Energy-Aware Hybrid Machine Learning on Edge Telemetry for Dynamic Resource Allocation and Carbon-Neutral Smart Factories

Authors

Ada John, Abilly Elly

Date; September 28, 2026

Abstract

The industrial sector's transition toward carbon neutrality requires intelligent systems capable of balancing energy efficiency with production performance in real time. This study addresses the gap in edge-deployable hybrid architectures that simultaneously optimize resource allocation and carbon footprint in smart manufacturing environments. We propose an Energy-Aware Hybrid Machine Learning (EA-HML) framework that integrates Temporal Convolutional Networks for energy demand forecasting, Random Forest classifiers for workload prioritization, and a Dueling Deep Q-Network controller for dynamic resource allocation on edge devices. The framework was validated on real-time telemetry data from discrete manufacturing operations, achieving 89.4% resource allocation accuracy while reducing energy consumption by 23.7% and operational costs by 19.2% compared to baseline methods. Statistical significance was confirmed ($p < 0.01$), with the hybrid architecture demonstrating superior convergence stability and edge deployment feasibility. These findings establish a replicable pathway for deploying energy-aware intelligence at the industrial edge, enabling factories to pursue carbon-neutral operations without sacrificing production throughput.

Keywords: Edge Computing, Hybrid Machine Learning, Energy-Aware Manufacturing, Dynamic Resource Allocation, Carbon-Neutral Factories

1. Introduction

1.1 Background

The global manufacturing sector accounts for approximately 54% of the world's total energy consumption and contributes significantly to industrial carbon emissions . As nations commit to climate neutrality targets, smart factories face mounting pressure to reconcile production efficiency with environmental sustainability. Industry 4.0 technologies—particularly Industrial Internet of Things (IIoT) sensors, edge computing, and machine learning—offer unprecedented opportunities for real-time energy optimization at the operational level .

Edge analytics has emerged as a critical enabler for latency-sensitive industrial applications where cloud-based processing introduces unacceptable delays . By moving computation closer to data sources, edge architectures enable millisecond-level decision-making for resource allocation, quality monitoring, and energy management. Recent advances in hybrid machine learning—combining forecasting models with reinforcement learning controllers—have demonstrated promising results for energy-efficient task offloading in edge environments . However, the integration of these techniques specifically for carbon-neutral factory operations on resource-constrained edge hardware remains underexplored.

1.2 Problem Statement

Existing approaches to industrial energy optimization typically employ either centralized cloud architectures that suffer from latency and bandwidth limitations, or single-model edge solutions that fail to capture the complex interplay between workload dynamics and energy consumption . Hybrid frameworks have been proposed for MEC environments , but these focus primarily on task offloading rather than holistic factory resource orchestration. Furthermore, most studies optimize for energy minimization alone, neglecting the carbon intensity dimension that varies with grid renewable penetration .

Recent work by Hasan et al. demonstrated edge analytics achieving 98.90% accuracy for process optimization, yet their approach did not incorporate dynamic energy-aware resource allocation or carbon-aware scheduling objectives. Similarly, Lyapunov-guided DRL approaches and Transformer-GNN frameworks have addressed MEC offloading but lack integration with production scheduling constraints and carbon accounting. No validated framework exists that simultaneously: (a) operates within edge computational constraints, (b) forecasts energy demand for proactive allocation, and (c) adapts resource scheduling to real-time carbon intensity signals.

1.3 Objectives of the Study

General objective: To design and validate an energy-aware hybrid machine learning framework for dynamic resource allocation in edge-enabled smart factories with carbon-neutral operational targets.

Specific objectives:

1. To identify key predictors of energy demand and carbon intensity in discrete manufacturing telemetry streams.
2. To design a hybrid architecture integrating TCN forecasting, RF classification, and Dueling DQN control for edge deployment.
3. To validate the framework's performance against baseline methods in terms of allocation accuracy, energy reduction, and carbon efficiency.

1.4 Research Questions

1. What combination of edge telemetry variables most accurately predicts short-term energy demand and carbon intensity in smart factory operations?
2. How does the proposed EA-HML framework compare to single-model and cloud-based approaches in resource allocation accuracy and energy savings?
3. What architectural and algorithmic considerations enable effective hybrid ML deployment on resource-constrained edge devices?

1.5 Significance of the Study

For practitioners: The framework provides a replicable deployment blueprint for factory operators seeking to implement edge-based energy intelligence without extensive cloud infrastructure investments.

For policymakers: Demonstrates that carbon-aware operational intelligence can be achieved at the edge, supporting distributed industrial decarbonization strategies.

For academic literature: Extends hybrid ML theory into the emerging domain of carbon-neutral manufacturing, establishing performance benchmarks for edge-deployable energy orchestration.

For future researchers: Provides a validated baseline and identifies open challenges in multi-objective optimization under edge constraints.

1.6 Scope and Limitations

This study focuses on discrete manufacturing environments with access to real-time telemetry from production equipment, environmental sensors, and grid carbon intensity signals. Validation employed a combination of historical operational data and simulated scenarios calibrated to realistic factory conditions. Key limitations include: (1) findings may not generalize to process industries with fundamentally different energy profiles, (2) carbon intensity forecasting assumes access to grid operator APIs with reasonable accuracy, and (3) edge hardware performance varies across deployment contexts.

2. Literature Review

2.1 Conceptual Review

Edge Telemetry in Smart Manufacturing. Edge telemetry refers to sensor data streams processed at or near the source of generation, enabling real-time analytics without cloud round-trips . In manufacturing contexts, telemetry encompasses equipment power consumption, operational states, environmental conditions, and production throughput metrics.

Hybrid Machine Learning. Hybrid ML architectures combine multiple model paradigms to leverage complementary strengths—typically forecasting models for anticipation, classifiers for decision-making, and reinforcement learning for sequential optimization .

Carbon-Neutral Factories. Facilities that achieve net-zero carbon emissions through a combination of renewable energy procurement, energy efficiency, and carbon offset strategies. Operational carbon neutrality requires real-time matching of energy consumption with renewable generation or low-carbon grid periods .

2.2 Theoretical Framework

The study integrates three theoretical foundations:

Lyapunov Optimization. This framework transforms long-term stochastic optimization into per-slot deterministic decisions, providing theoretical guarantees for queue stability and energy minimization . It guides the design of the resource allocation controller to ensure bounded energy consumption while maintaining production throughput.

Markov Decision Processes. The dynamic resource allocation problem is formulated as an MDP, enabling reinforcement learning solutions. State representation incorporates energy prices, carbon intensity, workload queues, and equipment states .

Prospect Theory. Applied to operator decision-making, prospect theory explains why factory managers may under-adopt energy optimization technologies despite positive expected value—loss aversion and status quo bias create adoption inertia .

2.3 Empirical Review

Hasan et al. proposed edge analytics using hybrid ML for smart manufacturing process optimization, achieving 98.90% accuracy compared to RF, LSTM, XGBoost, and KNN baselines. Their framework validated edge deployment feasibility but did not address energy-aware resource allocation or carbon objectives.

Li et al. introduced TCN-DQN integration for energy-intensive healthcare manufacturing, achieving MAPE of 1.83% for demand forecasting and 24.9% energy reduction. The study focused on HVAC and scheduling optimization without edge deployment constraints.

HyGEO-MEC combined Transformer-GNN with Dueling DQN for MEC task offloading, achieving 0.18-second average task delay and 84.27% decision alignment with optimal. The framework addressed latency and energy but not carbon intensity or production integration.

Yang et al. developed Lyapunov-guided multi-agent DRL for IIoT task offloading, demonstrating reduced energy consumption and queue stability. Their approach assumed cloud-edge collaboration rather than pure edge intelligence.

Wang et al. created Energy-Twin integrating CNN-GRU-BiLSTM with digital twin for industrial energy prediction, achieving 99.55% accuracy. The computational requirements likely exceed typical edge device capabilities.

2.4 Research Gap

No validated framework exists that integrates energy demand forecasting, carbon-aware resource allocation, and production scheduling within edge computational constraints for carbon-neutral factory operations. Existing hybrid approaches either optimize single objectives or require cloud resources incompatible with latency-sensitive industrial control. This study addresses this gap by designing and empirically validating an edge-deployable hybrid architecture that simultaneously pursues energy efficiency, production throughput, and carbon neutrality objectives.

3. Methodology

3.1 Research Design

This study employed a quantitative, design-based research approach combining retrospective telemetry analysis with prospective simulation validation. The design was selected because: (a) it enables rigorous performance quantification against baselines, (b) simulation allows controlled evaluation of carbon intensity scenarios not available in historical data, and (c) the dual-phase approach balances ecological validity with experimental control.

3.2 Study Area / Population

The study population comprises discrete manufacturing operations with the following characteristics: (1) continuous telemetry generation from production equipment, (2) hourly or sub-hourly energy consumption measurement, (3) access to grid carbon intensity signals, and (4) edge computing infrastructure. The sampling frame consisted of three manufacturing sites in the Asia-Pacific region representing electronics assembly, automotive components, and precision machining sectors.

3.3 Sample Size and Sampling Technique

Telemetry data were collected from 47 production machines across three facilities, yielding 2.3 million timestamped observations over a 12-month period (January–December 2025). Stratified sampling ensured representation across: shift patterns (day/night/weekend), production states (active/idle/maintenance), and seasonal variations in grid carbon intensity. The sample size provides sufficient statistical power (>0.90) for detecting medium effect sizes at $\alpha = 0.05$.

3.4 Data Collection Methods

Historical telemetry: Extracted from factory MES and SCADA systems, including: equipment power draw (kW), cycle counts, temperature and vibration sensors, production volume, and shift schedules.

Grid carbon intensity: Hourly marginal carbon intensity from regional grid operators, matched to factory location.

Simulated scenarios: Prospective carbon intensity profiles were generated using seasonal decomposition plus stochastic residual modeling, calibrated to historical grid patterns. Simulation enabled evaluation under future renewable penetration scenarios (30%, 50%, 70% renewable).

3.5 Research Instruments

Software stack: Python 3.11, PyTorch 2.1, scikit-learn 1.4, SimPy 4.1, NumPy, Pandas.

Preprocessing: Telemetry streams were resampled to 15-minute intervals, missing values imputed via forward-fill with mask indicators, and features normalized using robust scaling (median/IQR) to mitigate outlier effects.

Edge deployment target: NVIDIA Jetson Orin Nano (8GB), representing typical industrial edge hardware.

3.6 Validity and Reliability

Content validity: Feature selection was informed by domain expert consultation and prior literature on industrial energy analytics .

Predictive validity: Model performance was assessed through temporal cross-validation using held-out test periods.

Reliability: Three independent training runs with different random seeds assessed convergence stability (coefficient of variation < 5% for key metrics).

3.7 Data Analysis Techniques

Forecasting module: Temporal Convolutional Network (TCN) for 1–4 hour energy demand prediction. TCN was selected for its dilated causal convolutions enabling long-range dependency capture with linear computational complexity .

Classification module: Random Forest for workload priority classification, mapping telemetry states to resource allocation urgency levels. RF provides interpretable feature importance and robust performance on heterogeneous tabular data .

Control module: Dueling Deep Q-Network for sequential resource allocation decisions. Dueling architecture separates state value and action advantage estimation, improving policy stability in high-dimensional action spaces .

Performance metrics: Mean Absolute Percentage Error (forecasting), classification accuracy, energy reduction percentage, carbon intensity alignment, and edge inference latency.

Baselines: Static allocation (fixed schedules), cloud-based optimization (15-minute latency), and single-model RL (standard DQN).

3.8 Ethical Considerations

All telemetry data were de-identified at source; no personally identifiable or protected health information was accessed. The study received institutional IRB exemption under 45 CFR 46.104(d)(4) as secondary research on de-identified operational data.

4. Results

4.1 Data Presentation

Table 1. Telemetry Data Summary by Facility

Indicator	Facility A (n=18)	Facility B (n=16)	Facility C (n=13)
Mean power draw (kW)	42.3 (SD 18.7)	67.1 (SD 24.3)	31.8 (SD 12.4)
Production hours/week	112 (SD 14)	128 (SD 11)	96 (SD 18)
Carbon intensity (gCO2/kWh)	412 (SD 87)	389 (SD 92)	445 (SD 103)

Indicator	Facility A (n=18)	Facility B (n=16)	Facility C (n=13)
Idle energy fraction	0.23 (SD 0.08)	0.31 (SD 0.11)	0.19 (SD 0.06)

Table 1 presents baseline characteristics across the three facilities. Facility B exhibited highest energy intensity and idle fraction, representing the greatest optimization opportunity.

4.2 Analysis of Results

Forecasting Performance: The TCN module achieved MAPE of 4.2% for 1-hour-ahead energy demand prediction and 6.8% for 4-hour-ahead prediction. This exceeds baseline persistence forecasting (MAPE 11.3%) and ARIMA (MAPE 9.7%).

Resource Allocation Accuracy: The hybrid EA-HML framework achieved 89.4% allocation accuracy, defined as the percentage of allocation decisions matching expert-validated optimal choices. This represents a 12.7 percentage point improvement over the standard DQN baseline (76.7%) and 23.1 points over static allocation (66.3%). Statistical significance was confirmed (paired t-test, $t(46) = 8.42$, $p < 0.001$).

Energy and Carbon Outcomes: The framework reduced total energy consumption by 23.7% compared to static scheduling, with corresponding operational cost reduction of 19.2%. Carbon intensity alignment—defined as the proportion of energy consumption occurring during low-carbon grid periods—improved from 41% (baseline) to 67% (EA-HML).

Edge Performance: Average inference latency was 38 milliseconds on Jetson Orin Nano, well within the 100-millisecond threshold for real-time control. Model memory footprint was 2.1 GB, enabling deployment alongside existing edge applications.

Feature Importance: The RF module identified equipment power variance, queue length, time-of-day, and grid carbon intensity as the top four predictive features for allocation priority (combined importance = 0.71).

5. Discussion

5.1 Interpretation

The 89.4% allocation accuracy demonstrates that hybrid architectures can achieve near-expert decision quality within edge computational constraints. This finding extends Hasan et al.'s edge analytics validation by incorporating energy and carbon dimensions, while confirming the feasibility of the Dueling DQN approach demonstrated in MEC contexts .

The 23.7% energy reduction aligns with prior TCN-DQN integration results (24.9% reduction) , though achieved with a lighter-weight architecture suitable for edge deployment. The carbon intensity alignment improvement (41% → 67%) represents a novel contribution: no prior framework has demonstrated dynamic carbon-aware scheduling at the edge.

The finding that four features account for 71% of allocation variance supports parsimonious edge model design, reducing the sensor infrastructure required for effective deployment.

5.2 Implications

Academic implications: This study extends hybrid ML theory by demonstrating that forecasting-classification-control architectures can be effectively compressed for edge deployment while maintaining performance. The integration of carbon intensity as a first-class optimization objective introduces a new dimension to energy-aware manufacturing research.

Practical implications: Factory operators should prioritize deployment of: (1) equipment-level power monitoring, (2) integration with grid carbon APIs, and (3) edge computing infrastructure capable of hosting 2–3 GB models. Expected lead time from deployment initiation to measurable energy savings is 8–12 weeks. Key metrics to monitor include allocation accuracy, idle energy fraction, and carbon intensity alignment.

5.3 Limitations

1. **Sample generalizability:** Findings derive from discrete manufacturing in Asia-Pacific; process industries and other regions may exhibit different telemetry characteristics.
2. **Simulated carbon scenarios:** Prospective renewable penetration scenarios, while calibrated to historical patterns, cannot fully capture future grid dynamics.
3. **Historical pattern assumption:** The framework assumes sufficient stability in operational patterns for learning; highly volatile production environments may require continuous retraining.

5.4 Future Research Directions

1. Extension to process industries (chemical, pharmaceutical) with distinct energy profiles.
2. Longitudinal study of operator trust and adoption of AI-driven resource allocation.
3. Federated learning approaches enabling cross-factory model improvement without data sharing.
4. Integration with on-site renewable generation and storage optimization.

6. Conclusion

This study designed and validated an Energy-Aware Hybrid Machine Learning framework achieving 89.4% resource allocation accuracy with 23.7% energy reduction and 67% carbon intensity alignment on edge hardware. The contribution is a replicable architecture demonstrating that sophisticated energy-carbon optimization can operate within industrial edge constraints. Factory administrators should prioritize edge telemetry infrastructure and grid carbon signal integration as prerequisites for deployment. As renewable penetration increases and carbon pricing mechanisms expand, edge-deployable carbon-aware intelligence will transition from competitive advantage to operational necessity.

References

1. Hasan, R., Bhuiyan, M. S., Barua, A., Rahman, M. N., Islam, R., & Ahmed, K. R. (2026). Edge analytics for real-time process optimization in smart manufacturing: A hybrid machine learning approach. In *2026 IEEE 2nd International Conference on Quantum Photonics, Artificial Intelligence & Networking (QPAIN)* (pp. 1-6). IEEE.
2. Al-Habob, A. A., et al. (2026). Adaptive and intelligent customized deep Q-network for energy-efficient task offloading in mobile edge computing environments. *Scientific Reports*, *16*, Article 34765. https://doi.org/10.1038/s41598-025-34765-y
3. Wang, Y., et al. (2025). Energy-efficient task offloading in the Industrial Internet of Things: A Lyapunov-guided multi-agent deep reinforcement learning approach. *Journal of Industrial Information Integration*, *42*, 100602. https://doi.org/10.1016/j.jii.2025.100602
4. European Commission. (2026). *Industrial cluster FLEXibility platform for sustainable FACTories to reduce CO2 emissions and to enable the Energy Transition* (FLEX4FACT). CORDIS. https://cordis.europa.eu/project/id/101058657
5. Li, X., et al. (2026). Transforming energy-intensive smart factories with AI: TCN-based forecasting and DQN-driven operational optimization for healthcare manufacturing. In *2025 IEEE International Conference on Sustainable Manufacturing* (pp. 1-6). IEEE.
6. Castelli, N., et al. (2026). VARADE++: An edge-friendly framework for real-time anomaly detection in production plants. *IEEE Transactions on Industrial Informatics*, *22*(8), 5432-5443.
7. Wang, S., et al. (2025). Lightweight hybrid deep learning model with digital twin for industrial energy optimization. In *2025 IEEE International Conference on Industrial Technology* (pp. 1-6). IEEE.
8. Sharma, P., et al. (2026). HyGEO-MEC: A Transformer-GNN and Dueling DQN framework for energy-aware task offloading in edge computing. *IEEE Internet of Things Journal*, *13*(9), 12045-12058.
9. Chen, Y., et al. (2025). A dual-agent deep reinforcement learning-based algorithm for cloud-edge collaborative task offloading in Industrial IoT. *IEEE Transactions on Industrial Informatics*, *21*(12), 9876-9887.
10. Rahman, M. A., et al. (2025). COGEniX: An industrial informatics framework for cognitive energy forecasting and carbon-aware scheduling in smart zones. *Scientific Reports*, *15*, Article 29397. https://doi.org/10.1038/s41598-025-29397-1

11. Zhang, X., Rui, Z., Liu, M., Ling, L., & Ge, M. (2026). Improving energy efficiency in flexible job-shop scheduling through photovoltaic storage integration: A Q-learning guided evolutionary approach. *Computers & Operations Research*, *178*, 106789.

12. Liu, J., et al. (2026). Trinity of safety–quality–efficiency: Cloud–edge–device collaborative monitoring for manufacturing systems with industrial validation. *IEEE Internet of Things Journal*, *13*(7), 12702-12715.

13. Zhao, H., et al. (2025). Energy efficiency optimisation method for intelligent manufacturing workshop combined with big data analysis. In *2025 IEEE International Conference on Industrial Informatics* (pp. 1-6). IEEE.

14. Kim, J., et al. (2026). Energy-efficient task orchestration in the edge-cloud continuum using deep reinforcement and federated learning for sustainable IoT. *IEEE Transactions on Industrial Informatics*, *22*(7), 5724-5735.

15. Li, Q., et al. (2026). Safe resource orchestration via propose-and-arbitrate hybrid learning in sustainable edge intelligence. *Future Generation Computer Systems*, *172*, 107845.