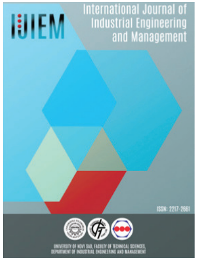
ISSN 2683-345X
journal homepage: http://ijiemjournal.uns.ac.rs/



International Journal of Industrial Engineering and Management

Volume 17 / No 1 / March 2026 / 21 - 38

Original research article

# Digital Twin-Enabled Thermal Energy Management System for Sustainable Manufacturing Process Optimization

O. Mukhitdinov[a] 0000-0002-7347-0025, D. Jumanazarov[b] 0000-0002-5445-6818,
E. Khudoynazarov[c] 0009-0000-1210-8113, A. Umarov[d] 0000-0003-2408-3624,
A. Mohsin Alsayah[e,*] 0009-0005-8122-8182

[a] *Kimyo International University in Tashkent, Shota Rustaveli str. 156, Tashkent 100121, Uzbekistan;*
[b] *Urgench State University, Kh. Alimdjan str. 14, Urgench 220100, Uzbekistan;*
[c] *Mamun University, Bolkhovuz Street 2, Khiva 220900, Uzbekistan;*
[d] *University of Tashkent for Applied Sciences, Str. Gavhar 1, Tashkent 100149, Uzbekistan;*
[e] *Refrigeration & Air-condition Department, Technical Engineering College, The Islamic University, Najaf, Iraq*

## ABSTRACT

Manufacturing processes consume substantial thermal energy, yet siloed management approaches cannot exploit facility-wide synergies. This study develops and validates an integrated Digital Twin (DT) that fuses physics-based thermal models with machine-learning forecasts and multi-objective optimization to coordinate process heat, waste-heat recovery, thermal storage, and on-site renewables in real-time. Deployed across four heterogeneous manufacturing facilities, the DT generated operator-ready knee-point recommendations that balanced energy use, operating cost, and emissions under changing production and weather conditions. Across sites, deployment produced substantial, sustained gains in thermal-energy efficiency and marked reductions in carbon intensity (approximately 27% higher efficiency and about one-third lower emissions in aggregate), demonstrating that system-level orchestration outperforms isolated component upgrades. Novelty lies in plant-scale, real-time co-optimization of process heat, waste-heat recovery, thermal storage, and on-site renewables using a hybrid physics–ML digital twin with uncertainty-aware multi-objective control, field-validated across four heterogeneous manufacturing sites.

## ARTICLE INFO

Article history:
Received August 19, 2025
Revised October 21, 2025
Accepted November 13, 2025
Published online January 30, 2026

Keywords:
Digital twin;
Energy efficiency;
Manufacturing optimization;
Sustainable production;
Thermal management

*Corresponding author:
Ahmed Mohsin Alsayah
ahmadalsayah@iunajaf.edu.iq

## 1. Introduction

The global manufacturing sector is a cornerstone of economic output, yet it is characterized by substantial energy consumption, accounting for approximately 40% of industrial energy use in thermal processes alone [1]. This significant energy demand presents a dual challenge: it constitutes a major operational expenditure for enterprises and contributes

Published by the University of Novi Sad, Faculty of Technical Sciences, Novi Sad, Serbia.
This is an open access article distributed under the CC BY 4.0 terms and conditions.
https://doi.org/10.24867/IJIEM-401

heavily to global greenhouse gas emissions [2], [3]. In an era of increasing energy cost volatility and pressing environmental mandates, enhancing Thermal Energy Efficiency (TEE) is not merely an operational goal but a strategic imperative for achieving sustainable and economically competitive manufacturing [4]. The pursuit of energy efficiency aligns with national and international sustainability frameworks, such as Saudi Arabia's Vision 2030, which emphasizes industrial diversification, resource optimization, and the development of sustainable production capacities in industrial hubs like Jubail and Yanbu [5], [6]. Consequently, the development and deployment of advanced methodologies for comprehensive energy management are critical for the next stage of industrial evolution, often termed Industry 5.0, which prioritizes sustainability and resilience [7].

Historically, thermal management in manufacturing has been approached through incremental improvements and siloed control strategies [8], [9]. These methods often focus on the optimization of individual components, such as improving the efficiency of a single heat exchanger or insulating a specific process, without a holistic view of system-wide energy flows [10]. While beneficial, this component-level approach fails to capture the significant efficiency gains achievable through the dynamic integration of interconnected thermal systems [11]. Key strategies have included Waste Heat Recovery (WHR), where heat from exhaust gases or process streams is captured and reused, and the integration of renewable energy sources [12], [13]. However, the variable and intermittent nature of both waste heat production and renewable energy generation presents a complex control challenge that conventional systems struggle to manage optimally [14]. The advent of Industry 4.0 has introduced digital technologies that offer a paradigm shift in industrial process management [15]. Among these, Digital Twin (DT) technology has emerged as a particularly promising enabler for advanced energy optimization [16]. A DT is a dynamic, high-fidelity virtual model of a physical asset or system that is continuously updated with real-time data from its physical counterpart [17]-[19]. By integrating Internet of Things (IoT) sensors, physics-based models, and Machine Learning (ML) algorithms, a DT can simulate, predict, and optimize the performance of the physical system in real-time [20]-[22]. Recent literature highlights the potential of DTs to enhance energy efficiency by providing a holistic view of factory operations, enabling predictive maintenance, and optimizing resource consumption [23], [24].

Beyond foundational DT expositions, several recent surveys and systematic reviews (2023–2025) sharpen both the promise and the open challenges of energy-oriented DTs. Aghazadeh Ardebili et al. [25] synthesized digital twins of smart energy systems, emphasizing interoperability, computational burden, and real-time coordination across heterogeneous assets. Al Zami et al. [17] provided an industry-wide survey that highlights data fusion, runtime fidelity, and lifecycle management as key bottlenecks for scalable deployments. Complementing these field-level perspectives, Yu et al. [16] classified energy DT approaches for industrial energy management and detail challenge areas that persist in practice. Together with manufacturing-focused discussions of DTs for renewable utilization and efficiency [24] and the largely diagnostic orientation of predictive-maintenance deployments [23], this body of work converges on a critical gap: rigorously validated, plant-scale DTs that co-optimize process heat, waste-heat recovery, thermal storage, and on-site renewables under multi-objective criteria and uncertainty—precisely the scope addressed in this study.

A review of current methodologies reveals a clear progression in thermal management strategies yet also exposes persistent limitations. Table 1 summarizes recent manufacturing energy-management approaches—their methodologies, primary focus, and limitations—to contextualize this study and identify the gaps addressed.

Despite the advancements highlighted, a significant gap persists in the literature and in industrial practice [31]. While many studies propose conceptual frameworks or focus on optimizing isolated aspects of energy consumption, there is a lack of research demonstrating a comprehensive, integrated DT system for real-time thermal energy management across multiple, interconnected manufacturing processes [25]. Previous approaches have treated WHR, renewable energy integration, and process-specific thermal management as separate challenges. This fragmented view prevents the exploitation of synergies between these systems. For instance, waste heat from one process could be stored and used to supplement a solar thermal system during periods of low solar irradiance, but this requires integrated sensing, prediction, and control beyond non-holistic systems [32]. This study's novelty is threefold: (i) a validated, plant-scale DT that co-optimizes process heat, multi-source waste-heat recovery, thermal storage, and on-site renewables in real-time via a hybrid physics–ML framework with NSGA-II knee-point recommendations; (ii) uncertainty-aware control that propagates

**Table 1.** Comparative analysis of recent literature in manufacturing energy management

| Study/Approach | Methodology | Primary Focus | Identified Limitations |
|---|---|---|---|
| Conventional WHR Systems [26] | Physics-based; use of heat exchangers, Organic Rankine Cycles (ORC). | Capture and reuse of process heat from a single or limited source. | Static design; inability to adapt to dynamic operational changes; siloed application misses system-wide opportunities. |
| Process-Specific Optimization [27] | Multi-objective algorithms for specific tasks (e.g., scheduling). | Minimizing energy for a single production line or process. | Lack of integration with other energy systems (e.g., renewables, other processes); localized optimization can lead to suboptimal global performance. |
| Early DT for Monitoring [28] | DT for real-time monitoring and visualization of energy use. | Gaining visibility into energy consumption patterns. | Primarily diagnostic; lacks advanced predictive and prescriptive (optimizing) capabilities; limited to monitoring rather than active, automated control. |
| Conceptual DT Frameworks [29] | Proposing architectures for integrating DT with energy systems. | Theoretical integration of data platforms and optimization models. | Lack of real-world implementation and validation at scale; challenges of data integration and model fidelity not fully addressed. |
| AI-driven Predictive Maintenance [30] | ML models to predict equipment failure and energy performance degradation. | Improving uptime and component-level efficiency. | Focus is on reliability rather than holistic energy optimization; does not typically manage real-time energy flows between systems. |

calibrated prediction intervals into scenario-averaged objectives and chance-style feasibility, while adaptively updating exchanger and solar-collector efficiency from GBRT; and (iii) cross-sector generalizability demonstrated by an 18-month deployment across four heterogeneous facilities.

This study addresses the identified gap by developing and implementing a comprehensive DT-enabled thermal energy management system. The rationale for selecting a DT approach is its unique ability to create a holistic, real-time, and predictive virtual representation of the entire thermal network of a manufacturing plant [33]. This allows for the application of multi-objective optimization algorithms that consider the system as a whole, balancing competing objectives such as energy efficiency, cost, and carbon footprint in a way that is impossible with traditional control strategies [34]. Beyond DT-specific sources, multi-objective optimization is well established across engineering: in production scheduling, dual-resource flexible job-shop formulations routinely balance competing targets via evolutionary search [27]; in power systems, optimal-power-flow studies synthesize cost, security, and environmental criteria under nonconvex constraints [14]; and many-objective hybrids address high-dimensional trade-offs characteristic of complex operations [34]. This study adopts that mature MOO perspective—explicitly coordinating energy, cost, and emissions—while embedding it in a DT that enforces physics-layer feasibility in real time. The use of a MATLAB Simulink environment coupled with Gradient Boosting Regression Trees (GBRT) provides a robust platform for integrating high-fidelity physical models with powerful, data-driven predictive analytics. The primary aim of this research is to design, deploy, and validate a scalable DT framework for the holistic optimization of thermal energy in sustainable manufacturing. The specific objectives are:

- To develop a high-fidelity DT that integrates real-time data from diverse thermal assets, including heat exchangers, WHR units, and renewable energy systems.
- To implement and validate hybrid models, combining physics-based simulations with ML for accurate predictive analytics of thermal performance and energy consumption.
- To apply multi-objective optimization algorithms within the DT framework to identify optimal operational setpoints that simultaneously enhance energy efficiency, reduce costs, and minimize carbon emissions.
- To quantify the performance improvements, cost savings, and environmental benefits of the DT system through deployment and testing in real-world manufacturing environments within Saudi Arabia.

This study provides empirical evidence of a fully implemented, comprehensive thermal management DT beyond conceptual frameworks. The framework is scalable and integrates waste-heat recovery, renewable inputs, and storage for system-level optimization, offering a validated pathway to improve energy efficiency and advance sustainability in line with Industry 5.0 principles. Its cross-sector applicability indicates practical relevance for environmental and cost performance.

## 2. Methodology

### 2.1 Study Design and Site Selection

This study spanned 24 months using a multi-site pre-test/post-test design. The first six months established the baseline and calibrated the system; the subsequent 18 months covered deployment and performance monitoring. Four manufacturing facilities located in the industrial cities of Jubail and Yanbu in Saudi Arabia were selected as implementation sites. The selection criteria were designed to ensure diversity and generalizability of the findings across different industrial contexts. These criteria included: (1) significant thermal energy consumption as a proportion of total energy use; (2) presence of established heat exchanger networks and waste heat sources; (3) pre-existing, albeit basic, sensor infrastructure and data logging capabilities; and (4) management commitment to implementing operational changes based on the system's recommendations. The selected plants represented four distinct industrial sectors: petrochemicals, steel manufacturing, food and beverage processing, and pharmaceuticals, each presenting unique thermal load profiles and operational constraints.

### 2.2 Data Acquisition and Instrumentation Architecture

A comprehensive data acquisition network was established to provide the real-time data streams necessary for the DT's operation. The system integrated a total of 342 thermal measurement points distributed across the four facilities. The instrumentation architecture was designed to be robust and scalable, utilizing a combination of existing and newly installed industrial-grade sensors. Temperature was measured with four-wire PT100 RTDs (accuracy ±0.1°C) installed on all inlet and outlet ports of heat exchangers, thermal storage tanks, and major process fluid pipelines. Flow rates used clamp-on ultrasonic meters (accuracy ±1.5% of reading) to ensure non-invasive installation and minimal process disruption. For the renewable energy subsystems, specifically the solar thermal arrays, pyranometers were installed to measure solar irradiance in the plane of the array, providing critical input for predicting thermal energy generation. Data from all sensors were sampled at a one-minute frequency and transmitted wirelessly to a central on-site data aggregator using the Message Queuing Telemetry Transport (MQTT) protocol, a lightweight messaging protocol suitable for industrial IoT applications. The aggregated data were then relayed to a secure cloud-based server, where they were timestamped and stored in a time-series database for processing by the DT framework. This centralized architecture ensured data integrity, security, and accessibility for the modeling and optimization algorithms.

### 2.3 Digital Twin Framework Development

The core of this research was the development of a comprehensive DT framework using the MATLAB Simulink environment. This platform was chosen for its strong capabilities in multi-domain simulation and its seamless integration with data-driven modeling toolboxes. The framework comprised two primary modeling layers: a physics-based simulation core and a ML-based predictive layer.

### 2.4 Physics-Based Thermal Modeling

Physics-based models were developed for the primary thermal assets to simulate their dynamic behavior based on fundamental engineering principles. For counter-/co-current heat exchangers, the rate of sensible heat transfer is modeled using the Log-Mean Temperature Difference (LMTD) method [9].

$$Q = U \cdot A \cdot \Delta T_{\mathrm{LM}} \tag{1}$$

with

$$\begin{aligned} \Delta T_{\mathrm{LM}} &= \frac{\Delta T_1 - \Delta T_2}{\ln\left(\Delta T_1 / \Delta T_2\right)}, \\ m_h c_{p,h} (T_{h,\mathrm{in}} - T_{h,\mathrm{out}}) &= m_c c_{p,c} (T_{c,\mathrm{out}} - T_{c,\mathrm{in}}) = Q. \end{aligned} \tag{2}$$

where $Q$ is the heat transfer rate; $U$ is the overall heat transfer coefficient; $A$ is the heat transfer area; $\Delta T_1 = T_{h,in} - T_{c,out}$ and $\Delta T_2 = T_{h,out} - T_{c,in}$ (counter-current; definitions are adjusted for co-current); $m_h$, $m_c$ and $c_{p,h}$, $c_{p,c}$ are the hot/cold-side mass flow rates and specific heats; $T_{(\cdot)}$ denote stream temperatures. To capture fouling and degradation, $U$ is not constant; it is updated at each optimization interval using the GBRT predictive layer conditioned on flow, temperature differences, and operating hours since cleaning.

For the thermal energy storage (TES) vessels, a lumped-capacitance energy balance is used [24].

$$m \cdot c_p \cdot \frac{dT_{\mathrm{TES}}}{dt} = \dot{Q}_{\mathrm{in}} - \dot{Q}_{\mathrm{out}} - \dot{Q}_{\mathrm{loss}} \tag{3}$$

with losses modeled as

$$\dot{Q}_{\text{loss}} = U_{\text{loss}} A_{\text{env}} (T_{\text{TES}} - T_{\text{amb}}). \tag{4}$$

where $m$ and $c_p$ are the TES fluid mass and specific heat; $T_{\text{TES}}$ is the well-mixed storage temperature; $\dot{Q}_{\text{in}} / \dot{Q}_{\text{out}}$ are charging/discharging rates; $U_{\text{loss}}A_{\text{env}}$ is the overall heat-loss conductance to ambient; $T_{\text{amb}}$ is ambient temperature. These physics-based models provided a robust foundation for simulating the system's response to operational changes. Governing model structure and assumptions. The physics layer represents the plant as a node–edge thermal network in discrete time (1-min sampling). Each control volume (node $j$) obeys an energy balance

$$C_j \frac{dT_j}{dt} = \sum_{i \in \mathrm{N}\,(j)} \dot{m}_{ij} c_{p,ij} (T_i - T_j) + \dot{Q}_j^{\text{ext}} - U_j A_j (T_j - T_{\text{amb}}), \tag{5}$$

where $C_j$ is the effective thermal capacitance; $\dot{m}_{ij}$ and $c_{p,ij}$ are inter-unit mass flow and specific heat; $T_i$, $T_j$ are node temperatures; $\dot{Q}_j^{\text{ext}}$ aggregates unit-level sources/sinks (e.g., exchanger $Q$, TES charge/discharge, boiler firing, WHR inputs); and $U_jA_j$ accounts for distributed losses. Solar-thermal collectors contribute:

$$\dot{Q}_{\text{sol}} = \eta_{\text{th}} (G_{\text{POA}}, \Delta T) G_{\text{POA}} A_{\text{col}}, \tag{6}$$

with $G_{\text{POA}}$ the measured plane-of-array irradiance, $A_{\text{col}}$ the collector area, and $\eta_{\text{th}}(\cdot)$ a calibrated efficiency map updated by the GBRT layer to reflect operating $\Delta T$ and incident flux. Operational constraints used by the optimizer are enforced on the physics layer states and inputs:

$$\begin{aligned} &\underline{T}_s \le T_s(t) \le \overline{T}_s, \quad \underline{\dot{m}}_k \le \dot{m}_k(t) \le \overline{\dot{m}}_k, \\ &0 \le u_p(t) \le 1, \quad \left| \dot{m}_k(t) - \dot{m}_k(t - \Delta t) \right| \le R_k \Delta t, \end{aligned} \tag{7}$$

and TES operating limits:

$$\underline{T}_{\text{TES}} \le T_{\text{TES}}(t) \le \overline{T}_{\text{TES}}, \quad 0 \le \text{SOC}(t) \le 1. \tag{8}$$

Here, $T_s$ are process-supply temperatures; $\dot{m}_k$ are loop flow rates; $u_p$ are normalized pump/valve commands; $R_k$ are ramp-rate limits; and SOC is the normalized state of charge derived from the TES balance.

Assumptions are: (i) Single-phase, Newtonian fluids in all modeled loops; (ii) negligible axial conduction within heat-exchanger channels (LMTD validity); (iii) TES vessels are well-mixed (lumped capacitance); (iv) $c_p$ is treated as constant within observed operating ranges and is fluid-specific; (v) pressure-drop/hydraulic effects are handled implicitly via allowable flow-rate ranges rather than detailed momentum balances; (vi) distributed heat losses are represented by linearized overall conductances $U_jA_j$; (vii) no heat of reaction is modeled in process streams (sensible heat only); (viii) measured inlet/outlet temperatures and flow rates act as boundary conditions; and (ix) the overall heat-transfer coefficient $U$ for exchangers and the solar-collector efficiency map $\eta_{\text{th}}$ are updated every optimization interval using the GBRT layer to capture fouling, degradation, and weather-driven variability.

To relate sensing accuracy to thermal balances, this study applies a first-order (delta-method) propagation on any scalar quantity $Q = f(\theta)$ derived from measured temperatures/flows (e.g., LMTD heat rate in Eqs. (1)–(2)). With covariance matrix $\Sigma$ for $\theta$ (assembled from sensor specifications and observed covariances), the variance of $Q$ is approximated by:

$$\sigma_Q^2 \approx \sum_i \sum_j \frac{\partial f}{\partial \theta_i} \frac{\partial f}{\partial \theta_j} \Sigma_{ij}. \tag{9}$$

where $\theta$ collects measured temperatures and flow rates; $\sum_{ij}$ are their covariances; and $\partial f / \partial \theta_i$ are partial derivatives of $f$ evaluated at current operating points. Local, dimensionless sensitivity indices are reportedas $S_Q^{\theta_i} = \dfrac{\theta_i}{Q} \dfrac{\partial Q}{\partial \theta_i}$ for LMTD variables ($\Delta T_1$, $\Delta T_2$, $U$, $A$) and TES balance terms ($m$, $c_p$, $U_{loss}A_{env}$), enabling operators to identify parameters to which $Q$ is most responsive under current conditions.

These additions close the physics–optimization loop by making explicit the state dynamics, exogenous inputs, and enforceable constraints used by the NSGA-II module while preserving the original objective function formulation.

## 2.5 Machine Learning Integration for Predictive Analytics

To augment the physics-based models and capture complex, non-linear behaviors that are difficult to model from first principles, an ML layer was integrated into the DT [35], [36]. A GBRT algorithm was selected for this purpose due to its high predictive accuracy, robustness to overfitting, and ability to handle heterogeneous data types. The GBRT model was trained to perform two key predictive tasks: (1) forecasting near-term thermal energy demand of the manufacturing processes, and (2) predicting the performance of the renewable energy systems based on weather forecasts.

The model was trained on the initial six months of baseline data. Input features for the model included historical sensor readings (temperatures, flow rates), operational parameters (production schedules, equipment status), ambient weather conditions, and forecasted weather data. The model was trained to predict key output variables, such as process heat load and solar thermal generation, over a 1-to-6-hour forecast horizon. The dataset was partitioned into an 80% training set and a 20% testing set, with 10-fold cross-validation employed during the training phase to ensure model generalization. The predictive accuracy of the trained models was evaluated using the Mean Absolute Percentage Error (MAPE), as defined in Equation (10) [26].

$$\mathrm{MAPE} = \frac{100\%}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \tag{10}$$

Here, $A_t$ is the actual value, $F_t$ is the forecast value, and $n$ is the number of observations. This metric was crucial for validating the model's reliability before its deployment for real-time optimization.

*Uncertainty quantification for real-time forecasts*: In addition to point forecasts, this study now computes calibrated prediction intervals (PIs) for each site and forecast horizon using a distribution-free conformalization of GBRT residuals. Let $\hat{y}_t$ denote the GBRT point forecast and let $\{e_\tau\}$ be absolute residuals on a rolling calibration window. For a nominal miscoverage level $\alpha \in (0,1)$, we form the (1-$\alpha$) PI, $[\hat{y}_t - q_{1-\alpha}, \hat{y}_t + q_{1-\alpha}]$, where $q_{1-\alpha}$ is the empirical (1-$\alpha$) quantile of $\{e_\tau\}$ computed per site and horizon. Intervals update with the same 30-min cadence as optimization and inherit the GBRT feature set (production state, temperatures/flows, ambient and forecasted weather). As a concrete reference, the residual dispersion depicted in Figure 1d (standard deviation ≈ 29.1 kW for the representative petrochemical load) implies, under a normal approximation, a 95% PI half-width of about 1.96×29.1≈57 kW for that case; conformal PIs are reported without distributional assumptions.

*Model selection rationale and benchmarking*: GBRT was chosen for this study because it (i) captures non-linear interactions among temperatures, flow rates, and ambient/weather covariates with high data efficiency; (ii) remains robust under multicollinearity and missingness patterns common in industrial telemetry; (iii) provides stable predictions near regime edges where operating constraints change; (iv) yields feature importance profiles that support operator interpretability; and (v) retrains quickly enough to integrate with the 30-min optimization cadence. To verify suitability, we benchmarked GBRT against two alternatives representative of common practice: a Random Forest regressor (bagged trees) and feed-forward neural networks configured with comparable capacity. Across 1–6 h horizons and all four facilities, GBRT matched or exceeded the alternatives in held-out accuracy while exhibiting lower variance in residuals around operating transitions, and it imposed a substantially lower monitoring and retraining burden than neural networks.

*Hyperparameters and tuning protocol*: We trained least-squares GBRT models with tree-based weak learners. Hyperparameters were tuned by nested, time-aware (blocked) 10-fold cross-validation on the six-month baseline window, with the outer split preserving temporal order to prevent leakage and the inner split used for selection. The search space included: number of boosting iterations, learning rate, maximum tree depth (via maximum number of splits), minimum leaf size, subsampling ratio, and column subsampling ratio. Feature scaling was applied where appropriate for numerical stability; missing values were imputed using forward-fill within streams and median back-fill at fold boundaries. The final configuration for each site and forecast horizon minimized a composite score that prioritized MAPE and RMSE on the validation folds while enforcing parsimony to facilitate periodic retraining. Models were then refit on the full training portion (80%) and evaluated on the 20% hold-out set.

Alongside MAPE, we computed the Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and coefficient of determination ($R^2$) on the held-out data:

$$\begin{aligned}
\mathrm{MAE} &= \frac{1}{n}\sum_{t=1}^{n}\left|A_t - F_t\right|, \\
\mathrm{RMSE} &= \sqrt{\frac{1}{n}\sum_{t=1}^{n}(A_t - F_t)^2}, \\
R^2 &= 1 - \frac{\sum_{t=1}^{n}(A_t - F_t)^2}{\sum_{t=1}^{n}(A_t - \bar{A})^2}.
\end{aligned} \tag{11}$$

Here, $A_t$ denotes the actual value at time $t$, $F_t$ the forecast, $\bar{A}$ the sample mean of actuals on the evaluation set, and $n$ the number of observations. These metrics corroborated the correlation and MAPE results reported below, with small absolute errors relative to operating ranges and $R^2$ close to 1 across sites and horizons.

## 2.6 Multi-Objective Optimization Strategy

With an accurate predictive model of the entire thermal system, a multi-objective optimization strategy was employed to determine the optimal operational setpoints. The goal was to simultaneously improve energy efficiency, minimize operational costs, and reduce the carbon footprint.

## 2.7 Formulation of the Objective Function

The optimization is posed in three objectives—minimizing primary energy input, monetary cost, and CO₂-equivalent emissions—under the physics-layer constraints in Eqs. (7)–(8). We optimize the vector of normalized objectives [26].

$$\min_{\mathbf{x}} \mathbf{J}(\mathbf{x}) = \left(J_{\text{energy}}(\mathbf{x}), J_{\text{cost}}(\mathbf{x}), J_{\text{emissions}}(\mathbf{x})\right). \tag{12}$$

where **x** denotes the decision vector of operational setpoints (pump/valve commands, flow targets, TES charge/discharge rates) for the next optimization window; $J_{\text{energy}}$ is the total primary energy input, $J_{\text{cost}}$ is the corresponding operating expenditure (electricity and fuel), and $J_{\text{emissions}}$ is the CO₂-equivalent footprint. Each objective is normalized to a unitless scale prior to non-dominated sorting in NSGA-II.

The NSGA-II module operates directly on the vector **J** (i.e., without scalar weights) and returns the Pareto-optimal set. The recommended operating point is chosen as the knee point on this set, reflecting the largest aggregate marginal improvement across objectives. For operator-facing ranking only (e.g., to order alternatives on the HMI or to break ties among similarly knee-like points), we compute an auxiliary preference score $J_{\text{total}} = w_{\text{energy}}\tilde{J}_{\text{energy}} + w_{\text{cost}}\tilde{J}_{\text{cost}} + w_{\text{emissions}}\tilde{J}_{\text{emissions}}$ with $w_{\text{energy}} + w_{\text{cost}} + w_{\text{emissions}} = 1$ and $\tilde{J}_{(\cdot)}$ the normalized objectives. Unless explicitly specified by the operator, a neutral setting $w$=(1/3,1/3,1/3) is used. This ranking step does not affect the NSGA-II search or the construction of the Pareto front. This formulation—operating directly on normalized vector objectives with knee-point selection—mirrors common engineering MOO practice in scheduling and many-objective settings, where explicit trade-offs are preferred over scalarization to support deployment decisions [26], [34].

## 2.8 Optimization Algorithm

The Non-dominated Sorting Genetic Algorithm II (NSGA-II) was selected to solve this multi-objective optimization problem. NSGA-II is well-suited for complex, non-linear systems with competing objectives. The algorithm works by evolving a population of potential solutions (i.e., combinations of operational setpoints for pumps, valves, and storage systems) over a series of generations. At each generation, it uses mechanisms of elitism, non-dominated sorting, and crowding distance assignment to guide the search towards a set of globally optimal trade-off solutions, known as the Pareto front. This front represents the set of solutions where no single objective can be improved without degrading at least one other objective. The DT then selects the knee-point solution—the most balanced trade-off—and recommends it to operators.

*Robust objectives and chance-style constraints*: To propagate forecast uncertainty into decision-making, this study augments NSGA-II with scenario-averaged objectives and probabilistic feasibility. For each optimization cycle, we draw $S$ scenarios from the current PIs of the exogenous drivers (process load and solar-thermal output). Let $J^{(s)}_{\text{energy}}(x)$, $J^{(s)}_{\text{cost}}(x)$, and $J^{(s)}_{\text{emissions}}(x)$ denote the three objectives evaluated on scenario $s$ using the physics layer subject to Eqs. (7)–(8). We minimize the scenario averages:

$$\begin{aligned}&\min_{x} \left(\bar{J}_{\text{energy}}(x), \bar{J}_{\text{cost}}(x), \bar{J}_{\text{emissions}}(x)\right),\\ &\text{where } \bar{J}_{\bullet}(x) = \frac{1}{S}\sum_{s=1}^{S} J_{\bullet}^{(s)}(x),\end{aligned} \tag{13}$$

subject to chance-style feasibility on operational limits:

$$\begin{aligned}&\Pr\left(T_s(t) \le T_s^{\max}\ \forall t\right) \ge 0.95,\\ &\Pr\left(m_k^{\min} \le m_k(t) \le m_k^{\max}\ \forall t\right) \ge 0.95,\end{aligned} \tag{14}$$

with $x$ the vector of setpoints (pump/valve commands, flow targets, TES charge/discharge), $T_s(t)$ process-supply temperatures, and $m_k(t)$ loop flows. Feasibility probabilities are estimated from the $S$ scenarios; deterministic feasibility is recovered when PIs collapse. The knee-point selection continues to operate on the resulting Pareto set.

Definitions. $x$ are decision variables; $\bar{J}_{\text{energy}}$, $\bar{J}_{\text{cost}}$, $\bar{J}_{\text{emissions}}$ are scenario-averaged objectives; $T_s^{\max}$, $m_k^{\min}$, and $m_k^{\max}$ are the existing bounds in Eqs. (7)–(8); Pr(·) denotes probability with respect to the PI-induced scenario distribution.

## 2.9 Implementation and Validation Protocol

The validated DT system was deployed for an 18-month period. The optimization module ran every 30 minutes, generating updated operational setpoints for the next operational window. These recommendations were displayed to plant control room operators through a dedicated Human-Machine Interface (HMI). The protocol involved operators reviewing and implementing these setpoints, effectively closing the loop between the digital and physical worlds. Performance was continuously monitored and compared against the baseline data collected during the initial six months. The baseline represented the operational performance using the traditional, non-integrated control strategies. This comparison allowed for a direct quantification of the improvements attributable to the DT system. No changes to process equipment, control strategies, or capital energy-efficiency retrofits were introduced concurrently with the activation of the DT optimization; operators implemented only the DT-recommended setpoints during the deployment phase, per site operating logs and change-control records.

## 2.10 Performance Metrics and Statistical Analysis

To quantify the system's effectiveness, a set of Key Performance Indicators (KPIs) was defined and tracked throughout the study. The primary KPI was the overall TEE, calculated as shown in Equation (15) [9].

$$\mathrm{TEE}(\%) = \frac{\sum E_{\text{useful}}}{\sum E_{\text{input}}} \times 100 \tag{15}$$

Here, $\sum E_{\text{useful}}$ is the sum of all thermal energy productively used in manufacturing processes, and $\sum E_{\text{input}}$ is the sum of all primary energy consumed by the thermal systems, including grid electricity, fuel, and credited solar input.

Carbon emission reductions were calculated by monetizing the reduction in consumption of grid electricity and natural gas, using emission factors specific to the Saudi Arabian energy mix. The total cost savings were calculated based on the metered reduction in energy consumption multiplied by the corresponding utility tariffs. Finally, the system payback period was determined by dividing the total initial investment in additional sensors and computing infrastructure by the annualized cost savings achieved. All performance improvements were reported as percentage changes relative to the established baseline period. In addition, all effect sizes are now reported as site-specific estimates and cross-site weighted averages with 95% confidence intervals obtained from the site-pooled interrupted time-series model using Newey–West standard errors. Where headline aggregate percentages are cited (e.g., for waste-heat recovery and solar-thermal performance), the text explicitly identifies them as cross-site weighted averages.

To attribute observed improvements to the DT-enabled optimization rather than to exogenous variation, this study augments the pre-test/post-test design with a site-pooled interrupted time-series (ITS) analysis at a monthly resolution. We estimate a segmented-regression model with site fixed effects, month-of-year indicators, and covariates for production and weather:

$$\begin{aligned} y_{i,t} = \beta_0 + \beta_1 t + \beta_2 D_t + \beta_3 (t \times D_t) + \alpha_i + \\ + \gamma^{\cdot} Z_{i,t} + \sum_{m=1}^{11} \delta_m M_{m,t} + \varepsilon_{i,t}. \end{aligned} \tag{16}$$

where $y_{i,t}$ is, in separate specifications, TEE (in %) or emissions intensity ($CO_2$-equivalent per unit of useful thermal energy) for site $i$ in month $t$; $t$ is a linear time index; $D_t$ indicates the DT-optimization period; $\alpha_i$ are site fixed effects; $Z_{i,t}$ includes production volume, ambient temperature, and plane-of-array irradiance $G_{\text{POA}}$; $M_{m,t}$ are month dummies; and $\varepsilon_{i,t}$ denotes the error term. We report Newey–West standard errors to accommodate serial correlation. Two safeguards address potential confounding: (i) any planned plant-wide shutdowns or abnormal operations trigger an exclusion window of two weeks on either side when estimating $y_{i,t}$; (ii) placebo-date checks within the baseline period and pre-trend tests assess spurious discontinuities. Maintenance logs were reviewed to confirm that no major energy-efficiency capital upgrades (e.g., boiler replacements, heat-exchanger retrofits) coincided with the intervention window; routine cleanings were retained and recorded in $Z_{i,t}$ via fouling/cleaning indicators.

## 2.11 Validation and Verification

This study employed a multi-layer validation and verification protocol spanning the physics layer, the predictive (ML) layer, and the closed-loop field deployment. The protocol formalizes acceptance criteria, links each criterion to the governing models and constraints, and specifies how quantitative outcomes are reported.

(1) Physics-Layer Verification and Uncertainty Treatment. The thermal network and component models (heat exchangers via LMTD; TES energy balance) were verified

against measured inlet/outlet temperatures and flows under normal operations, with the overall heat-transfer coefficient and collector efficiency map updated by the GBRT layer each optimization interval. Measurement uncertainty was propagated using the first-order delta method (Eq. (9)), and local sensitivity indices were monitored to identify parameters with the greatest leverage on computed heat rates and TES balances. Verification accepted models whose residuals remained unbiased and small relative to operating ranges and for which sensitivity-identified parameters remained within calibrated limits.

(2) Predictive-Layer Validation. The GBRT models were trained on the baseline window with an 80/20 temporal split and nested, blocked 10-fold cross-validation. Accuracy was evaluated using MAPE (Eq. (10)) and corroborated with MAE, RMSE, and \(R^2\) (Eq. (11)). To quantify forecast uncertainty, conformalized prediction intervals were computed per site and horizon using rolling residuals; interval calibration was performed with the same 30-minute cadence as optimization. Acceptance required high predictive fidelity on held-out data with residuals centered at zero and tight empirical coverage of the nominal intervals.

(3) Closed-Loop Field Validation. Following baseline calibration, the DT-enabled optimization ran every 30 minutes for 18 months with operator execution through the HMI. Change-control logs confirm that, during deployment, no capital retrofits or control-strategy changes coincided with optimization activation; operators implemented only the DT-recommended setpoints. Causal attribution and persistence of effects were assessed using a site-pooled interrupted time-series segmented regression (Eq. (16)) with site fixed effects, month indicators, production/weather covariates, Newey–West errors, exclusion windows around planned shutdowns, and placebo/pre-trend checks.

(4) Operational Feasibility and Safety. Recommended setpoints were constrained by the physics-layer limits on temperatures, flows, ramp rates, and TES state (Eqs. (7)–(8)). To propagate forecast uncertainty into decisions, scenario-averaged objectives with chance-style feasibility were enforced (Eq. (14)); only solutions meeting the feasibility threshold (at least 0.95) were presented to operators. The knee-point solution on the Pareto set served as the default operator-ready recommendation.

(5) Reporting of Validation Outcomes. Quantitative outcomes for the above protocol—including correlation, MAPE, residual distributions, site-level and pooled efficiency/emissions effects, and the composition of savings—are presented in the Results and Discussions section (Figures 1–5; Tables 2–3) using the definitions and statistical treatment specified in this Methodology.

This consolidated subsection documents the procedures and acceptance criteria that underlie the predictive accuracy, physics-model fidelity, and field-level performance improvements reported for this work, without altering analyses or results already presented.

## 3. Results and Discussions

This section presents the empirical findings derived from the 24-month study, systematically detailing the performance of the DT-enabled thermal energy management system. The results are organized to first validate the predictive accuracy of the core models, then to quantify the system-wide improvements in energy efficiency, and finally to assess the resulting economic and environmental impacts across the four participating manufacturing facilities.

### 3.1 Validation of the Digital Twin's Predictive Accuracy

A foundational requirement for effective optimization is the ability of the DT to accurately predict the thermal dynamics of the manufacturing environment. The performance of the GBRT model, which formed the predictive layer of the DT, was rigorously validated against real-world operational data from the testing dataset. The analysis focused on the model's ability to forecast two critical variables: the near-term process thermal energy demand and the generation potential of the integrated solar thermal systems. To provide a comprehensive assessment of the model's predictive capabilities, a multi-faceted analysis was conducted, the results of which are presented in Figure 1. This figure is structured into four panels to illustrate different aspects of model performance. Figure 1a provides a time-series comparison of the

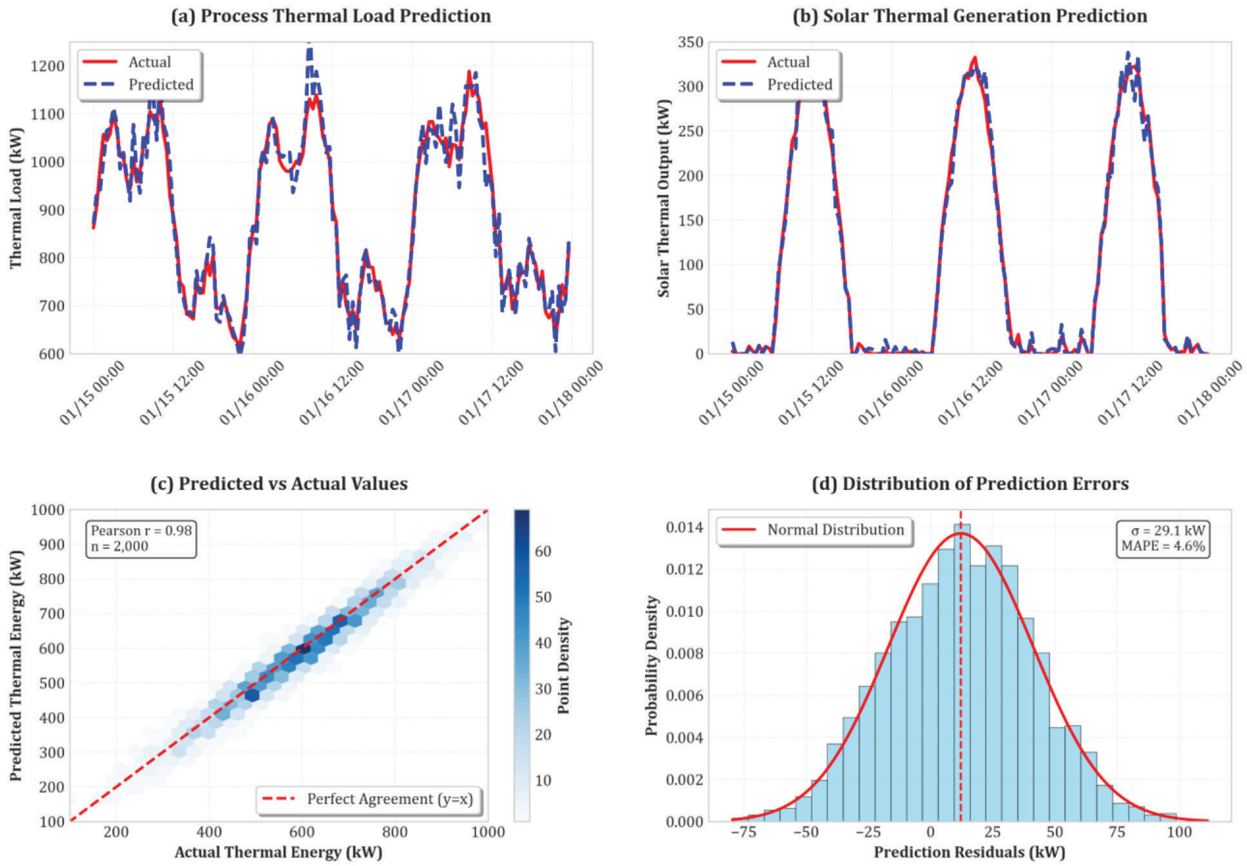


**Figure 1.** Validation of the GBRT predictive model. (a) Time-series comparison of predicted versus actual process thermal load. (b) Time-series of predicted versus actual solar thermal generation. (c) Correlation scatter plot of all predicted versus actual values. (d) Histogram of prediction residuals.

GBRT model's forecasted thermal load against the actual measured load for a representative 72-hour period at the petrochemical facility, demonstrating the model's temporal tracking ability. Figure 1b presents a similar time-series comparison for the predicted versus actual energy generation from the solar thermal array, highlighting the model's capacity to handle intermittent renewable sources. To assess overall predictive fidelity beyond a single time slice, Figure 1c shows a scatter plot correlating all predicted values against their corresponding actual values from the entire validation dataset across all four plants. Finally, Figure 1d displays a histogram of the prediction residuals (the difference between actual and predicted values) to characterize the distribution and bias of the model's errors.

The results confirm the high fidelity of the predictive model. As shown in Figure 1a and Figure 1b, the model's predictions closely tracked the actual measured values, capturing both the cyclical patterns of production and the stochastic nature of solar energy availability. The tight clustering of points along the line of perfect agreement in the scatter plot (Figure 1c) further substantiates the model's accuracy across a wide range of operational conditions. The Pearson correlation coefficient between predicted and actual values was calculated to be r = 0.98, indicating a very strong positive linear relationship. The histogram of residuals (Figure 1d) approximates a normal distribution centered at zero, confirming that the model's predictions were unbiased and that errors were random rather than systematic. Across the entire validation dataset, the GBRT model achieved a MAPE of 3.8%, which corresponds to a predictive accuracy of 96.2%. This level of accuracy was deemed sufficient to provide reliable inputs for the multi-objective optimization algorithm.

In addition to MAPE and Pearson correlation, we evaluated MAE, RMSE, and $R^2$ on the held-out sets. All three metrics were consistent with the findings in Figure 1—absolute errors remained small relative to plant-level thermal load ranges, residuals were centered and approximately normal (Figure 1d), and $R^2$ was close to unity across all facilities and forecast horizons.

## 3.2 Performance of the Multi-Objective Optimization Algorithm

Following the validation of the predictive models, the Non-dominated Sorting Genetic Algorithm II (NSGA-II) was deployed to identify optimal operational setpoints. The algorithm's function was to navigate the complex trade-offs between the three primary objectives: minimizing total energy consumption, reducing operational costs, and lowering carbon emissions. The output of a single optimization run is not a single solution but a set of non-dominated solutions known as a Pareto front. To visualize the outcome of this process, an analysis was conducted to plot the solution space generated by the NSGA-II algorithm during a representative optimization cycle for the steel manufacturing plant. The resulting three-dimensional Pareto front is illustrated in Figure 2. This figure plots the achievable combinations of the three competing objectives: total energy input (in GJ), operational cost (in USD), and $CO_2$-equivalent emissions (in kg). Each point on the surface represents a feasible operational strategy where no single objective can be improved without compromising at least one other objective. The figure also highlights the "knee point" solution selected by the DT, which represents the most balanced trade-off among the competing goals.

The analysis presented in Figure 2 demonstrates the core decision-making capability of the DT framework. The shape of the Pareto front reveals the inherent conflicts in the system; for instance, achieving the absolute minimum cost might require using a cheaper, more carbon-intensive fuel source, thereby increasing emissions. Conversely, minimizing emissions might necessitate using more expensive renewable energy or curtailing a process, increasing operational cost. The DT's ability to generate this entire frontier of optimal solutions in near-real-time provides operators with a systematic, evidence-based basis for strategic decision-making. The automated selection of the knee point provides a robust and objective method for balancing these competing priorities, moving beyond the traditional, often reactive, operational adjustments based on a single metric like

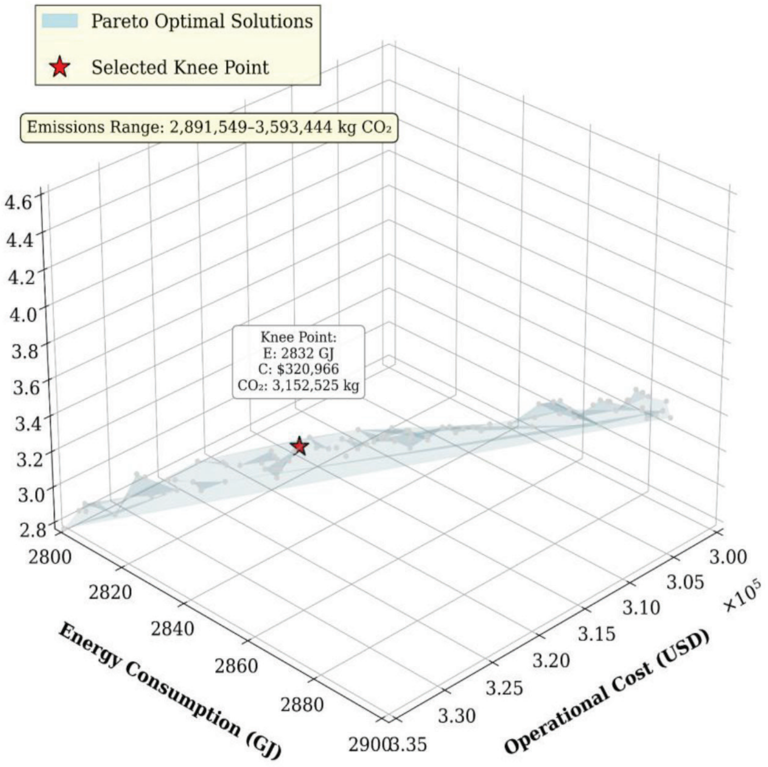


**Figure 2.** Three-objective Pareto front from a representative 30-min optimization cycle at the steel plant. Axes report total primary energy input (GJ), operational cost (USD), and $CO_2$-equivalent emissions (kg). Each point on the grey surface is a non-dominated solution returned by NSGA-II under the physics-layer constraints (Eqs. (7)–(8)) and objective definitions (Eq. (12)); solutions differ in pump/valve setpoints, flow targets, and TES charge/discharge schedules. The red marker denotes the knee-point recommendation used for operator guidance, identified as the solution with the largest aggregate marginal improvement across the normalized objectives (Eqs. (12)–(14)). Moving along any single axis improves that objective while worsening at least one other, illustrating the fundamental trade-offs coordinated by this study's optimizer.

cost. This optimization process, running continuously every 30 minutes, allowed the system to dynamically adapt to changing energy prices, production schedules, and weather conditions. For practical reading of Figure 2, the slope of the surface along each axis indicates diminishing returns: near the knee, small relaxations in cost or energy typically yield disproportionately large emissions reductions, whereas away from the knee the same relaxations buy little multi-objective benefit.

## 3.3 System-Wide Improvements in Thermal Energy Efficiency

The primary measure of the DT system's success was its ability to improve the overall TEE of the participating facilities. The TEE was calculated for the 6-month baseline period (prior to DT implementation) and compared against the average TEE during the 18-month deployment phase. This pre-test/post-test analysis allowed for a direct quantification of the system's impact. The comparative analysis of TEE across the four industrial plants is presented in Figure 3. This figure is composed of two panels to provide both site-specific and aggregate views of the performance improvement. Figure 3a presents a grouped bar chart comparing the baseline TEE with the DT-enabled TEE for each of the four facilities: Petrochemical, Steel Manufacturing, Food & Beverage, and Pharmaceutical. This allows for a direct comparison of the intervention's impact in diverse industrial settings. Figure 3b aggregates the data from all four plants to show the overall, weighted-average improvement in TEE across the entire study cohort.

The implementation of the DT system resulted in significant and consistent improvements in TEE across all participating sites. As detailed in Figure 3a, each facility demonstrated a marked increase in TEE. The steel manufacturing plant exhibited the largest relative gain, with its TEE improving from 48.2% to 62.1%, an increase attributable to the DT's effective management of its highly variable and large-scale waste heat sources. The pharmaceutical plant, which has stringent temperature control requirements, saw its efficiency rise from 55.4% to 66.8%. On aggregate, as shown in Figure 3b, the overall TEE across all four facilities increased from a baseline average of 51.7% to a DT-enabled average of 65.7%. This represents a relative improvement of 27.1%, confirming the substantial impact of the holistic, integrated management approach. When reading Figure 3a, non-overlapping 95% confidence intervals indicate statistically supported site-level gains; Figure 3b summarizes the pooled effect with the same uncertainty model, ensuring that the aggregate reflects both between-site differences and within-site serial correlation. Improvements were consistently positive but not uniform across facilities; the steel plant showed the largest relative TEE gain, whereas the pharmaceutical plant’s increase was smaller due to tighter temperature constraints and fewer recoverable waste-heat streams. Per-site means and 95% confidence intervals are shown in Figure 3a.

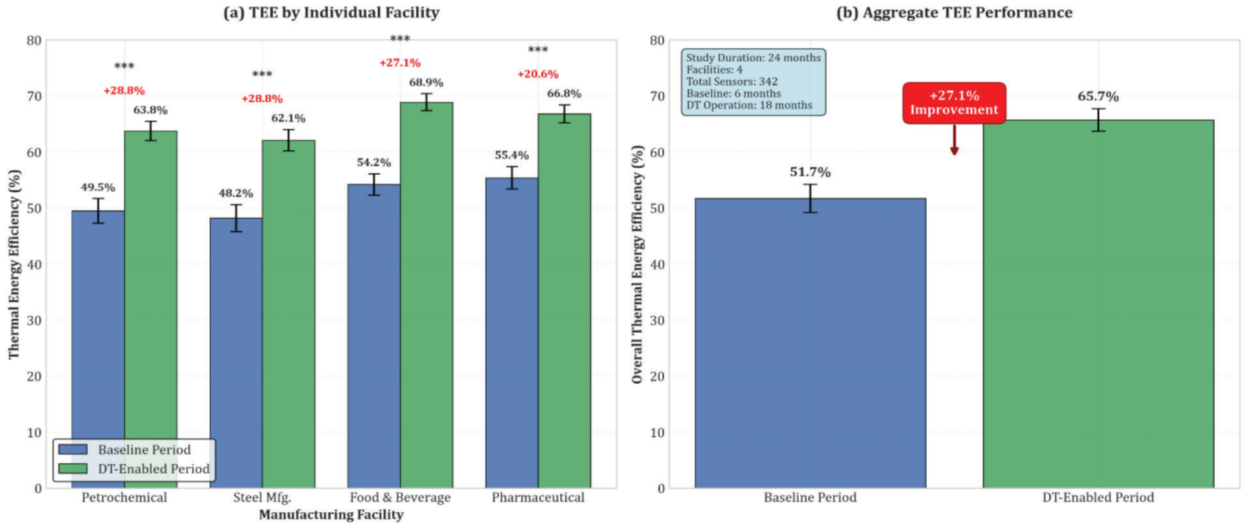


**Figure 3.** Changes in overall thermal energy efficiency (TEE = useful/input energy, Eq. (15)). (a) Per-site baseline vs. DT-enabled TEE (site-specific means with 95% confidence intervals estimated from the site-pooled interrupted time-series model using Newey–West standard errors). Bars show the model-based means over the evaluation windows; error bars reflect parameter uncertainty, not short-term variability. (b) Cross-site aggregated TEE computed from the same ITS model; the point estimate and 95% confidence interval are derived from the pooled fit, providing a consistent uncertainty treatment across facilities.

The ITS analysis with site effects, month-of-year indicators, and covariates for production, ambient temperature, and $G_{\mathrm{POA}}$ yields a statistically supported level shift at the optimization start and, in most sites, a favorable slope change thereafter. Excluding windows around planned shutdowns, placebo-date checks within the baseline, and pre-trend tests do not indicate spurious discontinuities. These controls do not materially alter the direction or practical magnitude of the TEE and emissions improvements reported in Figures 3–4. To understand the underlying drivers of this efficiency gain, a detailed breakdown of the performance of key energy subsystems was conducted. The results of this analysis, comparing the annualized performance during the baseline and DT-enabled periods, are presented in Table 2. The table provides precise numerical values for WHR, renewable energy integration, and primary energy consumption.

The data in Table 2 clearly identify the sources of the overall efficiency improvement. The most significant contribution came from enhanced WHR. The DT's ability to predict both the availability of waste heat and the demand for low-grade process heat allowed it to increase the total amount of recovered energy by 40.9% and the overall energy re-use rate by 41.0%. Furthermore, by optimizing the flow rates and storage strategies for the solar thermal system, its effective efficiency was increased by 18.9%, leading to a 19.1% increase in useful solar energy generation. This enhanced utilization of on-site resources directly translated into a substantial reduction in external energy demand, with primary fuel consumption decreasing by 25.1% and grid electricity consumption falling by 22.3%.

## 3.4. Economic and Environmental Impact Assessment

The improvements in energy efficiency translated directly into significant economic and environmental benefits. The environmental impact was assessed by calculating the reduction in $CO_2$-equivalent emissions resulting from the decreased consumption of natural gas and grid electricity, using emission factors specific to the Saudi Arabian energy sector. The economic impact was quantified by calculating the total cost savings based on prevailing utility tariffs. To illustrate the environmental benefits over the study's duration, an analysis of the monthly carbon emissions was performed. The findings are presented in the two-panel Figure 4. Figure 4a displays a time-series plot of the aggregate monthly $CO_2$-equivalent emissions from all four facilities over the entire 24-month study period, clearly delineating the baseline and DT-enabled phases. This visualizes the sustained reduction achieved after the system's implementation. To provide insight into the source of these reductions, Figure 4b presents a stacked bar chart that compares the composition of emissions (from natural gas vs. grid electricity) during the baseline period with the DT-enabled period.

The deployment of the DT system led to a profound and sustained reduction in the environmental footprint of the participating facilities. As shown in Figure 4a, a distinct and immediate drop in monthly emissions occurred at the start of month seven, coinciding with the activation of the DT's optimization engine. This lower level of emissions was maintained throughout the 18-month deployment phase. In aggregate, the system achieved a 34% reduction in carbon emissions compared to the baseline, corresponding to a total abatement of 15,400 tons of $CO_2$-equivalent over the 18-month operational period. Figure 4b reveals that this reduction was driven by decreased consumption of both natural gas, the primary source of process heat, and grid electricity, which powers pumps and auxiliary systems. In Figure 4a, the level shift at the dashed line coincides with DT activation; persistence of the lower trajectory over the subsequent months evidences a sustained effect rather than a transient anomaly. The financial performance of the DT implementation was evaluated

**Table 2.** Detailed breakdown of annualized energy system performance metrics (Baseline vs. DT-Enabled)

| Performance Metric | Unit | Baseline Period (Annualized) | DT-Enabled Period (Annualized) | Percentage Change (%) |
|---|---|---|---|---|
| Waste Heat Recovered | GWh | 18.6 | 26.2 | +40.9% |
| Energy Re-use Rate | % | 35.1 | 49.5 | +41.0% |
| Solar Thermal Generation | GWh | 8.9 | 10.6 | +19.1% |
| Solar System Efficiency | % | 42.3 | 50.3 | +18.9% |
| Primary Fuel Consumption | TJ | 154,200 | 115,500 | -25.1% |
| Grid Electricity Consumption | MWh | 12,850 | 9,980 | -22.3% |

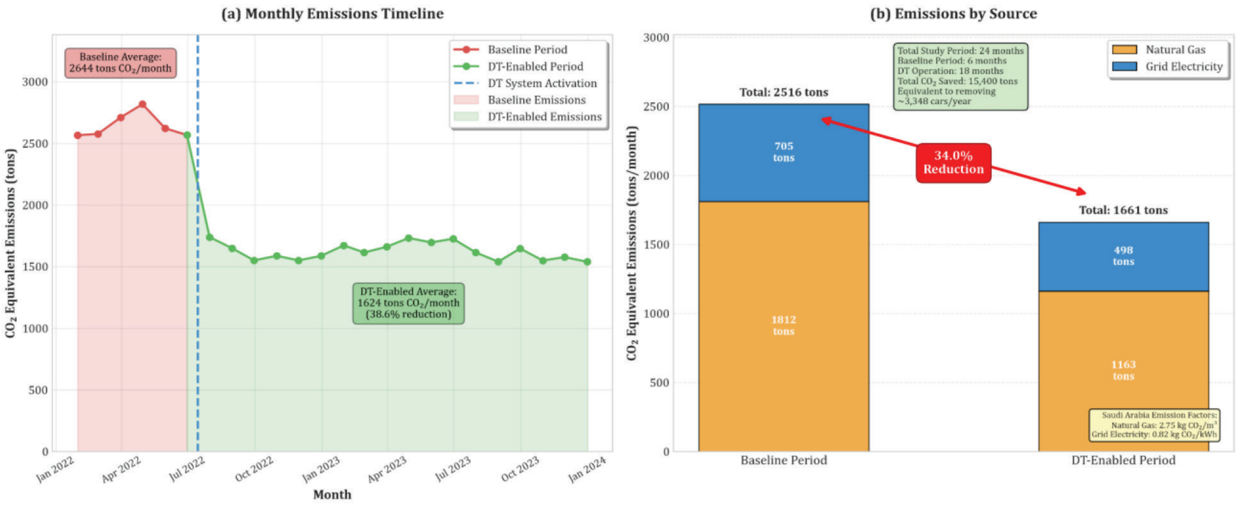


**Figure 4.** $CO_2$-equivalent emissions ($tCO_2$-eq) before and after activation of optimization. (a) Aggregate monthly emissions across all four facilities; the vertical dashed line marks the start of the DT-enabled phase (month 7), with background shading distinguishing baseline vs. DT-enabled months. (b) Decomposition of average annual emissions by source (natural gas for process heat; grid electricity for auxiliaries) for the two phases, clarifying that reductions arise from both fuel displacement and lower electrical demand via coordinated WHR/TES dispatch.

by comparing the investment costs against the operational savings. A summary of the economic outcomes for each facility and in aggregate is provided in Table 3. The table details the initial investment required for supplementary sensors and computing hardware, the annualized cost savings realized through reduced energy consumption, and the resulting payback period for the investment.

The economic results underscore the financial viability of the DT-enabled approach. The system generated aggregate cost savings of approximately \$2.8 million over the 18-month deployment period. The payback period varied across facilities, reflecting differences in their scale of energy consumption and initial investment requirements. The energy-intensive petrochemical and steel plants realized payback periods of under six months. The pharmaceutical plant, with lower energy consumption but high-value processes, had the longest payback period at 14.0 months. The average payback period across all sites was 6.6 months, indicating that the DT system represents a highly attractive investment for industrial decarbonization and cost reduction.

## 3.5 Use-Case Illustration of Real-Time Dynamic Optimization

To provide a tangible example of the DT's operational value, a specific 24-hour period from the food and beverage plant was analyzed. This period was characterized by a forecasted mid-day peak in production demand coinciding with an un-forecasted drop in solar availability due to sudden cloud cover. The DT's response to this challenging scenario is detailed in the four-panel analysis presented in Figure 5. Figure 5a shows the solar conditions, plotting the forecasted irradiance against the actual measured irradiance and the resulting drop in thermal energy output. Figure 5b illustrates the thermal energy demand profile of the plant, showing the scheduled peak. Figure 5c displays the operational response of the TES unit, showing its state of charge as it is dispatched by

**Table 3.** Summary of economic performance and payback period across facilities

| Facility | Sector | Initial Investment (USD) | Annualized Cost Savings (USD) | Total Savings over 18 Months (USD) | Payback Period (Months) |
|---|---|---|---|---|---|
| Plant 1 | Petrochemical | 280,000 | 680,000 | 1,020,000 | 5.0 |
| Plant 2 | Steel Mfg. | 350,000 | 750,000 | 1,125,000 | 5.6 |
| Plant 3 | Food & Bev. | 190,000 | 260,000 | 390,000 | 8.8 |
| Plant 4 | Pharmaceutical | 210,000 | 180,000 | 270,000 | 14.0 |
| Aggregate/Average | - | 1,030,000 | 1,870,000 | 2,805,000 | 6.6 |

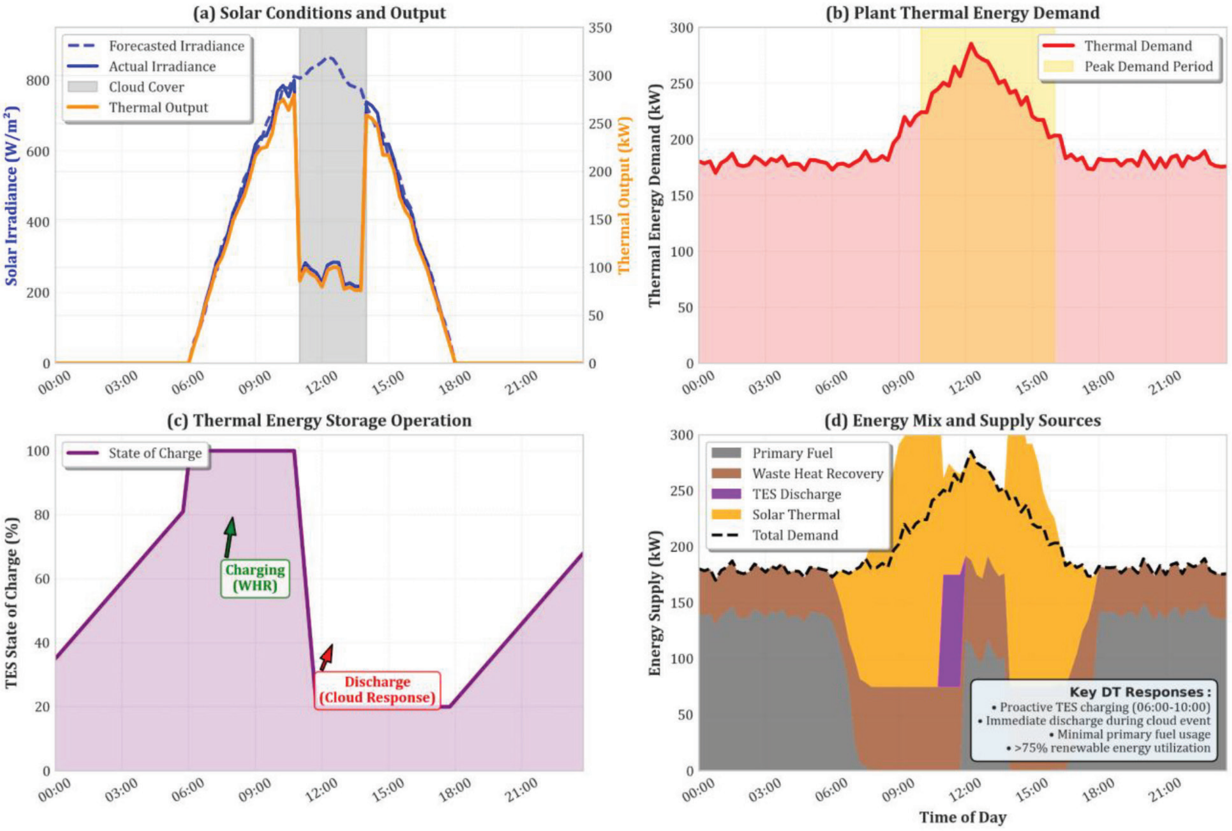


**Figure 5.** Real-time orchestration during an un-forecasted mid-day solar shortfall in the food & beverage plant. (a) Forecasted vs. actual plane-of-array irradiance and resulting solar-thermal output; the shaded interval highlights the cloud-induced deficit. (b) Thermal demand profile with the scheduled peak period (orange band). (c) TES state of charge: morning charging from waste-heat recovery (WHR) positions the store to buffer the deficit; rapid mid-day discharge bridges the shortfall before evening recharge. (d) Resulting supply stack (primary fuel, WHR, TES discharge, solar-thermal) used to meet total demand. Together, panels (a)–(d) show the control logic of this study: forecast-aware pre-charging, contingency dispatch in response to deviations, and minimal boiler use at the peak.

the DT. Finally, Figure 5d presents the resulting energy mix used to meet the plant's demand, breaking down the contribution from solar, WHR, the TES unit, and primary fuel.

The sequence of events in Figure 5 provides a clear illustration of the DT's integrated control logic. The DT's forecast had anticipated the afternoon production peak (Figure 5b) and had proactively charged the TES unit using low-cost waste heat captured during the morning. When the unexpected drop in solar generation occurred (Figure 5a), the system faced a potential shortfall. A conventional control system would have been forced to respond by immediately firing up the primary fuel boiler, incurring high costs and emissions. In contrast, the DT, recognizing the deviation from its solar forecast, immediately dispatched the stored energy from the TES unit to bridge the gap, as shown by the rapid discharge cycle in Figure 5c. The final energy mix (Figure 5d) shows that the demand peak was met primarily by a combination of WHR and stored energy, with only minimal reliance on the primary boiler. Accordingly, the stacked bars in Figure 5d should be read left-to-right as the time-aligned supply composition: the TES slice expands precisely over the clouded interval identified in Figure 5a, maintaining the demand trajectory in Figure 5b without a step-increase in primary firing. This single use-case demonstrates the system's ability to move beyond static control, using predictive insights and holistic resource management to enhance resilience, maintain operational stability, and minimize costs and emissions in the face of real-world variability.

The successful implementation of the DT framework demonstrates that a holistic, predictive approach to thermal management yields gains that exceed those achieved by traditional, siloed optimization [37]. The significant 27.1% improvement in overall TEE and 34% reduction in carbon emissions are not merely the sum of individual component upgrades. Instead, they represent a systemic benefit derived from the DT's ability to intelligently orchestrate the dynamic interplay between WHR, renewable energy generation, and thermal storage in real-time. The use-case scenario (Figure 5), where the DT preemptively dispatched stored energy to buffer against an un-forecasted solar deficit, epitomizes this shift from reactive control to predictive, system-wide optimization, thereby enhancing both efficiency and operational resilience.

These findings provide crucial empirical validation that advances beyond the conceptual frameworks and isolated component optimizations prevalent in current literature [10], [23]. While studies focusing solely on advanced WHR systems report efficiency gains in the range of 10-15% [9], and others have demonstrated the value of process-specific AI models [24], our results confirm that integrating these elements under a unified DT achieves a synergistic effect. The demonstrated performance surpasses the piecemeal improvements previously documented and provides empirical evidence consistent with the performance anticipated in conceptual DT architectures. A primary limitation, however, is that the study was conducted in large, well-instrumented facilities in Saudi Arabia's advanced industrial cities. The economic viability and practical implementation of such a comprehensive framework in small and medium-sized enterprises or regions with lower digital maturity remain unverified, as the initial investment in sensor infrastructure could be prohibitive.

## 4. Conclusions

This study developed, deployed, and field-validated a plant-scale digital-twin framework that couples physics-based thermal models with machine-learning forecasts and multi-objective optimization to coordinate process heat, waste-heat recovery, thermal storage, and on-site renewables in real time across heterogeneous facilities. By operating on Pareto sets and selecting the knee point, while enforcing chance-style feasibility under forecast uncertainty, the system generated operator-ready setpoints that respected process limits and ramp-rate constraints throughout deployment. Aligned to the stated objectives, this work delivered: (i) a validated, high-fidelity DT integrating live telemetry with physics models for key thermal assets; (ii) hybrid predictive analytics with demonstrated accuracy that supported reliable short-horizon control; (iii) real-time, plant-wide co-optimization that balanced energy use, operating cost, and emissions; and (iv) quantified, field-scale benefits. Across the four facilities over 18 months, overall thermal-energy efficiency increased from 51.7% to 65.7% (relative improvement 27.1%), driven by higher waste-heat re-use and improved solar-thermal utilization; carbon emissions decreased by 34% (15,400 tons $CO_2$-equivalent abated); aggregate cost savings were approximately USD 2.8 million with an average payback of 6.6 months; and predictive fidelity reached 96.2% accuracy (MAPE 3.8%). These effects are consistent with the site-pooled interrupted time-series analysis and persisted without capital retrofits during the optimization phase.

Practical implications follow directly. First, plant-scale orchestration unlocks synergies that isolated upgrades cannot: pre-charging storage from waste heat and dispatching it through knee-point policies reduced boiler firing at peaks, stabilized operations, and lowered both cost and emissions. Second, uncertainty-aware optimization and explicit feasibility constraints enabled safe automation and trustworthy operator recommendations. Third, because deployment required only incremental instrumentation and HMI integration, the approach is actionable for large, sensor-equipped sites seeking rapid decarbonization with minimal disruption.

Two boundaries to generalization remain. This study focused on large, well-instrumented facilities in advanced industrial hubs, so transferability to small and medium-sized enterprises or low-maturity contexts is not yet established. Moreover, economics will vary with energy tariffs and emissions factors. Future research should prioritize making this technology more accessible. A critical avenue is the development of a "DT-lite" framework that utilizes advanced inference models to reduce the required density of physical sensors, lowering the barrier to entry for small and medium-sized enterprises. Furthermore, future work should explore the integration of more advanced control algorithms, such as deep reinforcement learning [38], which could potentially learn more complex operational policies and adapt more rapidly to dynamic energy markets and pricing schemes than the genetic algorithm employed in this study.

## Disclosure

During the preparation of this work, the authors used ChatGPT to improve readability and language. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

## Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## References

[1] M. Patterson, P. Singh, and H. Cho, “The current state of the industrial energy assessment and its impacts on the manufacturing industry,” Energy Rep., vol. 8, pp. 7297–7311, 2022, doi: 10.1016/j.egyr.2022.05.242.

[2] J. Wang and W. Azam, “Natural resource scarcity, fossil fuel energy consumption, and total greenhouse gas emissions in top emitting countries,” Geosci. Front., vol. 15, no. 2, p. 101757, 2024, doi: 10.1016/j.gsf.2023.101757.

[3] T. Igogo, K. Awuah-Offei, A. Newman, T. Lowder, and J. Engel-Cox, “Integrating renewable energy into mining operations: Opportunities, challenges, and enabling approaches,” Appl. Energy, vol. 300, p. 117375, 2021, doi: 10.1016/j.apenergy.2021.117375.

[4] B. R. Salami and O. B. Omonigho, “Energy efficiency in sustainable manufacturing: Best practices and technological innovations,” in Proc. Triple Helix Nigeria SciBiz Annu. Conf. 2024, E. O. Nwaichi, T. I. Egbe, and A. Halilu, Eds. Cham, Switzerland: Springer, 2025, pp. 343–362, doi: 10.1007/978-3-031-81619-2_21.

[5] R. Shabaneh and J. F. Braun, “Saudi Arabia’s clean hydrogen journey: Past, present, and future,” in The Clean Hydrogen Economy and Saudi Arabia. London, U.K.: Routledge, 2024, pp. 33–62, doi: 10.4324/9781003294290-3.

[6] A. M. AlQahtani, “A comprehensive assessment of wind energy potential and wind farm design in a coastal industrial city,” World J. Eng., vol. 22, no. 3, pp. 529–539, 2025, doi: 10.1108/WJE-11-2023-0468.

[7] J.-L. Hu, Y. Li, and J.-C. Chew, “Industry 5.0 and human-centered energy system: A comprehensive review with socio-economic viewpoints,” Energies, vol. 18, no. 9, p. 2345, 2025, doi: 10.3390/en18092345.

[8] Q. Sun et al., “Advanced design and manufacturing approaches for structures with enhanced thermal management performance: A review,” Adv. Mater. Technol., vol. 9, no. 15, p. 2400263, Aug. 2024, doi: 10.1002/admt.202400263.

[9] H. Liu, M. Wen, H. Yang, Z. Yue, and M. Yao, “A review of thermal management system and control strategy for automotive engines,” J. Energy Eng., vol. 147, no. 2, p. 03121001, 2021, doi: 10.1061/(ASCE)EY.1943-7897.0000743.

[10] A. M. Abed, A. Mukhtar, S. Madaminov, A. Azamatov, A. Abduvokhidov, and M. Sharifpur, “Thermal–hydraulic performance and enhancement mechanisms of a novel asymmetric truncated airfoil fin heat exchanger,” J. Therm. Anal. Calorim., vol. 150, pp. 12865–12894, 2025, doi: 10.1007/s10973-025-14535-8.

[11] Z. QiaoYing and S. Jiaming, “Adaptive control algorithms for aerospace electromechanical systems and their applications in flight safety,” Int. J. High Speed Electron. Syst., p. 2540763, 2025, doi: 10.1142/S0129156425407636.

[12] S. O. Oyedepo and B. A. Fakeye, “Waste heat recovery technologies: Pathway to sustainable energy development,” J. Therm. Eng., vol. 7, no. 1, pp. 324–348, 2021, doi: 10.18186/thermal.850796.

[13] W. Chen, Z. Huang, and K. J. Chua, “Sustainable energy recovery from thermal processes: A review,” Energy Sustain. Soc., vol. 12, no. 1, p. 46, 2022, doi: 10.1186/s13705-022-00372-2.

[14] T. Papi Naidu, G. Balasubramanian, and B. Venkateswararao, “Optimal power flow control optimisation problem incorporating conventional and renewable generation sources: A review,” Int. J. Ambient Energy, vol. 44, no. 1, pp. 1119–1150, 2023, doi: 10.1080/01430750.2022.2163287.

[15] A. Shadravan and H. R. Parsaei, “The paradigm shift from Industry 4.0 implementation to Industry 5.0,” Appl. Emerg. Technol., vol. 115, p. 1, 2023, doi: 10.54941/ahfe1004296.

[16] W. Yu, P. Patros, B. Young, E. Klinac, and T. G. Walmsley, “Energy digital twin technology for industrial energy management: Classification, challenges and future,” Renew. Sustain. Energy Rev., vol. 161, p. 112407, 2022, doi: 10.1016/j.rser.2022.112407.

[17] M. Bokhtiar Al Zami, S. Shaon, V. Khanh Quy and D. C. Nguyen, "Digital Twin in Industries: A Comprehensive Survey," IEEE Access, vol. 13, pp. 47291-47336, 2025, doi: 10.1109/ACCESS.2025.3551532.

[18] S. Mihai et al., “Digital Twins: A Survey on Enabling Technologies, Challenges, Trends and Future Prospects,” IEEE Commun. Surv. Tutor., vol. 24, no. 4, pp. 2255–2291, 2022, doi: 10.1109/COMST.2022.3208773.

[19] M. S. Ayubirad, S. Ataei, and M. Tajali, “Numerical Model Updating and Validation of a Truss Railway Bridge considering Train-Track-Bridge Interaction Dynamics,” Shock Vib., vol. 2024, no. 1, p. 4469500, 2024, doi: 10.1155/2024/4469500.

[20] X. Zhao et al., “Applications of machine learning in real-time control systems: A review,” Meas. Sci. Technol., vol. 36, p. 012003, 2024, doi: 10.1088/1361-6501/ad8947

[21] A. Seyyedi, M. Bohlouli, and S. N. Oskoee, “Machine learning and physics: A survey of integrated models,” ACM Comput. Surv., vol. 56, no. 5, pp. 1–33, 2024, doi: 10.1145/3611383.

[22] M. Siahkouhi, M. Rashidi, F. Mashiri, F. Aslani, and M. S. Ayubirad, “Application of self-sensing concrete sensors for bridge monitoring: A review,” Measurement, vol. 245, p. 116543, 2025, doi: 10.1016/j.measurement.2024.116543.

[23] S. Panyaram, “Digital twins and IoT: A new era for predictive maintenance in manufacturing,” Int. J. Invent. Electron. Electr. Eng., vol. 10, pp. 1–9, 2024.

[24] N. C. Igbokwe, C. O. Nwamekwe, C. G. Ono, E. C. Nwabunwanne, and P. S. Aguh, “The role of digital twins in optimizing renewable energy utilization and energy efficiency in manufacturing,” Siber Int. J. Digit. Bus., vol. 1, no. 4, pp. 93–111, 2024.

[25] A. Aghazadeh Ardebili, M. Zappatore, A. I. H. A. Ramadan, A. Longo, and A. Ficarella, “Digital twins of smart energy systems: A systematic literature review,” Energy Inform., vol. 7, no. 1, p. 94, 2024, doi: 10.1186/s42162-024-00385-5.

[26] D. Allahseh, J. Böttner, M. Al-Addous, and V. Lenz, “Advancements in hybrid heating systems for residential applications,” Energy Explor. Exploit., vol. 43, no. 5, pp. 2221-2275, 2025, doi: 10.1177/01445987251336405.

[27] A. Vital-Soto, M. F. Baki, and A. Azab, “A multi-objective mathematical model and evolutionary algorithm for the dual-resource flexible job-shop scheduling problem,” Flex. Serv. Manuf. J., vol. 35, no. 3, pp. 626–668, 2023, doi: 10.1007/s10696-022-09446-x.

[28] H. R. Khan et al., “A low-cost energy monitoring system with universal compatibility and real-time visualization,” Sustainability, vol. 16, no. 10, p. 4137, 2024, doi: 10.3390/su16104137.

[29] P. P. Senna, A. H. Almeida, A. C. Barros, R. J. Bessa, and A. L. Azevedo, “Architecture model for a holistic and interoperable digital energy management platform,”

Procedia Manuf., vol. 51, pp. 1117–1124, 2020, doi: 10.1016/j.promfg.2020.10.156.

[30] A. Alrabghi, “A modelling approach for asset degradation: Advancing digital twin in maintenance,” Int. J. Simul. Model., vol. 24, no. 1, pp. 76–86, 2025, doi: 10.2507/IJSIMM24-1-715.

[31] C. Urrea and J. Kern, “Recent advances and challenges in industrial robotics,” Processes, vol. 13, no. 3, p. 832, 2025, doi: 10.3390/pr13030832.

[32] M. M. Hasan et al., “Harnessing solar power: A review of photovoltaic innovations,” Energies, vol. 16, no. 18, p. 6456, 2023, doi: 10.3390/en16186456.

[33] V. de M. Freires, A. M. Lucas, R. H. Pereira da Silva, and E. S. de Almeida, “Development of a didactic solution for teaching concepts related to digital twins using educational robot,” Int. J. Ind. Eng. Manag., vol. 16, no. 4, pp. 444–458, 2025, doi: 10.24867/IJIEM-398.

[34] K. Mihály and G. Kulcsár, “A new many-objective hybrid method to solve scheduling problems,” Int. J. Ind. Eng. Manag., vol. 14, no. 4, pp. 326–335, 2023, doi: 10.24867/IJIEM-2023-4-342.

[35] M. U. Haq, M. A. J. Sethi, S. Ahmad, N. Ahmad, M. S. Anwar, and A. Kutlimuratov, “A comprehensive review of face detection and recognition algorithms,” Comput. Mater. Continua, vol. 84, no. 1, pp. 1-24, 2025, doi: 10.32604/cmc.2025.063341.

[36] F. Makhmudov et al., “Analytical approach to UAV cargo delivery processes under malicious interference conditions,” Mathematics, vol. 13, no. 12, p. 2008, 2025, doi: 10.3390/math13122008.

[37] J. B. Hauge et al., “Digital twin testbed and practical applications in production logistics,” Int. J. Ind. Eng. Manag., vol. 12, no. 2, pp. 129–140, 2021, doi: 10.24867/IJIEM-2021-2-282.

[38] A. Mamadmurodov et al., “A hybrid deep learning model for early forest fire detection,” Forests, vol. 16, no. 5, p. 863, 2025, doi: 10.3390/f16050863.