

Digital Twin Frameworks for Predictive Maintenance in Smart Manufacturing

Sankaranarayanan R^{1*}, Chidambaram L² and Vasudevan S³

¹Department of Computational Intelligence, Netravati Institute of Edge and Cloud Intelligence, Nepal Email: ravi.sankaranarayanan@netravati.edu

ORCID: <https://orcid.org/4547-8673-1806-2366>

²Department of Computational Intelligence, Pamba Institute of Data-Driven Systems, Vietnam Email: lakshmi.chidambaram@pamba.edu

ORCID: <https://orcid.org/1869-5011-6983-7353>

³Department of Artificial Intelligence and Data Science, Manimala Center for Human-Computer Interaction, Nepal Email: sundar.vasudevan@manimala.edu

ORCID: <https://orcid.org/3096-3563-8272-1738>

ABSTRACT

Manufacturers have long faced a choice between reactive maintenance, which wastes money on unplanned downtime, and fixed-schedule preventive maintenance, which wastes good component life, and the digital twin has emerged as the organizing concept for a third option: a continuously synchronized virtual counterpart of a physical asset that can forecast a specific failure mode far enough in advance to act on it. This article synthesizes the literature on digital twin frameworks for predictive maintenance in smart manufacturing, tracing the concept from Grieves and Vickers' original physical-virtual-connection model through Kritzinger et al.'s influential three-tier taxonomy distinguishing a digital model, a digital shadow, and a true digital twin, into the reference architectures -- Lee et al.'s cyber-physical 5C pipeline, Tao and Zhang's shop-floor paradigm, and Lu et al.'s broader reference model -- that operationalize it on the factory floor. We review the physics-based, data-driven, and hybrid methods used to turn a digital twin into a working prognosis, and we examine the comparatively rare field evaluations, including a recent CNC-machining framework reporting R^2 above 0.94 for two prediction tasks and an 18-month steel-industry field trial with a measured, transparent detection record, that ground the field's architectural claims in evidence. We conclude that the conceptual and architectural foundations are mature and convergent across independent research groups, while the empirical base for sustained field performance, the closing of the automated loop that separates a digital twin from a digital shadow, and the integration of asset-level prognoses into system-level production planning remain open engineering problems.

Keywords: digital twin; predictive maintenance; smart manufacturing; cyber-physical systems; prognostics and health management

Citation: Sankaranarayanan R, Chidambaram L and Vasudevan S [2022]. Digital Twin Frameworks for Predictive Maintenance in Smart Manufacturing. DOI: <https://doi.org/10.5281/zenodo.22843838>

Copyright: © 2022 by the authors. Open access under CC BY 4.0 license.

Article Information: Received: 2 September 2021 Accepted: 7 December 2021 Published: 14 March 2022

Article Class: Research Article

1. Introduction

A modern production line generates more data in a single shift than most factories generated in a year two decades ago -- vibration traces off every spindle bearing, temperature and current draw from every drive, machine-vision frames from every inspection station, and a stream of PLC and SCADA tags recording the state of the process almost continuously. What manufacturers have struggled with is not the volume of that data but what to do with it: how to turn a torrent of sensor readings into an actionable answer to the question maintenance engineers have always asked -- will this machine fail, and when. The dominant paradigm for decades was reactive maintenance (fix it when it breaks) supplemented by preventive maintenance (replace it on a fixed schedule whether it needs it or not), and both approaches waste money in opposite directions: reactive maintenance costs unplanned downtime and collateral damage, while preventive maintenance throws away good component life and still misses failures that do not follow the assumed schedule.

The digital twin has emerged, over roughly the last decade, as the organizing concept that manufacturing has settled on to close this gap. The term itself predates most of the technology needed to realize it: Grieves and Vickers trace the idea to a 2002 product-lifecycle-management presentation and formalize it as a conceptual model consisting of a physical entity, a virtual entity, and the data connections that tie the two together bidirectionally, explicitly framing the virtual counterpart's job as anticipating and mitigating "unpredictable, undesirable emergent behavior" in the physical system before it happens [1]. What has changed since then is not the concept but the plumbing: cheap sensors, industrial IoT connectivity, cloud and edge compute, and machine learning models capable of learning degradation patterns from historical run-to-failure data have together made it possible to build virtual replicas that are not static CAD models but live, continuously updated digital counterparts of physical assets -- and predictive maintenance, forecasting a specific failure mode far enough in advance to act on it, is the application that has attracted the most sustained industrial and academic attention.

Lee, Bagheri, and Kao's cyber-physical systems (CPS) architecture for Industry 4.0-based manufacturing gave this convergence its first widely cited systems

framework, proposing a 5C (connection, conversion, cyber, cognition, configuration) pipeline that moves raw sensor data through progressively higher levels of abstraction until it becomes a decision a machine or a human can act on, with self-comparison against a fleet of similar machines as the mechanism for turning individual sensor readings into a prognosis [2]. This architecture, published in the same period that "Industry 4.0" was becoming a policy buzzword in Germany and elsewhere, gave the digital twin concept a concrete home in manufacturing systems engineering rather than leaving it as a product-lifecycle abstraction, and much of the architectural work synthesized in this article builds, explicitly or implicitly, on the 5C pattern it proposed.

This article synthesizes the literature on digital twin frameworks for predictive maintenance in smart manufacturing. We are not reporting new experiments; we are pulling together what a decade of foundational, architectural, and applied research has established about how these systems are built, what they have actually demonstrated in deployment, and where the evidence runs out. We organize the discussion around five themes: the conceptual and architectural foundations of the digital twin idea as it moved into manufacturing; the reference architectures and shop-floor implementations that operationalize it; the data-driven and physics-based methods that turn a digital twin into a predictive-maintenance capability; the empirical evidence -- still comparatively thin -- from real industrial deployments; and the open challenges that the literature itself flags as unresolved. Throughout, we try to be precise about which claims are backed by a fielded deployment with measured outcomes and which remain, however well-argued, proposals validated only in simulation or on a laboratory testbed. That distinction matters more in this literature than in most, because "digital twin" has become enough of a marketing term that separating the demonstrated capability from the aspiration requires reading past the abstract into the actual reported results.

2. Foundations of the Digital Twin Concept

From a Conceptual Model to a Manufacturing Discipline

The conceptual core of the digital twin, as Grieves and Vickers describe it, is deliberately simple: a physical object, a virtual model of that object, and a data

connection that keeps the two synchronized so that the virtual model can be used to explore, predict, and optimize the physical object's behavior without disturbing it [1]. What makes this more than a restatement of "simulation" is the emphasis on the live, bidirectional data link -- a traditional finite-element or discrete-event simulation is built once and validated against a snapshot of reality, while a digital twin is continuously re-synchronized as the physical asset ages, is repaired, or drifts from its as-designed configuration. That continuous synchronization is what, in principle, lets a digital twin track a specific physical asset's actual degradation trajectory rather than a generic population-average failure curve.

Kritzinger, Karner, Traar, Henjes, and Sihn's categorical literature review made an influential and often-cited contribution to disciplining this concept: they proposed a three-tier taxonomy distinguishing a digital model (no automated data flow between physical and virtual, essentially a conventional simulation), a digital shadow (automated one-way data flow from physical to virtual, so the virtual model reflects reality but cannot act back on it), and a digital twin proper (automated two-way data flow, so the virtual model's outputs can feed back into and change the physical process) [3]. This distinction turns out to matter enormously for predictive maintenance specifically, because a large share of what the industrial literature calls a "digital twin" is, by Kritzinger et al.'s own criterion, actually a digital shadow: a model that ingests live sensor data and produces a prognosis, but where the resulting maintenance action is still executed by a human rather than by an automated closed loop back to the physical asset. We adopt this distinction throughout and flag where the literature we discuss falls on either side of it.

Jones, Snider, Nassehi, Yon, and Hicks conducted a systematic literature review specifically aimed at reconciling the many divergent definitions of "digital twin" that had proliferated by the end of the 2010s, extracting the characterising attributes that recur across the reviewed corpus (physical entity, virtual entity, connection, data, and the various definitions' differing views on synchronization fidelity and automation level) and synthesizing them into a single, more precise characterising framework [4]. Attaran and Celik's more recent survey of digital twin benefits, use cases, challenges, and opportunities across industries reaches

a similar conclusion from a practitioner-facing angle, noting that organizations adopting digital twin technology frequently report benefit categories -- reduced downtime, improved product quality, faster time-to-market -- without always specifying which tier of Kritzinger et al.'s taxonomy their system actually occupies, which the authors argue complicates cross-organization benchmarking of reported digital twin value [5]. Their review is a useful corrective for a field prone to definitional drift, and it is worth noting explicitly, as they do, that a meaningful fraction of "digital twin" papers in the manufacturing literature describe systems that do not meet even the more permissive definitions under scrutiny -- a methodological caution that applies to some of the predictive-maintenance literature discussed later in this article as well.

Cyber-Physical Systems and the Industry 4.0 Context

The digital twin idea did not develop in isolation from the broader cyber-physical systems and Industry 4.0 literature; if anything, it is best understood as the concrete instantiation those broader programs needed. Lee, Bagheri, and Kao's 5C architecture -- connection (acquire reliable sensor data), conversion (turn raw data into meaningful information, for instance through feature extraction), cyber (aggregate information across a fleet of machines into a shared, comparable representation), cognition (generate an integrated, prioritized diagnosis or prognosis for human or machine decision-making), and configuration (apply corrective or preventive action back to the physical machine) -- gives the vague aspiration of "connecting the physical and the virtual" a concrete pipeline with a clear division of responsibility at each stage [2]. Read against Kritzinger et al.'s taxonomy, the 5C architecture's cognition and configuration layers are precisely what separates a digital shadow from a true digital twin: cognition without an automated configuration loop back to the physical asset is a sophisticated digital shadow, however elaborate its cyber-layer analytics.

Tao, Cheng, Qi, Zhang, Zhang, and Sui extended the digital twin concept explicitly into product design, manufacturing, and service, proposing that a digital twin should not merely mirror a machine's operating state but should integrate physical, virtual, and service data throughout the entire product lifecycle -- design, manufacturing process planning, and in-service

monitoring -- using big-data fusion techniques to keep all three representations consistent with one another [6]. Their framing is useful for predictive maintenance because it situates the maintenance problem inside a larger loop: a digital twin that has tracked an asset's actual manufacturing tolerances and its actual load history since commissioning is, in principle, positioned to make a far more asset-specific remaining-useful-life estimate than a model built solely from as-designed specifications, though realizing that promise in practice requires exactly the kind of continuous data integration that is often the hardest part of a deployment.

3. Digital Twin Architectures for Smart Manufacturing

Shop-Floor Reference Architectures

Moving from concept to a buildable system requires an explicit architecture, and Tao and Zhang's digital twin shop-floor (DTS) proposal is among the earliest and most cited attempts to specify one concretely for manufacturing [7]. Their architecture defines four core elements -- a physical shop floor, a virtual shop floor, shop-floor service systems, and shop-floor digital twin data -- and argues that the data layer specifically, unifying physical sensor data, virtual simulation data, and service-system data into a single fused representation, is what makes the other three elements usable together rather than as separate, siloed systems. This data-centric framing anticipates a recurring theme in later work: the architectural challenge of a manufacturing digital twin is less about the fidelity of any individual model and more about the data-integration layer that keeps disparate models, sensors, and control systems mutually consistent.

Lu, Liu, Wang, Huang, and Xu's reference model extends this shop-floor-level thinking into a broader connotation for digital-twin-driven smart manufacturing, organizing the field around five dimensions -- physical space, virtual space, digital twin data, connection, and service -- and, notably, conducting a systematic survey of what applications the existing literature had actually demonstrated versus merely proposed at the time of writing [8]. Their honest accounting is useful context for this article: even by 2020, a substantial share of the surveyed "digital twin manufacturing" literature described architectures and frameworks rather than deployed, evaluated systems, with predictive maintenance and process monitoring cited as the

application areas with the most mature reported implementations relative to more ambitious goals like fully autonomous self-organizing production.

Table 1: Representative digital twin and predictive-maintenance contributions discussed in this article, spanning quantitative applied results and structural/taxonomic contributions, with the effect or finding each cited source reports. [19,23,3,4,14].

Contribution	Source	Domain / Setting	Reported Result
Random forest surface-roughness prediction	Khan et al. [19]	CNC machining digital twin, 30 runs	$R = 94.2\% \pm 2.4\%$, $RMSE = 0.017 \pm 0.003$
XGBoost power-consumption prediction	Khan et al. [19]	CNC machining digital twin, 30 runs	$R = 98.9\% \pm 0.5\%$, $RMSE = 42.65 \pm 8.93$
Digital-twin deviation detection, field trial	Hassan et al. [23]	Steel-industry rotating machinery, 18 months	33 of 47 flagged events correctly linked to a maintenance need; 1 false alarm
Three-tier digital twin taxonomy	Kritzinger et al. [3]	Cross-industry categorical literature review	Automated-data-flow criterion now used as the field's standard classification
Characterising-attributes synthesis	Jones et al. [4]	Cross-domain digital twin literature	Consolidated divergent definitions into one characterising framework
Closed-loop fault-diagnosis architecture	Wang et al. [14]	Rotating-machinery fault diagnosis	Demonstrated bidirectional virtual-physical mapping for real-time fault localization

Negri, Fumagalli, and Macchi's review takes a complementary angle, cataloguing the specific roles a digital twin can play inside a broader cyber-physical production system -- as a simulation tool for what-if analysis, as a monitoring and diagnostic tool tracking deviations from expected behavior, and as a decision-support tool feeding optimization or scheduling algorithms -- and observing that most reviewed implementations specialize in one of these roles rather than realizing all of them simultaneously in a single deployed system [9]. For predictive maintenance specifically, this maps onto the monitoring-and-diagnostic role most directly, though the more ambitious formulations in the literature also fold in the decision-support role by using the digital twin's prognosis to drive maintenance scheduling and spare-parts logistics rather than merely flagging an anomaly for a human to interpret.

Comparative Framing Against Big Data and Industry 4.0

Qi and Tao's "360 degree comparison" between digital

twin and big-data-driven approaches to smart manufacturing is a useful piece for situating predictive maintenance methodologically, because it makes explicit a distinction that is often blurred in applied papers: a purely data-driven, big-data approach learns a predictive model directly from historical sensor data without an explicit physics-based representation of the asset, while a digital-twin approach maintains an explicit virtual model -- physics-based, data-driven, or hybrid -- of the specific physical asset, and uses live data to keep that model synchronized rather than to train a population-level statistical model from scratch [10]. Their comparison argues the two approaches are complementary rather than competing: a digital twin's physics-based component supplies interpretability and extrapolation beyond the observed operating envelope, while big-data-driven components supply pattern-recognition capability that is difficult to encode as first-principles physics, and the strongest recent predictive-maintenance systems, discussed further below, increasingly combine both.

Rasheed, San, and Kvamsdal's survey approaches digital twins specifically from a modeling-methods perspective, laying out a spectrum from purely physics-based digital twins, through hybrid twins that combine physics-based models with data-driven correction terms (often to compensate for model-form error or unmodeled effects), to purely data-driven digital twins that learn the input-output mapping directly from sensor data without an explicit governing-equation model at all [11]. Their survey is candid that purely physics-based digital twins, while the most interpretable and the most capable of extrapolating to conditions outside the training data, are frequently too computationally expensive to run in real time on live sensor streams, which is precisely the practical pressure that has pushed much of the predictive-maintenance literature toward hybrid and data-driven digital twin formulations -- a tension we return to in the next section.

4. Data-Driven and Physics-Based Methods for Predictive Maintenance

Physics-Based and Hybrid Prognostic Models

The physics-based end of the predictive-maintenance spectrum has its clearest early articulation in aerospace structural-health monitoring, a domain that adopted digital-twin-style thinking earlier than discrete-parts manufacturing because the cost of an undetected

structural failure is so severe. Tuegel, Ingrassia, Eason, and Spottswood's proposal to "reengineer aircraft structural life prediction using a digital twin" is widely credited as one of the earliest concrete applications of the digital twin concept to a maintenance problem specifically, arguing for an aircraft-specific structural model, continuously updated with flight-recorder and inspection data over that particular airframe's service life, that would replace fleet-wide, statistically averaged life-prediction methods with an individualized crack-growth and fatigue-damage prognosis for each tail number [12]. The proposal is explicitly forward-looking rather than a report of a fielded system, but it set out the individualized-prognosis argument that the rest of the predictive-maintenance literature has, in various forms, continued to make: a population-average maintenance schedule is conservative for most assets and dangerously optimistic for the unlucky few, and an asset-specific model can in principle do better than either extreme.

Li, Mahadevan, Ling, Choe, and Wang's dynamic Bayesian network framework for aircraft wing health monitoring is a direct technical response to that aspiration, fusing a physics-based crack-growth model with in-service sensor data (strain gauges and acoustic emission sensors monitoring the wing structure) inside a Bayesian filtering framework that updates the estimated damage state, and its uncertainty, as new sensor readings arrive [13]. The Bayesian formulation is architecturally significant for predictive maintenance because it treats the digital twin's internal state as a probability distribution over possible damage conditions rather than a single point estimate, which lets the system express calibrated uncertainty about a prognosis -- a property that purely deterministic physics simulations and many purely data-driven machine-learning models do not provide by default, and one increasingly identified in the broader digital-twin literature as important precisely because a maintenance decision that overstates its own certainty is more dangerous than one that is simply less accurate.

Wang, Ye, Gao, Li, and Zhang's digital twin architecture for rotating-machinery fault diagnosis brings this hybrid physics-plus-data approach into discrete-parts smart manufacturing specifically, proposing a closed-loop architecture in which a physics-informed virtual model of a rotating asset (a spindle or bearing assembly, in their reported implementation) is continuously updated from

vibration and operational sensor data, used to localize and classify developing faults, and then used in turn to update maintenance and operational decisions back on the physical machine [14]. Their architecture is a clear instance of Kritzing et al.'s strict definition of a digital twin (bidirectional, automated data flow) applied to a fault-diagnosis and, by extension, predictive-maintenance problem, and it is one of the more architecturally complete implementations discussed in the manufacturing digital-twin literature relative to the many proposals that stop at the diagnostic, digital-shadow stage. Zheng, Yang, and Cheng's application framework for digital twins, illustrated through a case study on a machining process, follows a broadly similar closed-loop pattern -- a physical-space model, a virtual-space model, and a connection layer that lets diagnostic outputs feed back into physical-process adjustments -- and their reported case study is a useful additional data point that this closed-loop pattern generalizes across more than one manufacturing sub-domain rather than being specific to rotating machinery [15].

Purely Data-Driven and Machine-Learning Approaches

A separate and much larger body of work has approached predictive maintenance primarily as a machine-learning problem, largely independent of the digital-twin framing, and it is essential context because much of the algorithmic substance inside a modern manufacturing digital twin's cognition layer comes directly from this literature. Carvalho, Soares, Vita, Francisco, Basto, and Alcalá's systematic literature review surveys the landscape of machine-learning methods that have been applied to predictive maintenance across industries, examining which algorithm families (traditional statistical and machine-learning methods, and the growing share of deep-learning approaches) have been used for which sub-tasks -- fault detection, fault diagnosis, and remaining-useful-life (RUL) estimation being the three most common framings in the surveyed corpus [16]. Their review is a useful reminder that "predictive maintenance" is not a single prediction task but at least three distinct ones with different data requirements and different failure costs when the model is wrong, and that a manufacturing digital twin's practical value depends on which of these three tasks its cognition layer is actually built to solve.

Zonta, da Costa, da Rosa Righi, de Lima, da Trindade, and Li's systematic literature review of predictive maintenance in Industry 4.0 covers overlapping ground from a slightly different angle, situating the machine-learning methods literature inside the broader Industry 4.0 technology stack -- IoT connectivity, cloud and edge computing, and big-data analytics platforms -- that has to be in place before a predictive-maintenance model, however accurate in a research setting, can actually run against a live production line [17]. Their review's emphasis on this infrastructural dependency echoes a point made from the architecture side by Lu et al.: the gap between a predictive-maintenance algorithm validated on a benchmark dataset and one running reliably against noisy, intermittently missing, real production-line sensor data is often the larger practical obstacle, more than any remaining gap in model accuracy on clean data. Errandonea, Beltrán, and Arrizabalaga's literature review focused specifically on digital twins for maintenance (as distinct from the broader Industry 4.0 predictive-maintenance literature surveyed by Zonta et al. and Carvalho et al.) reaches a compatible conclusion from the maintenance-management side, arguing that the maintenance discipline's own established frameworks -- condition-based maintenance, reliability-centered maintenance -- provide a more mature methodological vocabulary for evaluating a digital twin's maintenance value than the digital-twin literature has generally adopted on its own, and that closer integration between the two research communities would benefit both [18].

Khan, Khan, Khan, Mollan, Morkvenaite-Vilkonciene, and Pandey's 2025 data-driven digital twin framework for predictive maintenance of smart manufacturing systems is among the more recent and more quantitatively reported applied contributions to this literature, training and comparing several regression models against sensor data from a CNC machining process to predict two distinct process outcomes relevant to tool and machine condition: surface roughness and power consumption [19]. Averaged across thirty independent runs, their best-performing model for surface-roughness prediction, a random forest regressor, achieved a coefficient of determination (R) of 94.2% ($\pm 2.4\%$) with a root-mean-square error of 0.017 (± 0.003), while their best-performing model for power-consumption prediction, an XGBoost regressor, achieved a substantially higher R of 98.9% ($\pm 0.5\%$) with

an RMSE of 42.65 (± 8.93) [19]. The gap between the two tasks' achievable accuracy is itself informative: power consumption is a more directly and cleanly instrumented electrical quantity, while surface roughness is a downstream physical outcome influenced by more latent variables (tool wear state, vibration, material inhomogeneity), and the harder of the two tasks is arguably the more diagnostically useful one for anticipating a specific failure mode, which is a tension the broader predictive-maintenance literature returns to repeatedly: the easiest signals to predict accurately are not always the ones that matter most for maintenance decisions.

Fault Diagnosis, Application Reviews, and the Diversity of Deployed Use Cases

Cimino, Negri, and Fumagalli's review of digital twin applications in manufacturing provides a useful cross-sectional view of where, across the broader manufacturing digital-twin literature (not limited to predictive maintenance specifically), applications actually cluster: production planning and control, process and quality monitoring, and prognostics and health management emerge as the three most frequently reported application categories, with prognostics and health management -- essentially predictive maintenance under a different name -- consistently among the most mature in terms of the number of reported implementations with at least some empirical validation [20]. Their review's methodology, cataloguing applications by the specific digital twin function each reported system performs rather than by industry sector alone, is a useful complement to the more definitional reviews discussed earlier, because it grounds the architectural and taxonomic discussion in what practitioners have actually built.

Bécue, Maia, Feeken, Borchers, and Praça's digital twin concept for factory resilience and optimization broadens the predictive-maintenance framing further, proposing that a factory-level digital twin's value is not confined to forecasting individual component failures but extends to modeling how a local disruption -- a machine failure, a supply interruption, or, in their specifically emphasized case, a cybersecurity incident -- propagates through an interconnected production system, and using that propagation model to plan resilient responses before the disruption occurs rather than reacting after the fact [21]. Their broader framing is a useful counterweight to the

asset-level focus of most of the predictive-maintenance literature discussed above: an individual machine's remaining useful life is only actionable maintenance information once it is combined with a system-level understanding of how that machine's failure would actually affect the rest of the production line, spare-parts availability, and scheduled orders, and relatively few of the papers synthesized in this article integrate the asset-level prognosis with that system-level context as explicitly as Bécue et al. argue it needs to be integrated.

5. Case Evidence: Digital Twins in Industrial Maintenance Practice

Despite the volume of architectural and methodological literature, field-deployed digital twin predictive-maintenance systems with rigorously measured, reported outcomes remain comparatively rare, which makes the handful of papers that do report them disproportionately valuable for grounding the field's claims in evidence rather than aspiration. Aivaliotis, Georgoulas, and Chrysosouris's early treatment of digital twins for predictive maintenance is one of the more concrete methodological contributions aimed specifically at this evidentiary gap, proposing a physics-based simulation methodology for estimating a component's remaining useful life by continuously re-calibrating a simulation model of the component's degradation against measured operational load data, rather than relying on a fixed, pre-computed degradation curve [22]. Their approach sits squarely in the hybrid, physics-plus-data category discussed above, and is notable for treating the calibration procedure itself -- how much and how often the simulation model's parameters should be updated as new operational data arrives -- as a first-class methodological question rather than an implementation detail, since under-calibration leaves the model stale relative to the asset's actual condition while over-calibration risks the model simply tracking sensor noise.

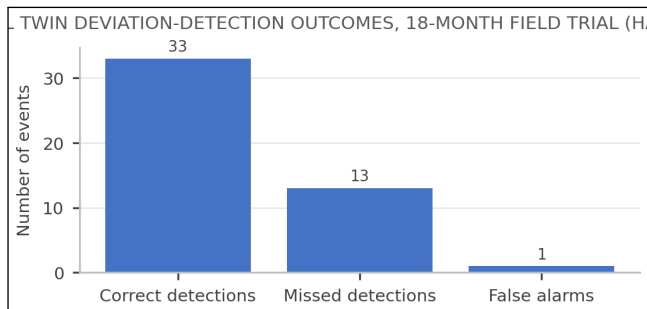


Figure 1: Outcomes of 47 flagged deviation events during an 18-month field trial of a digital twin maintenance system on steel-industry rotating machinery -- one of the few reported field evaluations with measured, transparent detection outcomes in this literature [23].

Hassan, Svadling, and Björsell's report on implementing digital twins for maintenance in an industrial steel-processing context is, to our knowledge, among the more methodologically transparent field evaluations in the recent literature, and its results are worth reporting in some detail precisely because they are more measured and more candid about limitations than most of the architectural literature discussed above [23]. Over an 18-month field trial comparing digital twin detection results against maintenance and control-system logs for rotating machinery, the authors report 47 total flagged deviation events, of which 33 were correctly linked to a genuine underlying maintenance need, 13 were missed detections (a real maintenance need that the digital twin did not flag in time), and 1 was judged a genuine false alarm, with an additional 10 flagged events attributable to legitimate operator parameter changes rather than to either a true maintenance need or a model error [23]. The authors are explicit that, given the several hundred individual units examined over the trial period, a single false alarm is an operationally acceptable rate, but they are equally explicit that the missed-detection rate and the parameter-change-induced false positives represent real, unresolved practical challenges rather than a solved problem, and they note candidly that they did not attempt to use the most sophisticated available algorithms at every stage of the pipeline, meaning the reported detection performance should be read as a floor on what a more carefully tuned implementation might achieve rather than a ceiling.

Read together, these two field-oriented papers illustrate a useful methodological contrast with the primarily architectural and taxonomic literature discussed in the previous two sections: Aivaliotis et al. focus on the calibration methodology needed to keep a physics-based digital twin synchronized with a specific

asset's actual condition, while Hassan et al. report what actually happens when such a system is run against real production data for a year and a half, including its failure modes. Both papers, notably, are considerably more cautious in their claims than the framework and architecture papers surveyed earlier in this article, which is itself an informative pattern: the parts of this literature closest to an actual fielded deployment tend to report the most measured, hedged conclusions, while the parts furthest from deployment -- conceptual frameworks, reference architectures, and taxonomies -- tend to make the broadest claims about what digital twins can achieve. Table 1 places several of the quantitative and structural contributions discussed across this article side by side to make this pattern, and the underlying evidence for each, more concrete.

6. Open Challenges

Several recurring gaps in this literature are worth naming plainly rather than leaving implicit. First, and most fundamentally, the definitional looseness that Kritzing et al. and Jones et al. both document empirically means that a nontrivial share of published "digital twin predictive maintenance" work describes systems that are, by a strict reading of the field's own proposed definitions, digital shadows rather than true digital twins -- systems that diagnose and predict but do not close the loop back to automated action on the physical asset [3,4]. This is not merely a terminological quibble: a digital shadow still requires a human in the loop to interpret the prognosis and schedule the maintenance action, which reintroduces exactly the kind of latency and human-judgment variability that the more ambitious digital-twin vision was meant to reduce, and readers evaluating a claimed "digital twin" system's practical maturity should check explicitly whether the reported architecture actually closes that loop.

Second, evaluation realism remains a genuine limitation across much of the literature synthesized here. Many of the architectural and reference-model papers -- including foundational contributions like Tao and Zhang's shop-floor architecture and Lu et al.'s reference model -- are, by their own authors' account, proposals validated on illustrative or laboratory-scale examples rather than reports of sustained, measured field deployments [7,8]. Hassan et al.'s field study is valuable precisely because it is a rare exception, but a single 18-month case study in one industrial sector does not

establish how representative its 70%-plus correct-detection rate and low false-alarm rate are of digital twin maintenance systems more broadly, particularly given how candid the authors themselves are that their algorithmic choices were not optimized for maximum accuracy [23]. The field would benefit from more studies reporting comparably transparent, multi-month field results with explicit accounting of missed detections and false alarms rather than accuracy figures computed only on held-out historical data.

Third, the tension Rasheed et al. identify between physics-based interpretability and real-time computational feasibility remains substantially unresolved [11]. Hybrid approaches like Li et al.'s Bayesian-filtering framework and Aivaliotis et al.'s continuously recalibrated simulation methodology represent genuine progress on this front, but both require nontrivial domain-specific physics modeling effort per asset class, which limits how quickly such approaches can be generalized across the heterogeneous mix of equipment found on a typical factory floor [13,22]. Purely data-driven approaches like Khan et al.'s regression-based framework sidestep this modeling burden and can be retrained more readily for a new asset class, but at the cost of the extrapolation guarantees and physical interpretability a physics-based or hybrid model provides, and of requiring a substantial volume of representative historical failure data that many individual manufacturing sites, particularly for less common or newer equipment, simply do not yet have [19].

Fourth, the system-level integration challenge that Bécue et al. raise -- connecting an individual asset's prognosis to the broader production system's scheduling, spare-parts logistics, and resilience planning -- is addressed by relatively few of the papers surveyed here as an explicit, first-class design concern rather than an assumed downstream consequence of having an accurate prognosis in the first place [21]. Most of the predictive-maintenance-specific literature, understandably, concentrates its effort on the harder and more tractable problem of getting the prognosis itself right; the practical value of that prognosis to a plant manager, however, depends substantially on how well it is integrated into the maintenance-scheduling and production-planning systems that actually decide when and how a flagged asset gets serviced, and this

integration layer is, in our reading, one of the more under-studied parts of the pipeline relative to its practical importance.

Finally, data quality and availability remain a quietly persistent constraint across nearly every paper discussed in this article, architectural and applied alike. Zonta et al.'s and Carvalho et al.'s reviews both note, in different ways, that the sensor infrastructure, historical labeled-failure data, and reliable connectivity a predictive-maintenance system needs are often the actual bottleneck to deployment, more than any remaining gap in modeling sophistication [17,16]. This is consistent with the pattern observed in the case-evidence section above: the field-deployed systems that do report results tend to be candid that data quality, sensor placement, and the labor-intensive process of building a reliable ground-truth maintenance log against which to validate the model were larger practical obstacles than the choice of prediction algorithm itself.

7. Conclusion

The literature reviewed here supports a reasonably specific and, we think, defensible claim about digital twin frameworks for predictive maintenance in smart manufacturing: the conceptual foundation is mature and reasonably well agreed upon, following Grieves and Vickers's original physical-virtual-connection formulation as refined and disciplined by Kritzing et al.'s automated-data-flow taxonomy and Jones et al.'s characterising synthesis [1,3,4], and the architectural patterns for realizing it in a manufacturing context -- from Lee et al.'s 5C cyber-physical systems pipeline through Tao and Zhang's shop-floor architecture to Lu et al.'s broader reference model -- have converged on a recognizably similar structure across independently developed proposals [2,7,8]. What remains considerably less mature is the empirical evidence base for how well these systems perform once deployed against real, messy production data over a sustained period: the small number of field-evaluated systems discussed here, notably Aivaliotis et al.'s recalibration methodology and Hassan et al.'s 18-month field trial, report genuinely useful but modest and carefully hedged results, not the transformative accuracy figures that framework papers sometimes imply are achievable [22,23].

The methodological trajectory of the field, in our reading, is toward hybrid architectures that combine an explicit physics-based or simulation-based component

with data-driven machine-learning models, following the pattern Rasheed et al. and Qi and Tao both describe as complementary rather than competing [11,10], applied to specific, well-instrumented sub-problems -- rotating-machinery fault diagnosis, tool-condition and surface-quality prediction, aircraft structural health monitoring -- where the physics is reasonably well understood and the relevant sensor data is comparatively easy to obtain reliably [14,19,13,12]. Extending this pattern to the full diversity of equipment on a typical factory floor, closing the automated loop from prognosis back to physical action so that more of these systems qualify as true digital twins rather than digital shadows, and integrating asset-level prognoses into system-level production and resilience planning as Bécue et al. and Negri et al. both argue is necessary, remain the field's most consequential open engineering problems [21,9]. None of these gaps are reasons for skepticism about the underlying pattern -- if anything, the consistency with which independent research groups across aerospace, discrete-parts manufacturing, and process industries have converged on structurally similar hybrid digital twin designs is itself evidence that the pattern is architecturally sound -- but the honest state of the evidence, as of the literature reviewed here, is that digital twin predictive maintenance is a well-motivated and increasingly well-demonstrated engineering discipline still working through the gap between compelling architecture and consistently measured field performance, rather than a solved problem ready for turnkey deployment across arbitrary manufacturing contexts.

8. References

- Grieves, M., Vickers, J. (2016) Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems. In: Transdisciplinary Perspectives on Complex Systems. Springer International Publishing, 85-113. DOI: [10.1007/978-3-319-38756-7_4](https://doi.org/10.1007/978-3-319-38756-7_4)
- Lee, J., Bagheri, B., Kao, H.A. (2015) A Cyber-Physical Systems Architecture for Industry 4.0-Based Manufacturing Systems. *Manufacturing Letters*, 3: 18-23. DOI: [10.1016/j.mfglet.2014.12.001](https://doi.org/10.1016/j.mfglet.2014.12.001)
- Kritzinger, W., Karner, M., Traar, G., Henjes, J., Sihn, W. (2018) Digital Twin in manufacturing: A categorical literature review and classification. *IFAC-PapersOnLine*, 51(11): 1016-1022. DOI: [10.1016/j.ifacol.2018.08.474](https://doi.org/10.1016/j.ifacol.2018.08.474)
- Jones, D., Snider, C., Nassehi, A., Yon, J., Hicks, B. (2020) Characterising the Digital Twin: A systematic literature review. *CIRP Journal of Manufacturing Science and Technology*, 29: 36-52. DOI: [10.1016/j.cirpj.2020.02.002](https://doi.org/10.1016/j.cirpj.2020.02.002)
- Attaran, M., Celik, B.G. (2023) Digital Twin: Benefits, use cases, challenges, and opportunities. *Decision Analytics Journal*, 6: 100165. DOI: [10.1016/j.dajour.2023.100165](https://doi.org/10.1016/j.dajour.2023.100165)
- Tao, F., Cheng, J., Qi, Q., Zhang, M., Zhang, H., Sui, F. (2018) Digital twin-driven product design, manufacturing and service with big data. *International Journal of Advanced Manufacturing Technology*, 94: 3563-3576. DOI: [10.1007/s00170-017-0233-1](https://doi.org/10.1007/s00170-017-0233-1)
- Tao, F., Zhang, M. (2017) Digital Twin Shop-Floor: A New Shop-Floor Paradigm Towards Smart Manufacturing. *IEEE Access*, 5: 20418-20427. DOI: [10.1109/ACCESS.2017.2756069](https://doi.org/10.1109/ACCESS.2017.2756069)
- Lu, Y., Liu, C., Wang, K.I.K., Huang, H., Xu, X. (2020) Digital Twin-driven smart manufacturing: Connotation, reference model, applications and research issues. *Robotics and Computer-Integrated Manufacturing*, 61: 101837. DOI: [10.1016/j.rcim.2019.101837](https://doi.org/10.1016/j.rcim.2019.101837)
- Negri, E., Fumagalli, L., Macchi, M. (2017) A Review of the Roles of Digital Twin in CPS-based Production Systems. *Procedia Manufacturing*, 11: 939-948. DOI: [10.1016/j.promfg.2017.07.198](https://doi.org/10.1016/j.promfg.2017.07.198)
- Qi, Q., Tao, F. (2018) Digital Twin and Big Data Towards Smart Manufacturing and Industry 4.0: 360 Degree Comparison. *IEEE Access*, 6: 3585-3593. DOI: [10.1109/ACCESS.2018.2793265](https://doi.org/10.1109/ACCESS.2018.2793265)
- Rasheed, A., San, O., Kvamsdal, T. (2020) Digital Twin: Values, Challenges and Enablers From a Modeling Perspective. *IEEE Access*, 8: 21980-22012. DOI: [10.1109/ACCESS.2020.2970143](https://doi.org/10.1109/ACCESS.2020.2970143)
- Tuegel, E.J., Ingraffea, A.R., Eason, T.G., Spottswood, S.M. (2011) Reengineering Aircraft Structural Life Prediction Using a Digital Twin. *International Journal of Aerospace Engineering*, 2011: 1-14. DOI: [10.1155/2011/154798](https://doi.org/10.1155/2011/154798)
- Li, C., Mahadevan, S., Ling, Y., Choe, S., Wang, L. (2017) Dynamic Bayesian Network for Aircraft Wing Health Monitoring Digital Twin. *AIAA Journal*, 55(3): 930-941. DOI: [10.2514/1.J055201](https://doi.org/10.2514/1.J055201)
- Wang, J., Ye, L., Gao, R.X., Li, C., Zhang, L. (2019) Digital Twin for rotating machinery fault diagnosis in smart manufacturing. *International Journal of Production Research*, 57(12): 3920-3934. DOI: [10.1080/00207543.2018.1552032](https://doi.org/10.1080/00207543.2018.1552032)
- Zheng, Y., Yang, S., Cheng, H. (2019) An application framework of digital twin and its case study. *Journal of Ambient Intelligence and Humanized Computing*, 10: 1141-1153. DOI: [10.1007/s12652-018-0911-3](https://doi.org/10.1007/s12652-018-0911-3)
- Carvalho, T.P., Soares, F.A.A.M.N., Vita, R., Francisco, R.P.,

- Basto, J.P., Alcalá, S.G.S. (2019) A systematic literature review of machine learning methods applied to predictive maintenance. *Computers & Industrial Engineering*, 137: 106024. DOI: [10.1016/j.cie.2019.106024](https://doi.org/10.1016/j.cie.2019.106024)
17. Zonta, T., da Costa, C.A., da Rosa Righi, R., de Lima, M.J., da Trindade, E.S., Li, G.P. (2020) Predictive maintenance in the Industry 4.0: A systematic literature review. *Computers & Industrial Engineering*, 150: 106889. DOI: [10.1016/j.cie.2020.106889](https://doi.org/10.1016/j.cie.2020.106889)
18. Errandonea, I., Beltrán, S., Arrizabalaga, S. (2020) Digital Twin for maintenance: A literature review. *Computers in Industry*, 123: 103316. DOI: [10.1016/j.compind.2020.103316](https://doi.org/10.1016/j.compind.2020.103316)
19. Khan, T., Khan, U., Khan, A., Mollan, C., Morkvenaite-Vilkonciene, I., Pandey, V. (2025) Data-Driven Digital Twin Framework for Predictive Maintenance of Smart Manufacturing Systems. *Machines*, 13(6): 481. DOI: [10.3390/machines13060481](https://doi.org/10.3390/machines13060481)
20. Cimino, C., Negri, E., Fumagalli, L. (2019) Review of digital twin applications in manufacturing. *Computers in Industry*, 113: 103130. DOI: [10.1016/j.compind.2019.103130](https://doi.org/10.1016/j.compind.2019.103130)
21. Bécue, A., Maia, E., Feeken, L., Borchers, P., Praça, I. (2020) A New Concept of Digital Twin Supporting Optimization and Resilience of Factories of the Future. *Applied Sciences*, 10(13): 4482. DOI: [10.3390/app10134482](https://doi.org/10.3390/app10134482)
22. Aivaliotis, P., Georgoulas, K., Chrysosouris, G. (2019) The use of Digital Twin for predictive maintenance in manufacturing. *International Journal of Computer Integrated Manufacturing*, 32(11): 1067-1080. DOI: [10.1080/0951192X.2019.1686173](https://doi.org/10.1080/0951192X.2019.1686173)
23. Hassan, M., Svadling, M., Björnell, N. (2023) Experience from implementing digital twins for maintenance in industrial processes. *Journal of Intelligent Manufacturing*, 35: 875-884. DOI: [10.1007/s10845-023-02078-4](https://doi.org/10.1007/s10845-023-02078-4)

9. Declarations

Funding

This literature synthesis received no dedicated grant funding from public, commercial, or not-for-profit agencies.

Conflict of Interest

The authors declare that they have no competing interests relevant to the content of this article.

Data Availability Statement

This article is a literature synthesis and did not generate new primary data; all findings discussed are drawn from the cited publications, and readers should consult those original sources for their underlying data and code.

Ethical Approval

Not applicable. This work did not involve human participants, animal subjects, or the collection of primary experimental data.