



Multi-Objective Optimization of Smart Sustainable Manufacturing Using Hybrid Machine Learning and Metaheuristic Algorithms: A Comparative Computational Study

Taha Muftah Abuali *

Department of Mechanical Engineering, Collage of Technical Sciences, Bani Walid, Libya

* Corresponding Author: abualitaha70@gmail.com

التحسين متعدد الأهداف للتصنيع الذكي المستدام باستخدام خوارزميات التعلم الآلي والاستدلالية
الهجينة: دراسة حسابية مقارنة

طه مفتاح ابو علي *

قسم الهندسة الميكانيكية والنفطية، كلية العلوم التقنية بني وليد، ليبيا

Date of Submission: 10-11-2025

Date of acceptance: 22-01-2026

Date of publishing: 05-02-2026



Copyright: © 2026 by the authors. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract

Manufacturers must increasingly optimize throughput, energy, emissions, quality, reliability, and worker-centered constraints at the same time. Machine learning (ML) predicts complex system behavior, while metaheuristics search large, discontinuous, and multi-objective decision spaces; their hybridization can therefore connect perception with action. This article synthesizes exactly 20 recent and authoritative sources - 11 peer-reviewed studies, five institutional or standards-based sources, and four official company cases - and develops a reproducible analytical demonstrator. The computational study represents a synthetic cell with 30 jobs, six heterogeneous machines, three eligible machines per job, three operating modes, time-varying carbon intensity, and machine-health effects. A scalar genetic algorithm (GA), NSGA-II, and an Extra-Trees-assisted NSGA-II were compared over eight independent runs under an equal budget of 396 exact simulator evaluations. The hybrid screened four times as many candidate offspring while retaining 15% random exploration and exactly re-evaluating every selected candidate. Its mean weighted compromise score was 0.977, compared with 0.990 for plain NSGA-II and 0.989 for GA; the paired hybrid-versus-NSGA-II score difference was significant (Wilcoxon $p = 0.039$). Relative to energy-aware list dispatch, the selected hybrid operating point reduced electricity by 6.13%, modeled CO₂ by 5.42%, and quality/reliability risk by 10.47%, while increasing risk-adjusted makespan by 5.10%. These are model results, not plant claims. The evidence and demonstrator jointly indicate that hybrid methods are most valuable when evaluations are expensive, objectives conflict, and decisions must adapt to changing production conditions. Their responsible industrial use requires exact feasibility checks, model-drift monitoring, lifecycle-aware sustainability accounting, human approval for consequential actions, and governance aligned with manufacturing standards.

Keywords: smart manufacturing; sustainable manufacturing; machine learning; metaheuristics; NSGA-II; digital twin; predictive maintenance; multi-objective optimization; Industry 5.0.

الملخص

يتعين على الشركات الصناعية بصورة متزايدة تحسين الإنتاجية، واستهلاك الطاقة، والانبعاثات، والجودة، والموثوقية، والقيود المرتبطة بالعاملين في الوقت نفسه. ويسهم التعلم الآلي (ML) في التنبؤ بسلوك الأنظمة المعقدة، في حين تبحث

الخوارزميات الاستدلالية (Metaheuristics) في فضاءات قرار كبيرة ومتقطعة ومتعددة الأهداف؛ ومن ثم، يمكن أن يؤدي دمجها في إطار هجين إلى الربط بين الإدراك واتخاذ القرار والتنفيذ.

تقدم هذه المقالة تركيبًا تكامليًا لـ 20 مصدرًا حديثًا وموثوقًا، تتضمن 11 دراسة محكّمة، وخمسة مصادر مؤسسية أو قائمة على المعايير، وأربع حالات صناعية رسمية لشركات، كما تطور نموذجًا توضيحيًا تحليليًا قابلاً لإعادة الإنتاج. تمثل الدراسة الحاسوبية خلية تصنيع اصطناعية تتكون من 30 مهمة، وست آلات غير متجانسة، وثلاث آلات مؤهلة لكل مهمة، وثلاثة أنماط تشغيل، وكثافة كربونية متغيرة بمرور الوقت، وتأثيرات مرتبطة بحالة الآلات وصحتها التشغيلية. وتمت مقارنة الخوارزمية الجينية القياسية (GA)، وخوارزمية الفرز الجيني غير المهيمن من الجيل الثاني (NSGA-II)، وخوارزمية NSGA-II مدعومة بخوارزمية Extra-Trees، وذلك عبر ثماني عمليات تشغيل مستقلة وباستخدام ميزانية متساوية قدرها 396 تقييمًا دقيقًا للمحاكي.

قامت الخوارزمية الهجينة بفحص عدد من المرشحين الجدد يعادل أربعة أضعاف ما تم فحصه في الأساليب الأخرى، مع الاحتفاظ بنسبة 15% للاستكشاف العشوائي، وإعادة التقييم الدقيق لكل مرشح تم اختياره. وقد بلغ متوسط درجة الحل التوافقي الموزونة للخوارزمية الهجينة 0.977، مقارنةً بـ 0.990 للخوارزمية NSGA-II التقليدية و 0.989 للخوارزمية الجينية GA. كما كان الفرق المقترن في درجة الأداء بين الخوارزمية الهجينة و NSGA-II ذا دلالة إحصائية، حيث بلغت قيمة Wilcoxon $p = 0.039$.

وبالمقارنة مع جدولة القوائم المعتمدة على الطاقة، حققت نقطة التشغيل المختارة باستخدام الخوارزمية الهجينة انخفاضًا قدره 6.13% في استهلاك الكهرباء، و 5.42% في انبعاثات ثاني أكسيد الكربون (CO₂) المحسوبة، و 10.47% في مخاطر الجودة والموثوقية، مقابل زيادة قدرها 5.10% في زمن الإنجاز (Makespan) المعدل حسب المخاطر. وتُعد هذه النتائج نتائج مستمدة من النموذج الحسابي وليست ادعاءات أو نتائج مثبتة في مصنع فعلي. وتشير الأدلة والنموذج التوضيحي مجتمعين إلى أن الأساليب الهجينة تكون أكثر قيمة عندما تكون عمليات التقييم مكلفة حسابيًا، والأهداف متعارضة، والقرارات بحاجة إلى التكيف مع ظروف الإنتاج المتغيرة.

ويتطلب الاستخدام الصناعي المسؤول لهذه الأساليب إجراء فحوصات دقيقة لقابلية التنفيذ والجدوى، ومراقبة انحراف النموذج (Model Drift)، ومراعاة الاستدامة على أساس دورة الحياة الكاملة، ووجود موافقة بشرية على الإجراءات ذات العواقب المهمة، وتطبيق حوكمة متوافقة مع معايير التصنيع.

الكلمات المفتاحية: التصنيع المستدام؛ التعلم الآلي؛ الخوارزميات الاستدلالية؛ NSGA-II؛ التوأم الرقمي؛ الصيانة التنبؤية؛ التحسين متعدد الأهداف؛ الصناعة 5.0.

1. Introduction

Manufacturing sustainability is no longer a single-variable energy exercise. A production decision can improve utilization while increasing peak electricity demand; reduce cycle time while accelerating tool wear; or lower direct emissions while shifting burdens to material waste, data infrastructure, or suppliers. The International Energy Agency's Energy Efficiency 2024 report treats industry as a central sector in the global efficiency transition and emphasizes electrification, flexibility, investment, and sector-specific policy rather than isolated equipment upgrades [1]. At the same time, the European Commission's Industry 5.0 framing broadens the objective from shareholder value alone to sustainable, human-centric, and resilient stakeholder value [2]. The technical problem is therefore multi-objective, dynamic, constrained, and socio-technical.

Two families of methods address complementary parts of this problem. ML can estimate remaining useful life, defect probability, energy demand, processing time, and other responses that are difficult to encode analytically. Metaheuristics - including genetic algorithms, NSGA-II, particle swarm optimization, artificial bee colony, grey wolf optimization, and memetic algorithms - can search nonconvex spaces containing discrete sequences, machine assignments, operating modes, and maintenance decisions. A hybrid system uses learned information to accelerate or guide search, uses metaheuristics to improve ML models, or combines both directions in a closed decision loop.

Digital twins provide the operational context for this loop. ISO 23247-1 defines general principles and requirements for manufacturing digital-twin frameworks [3]. Yet a technically functional twin is not automatically trustworthy. NIST's AI Risk Management Framework organizes AI governance around the functions Govern, Map, Measure, and Manage, with

attention to validity, reliability, safety, security, resilience, accountability, transparency, and explainability [4]. These properties are especially important when an optimizer can change schedules, set-points, maintenance timing, or human work allocation.

Recent scholarship confirms rapid growth in ML-metaheuristic integration and distinguishes ML-assisted metaheuristics from metaheuristic-assisted ML [5]. However, four gaps remain. First, much of the evidence is distributed across machining, scheduling, maintenance, additive manufacturing, and energy management, making transfer conditions hard to see. Second, algorithm papers often compare solution quality without imposing an equal budget of expensive physical or digital-twin evaluations. Third, sustainability is frequently reduced to one operational energy measure, while carbon timing, quality loss, reliability, and human factors remain external. Fourth, corporate case pages report valuable deployment evidence but rarely provide counterfactual designs or enough detail to attribute effects to one algorithmic component.

1.1 Aim and contributions

This article addresses those gaps through a bounded integrative review and a transparent computational study. It contributes: (i) a taxonomy of hybrid roles; (ii) a synthesis of 11 recent peer-reviewed studies and five authoritative institutional sources; (iii) a critical comparison of four company implementations; (iv) an equal-exact-evaluation comparison of GA, NSGA-II, and ML-assisted NSGA-II; and (v) an implementation roadmap for companies, engineers, operators, sustainability teams, and public or professional stakeholders.

1.2 Research questions

RQ1. Where does hybridization create operational value?

The review examines whether ML is most useful as a surrogate evaluator, predictive-maintenance model, search-policy controller, parameter tuner, or decision-support layer, and how that role changes across scheduling, machining, additive manufacturing, and utilities.

RQ2. What is gained under an equal exact-evaluation budget?

The demonstrator tests whether surrogate screening improves the weighted operational compromise when plain NSGA-II receives the same number of exact simulator evaluations.

RQ3. What conditions make deployment responsible and scalable?

The synthesis evaluates data readiness, exact verification, digital-twin integration, workforce design, sustainability accounting, cybersecurity, and governance requirements.

2. Conceptual foundation

2.1 Three forms of hybridization

Hybridization is not one algorithm. It is an allocation of learning and search responsibilities. Table 1 separates three practically important forms. The distinction matters because a system that uses a metaheuristic only to tune an ML model has a different safety boundary from a system that lets an ML model guide production schedules in real time.

Table 1. Taxonomy of ML-metaheuristic hybridization in manufacturing

Form	Mechanism	Manufacturing example	Primary value	Principal risk
ML-assisted metaheuristic	ML predicts fitness, constraints, promising regions, operators, or parameter settings.	Surrogate-assisted NSGA-II screens schedules before exact simulation.	Fewer expensive evaluations; faster adaptation.	Surrogate bias can exclude useful regions.
Metaheuristic-assisted ML	A metaheuristic selects features, tunes	GA tunes a defect classifier or energy forecaster.	Improved predictive performance	Search may overfit

	hyperparameters, designs architectures, or balances loss terms.		and compact models.	validation data.
Bidirectional closed loop	Prediction and search update one another through telemetry, twin evaluation, and verified execution.	RUL prediction changes maintenance windows; outcomes retrain the predictor.	Continuous decision improvement under drift.	Feedback can amplify error without gates.

Source: synthesis by the author based on the hybrid taxonomy in [5] and the manufacturing evidence reviewed below.

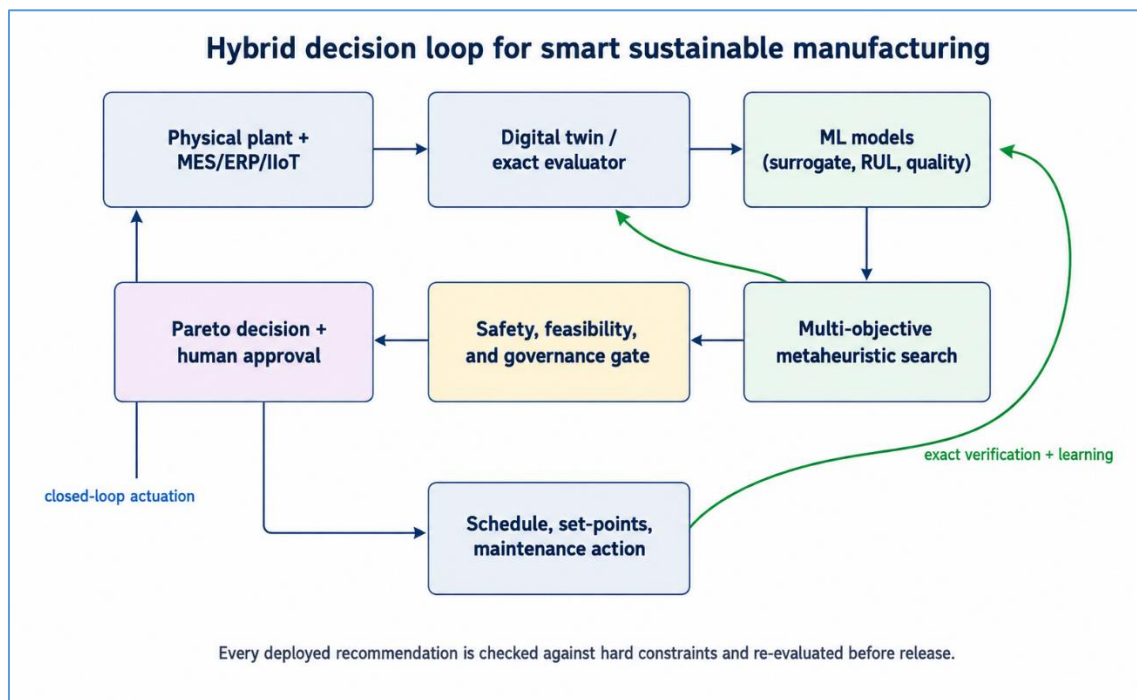


Figure 1. Closed-loop architecture for hybrid smart sustainable manufacturing

Figure 1 locates the hybrid algorithm inside a broader control and governance system. Plant, MES, ERP, quality, energy, and maintenance data update a digital twin or exact evaluator. ML models estimate expensive responses and guide a multi-objective search. Candidate decisions then pass hard feasibility, safety, and governance checks before a human or authorized controller releases them. Outcomes return as telemetry. This architecture prevents the surrogate from becoming the final authority: it screens possibilities, while the trusted evaluator and operational gate determine release.

2.2 Multi-objective sustainability

A decision vector x may contain job priorities, machine assignments, cutting conditions, batch sizes, maintenance timing, or energy modes. A minimum formulation can be written as:

$$\min F(x) = [T(x), E(x), R(x), C(x)] \quad (1)$$

where T is throughput loss or makespan, E is energy use, R is a quality-reliability risk measure, and C is carbon impact. Constraints define precedence, machine eligibility, labor and material availability, process limits, maintenance windows, safety envelopes, and delivery commitments. A Pareto solution is one for which no objective can be improved without

worsening another. This is more informative than collapsing objectives too early, because stakeholders can see whether a claimed sustainability gain is purchased through unacceptable delay or risk.

Sustainability also requires a declared system boundary. Operational electricity is useful but incomplete. A mature model should progressively include time-varying grid carbon, material yield and scrap, maintenance-induced life extension, water, embodied impacts, transport, worker exposure, and rebound effects. Not every pilot can model all dimensions, but omitted dimensions should be disclosed rather than implicitly treated as zero.

2.3 Why the hybrid can outperform either component alone

Pure ML predicts outcomes but does not reliably solve large combinatorial decision problems with hard constraints. Pure metaheuristics search flexibly but may require thousands of expensive simulations or experiments. In the hybrid, an ML surrogate ranks many candidates cheaply; a metaheuristic maintains exploration and trade-off diversity; an exact model validates selected solutions; and telemetry corrects the next model. The expected advantage is greatest when exact evaluations dominate cost, the response surface is learnable enough for ranking, and operational conditions change more slowly than the update cycle. The advantage shrinks when exact evaluation is already cheap, constraints can be solved exactly, the data distribution changes abruptly, or a weak surrogate cannot preserve candidate ordering.

3. Review of recent literature

3.1 Digital twins and predictive maintenance

Chen et al. review the role of ML in digital-twin-driven predictive maintenance, covering fault diagnosis, degradation estimation, remaining useful life, and decision support [6]. Their synthesis clarifies that the twin contributes context and feedback, while ML contributes inference. The unresolved challenge is to connect prediction uncertainty to maintenance and production decisions instead of treating a point estimate as certain.

Shoorkand, Noureldath, and Hajji develop a rolling-horizon hybrid deep-learning approach that coordinates predictive maintenance and production planning for multi-state systems [7]. This moves beyond isolated health prediction: the predicted state affects lot-sizing, maintenance, and demand satisfaction. The study demonstrates the value of integration, although the transfer of learned degradation behavior across plants and machine families remains a practical question.

Billey and Wuest provide an unusually concrete energy digital-twin case for a heating tunnel, reporting energy savings of up to 40% in the studied setting [8]. The result shows why exact or high-fidelity evaluation matters: plant behavior, timing, and operating rules can create savings that a static nameplate calculation would miss. It also illustrates a limitation of many studies - a strong single-cell result does not by itself establish cross-site generalizability.

3.2 Energy and process-parameter optimization

Ha, Lee, and Hwangbo present ML-based utility optimization in large-scale manufacturing plants [9]. Their three cases report a 2%-11% reduction in utility consumption, 6%-10% economic improvement, and approximately 140-fold data-volume reduction in the data-transformation case. These results are industrially meaningful because they link operational guidance, economic feasibility, and data engineering. The optimization is nevertheless plant-specific, and lifecycle or workforce effects are not the central outcome.

Yang et al. embed hybrid code-based and data-driven resource models inside NSGA-II to minimize time, energy, and material for material-extrusion 3D printing [10]. Validation on two machines strengthens external validity relative to a single simulated device. The study also demonstrates a transferable pattern: combine process code, measured behavior, and multi-objective search rather than relying on a single global regression.

In turning, Nguyen et al. combine multiple ML models with NSGA-II to minimize surface roughness and cutting force for AISI 4340 steel [11]. Their validation reports mean absolute percentage errors of 9.41% for surface roughness and 3.55% for cutting force at selected Pareto points. A related study by Nguyen et al. optimizes surface roughness and tool wear for Ti6Al4V turning [12]. Together, the studies show how ML can convert sparse experiments into continuous objective functions for Pareto search. Their main boundary is that optimal parameters remain tied to the tested machine-tool-material configuration.

3.3 Energy-aware scheduling and learning-guided search

Hu et al. propose a matheuristic and learning-oriented multi-objective artificial bee colony algorithm for energy-aware flexible assembly job-shop scheduling [13]. The contribution is not merely a new swarm variant: mathematical structure and learning inform search decisions in a domain with precedence, assembly, and energy constraints. Chang, Jia, and Ren similarly use reinforcement learning to enhance a memetic algorithm for multi-objective flexible job-shop scheduling toward Industry 5.0 [14]. These approaches represent performance-oriented ML assistance, where the learning mechanism selects or adapts search behavior.

Zhang, Zheng, and Ahmad develop an improved grey wolf optimizer for a flexible job-shop-type remanufacturing system, jointly reducing completion time and energy through disassembly, reprocessing, and reassembly decisions [15]. The remanufacturing context is important: sustainability depends on coordinating several process stages, not only optimizing one new-production machine. However, the algorithm is metaheuristic rather than an ML-metaheuristic hybrid; it therefore serves as a strong comparator and shows what domain-designed search can already achieve without learned screening.

Table 2. Evidence matrix for the 11 peer-reviewed sources

Ref.	Setting	Hybrid mechanism	Reported outcome focus	Primary limitation
[5]	Cross-domain review	ML-assisted MH; MH-assisted ML	Taxonomy and publication trend	Broad synthesis; no manufacturing-only effect estimate
[6]	Predictive maintenance	Digital twin + ML	Diagnosis, RUL, maintenance decisions	Review; deployment evidence varies
[7]	Multi-state production	Hybrid DL + rolling-horizon planning	Integrated cost and demand satisfaction	Transfer and uncertainty need plant validation
[8]	Heating-tunnel cell	Energy digital twin	Up to 40% energy saving	Single application and site
[9]	Large-scale utilities	ML-based feasible-region optimization	2%-11% utility reduction; 6%-10% economic gain	Plant-specific cases
[10]	Material-extrusion 3DP	Hybrid process model + NSGA-II	Time, energy, and material Pareto set	Two machines; process-specific calibration
[11]	AISI 4340 turning	ML surrogates + NSGA-II	Roughness and cutting-force trade-off	Configuration-specific; small experimental design
[12]	Ti6Al4V turning	ML surrogates + NSGA-II	Roughness and tool-wear trade-off	Dry-turning boundary

[13]	Assembly job shop	Matheuristic + learning-oriented ABC	Makespan and energy scheduling	Benchmark-centered evidence
[14]	Flexible job shop	RL-enhanced memetic algorithm	Industry 5.0 multi-objective scheduling	Complexity and policy interpretability
[15]	Remanufacturing	Improved grey wolf optimization	Completion-time and energy reduction	No learned surrogate; comparator role

Source: structured synthesis of [5]-[15]. Quantitative outcomes are reported only when the source provides a traceable value.

3.4 Industrial and company evidence

Academic results show feasibility, while factory cases show organizational scale. The World Economic Forum's 2025 Global Lighthouse Network report synthesizes lessons from 36 newly recognized sites and emphasizes capability building, frontline engagement, and end-to-end sustainability rather than isolated proofs of concept [16]. The report describes average reductions of approximately 30% in material waste and 25% in energy and water consumption across the relevant lighthouse evidence. These are portfolio-level reported outcomes, not estimates of one algorithm's causal effect.

At Siemens' Erlangen electronics factory, the company reports that AI, digital twins, and robotics contributed to a 69% productivity increase and 42% energy-consumption reduction over four years [17]. Schneider Electric reports that its Shanghai plant used ML-enabled prototyping, smart planning, and generative-AI maintenance to reduce make-to-order lead time by 67% and increase labor productivity by 82%; its Monterrey plant reported approximately 30% lower energy and water consumption alongside lower cost, defects, and lead time [18].

Bosch's 2025 annual-report case describes a Shopfloor Agent intended to preserve troubleshooting knowledge and shorten downtime; Bosch reports around EUR 850,000 annual savings per plant and an external pilot claim of up to tenfold faster failure resolution when sufficiently trained [19]. BMW's AIQX platform integrates cameras, sensors, and AI to provide real-time quality feedback, detect assembly anomalies, and reduce rework, although its public page does not disclose a controlled percentage effect [20]. These cases demonstrate breadth - energy, scheduling, maintenance, knowledge, and quality - but the metrics remain company-reported and often bundle hardware, software, process redesign, and workforce initiatives.

Table 3. Industrial evidence and interpretation of official company reports

Source	Technology focus	Reported outcome	Transferable lesson	Evidence caveat
WEF GLN [16]	Multi-site transformation portfolio	About 30% material-waste and 25% energy/water reductions on average in cited portfolio evidence	Scale, capability building, workforce	Portfolio synthesis; bundled interventions
Siemens [17]	AI + digital twin + robotics	69% productivity increase; 42% energy reduction over four years	Factory-wide digital architecture	Company-reported attribution
Schneider [18]	ML planning, inspection,	Shanghai: 67% lead-time reduction, 82%	End-to-end operations	Multiple technologies

	maintenance, energy systems	labor-productivity gain; Monterrey: about 30% lower energy/water		changed together
Bosch [19]	Agentic troubleshooting and knowledge reuse	About EUR 850,000 annual savings per Bosch plant; pilot reports up to 10x faster resolution	Human-AI support and downtime	Economic claim; limited sustainability boundary
BMW [20]	AIQX sensor and vision quality platform	Real-time anomaly feedback and less rework; no public controlled percentage	Quality-at-source and operator feedback	Quantitative baseline not disclosed

Source: [16]-[20]. Corporate values are reproduced as reported outcomes and should not be interpreted as independently audited causal estimates.

3.5 Synthesis and research gap

The literature converges on five findings. First, learned models are most useful when they replace an expensive response calculation, not when they merely add complexity. Second, multi-objective search is essential because energy, throughput, quality, maintenance, and human workload conflict. Third, digital twins and rolling horizons create value by updating decisions rather than optimizing a static snapshot. Fourth, domain knowledge - process code, mathematical constraints, machine eligibility, or expert rules - remains necessary for feasibility. Fifth, industrial gains generally come from a transformation bundle, so algorithm-only attribution is rarely justified.

The principal unresolved issue is trustworthy search under limited exact evaluation. A surrogate may be accurate on average but wrong in the narrow high-value region selected by an optimizer. Few studies publish both surrogate fidelity by objective and the performance of a plain metaheuristic under the same exact-evaluation budget. The demonstrator below targets that gap while maintaining a clear limitation: it is synthetic and explanatory, not a substitute for factory validation.

4. Study design and methods

4.1 Evidence protocol

A bounded integrative review was conducted with a cutoff of 5 August 2026. The evidence window was 2021-August 2026. Sources were included when they directly addressed (a) ML-metaheuristic integration; (b) sustainable manufacturing optimization, scheduling, energy, quality, or maintenance; (c) manufacturing digital twins and AI governance; or (d) a traceable official factory implementation. Peer-reviewed journal articles were preferred for methods and results; official standards, international organizations, and first-party company pages were used for architecture, governance, and deployment cases. Unsourced aggregators, duplicate summaries, and non-manufacturing optimization papers were excluded. The final reference set was deliberately controlled to exactly 20 sources, as required. This is an integrative evidence synthesis rather than a PRISMA-complete systematic review or meta-analysis.

4.2 Analytical demonstrator

The computational component models a heterogeneous single-operation manufacturing cell. It is designed to expose trade-offs and evaluation-budget behavior without presenting synthetic outputs as empirical factory data. Each candidate contains a continuous random-key priority,

one eligible-machine selector, and one operating-mode selector for every job. Decoding converts the keys into a feasible machine assignment and sequence.

Table 4. Configuration of the synthetic manufacturing-cell demonstrator

Element	Specification
Jobs and machines	30 independent jobs; 6 heterogeneous machines
Eligibility	Each job can run on 3 of the 6 machines
Operating modes	Eco, standard, and fast, with different duration, power, and quality-risk multipliers
Machine state	Speed, power, idle-power, and health factors vary by machine
Environmental signal	Deterministic time-of-use carbon-intensity curve
Decision variables	Priority key, eligible-machine key, and mode key for each job (90 variables)
Objectives	Risk-adjusted makespan, electricity, quality/reliability risk, and CO2
Baseline	Energy-aware longest-job-first list dispatch with standard mode
Algorithms	Scalar GA; NSGA-II; Extra-Trees-assisted NSGA-II
Budget	Population 36; 10 generations; 396 exact evaluations per method
Replications	8 independent algorithm seeds; one fixed manufacturing instance

Source: authors' analytical design. All values are synthetic and reproducible from the stated rules and fixed seeds.

4.3 Objective calculation

Processing duration depends on the job, machine-speed factor, operating mode, and an accumulated thermal-load term. Electricity combines active processing and idle consumption until the cell makespan. CO2 weights electricity by the modeled intensity at the midpoint of each operation. Quality/reliability risk combines job sensitivity, mode, machine health, thermal sequence, and expected downtime exposure. The baseline value of each objective is used for normalization.

$$S(x) = 0.40 T/T0 + 0.32 E/E0 + 0.18 R/R0 + 0.10 C/C0 \quad (2)$$

The weights in Equation 2 are illustrative stakeholder preferences, not universal sustainability weights. Throughput receives 0.40, electricity 0.32, quality/reliability risk 0.18, and carbon 0.10. The carbon weight is lower because carbon is partly correlated with electricity in this demonstrator; a plant with dynamic renewable procurement or a binding carbon target should choose different weights. The full Pareto set is retained even when Equation 2 selects one operating point.

4.4 Algorithm comparison

The GA optimizes the scalar score directly. NSGA-II uses non-dominated sorting, crowding distance, blend crossover, mutation, and elitist environmental selection across the first three operational objectives, while carbon is retained as a reported measure. The hybrid uses the same NSGA-II structure but trains an Extra-Trees multi-output surrogate on every exact evaluation. In each generation it produces four times the ordinary candidate pool, ranks candidates by predicted objectives, preserves 15% random exploration, and exactly evaluates only the selected population-sized subset. Thus both NSGA-II variants receive the same 396 exact evaluations; the hybrid spends additional inexpensive model predictions.

Every candidate admitted to the population is exactly re-evaluated. This safeguard is fundamental: predicted feasibility or predicted objective values alone never authorize a result. Plain NSGA-II is the principal ablation because removing the ML screener leaves the same representation, variation operators, population, generations, and exact-evaluation budget.

4.5 Analysis and statistical treatment

For each run, the operating point with the lowest Equation 2 score on the final non-dominated set was reported. Mean and sample standard deviation summarize the eight replications. Hybrid and plain NSGA-II were paired by random seed and compared with a two-sided Wilcoxon signed-rank test. With only eight runs, p-values are interpreted cautiously and no multiple-comparison claims are made. Convergence curves show the mean best score and one standard deviation. Surrogate fidelity was evaluated on 120 newly sampled held-out schedules per run using R-squared for each objective.

5. Analytics and comparative results

5.1 Convergence under equal exact evaluations

Figure 2 shows the central budget test. All methods start from the same dispatch-containing random population and receive 396 exact evaluations. The hybrid improves earlier and finishes with a lower mean best score than plain NSGA-II. The widening bands reflect stochastic variation and the small number of runs, so the curve should be read as comparative evidence for this instance rather than a universal convergence law.

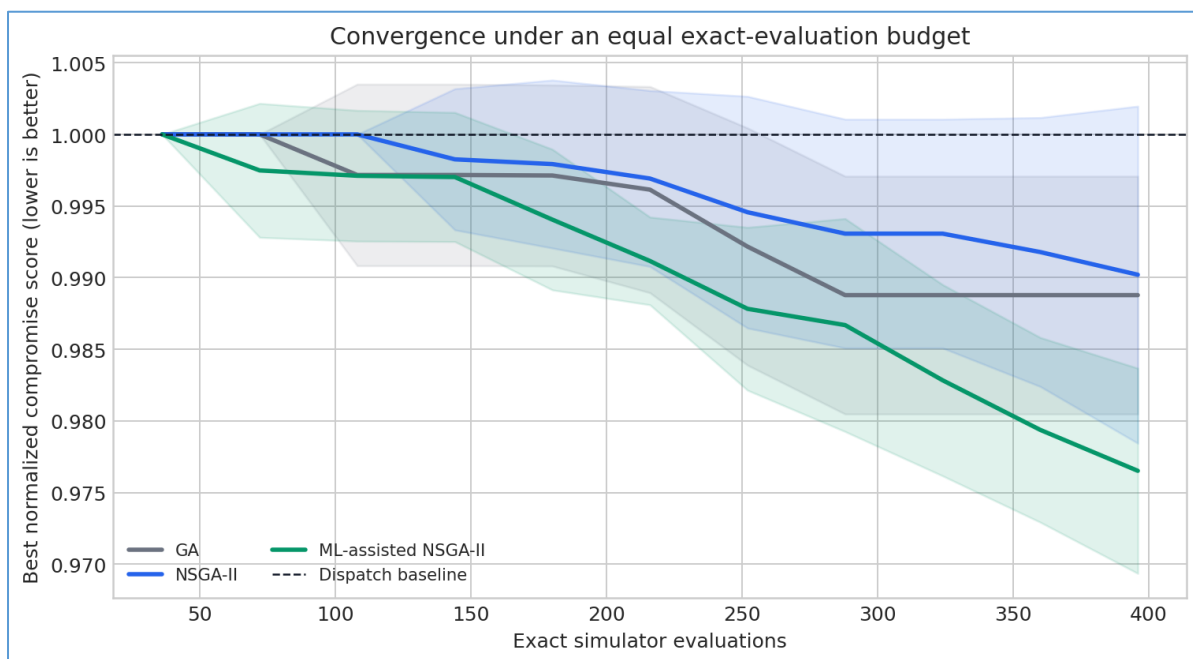


Figure 2. Convergence under an equal exact-simulator budget

Source: authors' synthetic demonstrator; mean +/- one sample standard deviation across 8 runs.

5.2 Pareto structure

The representative projection in Figure 3 illustrates why a single 'best algorithm' statement is incomplete. The dispatch baseline is fast but energy-intensive. Both NSGA-II variants discover slower, lower-energy schedules. The hybrid supplies several points in the central region where energy is reduced without moving to the extreme longest schedules. Because the actual search also considers risk, points that appear dominated in this two-dimensional projection can remain non-dominated in the full objective space.

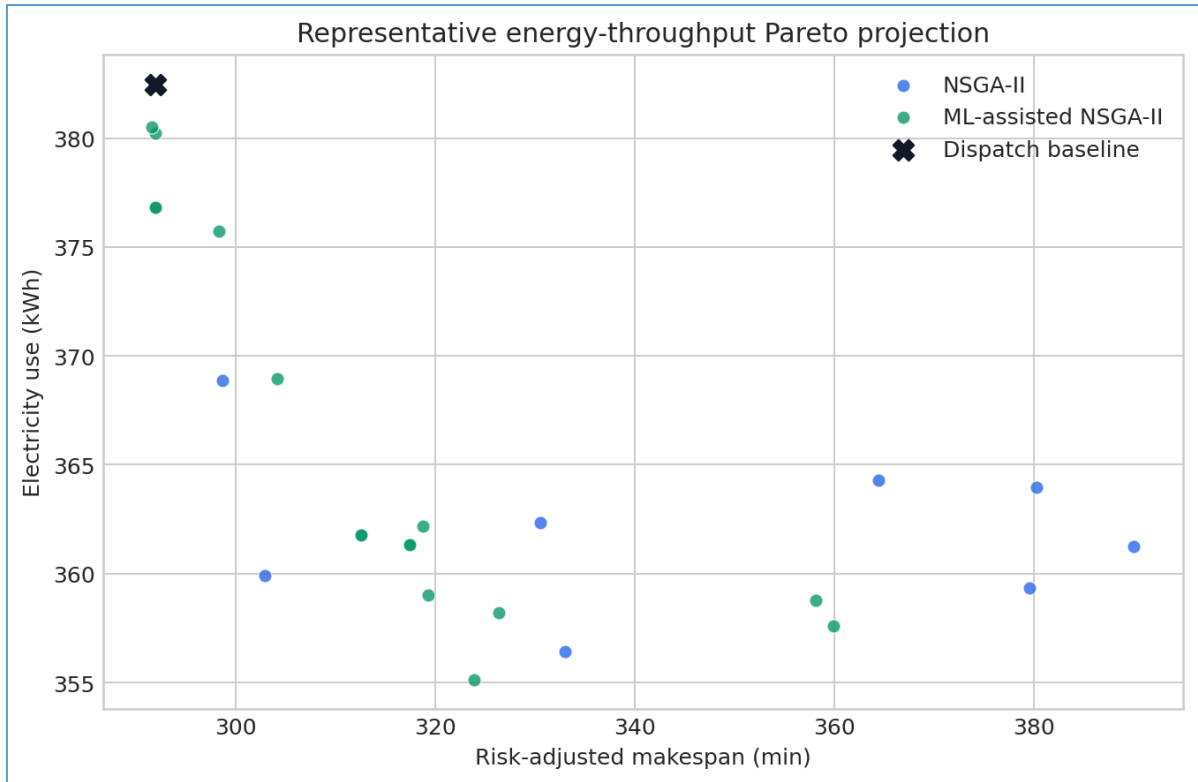


Figure 3. Representative energy-throughput Pareto projection

Source: authors' synthetic demonstrator, representative seed 4101. The plotted plane is a projection of a higher-dimensional non-dominated set.

5.3 Comparative operating points

Table 5 reports the consistent weighted operating rule. The hybrid's mean score of 0.977 was 1.38% lower than plain NSGA-II and 1.24% lower than GA. The paired score difference against NSGA-II was statistically significant at $p = 0.039$. Individual hybrid-versus-NSGA-II outcomes did not reach $p < 0.05$: makespan was 2.21% higher ($p = 0.250$), electricity 3.26% lower ($p = 0.109$), risk 5.81% lower ($p = 0.109$), and CO2 2.81% lower ($p = 0.148$). The result is therefore a better weighted compromise in this configuration, not dominance on every metric.

Table 5. Comparative outcomes for the synthetic manufacturing cell

Method	Makespan (min)	Energy (kWh)	Risk index	CO2 (kg)	Score
Energy-aware dispatch	292.0	382.5	30.02	123.99	1.000
GA	291.8 +/- 7.0	374.9 +/- 6.7	29.54 +/- 0.58	121.75 +/- 1.92	0.989 +/- 0.008
NSGA-II	300.2 +/- 10.0	371.1 +/- 12.9	28.53 +/- 1.66	120.65 +/- 4.00	0.990 +/- 0.012
ML-assisted NSGA-II	306.9 +/- 14.0	359.0 +/- 12.4	26.88 +/- 1.71	117.27 +/- 3.48	0.977 +/- 0.007

Source: authors' synthetic demonstrator. Dispatch is one deterministic run; algorithm rows are mean +/- sample standard deviation across 8 runs. Lower is better for every metric.

5.4 Baseline-relative trade-offs

Relative to energy-aware dispatch, the hybrid reduced electricity by 6.13%, modeled CO₂ by 5.42%, and the quality/reliability risk index by 10.47%, while risk-adjusted makespan increased by 5.10%. Figure 4 makes the opportunity cost visible. A decision maker who assigns a much larger weight to immediate throughput would probably choose the GA or a different Pareto point; a plant facing energy limits, carbon constraints, or expensive scrap could rationally accept the delay. The analytical lesson is that sustainability improvements should be reported with the displaced objective, not as a free gain.

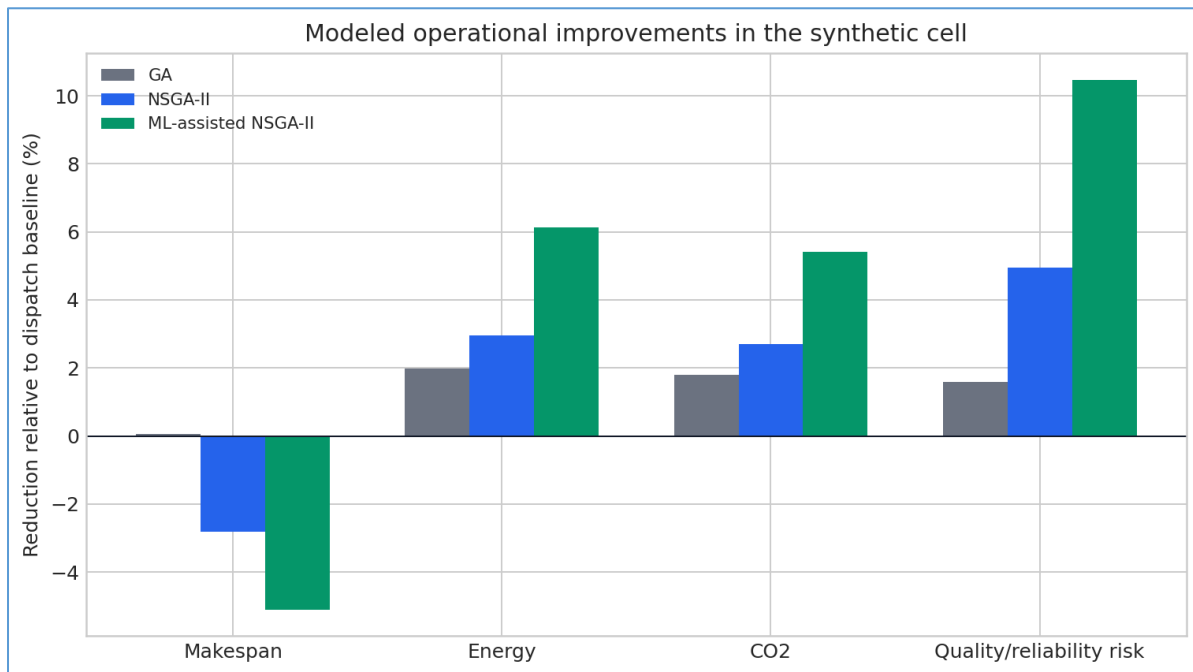


Figure 4. Modeled changes relative to energy-aware dispatch

Source: authors' synthetic demonstrator. Positive bars are reductions; negative makespan bars indicate an increase in completion time.

5.5 Surrogate fidelity

Held-out R-squared averaged 0.936 for makespan, 0.718 for energy, 0.180 for quality/reliability risk, and 0.758 for CO₂. The strong makespan and moderate energy/carbon ranking explain why screening can help. The weak risk model is equally informative: average predictive performance is objective-dependent, and a hybrid should not silently substitute a weak surrogate for exact risk evaluation. In this study all selected candidates were exact-evaluated, and 15% random exploration limited premature exclusion. In a plant, low risk-model fidelity would justify additional sensing, feature engineering, uncertainty-aware acquisition, or removal of that surrogate from release-critical decisions.

6. Discussion

6.1 Answer to RQ1: where hybridization creates value

The review and demonstrator agree that hybridization creates the clearest value at the expensive-evaluation boundary. Examples include a digital twin that simulates energy and thermal behavior, a physical experiment that measures roughness or tool wear, a maintenance model that updates from sensor data, or a production plan that must be recomputed as health and demand change. ML is valuable when it ranks or approximates these evaluations; metaheuristics are valuable when decisions remain discrete, constrained, nonlinear, and multi-objective. The combination is less compelling for a small deterministic problem that a

mathematical optimizer can solve exactly or for a low-dimensional response that can be sampled cheaply.

6.2 Answer to RQ2: what was gained under equal budget

Under an equal 396-evaluation budget, the hybrid improved the final weighted compromise relative to plain NSGA-II by 1.38%, with $p = 0.039$ across paired seeds. The magnitude is modest, which is credible for an already competent optimizer and a small data regime. The hybrid did not dominate: its selected point was slower than the plain NSGA-II mean but used less energy and carried lower modeled risk. This supports a conditional conclusion. Surrogate assistance can buy a better compromise by screening more candidates, but the benefit depends on stakeholder weights, surrogate ranking quality, and exact verification.

6.3 Relation to published and corporate evidence

The modeled reductions are smaller than several reported factory transformations. This difference should not be treated as a failure. Siemens, Schneider, and the Lighthouse cases combine data infrastructure, automation, equipment upgrades, process redesign, and workforce programs [16]–[18]. Bosch and BMW similarly embed AI in knowledge and quality systems rather than applying a single optimizer [19], [20]. An algorithm-only demonstrator should not be expected to reproduce transformation-level effects. Instead, it isolates one mechanism: ML screening can allocate a limited evaluation budget more effectively.

6.4 Sustainability interpretation and rebound

Operational energy and carbon savings can be offset by higher output, more computing, accelerated equipment replacement, upstream materials, or schedule changes that shift rather than remove burdens. A credible deployment should report absolute and intensity metrics, the counterfactual production volume, grid-carbon timing, scrap and rework, and the energy used for training and inference. It should also monitor whether improved efficiency causes the plant to consume more total resources through expanded production. The Industry 5.0 perspective adds worker well-being and resilience to this boundary [2].

6.5 Human factors, safety, and governance

The company cases show that operators are not peripheral. Schneider attributes scale partly to workforce adoption; Bosch uses AI to make troubleshooting knowledge accessible; BMW returns quality feedback to employees in real time [18]–[20]. A human-centered hybrid should explain the trade-off set, identify the data and constraints driving a recommendation, preserve operator override, and record who approved consequential actions. Safety constraints must remain hard constraints, not weighted objectives that can be traded away. Governance should map the use case, measure model and system performance, manage drift and incidents, and assign accountable owners in line with NIST AI RMF [4].

6.6 Limitations

The literature synthesis is intentionally bounded to 20 sources and is not exhaustive. Company figures are first-party or WEF-compiled and are not independently re-estimated. The computational instance is synthetic, uses one fixed plant configuration, and simplifies precedence, labor, inventory, tariffs, maintenance intervention, and lifecycle impacts. Only eight stochastic runs were used, limiting statistical power. The weighted score is illustrative; different preferences select different Pareto points. The surrogate's low risk R-squared demonstrates that not all objectives were equally learnable. Finally, wall-clock savings were not claimed because the exact simulator in this demonstrator is computationally cheap; evaluation count is used as a proxy for settings where physical experiments or high-fidelity twins are expensive.

7. Implications for companies, engineers, and stakeholders

Implementation should proceed as a governed capability rather than a one-off algorithm contest. Table 6 assigns practical responsibilities and acceptance evidence. The sequence starts

with a bounded decision and measured baseline, then adds a trusted evaluator, an offline Pareto pilot, supervised shadow operation, and only later closed-loop control for low-consequence decisions.

Table 6. Implementation roadmap by stakeholder

Stakeholder	Primary action	Minimum acceptance evidence	Critical caution
Executive sponsor / plant manager	Choose one decision with material cost, energy, quality, or downtime value; approve trade-off weights.	Baseline and target signed by operations and sustainability; named accountable owner.	Avoid broad 'AI factory' programs without a decision owner.
Manufacturing / process engineer	Encode feasibility, precedence, operating envelopes, changeovers, and fail-safe defaults.	All released candidates pass deterministic checks and engineering review.	Never allow a surrogate to soften a safety limit.
Data / ML engineer	Build lineage, drift checks, uncertainty estimates, and objective-specific validation.	Hold-out and high-value-region performance; retraining trigger; rollback package.	Average accuracy can hide optimizer-selected errors.
Optimization engineer	Select representation, baseline, evaluation budget, diversity metric, and exact verification policy.	Equal-budget benchmark; ablation against non-ML search; reproducible seeds.	Do not compare algorithms with unequal exact calls.
Sustainability / finance	Define energy, carbon, water, scrap, cost, and lifecycle boundaries; value rebound and avoided loss.	Absolute and intensity KPIs with production-normalized counterfactuals.	Do not count shifted burdens as reductions.
Operators / maintenance	Co-design constraints and interfaces; test explanations, overrides, and workload effects.	Shadow-mode acceptance; documented override; training and feedback channel.	Human approval must be usable, not ceremonial.
Quality, safety, OT cyber, regulators	Threat-model data and actuation paths; control access; audit decisions and incidents.	Signed release criteria, logs, access control, incident and model-change records.	Treat optimizer output as an operational change.

Source: authors' implementation synthesis, aligned with ISO 23247-1 [3], NIST AI RMF [4], and the workforce lessons in [16]-[20].

7.1 Recommended staged deployment

Stage 0 - decision framing: define the controllable decision, hard constraints, baseline, sustainability boundary, and accountable owner. Stage 1 - measurement: establish data quality, energy and quality metering, machine-state context, and a versioned evaluator. Stage 2 - offline Pareto pilot: compare dispatch, exact or mathematical methods where feasible, a plain metaheuristic, and the hybrid under equal exact-evaluation budgets. Stage 3 - shadow mode:

generate recommendations without actuation and measure feasibility, regret, operator acceptance, and drift. Stage 4 - bounded execution: automate only low-consequence actions inside a verified safety envelope, with rollback and periodic review. Higher-consequence maintenance or process set-points should retain qualified human approval until evidence supports a more autonomous safety case.

7.2 Engineering KPI set

A minimum technical dashboard should include makespan or throughput, energy per good unit, time-matched carbon, peak demand, scrap/rework, predicted and realized downtime, service level, compute and simulation cost, exact-evaluation count, Pareto hypervolume or diversity, constraint violations, surrogate error by objective, uncertainty coverage, drift, override rate, and operator workload. Business reporting should separate model performance, decision quality, and realized plant impact; they are not interchangeable.

8. Conclusions

Hybrid ML and metaheuristic algorithms offer a practical bridge between data-driven prediction and constrained multi-objective action. The recent literature demonstrates credible patterns in predictive maintenance, utility optimization, additive manufacturing, machining, scheduling, and remanufacturing. Factory cases show that measurable gains are possible at scale, but they also show that value usually comes from a bundle of digital infrastructure, process redesign, equipment, and people - not from one algorithm in isolation.

In the equal-budget synthetic study, ML-assisted NSGA-II achieved a mean weighted score of 0.977 versus 0.990 for plain NSGA-II and 0.989 for GA. Relative to energy-aware dispatch, it reduced electricity by 6.13%, CO₂ by 5.42%, and modeled quality/reliability risk by 10.47%, while increasing makespan by 5.10%. The paired weighted-score improvement over plain NSGA-II was significant at $p = 0.039$, but individual objectives were not. The result supports a precise conclusion: surrogate screening can improve the chosen compromise when exact evaluations are scarce, but it does not remove trade-offs and should not bypass exact validation. For companies and engineers, the recommended path is to begin with one bounded decision, publish the objective weights and system boundary, benchmark against strong non-ML methods under equal budgets, validate every selected candidate, monitor error by objective, and deploy through shadow mode with meaningful operator participation. Future research should add uncertainty-aware acquisition, dynamic and multi-site benchmarks, lifecycle assessment, worker-centered objectives, cybersecurity stress tests, and independently audited industrial trials. Smart sustainable manufacturing will be credible when optimization performance, environmental accounting, human agency, and operational safety are treated as one engineering system.

References

- [1] International Energy Agency, Energy Efficiency 2024. Paris, France: IEA, 2024. Available: [official source page](#)
- [2] European Commission, Directorate-General for Research and Innovation, Industry 5.0: Towards a Sustainable, Human-Centric and Resilient European Industry. Luxembourg: Publications Office of the European Union, 2021. Available: [official source page](#)
- [3] International Organization for Standardization, ISO 23247-1:2021, Automation Systems and Integration - Digital Twin Framework for Manufacturing - Part 1: Overview and General Principles. Geneva, Switzerland: ISO, 2021. Available: [official source page](#)
- [4] E. Tabassi, Artificial Intelligence Risk Management Framework (AI RMF 1.0), NIST AI 100-1. Gaithersburg, MD, USA: National Institute of Standards and Technology, 2023. DOI: [10.6028/NIST.AI.100-1](#)

- [5] R. Zhang, J. Wang, C. Liu, K. Su, H. Ishibuchi, and Y. Jin, 'Synergistic integration of metaheuristics and machine learning: latest advances and emerging trends,' *Artificial Intelligence Review*, vol. 58, art. 268, 2025. DOI: [10.1007/s10462-025-11266-y](https://doi.org/10.1007/s10462-025-11266-y)
- [6] C. Chen, H. Fu, Y. Zheng, F. Tao, and Y. Liu, 'The advance of digital twin for predictive maintenance: The role and function of machine learning,' *Journal of Manufacturing Systems*, vol. 71, pp. 581-594, 2023. DOI: [10.1016/j.jmsy.2023.10.010](https://doi.org/10.1016/j.jmsy.2023.10.010)
- [7] H. D. Shoorkand, M. Nourelfath, and A. Hajji, 'A hybrid deep learning approach to integrate predictive maintenance and production planning for multi-state systems,' *Journal of Manufacturing Systems*, vol. 74, pp. 397-410, 2024. DOI: [10.1016/j.jmsy.2024.04.005](https://doi.org/10.1016/j.jmsy.2024.04.005)
- [8] A. Billey and T. Wuest, 'Energy digital twins in smart manufacturing systems: A case study,' *Robotics and Computer-Integrated Manufacturing*, vol. 88, art. 102729, 2024. DOI: [10.1016/j.rcim.2024.102729](https://doi.org/10.1016/j.rcim.2024.102729)
- [9] B. Ha, H. Lee, and S. Hwangbo, 'Towards sustainable energy efficiency: Data-driven optimization in large-scale plants using machine learning applications,' *Energy*, vol. 331, art. 137059, 2025. DOI: [10.1016/j.energy.2025.137059](https://doi.org/10.1016/j.energy.2025.137059)
- [10] J. Yang, F. Wang, Y. Dun, et al., 'Prediction-based multi-objective optimization method for 3D printing resource consumption,' *The International Journal of Advanced Manufacturing Technology*, vol. 134, pp. 1805-1843, 2024. DOI: [10.1007/s00170-024-14143-0](https://doi.org/10.1007/s00170-024-14143-0)
- [11] A.-T. Nguyen, V.-H. Nguyen, T.-T. Le, and N.-T. Nguyen, 'A hybridization of machine learning and NSGA-II for multi-objective optimization of surface roughness and cutting force in AISI 4340 alloy steel turning,' *Journal of Machine Engineering*, vol. 23, no. 1, pp. 133-153, 2023. DOI: [10.36897/jme/160172](https://doi.org/10.36897/jme/160172)
- [12] V.-H. Nguyen, T.-T. Le, M.-V. Le, H.-D. Minh, and A.-T. Nguyen, 'Multi-objective optimization based on machine learning and non-dominated sorting genetic algorithm for surface roughness and tool wear in Ti6Al4V turning,' *Machining Science and Technology*, vol. 27, no. 4, pp. 380-421, 2023. DOI: [10.1080/10910344.2023.2235610](https://doi.org/10.1080/10910344.2023.2235610)
- [13] Y. Hu, L. Zhang, Z. Zhang, Z. Li, and Q. Tang, 'Matheuristic and learning-oriented multi-objective artificial bee colony algorithm for energy-aware flexible assembly job shop scheduling problem,' *Engineering Applications of Artificial Intelligence*, vol. 133, art. 108634, 2024. DOI: [10.1016/j.engappai.2024.108634](https://doi.org/10.1016/j.engappai.2024.108634)
- [14] X. Chang, X. Jia, and J. Ren, 'A reinforcement learning enhanced memetic algorithm for multi-objective flexible job shop scheduling toward Industry 5.0,' *International Journal of Production Research*, vol. 63, no. 1, pp. 119-147, 2025. DOI: [10.1080/00207543.2024.2357740](https://doi.org/10.1080/00207543.2024.2357740)
- [15] W. Zhang, Y. Zheng, and R. Ahmad, 'An energy-efficient multi-objective scheduling for flexible job-shop-type remanufacturing system,' *Journal of Manufacturing Systems*, vol. 66, pp. 211-232, 2023. DOI: [10.1016/j.jmsy.2022.12.008](https://doi.org/10.1016/j.jmsy.2022.12.008)
- [16] World Economic Forum, *Global Lighthouse Network: The Mindset Shifts Driving Impact and Scale in Digital Transformation*. Geneva, Switzerland: World Economic Forum, 2025. Available: [official source page](#)
- [17] Siemens, 'Siemens factory in Erlangen named Digital Lighthouse Factory,' Oct. 17, 2024. Available: [official source page](#)
- [18] Schneider Electric, 'World Economic Forum recognizes Schneider Electric Shanghai, China and Monterrey, Mexico factories as new Lighthouses,' Oct. 7, 2024. Available: [official source page](#)
- [19] Robert Bosch GmbH, 'The Bosch Shopfloor Agent: Artificial intelligence for competitive factories,' in *Bosch Annual Report 2025*, 2026. Available: [official source page](#)

- [20] BMW Group, 'How AI is revolutionising production,' Nov. 27, 2023. Available: [official source page](#)
- [21] Abdulgader Alsharif, Abdussalam Ali Ahmed, Omar Ahmed Mohamed, & Taha Muftah Abuali. (2026). Hybrid Machine Learning Approaches for Accurate Solar Energy Forecasting from Real-World Weather Data. Libyan Journal of Health, Science, and Development (LJHSD), 2(1), 09-17. <https://ljhsd.org.ly/index.php/ljhsd/article/view/10>
- [22] Optimizing Efficiency: A Comprehensive Overview of Lean Manufacturing Techniques and Their Impact on Industry. (2025). 27-18, 4(1), مجلة الباحث للعلوم التطبيقية. <https://albaitjas.com.ly/index.php/albahit/article/view/39>

Compliance with ethical standards*Disclosure of conflict of interest*

The authors declare that they have no conflict of interest.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of **DMAHS** and/or the editor(s). **DMAHS** and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content