conducted across timing-only detection, optical-only detection, thermal-only detection, dual-modality fusion, and full multimodal fusion.

Dataset

The experimental data used in this study were obtained from the in-situ optical and thermal monitoring dataset developed by Adhami et al. [15]. The dataset consists of welding recordings in MP4 and RAVI formats, covering aluminum, steel, stainless steel, and battery tab welding scenarios. File sizes range from 0.54 MB to 125.17 MB, reflecting variability in recording duration and acquisition format.

For clarity, the recording duration for each dataset is defined as

$$\text{Duration} = \frac{\text{Number of Samples}}{\text{Sampling Rate}} \tag{5}$$

with a fixed sampling rate of 100 MHz. The duration values in Tables 1 and 2 use the per-dataset recording duration values reported in the runtime table.

TABLE 1. Performance and system metrics: timing-only versus full hybrid fusion

Dataset	Size (MB)	Duration (s)	Timing-Only			Full Hybrid (O+T+Time)		
			Acc	Lat (ms)	TP (MB/s)	Acc	Lat (ms)	TP (MB/s)
Tab Welding	3.55	25.1624	0.9994	1.387	0.26	0.9998	0.900	0.26
Laser Welds	2.17	25.2757	0.9997	1.417	0.15	0.9998	0.907	0.16
Steel No Glass 1	5.46	25.2033	0.9991	1.378	0.40	0.9999	0.908	0.40
Steel No Glass 2	3.48	25.2543	0.9994	1.471	0.24	0.9998	0.913	0.25
Battery Tab Weld	52.60	25.3799	0.9983	1.436	3.66	0.9995	0.920	3.81
Thermal Test Weld Aluminium	125.17	25.0533	0.9993	1.440	8.69	0.9998	0.901	9.27
Test Weld Recording	0.54	25.1084	0.9993	1.418	0.04	0.9999	0.906	0.04
Filter Weld 2	96.62	24.9773	0.9990	1.505	6.42	0.9998	0.897	7.18
Weld 02	87.63	25.0752	0.9991	1.403	6.25	0.9998	0.898	6.51
Weld 01 Stainless	100.59	25.4877	0.9993	1.476	6.82	0.9999	0.926	7.24
Thermal Test Weld Steel 1	115.47	25.5788	0.9987	1.416	8.15	0.9978	0.929	8.29
Thermal Test Weld Steel 2	43.11	25.4334	0.9999	1.447	2.98	0.9986	0.933	3.08

Evaluation Baseline Versus Full

Performance was evaluated using Accuracy, Precision, Recall, F1-score, per-block inference latency (ms/block), and throughput (MB/s). All datasets were processed directly on the edge platform to assess both detection performance and real-time feasibility. Comparative analysis was conducted between the Timing-Only baseline and the proposed Full Hybrid (Optical + Thermal + Timing) configuration.

Table 1 presents the per-file performance and system metrics. Across nearly all datasets, the Full Hybrid configuration achieves equal or higher accuracy compared to the Timing-Only baseline. For example, in *Tab Welding*, accuracy improves from 0.9994 to 0.9998, while inference latency decreases from 1.387 ms to 0.900 ms per block. Similar improvements are observed for *Laser Welds*, *Steel No Glass 1*, and *Filter Weld 2*, where the hybrid model consistently achieves both higher accuracy and lower latency.

Importantly, the hybrid configuration does not introduce computational overhead. Measured per-block inference latency is consistently lower for the Full Hybrid model across all datasets, with reductions ranging from approximately 30% to 40% compared to the Timing-Only baseline. Throughput correspondingly improves across most files, demonstrating that multimodal fusion enhances detection performance without sacrificing real-time execution.

Under thermal stress conditions, such as *Thermal Test Weld Steel 1* and *Thermal Test Weld Steel 2*, accuracy differences between the two models are minimal, indicating stable behavior under dynamic process variability. Overall, the results demonstrate that the lightweight multimodal fused hybrid framework improves detection robustness while simultaneously reducing inference latency and maintaining stable throughput across diverse manufacturing scenarios.

System Configurations and Performance

The evaluated configurations represent progressively richer sensing strategies for anomaly detection. The Timing-Only configuration relies exclusively on execution latency measurements, whereas Optical-Only and Thermal-Only utilize individual physical-layer sensing channels. The dual-modality configurations combine computational and physical signals, while the Full configuration integrates all three modalities to capture cross-sensor correlations.

From a performance perspective, the Full configuration achieves the highest overall detection coverage. Both Timing + Optical and Timing + Thermal also show substantial improvements over single-modality approaches. In contrast, Optical-Only and Thermal-Only configurations exhibit noticeably lower standalone robustness, demonstrating that individual physical modalities alone are insufficient for reliable anomaly detection.

Overall, the results demonstrate that multimodal fusion, particularly the integration of timing and thermal signals, provides the strongest performance–efficiency balance. The Full configuration delivers maximal detection coverage with minimal latency and stable throughput, confirming its suitability for real-time edge deployment.

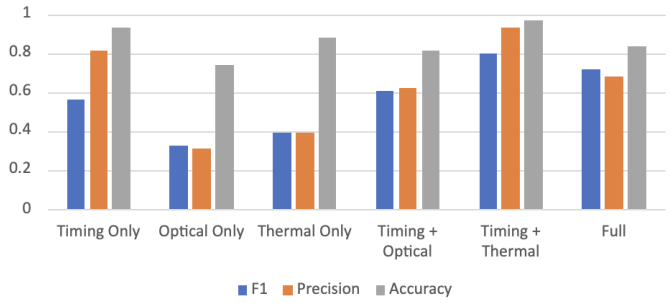

FIGURE 2. F1-score, precision, and accuracy comparison for the Filter Weld 2 dataset.

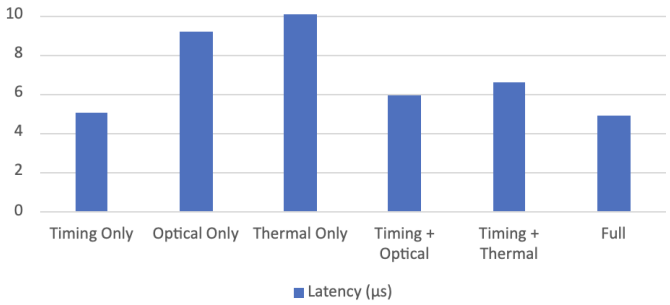

FIGURE 3. Latency comparison for the Filter Weld 2 dataset.

Configuration-Level Performance and System Metrics

Figures 2, 3, and 4 illustrate the configuration-level performance and system metrics for the Filter Weld 2 dataset across all evaluated configurations.

From a performance perspective, the Full configuration achieves the highest F1-score, precision, and accuracy, indicating the strongest overall detection capability. Both Timing + Optical and Timing + Thermal also show substantial improvements over single-modality approaches, with nearly identical performance. In contrast, Optical-Only and Thermal-Only configurations exhibit noticeably lower F1-score and precision values, demonstrating that individual physical modalities alone are insufficient for robust detection.

Configuration Matrix

Table 2 presents a detailed confusion matrix comparison between the Timing-Only baseline and the proposed Hybrid Fused framework across multiple industrial datasets of varying sizes. The results consistently demonstrate a significant reduction in False Negatives when multimodal fusion is applied. In the Tab Welding dataset, for example, FN decreases from 2 under the baseline to 0 with the hybrid approach. Similar reductions are observed across the remaining datasets, directly increasing true detection coverage.

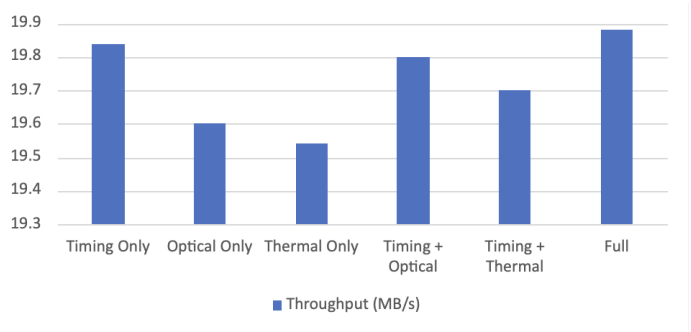

FIGURE 4. Throughput comparison for the Filter Weld 2 dataset.

Importantly, this improvement in recall does not introduce excessive False Positives. Across all datasets, FP remains relatively low compared to the total number of evaluated blocks, preserving operational reliability. True Negative counts remain stable, indicating that the hybrid model maintains normal process recognition while improving anomaly sensitivity.

Hardware Results

Industrial manufacturing systems increasingly rely on edge devices to perform real-time monitoring and anomaly detection directly at the machine level. These edge platforms must satisfy strict latency, memory, and throughput constraints while maintaining deterministic execution.

To evaluate deployment feasibility, the proposed Optical + Thermal + Timing Hybrid model was synthesized for the FPGA fabric of the PYNQ-Z2 platform and executed on its embedded ARM processor for runtime evaluation.

FPGA Resource Utilization Post-synthesis results confirm a compact hardware footprint. The hybrid detection logic occupies only 155 LUTs (0.29%) and 129 flip-flops (0.12%) of available resources, ensuring compatibility with existing industrial control logic.

Theoretical Hardware Performance The design was constrained at 100 MHz. With a fully pipelined architecture and initiation interval of one cycle, the detector processes one 128-bit block per clock cycle:

$$T = \frac{1}{100\ \text{MHz}} = 10\ \text{ns} \tag{6}$$

The corresponding theoretical throughput is:

$$\text{Throughput}_{\text{HW}} = 100 \times 10^6 \times 128\ \text{bits/s} = 12.8\ \text{Gbps} \tag{7}$$

which is equivalent to approximately 1600 MB/s.

TABLE 2. Confusion matrix comparison: timing-only versus full hybrid fusion

Dataset	Size (MB)	Duration (s)	Timing-Only				Full Hybrid (O+T+Time)			
			FN	FP	TN	TP	FN	FP	TN	TP
Tab Welding	3.55	25.1624	2	2	6,314	682	0	2	9,434	1,064
Laser Welds	2.17	25.2757	0	2	6,289	709	0	2	9,441	1,057
Steel No Glass 1	5.46	25.2033	4	2	6,350	644	0	1	9,420	1,079
Steel No Glass 2	3.48	25.2543	2	2	6,291	705	0	2	9,428	1,070
Battery Tab Weld	52.60	25.3799	10	2	6,291	697	0	5	9,436	1,059
Thermal Test Weld Aluminium	125.17	25.0533	3	2	6,336	659	0	2	9,455	1,043
Test Weld Recording	0.54	25.1084	3	2	6,312	683	0	1	9,473	1,026
Filter Weld 2	96.62	24.9773	5	2	6,272	721	0	2	9,434	1,064
Weld 02	87.63	25.0752	4	2	6,298	696	0	2	9,465	1,033
Weld 01 Stainless	100.59	25.4877	3	2	6,271	724	0	1	9,415	1,084
Thermal Test Weld Steel 1	115.47	25.5788	7	2	6,318	673	0	23	9,440	1,037
Thermal Test Weld Steel 2	43.11	25.4334	0	1	6,299	700	0	15	9,405	1,080

TABLE 3. Runtime performance of full hybrid (optical + thermal + timing) on edge platform

Dataset	Size (MB)	Inference Latency (μs/block)	End-to-End Latency (sec)	ARM Throughput (MB/s)	HW Throughput (MB/s)	Memory (MB)
Tab Welding	3.55	900	25.1624	0.26	1600	9.63
Laser Welds	2.17	907	25.2757	0.16	1600	9.64
Steel No Glass 1	5.46	908	25.2033	0.40	1600	9.63
Steel No Glass 2	3.48	913	25.2543	0.25	1600	9.66
Battery Tab Weld	52.60	920	25.3799	3.81	1600	9.61
Thermal Test Weld Aluminium	125.17	901	25.0533	9.27	1600	9.62
Test Weld Recording	0.54	906	25.1084	0.04	1600	9.79
Filter Weld 2	96.62	897	24.9773	7.18	1600	9.66
Weld 02	87.63	898	25.0752	6.51	1600	9.63
Weld 01 Stainless	100.59	926	25.4877	7.24	1600	9.64
Thermal Test Weld Steel 1	115.47	929	25.5788	8.29	1600	9.62
Thermal Test Weld Steel 2	43.11	933	25.4334	3.08	1600	9.60

Measured Edge Runtime Performance Measured inference latency on the ARM processor is reported in microseconds per block. Table 3 summarizes per-block inference latency, end-to-end processing time, throughput, and memory footprint for each dataset.

Although the theoretical FPGA latency is 10 ns, measured ARM inference latency ranges between approximately 897–933 μs per block, reflecting software execution and I/O overhead. End-to-end latency scales proportionally with dataset size. Measured throughput ranges from 0.04 MB/s to 9.27 MB/s depending on dataset characteristics.

The significant difference between theoretical FPGA throughput and measured ARM throughput reflects software processing constraints rather than limitations of the hybrid detection logic itself. The synthesis results demonstrate that the core detection module can operate at wire speed when fully mapped to programmable logic.

Power Analysis Vivado power estimation reports total on-chip power consumption of 0.141 W, with 0.036 W dynamic power. The low power overhead confirms suitability for continuous edge deployment.

CONCLUSION

Manufacturing systems are increasingly interconnected, sensor-rich, and dependent on embedded edge platforms for real-time monitoring and secure operation. While this connectivity enhances process visibility and productivity, it also expands the attack surface for coordinated physical–cyber threats. Traditional security mechanisms, particularly timing-only anomaly detection, are insufficient in environments where benign process variability and malicious perturbations produce overlapping execution signatures. This work addresses this limitation by proposing a lightweight multimodal fused hybrid framework that integrates statistical thresholding with machine learning to jointly analyze execution timing, optical emissions, and thermal behavior.

Experimental results demonstrate that multimodal fusion significantly improves anomaly detection robustness across diverse industrial datasets. Compared to timing-only monitoring, the proposed hybrid model consistently reduces false negatives, in several cases eliminating missed detections entirely, while maintaining low and controlled false positive rates. Configuration-level analysis further shows that combining timing and thermal signals provides a strong balance between precision and recall, while full multimodal fusion maximizes detection coverage. Importantly, these gains are achieved without introducing prohibitive computational overhead.

From a systems perspective, this work demonstrates that secure physical–cyber monitoring can be implemented within the constraints of industrial edge platforms. FPGA synthesis results show that the hybrid detector occupies less than 0.3% of available programmable logic and requires no BRAM or DSP resources, preserving capacity for control and communication tasks. Theoretical analysis indicates deterministic 10 ns latency and 12.8 Gbps throughput at a 100 MHz clock, while measured edge execution on the PYNQ platform confirms practical feasibility with approximately 900 μs per-block inference latency, stable memory usage (about 9.6 MB), and throughput up to 9.27 MB/s. These results collectively demonstrate that multimodal security monitoring can coexist with real-time industrial control workloads on resource-constrained devices.

Beyond empirical performance, the key contribution of this work is its architectural insight: effective physical–cyber attack detection in manufacturing requires modeling *cross-modal consistency* between execution-layer behavior and process-layer sensing. By capturing relationships across modalities, the proposed framework improves resilience against both overt faults and subtle coordinated perturbations. The lightweight design further ensures that enhanced security does not compromise real-time responsiveness or energy efficiency, which are critical requirements in industrial deployments.

Future work will focus on improving both performance and generality. First, migrating the full inference pipeline from the ARM processor to programmable logic will reduce the gap between theoretical and observed throughput. Second, controlled adversarial experiments will be conducted to evaluate robustness against coordinated cross-layer attacks, including combined timing and sensing perturbations. Third, adaptive thresholding and drift-aware learning strategies will be explored to maintain performance under long-term process evolution. Finally, extending the framework to additional manufacturing domains will enable broader validation of multimodal fusion as a scalable defense strategy.

Generalization Across Manufacturing Processes Although this work focuses on welding datasets, the proposed framework is inherently process-agnostic. It operates on abstracted multimodal features, timing, thermal, and optical signals, that are common across a wide range of manufacturing systems, including additive manufacturing, machining, and robotic assembly. The underlying principle of cross-modal consistency depends on relationships between sensing and computational behavior rather than process-specific characteristics.

Future work will evaluate the framework across diverse manufacturing processes and datasets to quantify generalization performance under varying operational conditions and attack scenarios.

Acknowledgments

This research was supported by the United States' National Science Foundation (NSF), USA, Grants Nos.CCF-2503055, and CPS-2237696. This work is also partially supported by the National Institute of Standards & Technology, United States, under grant number 70NANB23H030, as well as the South Carolina Space Grant Consortium under grants 21-117-RID RGP-SC-009, 25-073-REAP-SC-001, and 25-073-BTC-SC-001. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation, the South Carolina Space Grant Consortium, or the National Institute of Standards & Technology.

REFERENCES

[1] Zolfagharian, A., Jin, L., Ge, Q., Liao, W.-H., Díaz Lantada, A., Franco Martínez, F., Zhang, T., Liu, T., Wang, C. C. L., Mosallanejad, M. H., et al., 2026. "Roadmap on

artificial intelligence-augmented additive manufacturing". *Advanced Intelligent Systems*, p. e202500484.

[2] Fu, Y., Downey, A. R. J., Yuan, L., Huang, H.-T., and Ogunniyi, E. A., 2025. "Simulation-in-the-loop additive manufacturing for real-time structural validation and digital twin development". *Additive Manufacturing,* ***98***, p. 104631.

[3] Everton, S. K., Hirsch, M., Stravroulakis, P., Leach, R. K., and Clare, A. T., 2016. "A review of in-situ monitoring and control of powder bed fusion additive manufacturing". *Materials & Design,* ***95***, pp. 431–445.

[4] Reinhart, T., et al., 2024. "Side-channel analysis of encrypted in-situ monitoring for smart additive manufacturing". *IEEE Transactions on Industrial Informatics*. Publication details not available in current bibliography.

[5] Humayed, A., Lin, J., Li, F., and Luo, B., 2017. "Cyber-physical systems security—a survey". *IEEE Internet of Things Journal,* ***4***(6), pp. 1802–1831.

[6] Tuptuk, N., and Hailes, S., 2018. "Security of smart manufacturing systems". *Journal of Manufacturing Systems,* ***47***, pp. 93–106.

[7] Chinnasami, N., Stahle-Smith, R., and Karakchi, R., 2025. "Poster: Hybrid monitoring for side-channel security in edge socs". In Proceedings of the Tenth ACM/IEEE Symposium on Edge Computing.

[8] Kocher, P. C., 1996. "Timing attacks on implementations of diffie-hellman, rsa, dss, and other systems". In Advances in Cryptology—CRYPTO'96, Springer, pp. 104–113.

[9] Schneider, T., 2019. "Side-channel attacks on hardware and software implementations of aes". *Journal of Cryptographic Engineering,* ***9***(4), pp. 295–306.

[10] Hasan, M. A., Kabir, M. H., and Hasan, M. M., 2021. "Fault injection attacks on aes: A review". *Information Security Journal,* ***30***(2), pp. 67–79.

[11] Li, Y., et al., 2019. "A comprehensive survey on fault attacks and countermeasures for cryptographic devices". *IEEE Access,* ***7***, pp. 86827–86850.

[12] Zia, S., and Bibi, N., 2024. "Enhanced anomaly detection in iot: A transformer based approach for multivariate time series data". In 2024 International Conference on Engineering & Computing Technologies (ICECT), IEEE, pp. 1–6.

[13] Rashvand, N., Tabkhi, H., et al., 2024. "Enhancing automatic modulation recognition for iot applications using transformers". *IoT,* ***5***(2), pp. 214–230.

[14] Chinnasami, N., Stahle-Smith, R., and Karakchi, R., 2025. "Ml-enhanced aes anomaly detection for real-time embedded security". *arXiv preprint arXiv:2507.04197*.

[15] Adhami, M., Garcia-Sandoval, M., Hoang, T., Whetham, M., Sun, C., Downey, A. R. J., Fu, Y., and Yuan, L., 2026. "Design of vacuum- and pressure-compatible optical and thermal camera modules for laser powder bed fusion". In Proceedings of SPIE Smart Structures + Nondestructive Evaluation (SPIE SHM), SPIE.