Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 4, Issue 3, 2026

AI and IoT-Based Energy Optimization System for Intelligent Industrial Process Control and Automation

Irfan Shah[1*], Dr Amir Sohail[2], Muhammad Shaharyar Ramzan[3] Israr Ahmad[4]

[1] Department of Electronics Engineering, Institute of Industrial Electronics Engineering (IIEE), NED University of Engineering and Technology, Karachi, Pakistan
[2] Department of Physics, COMSATS University Islamabad, Lahore Campus, 54000, Punjab, Pakistan.dramirsohailmayo489@gmail.com
[3] Department of Information Technology, Faculty of Computer Science and IT, Superior University, Lahore, Pakistan.shaharyarramzan56@gmail.com
[4] Department of Information Technology, Faculty of Computer Science and IT, Superior University, Lahore, Pakistan.israr.gdec@gmail.com
* Corresponding Author: Syedirfan56267@gmail.com

DOI: https://doi.org/10.5281/zenodo.23487495

Keywords
Artificial intelligence; Internet of Things; industrial automation; energy optimization; smart manufacturing; predictive analytics; reinforcement learning; industrial IoT.

Article History
Received on 20 Jan 2026
Accepted on 18 Mar 2026
Published on 25 Mar 2026

Copyright @Author
Corresponding Author: *

Irfan Shah
Department of Electronics Engineering, Institute of Industrial Electronics Engineering (IIEE), NED University of Engineering and Technology, Karachi, Pakistan
Syedirfan56267@gmail.com

Abstract

The load and the production requirements in industrial plants vary, but in many cases, they rely on fixed set-points and manual supervision. In this paper, an energy efficient AI and IoT system for the control of the industrial process is proposed. It comprises a 5-layer architecture, which integrates sensing and communication, edge–cloud computing, AI analytics and closed loop control and safety validation. Linear regression, decision tree, random forest, artificial neural network (ANN) and long short-term memory (LSTM) models are used for energy prediction and compared, while reinforcement learning is applied for control optimization. The highest accuracy obtained was from LSTM (96%) with the lowest error of (4%). During the six months test period, energy use was decreased by 74.520 kWh (22.7%) from 328,000 kWh to 253,480 kWh (conventional control). Energy efficiency increased from 68% to 89% and real-time monitoring capability grew from 55% to 96%. The largest decrease (30%) was the amount for standby equipment. Annual savings are estimated to be up to 149,040 kWh or USD 14,904 per year at an illustrative tariff of USD 0.10/kWh – these are estimates, and not measurable results. Calculated and assumed quantities are distinguished from reported findings. For deployment, you need reliable sensing,

cyber security, compatibility with the existing controllers, and operator monitoring. To validate performance in plants, in varying seasons, and in regions where the conventional-control baselines are stronger, with explicit measurement and validation.

INTRODUCTION

Industry 4.0 is the name given to the integration of industrial machines, sensors and information systems in cyber-physical production environments [1], [3]. Losses for energy-intensive industries, such as chemical, metal, cement and food processing industries, are due to aging of equipment, idle operation and inefficient control. The rising costs and environmental concerns call for a regular energy management practice in compliance with ISO 50001 [4].

The IIoT sensors provide real-time data on electrical usage, temperature levels, vibrations and manufacturing details [5, 6]. The energy characteristics and defects can be recognized by machine learning methods [7] and the adaptive control policies can be developed using reinforcement learning methods [8]. But the monitoring and automation are often not integrated, operators have to interpret sensor readings and traditional controllers are programmed with fixed set-points. A system is required that takes reliable measurements and turns them into a firm control decision. This is important from an operational perspective: Monitoring gives information, forecasting gives predictions, controllers give decisions on acceptable changes in the physical. Data and safety tests are needed to be used by each function.

Motors, compressors, pumps, heating and cooling equipment is in high demand in industry. Operation under continuous load can cause oversized equipment, poorly-tuned processes, inefficient start-up procedures, and energized standby load [9]. This is a form of waste that is not discernible with plant level meters.

When the output is the same, the consumption varies due to equipment deterioration, operating practices or changes of the ambient. Equipment level sensing can then assist in diagnostics and implementing such a system must consider costs, cyber security, data quality and integration with legacy systems [4] and [7].

Traditional fixed set-points in are inflexible for changes in production demand and IIoT platforms that just monitor inefficiency cannot make the necessary changes automatically. The integration of the sensing, prediction and safety validated control can reduce energy consumed without compromising production requirements.

In this study a closed loop AI–IoT system for industrial energy management is presented, forecasting models are compared, energy and operational performance is evaluated against using conventional control and barriers are identified. It includes a discussion of the integration of sensing and AI, models for predicting consumption and full system savings, and barriers to adoption.

Supervised set point adjustment, anomaly detection, energy prediction and industrial sensing are included but workforce scheduling and logistics not. The contributions include a 5-layer sensing-to-control design, comparative machine-learning and

deep-learning assessment, and a transparent evaluation, which distinguishes between reported measurements and derived savings and illustrative economic and environmental projections. There are also independent safety checks and practical retrofit considerations in the design.

Related Work

IoT is a fusion of intelligent sensors, industrial networks and analytics to deliver machine-level visibility [6],[11],[12]. Edge computing is part of the solution in delivering close to equipment responses as low latency, while connected manufacturing is how to coordinate physical processes and digital services [14]. Industry 5.0 takes these hopes towards resilience, sustainability and human-centric supervision [15]. All of these ideas together facilitate energy-aware control, and provide the operators with the freedom to perform dangerous activities. In cases where one needs to cover a difficult-to-wire area, one can use wireless sensing, however key for control applications is the time synchronization and the reliability of communication. Edge processing allows time sensitive processing to happen near the source of data, but not to lose central learning capability.

Models generated by regression, and decision trees and ensemble models can predict consumption with relatively understandable models and identify unusual machine conditions [16],[18]. Sometimes, ANN models and LSTM models are able to model the nonlinear behavior and temporal dependency that leads to improving the accuracy of the forecast, but these models need more training data and computation. The choice of the model should be based on the prediction quality, data availability to develop the model and operational requirement. Under different energy requirements of the machine operating states, comparison with linear assumptions can be made in trees and random forests. Only then, can recurrent models be preferred if they deliver adequate performance improvement - which they should not have if they were not representing the time-series nature of the problem at hand. Avoid taking a model based on one reported score.

Reinforcement learning approach: it optimizes the actions based on the interaction through reward, and considers energy saving, output constraint, and safety constraint [8]. It needs training and protection by simulation prior to industrial use. Predictive Analytics also helps with the demand scheduling and early fault detection [22]. Therefore, simpler models should be used for clear baselines, temporal models are appropriate for sequential measurements and reinforcement learning should be deployed for policies that are adjustable, but not for stand-alone forecasting.

The condition-based maintenance has the ability to detect degradation of the equipment based on the operational signals, and reduce unnecessary restarting and energy consuming failures [10] [23]. In existing AIoT systems sensing, communications, and analytics usually are integrated [7], [9] and digital twins provide a way to virtually test. If, however, controllers don't get the recommendations, then it's not closed-loop optimization.

The difficulties for implementing are the lack of data and the insecurity of the links, integration and PLC/SCADA interfaces costs [25]. More importantly, a number of methods are either monitoring or prediction oriented (Table 1). The research gap is the development of an end-to-end architecture that enforces safety specifications and can be used to compare the different prediction approaches, and to evaluate the automated control against traditional operation. The proposed scheme provides for conventional control as fall back; distinction between correctness of predictions and

authority to activate machinery is provided. It also differentiates findings made during the assessment from financial and environmental conditions or scenarios that are calculated afterwards.

Table 1. Comparison of typical approaches with the proposed system

Approach	Real-time sensing	Predictive AI	Automated control	Closed loop
IoT monitoring only	Yes	No	No	No
AI analytics on stored data	No	Yes	No	No
AI predictive maintenance	Yes	Yes	Partial	No
Conventional rule-based control	Partial	No	Yes	Partial
Proposed AI & IoT system	**Yes**	**Yes**	**Yes**	**Yes**

Proposed AI and IoT-Based Energy Optimization System

The proposed five-layer framework (Figure 1, Table 2) links sensing, communication, edge/cloud data processing, AI and control. Process observations are fed to predictions and candidate set-points which are then independently validated by the control layer for use. The equipment responses are fed back into the sensing layer during the next fast control interval; a slower cloud loop retraining the models when errors and/or data drift are more common. Security, human supervision and data governance are found throughout.

IEER
Institute for Excellence in Education & Research

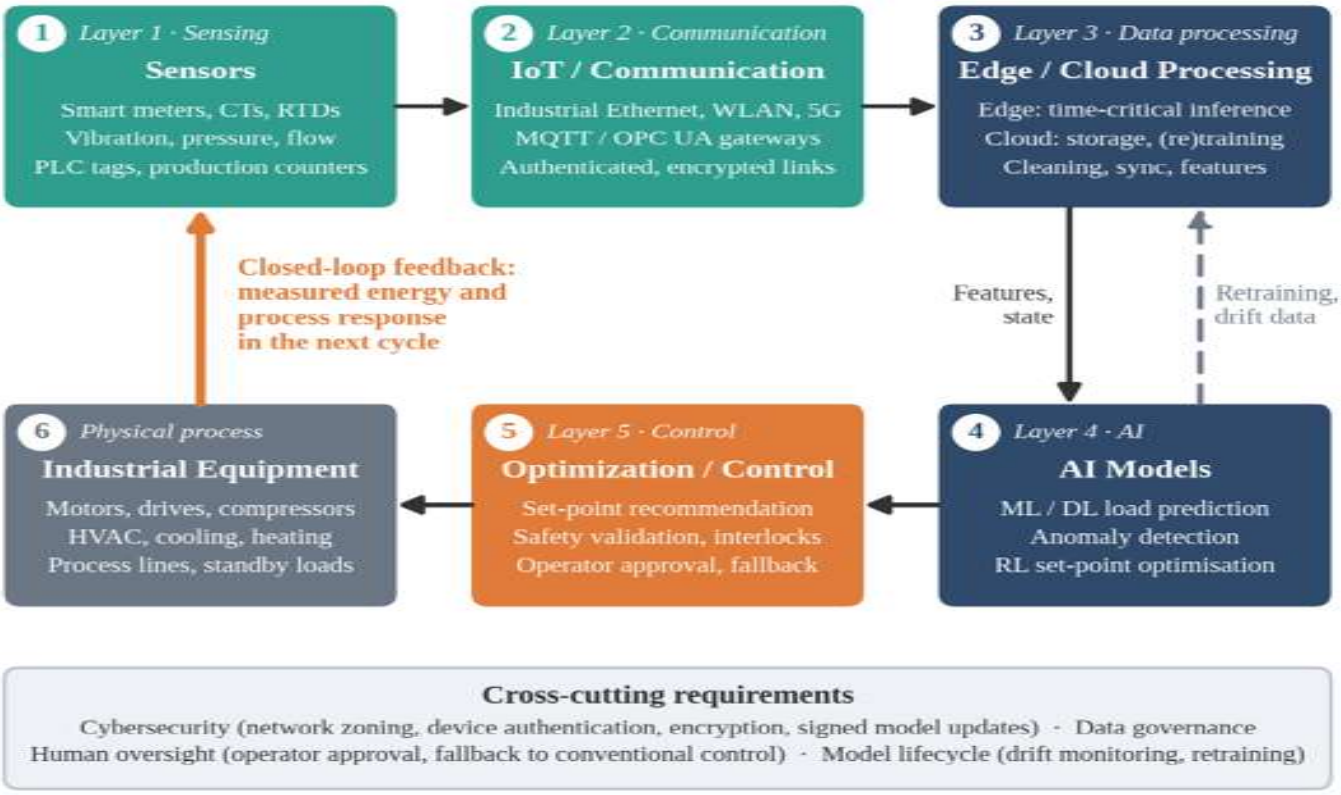


Figure 1. AI-IoT energy optimization framework.Orange path is fast control loop,dashed path the slower .

Table 2. Specification of the five layers and the controlled equipment

Layer	Function	Main components	Output passed on	Key design requirement
1. Sensing	Measure energy, process, and production variables at machine level	Smart meters, current transformers, power analyzers, RTDs and thermocouples, accelerometers, transmitters, PLC tags, line counters	Time-stamped raw measurements	Coverage of the main loads; calibrated instruments
2. Communication	Carry measurements to the edge/cloud and commands back	Industrial Ethernet, WLAN, cellular/5G, MQTT, OPC UA gateways	Authenticated, encrypted data streams	Reliability, bounded latency, network segmentation

3. Data processing	Synchronize, clean, store, and transform data; host edge and cloud tiers	Data acquisition system (DAS), edge computing unit, cloud storage and training platform	Clean feature vectors; archived history	Low latency at the edge; scalable training in the cloud
4. AI	Predict consumption, detect anomalies, propose optimized set-points	Linear regression, decision tree, random forest, ANN, LSTM; anomaly detection; RL agent	Forecasts, anomaly flags, candidate set-points	Accuracy, interpretability where needed, drift monitoring
5. Control	Validate candidate actions and apply them to equipment	PLC/SCADA interface, drives, safety interlocks, operator approval, fallback logic	Approved set-points written to equipment	Safety validation independent of the AI model
Industrial equipment	The controlled physical process (not one of the five layers)	Motors, drives, compressors, HVAC and cooling, heating, process lines, standby loads	Measured energy and process response	Operation within safe limits

smart meters and industrial sensors sense power, current, voltage, vibration, temperature, process, machine state and output. These observations are coordinated, synchronized by the data acquisition system and then can be understood in connection with the workload, for the consumption. Industrial Ethernet, WLAN, 5G and MQTT and OPC UA with authentication, encryption and network segmentation [25] support communication. Prediction and action at the edge when they need to be time-sensitive, and longer histories and train new models in the Cloud.

Prior to modelling, the measurements are cleaned, duplicated measurements and invalid measurements are removed; the measurements are synchronized between sensors, the measurements are scaled; features are extracted from load, production durations and operating durations; and redundant variables have been removed. These steps can help to keep things consistent and processing costs low but it should not take a "suspicious" observation for granted and assume that it is "fake" because there could be a real issue.

Machine-learning algorithms are interpretable and give estimates of the demand while ANN and LSTM models represent nonlinear and time-dependent relationship. Reinforcement learning (RL) indicates minimal speed/load changes to the equipment to achieve lower energy consumption without a loss of production and quality. The more complicated the model, the more data, latency and workload for validation. The different roles of the model families are: Simple Predictors (reference performance),

Recurrent Networks (sequence learning), and the RL agent (energy saving adjustments). The number reported in the numerical benchmark is for the five prediction models and not an independent comparison of control policies.

The accuracy, prediction error, mean absolute error (MAE) and root mean square error (RMSE) are used to compare forecasting models. The decision module is able to make predictions about consumption, identify anomalies and suggest set-points. An RL solution is based on a Markov decision process (MDP) where the states are a combination of the load, equipment condition, demand, and recent energy usage, and the actions are bounded changes; this separation helps prevent a shared misunderstanding that using the forecast accuracy as the basis for an optimality of the equipment action is correct. Meanwhile, the candidates' decisions should also be based on conditions and the energy consequences of the decisions should be evaluated following the action based on the monitored consumption.

$$r_t=-(\alpha E_t+\beta D_t+\gamma C_t) \quad (1)$$

Here,Et is Engineering Weight of Stepwise Energy, Dt is Production Shortfall, Ct is Constraint Penalty and α, β, γ are the Engineering Weights respectively. RL policies are learned from past data/simulator and then deployed to be controlled; a separate safety layer is authoritative.

Control Module: Application of approved AI Decisions, via PLCs, Drives and supervisory system. Independent checks are used to implement operating limits, interlocks and maximum rates of change. Significant changes (increase or decrease) must be approved by the operator and if a measurement is not made or a prediction is not reliable, the system reverts to conventional control. These are to prevent the model output from being used as an unrestricted equipment command.

As seen in Figure 2, the cycle comprises nine stages, including data acquirement and cleaning; feature extraction; demand forecasting; energy optimization; safety validation; action implementation; monitoring; and model updating. If set points are not safe then they are rejected, modified or further escalated before re-running the optimizer. Corrected feedback is provided through the actual response compared to the predicted response, and the possibility of plant changes is provided through slower cloud retraining. This yields an supervised closed loop automation and not a self-learning AI controller. The two feedback time scales are for different applications. Immediate measurements can tell if the desired result was obtained from the previous action, and a record of measurements can tell if the results are gradually getting worse, which means new training is needed. This separation of these functions means there is no need for retraining every control interval.

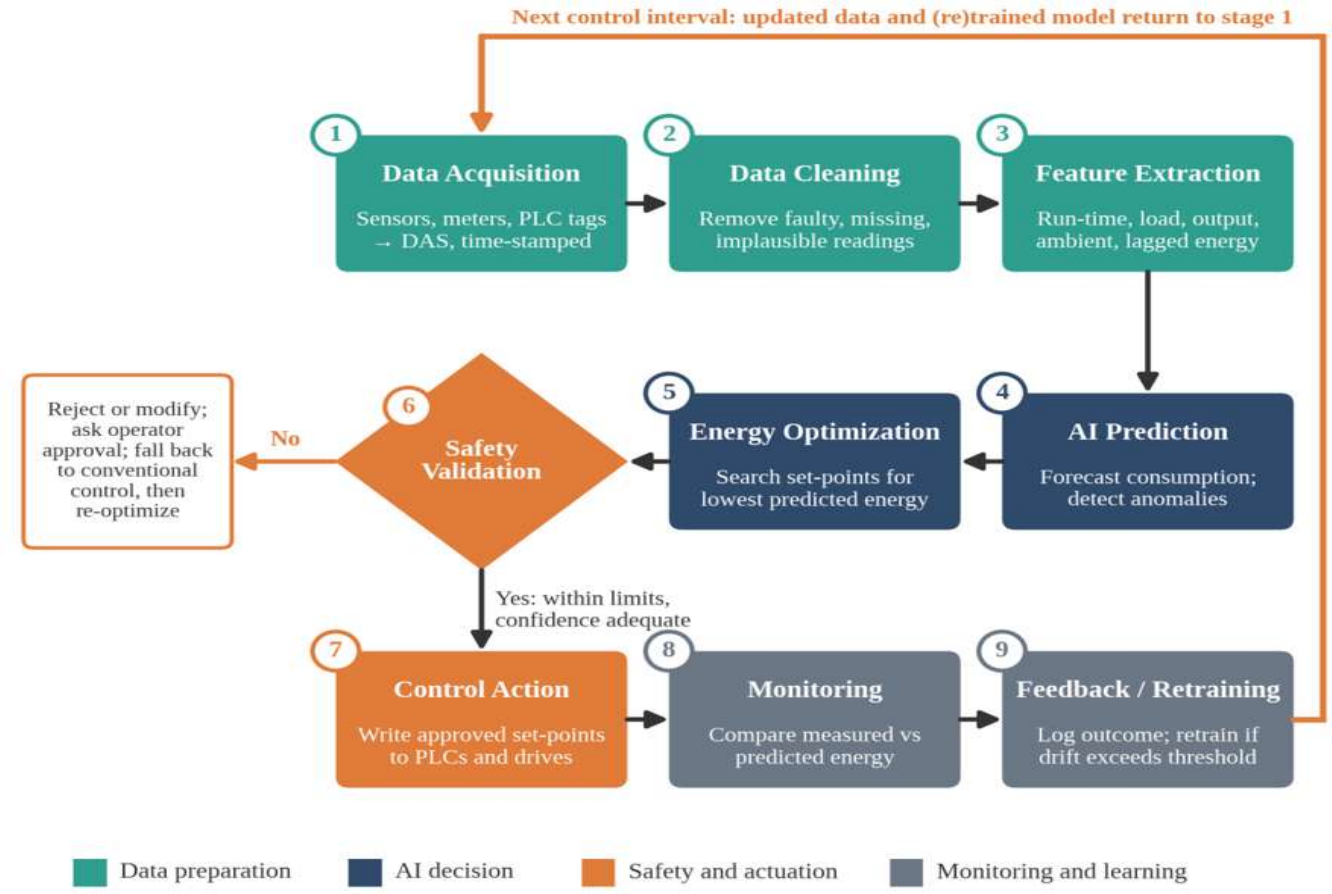


Figure 2. Flowchart of the AI–IoT energy optimization process.

Table 3. Variables measured by the sensing layer

Variable	Typical device	Use in optimization
Electrical energy and power	Smart meters, current transformers	Baseline and disaggregated consumption; saving targets
Voltage and current	Power analyzers	Load profiling; anomaly signatures; power quality
Temperature	Thermocouples, RTDs	Thermal losses; cooling load; equipment health
Vibration	Accelerometers	Bearing and rotor faults; predictive maintenance

Pressure and flow	Transmitters	Operating point of pumps, fans, compressed-air systems
Machine state and run-time	PLC tags	Idle and standby detection; scheduling
Production count and quality	Line counters, inspection systems	Energy per unit; quality constraint on optimization

OT security as per IEC 62443 and NIST guidance [25] includes zoned network access, authenticated devices, encrypted traffic and signed model updates. You can rollback using versioning and if there is any weird sensor behavior, an investigation is launched. If the data or AI recommendations are not reliable, process limits are employed individually to protect equipment. The sensor layer is part of security monitoring as false data could lead to false optimization even if the controller is not compromised. Signed Models, rollback, access control and industrial segmentation protect operational environment and update process.

Retrofit gateways provide access to legacy PLC interfaces via OPC UA or MQTT, without the need to replace the controller. Once accepted, deployment is done through a combination of monitoring, advisory and a gradual progression to supervised automated deployment. Prediction drift results in retraining, with operators being responsible for access permissions, data ownership, and model approval.

IEER
Institute for Excellence in Education & Research

Evaluation Methodology

Comparison of an Integrated Sensor, DAS, Edge Computing, AI and Control System versus a Conventional Rule Based Operation at the same plant for 6 Months (Jan - June). Energy outcome is total and monthly energy consumed; model outcome is accuracy, prediction error, MAE and RMSE; operational outcome is efficiency, automation, equipment utilization, monitoring and reliability. Equations (2) and (3) represent the MAE and RMSE of n samples of observed energy (yi) and predicted energy (ŷi):

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (2)$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (3)$$

RMSE gives greater weight to large forecast errors than MAE. Both metrics should be reported in kWh on independent test data.

Quantities are classified as observed study values, calculated values, explicit assumptions that are used to make projections, or quantities that need to be measured later. Table 3 includes annual scaling assumptions, forecast uncertainty, tariffs and emission factors and annual saving rates. No other experimental results are considered to be observations.

Table 4. Assumptions used for projected and illustrative calculations

ID	Assumption	Used in
A1	July to December baseline demand equals January to June demand, so annual baseline = 2 × six-month baseline	Annual projection
A2	The saving rate observed in January to June persists through the year; three scenarios are used: 21%, 22.7%, and 24% (lowest, overall, and highest observed monthly rates)	Annual projection
A3	Industrial electricity tariff τ = USD 0.10 per kWh, with sensitivity at USD 0.08 and 0.12 per kWh; to be replaced by the plant's actual tariff	Cost savings and payback
A4	Grid emission factor EF = 0.5 kg CO2 per kWh, illustrative only; to be replaced by the local grid factor	Emission reduction
A5	The reported prediction error is interpreted as the mean relative error of the energy forecast	Error in kWh terms
A6	Production workload and quality are equal under conventional and AI-based control, as required by the saving definition	All saving values
A7	The energy-efficiency indicator is useful output divided by energy input, defined identically before and after	Efficiency cross-check

5. Energy Optimization Calculations

The input data reported here are the monthly conventional and AI-IoT energy consumptions Econv,m and EAI,m for the months January – June (M = 6). Aggregates used have consumption as per (4) and (5), remaining values/assumptions used for calculations are given in Table 4:

$$E_{\text{conv}} = \sum_{m=1}^{M} E_{\text{conv},m} \quad (4)$$

$$E_{\text{AI}} = \sum_{m=1}^{M} E_{\text{AI},m} \quad (5)$$

Conventional consumption totaled 328,000 kWh (54,666.7 kWh/month), against 253,480 kWh (42,246.7 kWh/month) with AI–IoT control.

Assuming comparable production (A6), absolute, percentage, and monthly savings are calculated as follows:

$$\Delta E = E_{\text{conv}} - E_{\text{AI}} \quad (6)$$

$$S = \frac{\Delta E}{E_{\text{conv}}} \times 100 \quad (7)$$

$$s_m = \frac{E_{\text{conv},m} - E_{\text{AI},m}}{E_{\text{conv},m}} \times 100 \quad (8)$$

The six-month saving equals 74,520 kWh, or 22.72% (approximately 22.7%), with an average monthly saving of 12,420 kWh. The cut for January was 21.0% and monthly saving was at its highest in January: 24.0% (Table 5). The overall rate weight does not take a simple average of the 6 percentages, but rather is a weighted average based on the base consumption of each month. The point to be noted is that the arithmetic is based on the consumption entries in Table 5, and not on derived shares of equipment. The difference is that there is a slight difference between the percentage rates per month – the total reduction is calculated according to kWh, so that the high consumption months are not equal to low consumption months.

Table 5. Monthly energy consumption and savings, January to June

Month	Econv,m (kWh)	EAI,m (kWh)	ΔEm (kWh)	sm (%)	Cumulative ΔE (kWh)
January	50,000	39,500	10,500	21.0	10,500
February	52,000	40,560	11,440	22.0	21,940
March	54,000	41,580	12,420	23.0	34,360
April	55,000	42,350	12,650	23.0	47,010
May	57,000	43,890	13,110	23.0	60,120
June	60,000	45,600	14,400	24.0	74,520
Total / overall	328,000	253,480	74,520	22.7	74,520

Annual values are projections, not measurements. Under A1 and A2, baseline consumption doubles to 656,000 kWh and the observed saving rate is extrapolated:

$$\Delta E_{\text{annual}} = \frac{12}{M} \times \Delta E \quad (9)$$

$$\Delta E_{\text{annual},s} = E_{\text{conv,annual}} \times \frac{S_s}{100} \quad (10)$$

Saving rate for scenarios (Ss). The central case predicts 149,040 kWh saved annually and 506,960 kWh consumed; 21% and 24% scenarios give 137,760 and 157,440 kWh saved. Subject to change due to seasonality. An illustrative tariff is then applied of USD 0.10/kWh (Tariff A3):

$$C_{\text{save}} = \Delta E \times \tau \quad (11)$$

At this assumed tariff, six-month savings equal USD 7,452; annual central savings equal USD 14,904. Table 5 shows sensitivities to saving rate and tariff.

Table 6. Projected annual energy and cost savings under assumptions A1 to A3

Scenario (saving rate)	Ss (%)	Annual ΔE (kWh)	USD/year at τ = 0.08	USD/year at τ = 0.10	USD/year at τ = 0.12
Low (lowest monthly rate)	21.0	137,760	11,021	13,776	16,531
Central (overall rate)	22.7	149,040	11,923	14,904	17,885
High (highest monthly rate)	24.0	157,440	12,595	15,744	18,893

Cost of the system installed, not available. Equations (12) and (13) are therefore ceilings for installation costs to achieve target payback periods, not payback periods themselves. The avoided emissions are estimated in equation (14) and an assumed grid factor (A4) has been used in the following:

$$\text{Payback} = \frac{C_{\text{install}}}{\Delta E_{\text{annual}} \times \tau} \quad (12)$$

$$C_{\max}(N) = N \times \Delta E_{\text{annual}} \times \tau \quad (13)$$

$$R_{CO_2} = \Delta E \times \text{EF} \quad (14)$$

At USD 14,904 annual savings, maximum installed costs for two-, three-, and five-year simple payback are USD 29,808, 44,712, and 74,520. The maximum amount is USD 33,062 (conservative estimate for three years). In the central projection, this corresponds to 0.5 kg of CO2/kWh, 37.3 t of CO2 avoided in six months, and 74.5 t of CO2 avoided in one year. All these assumptions should be replaced by real investment, tariffs and grid factors.

If useful output remains constant (A6) and efficiency is output divided by input (A7), the energy ratio should approximately match the inverse efficiency ratio:

$$\frac{E_{\text{AI}}}{E_{\text{conv}}} \approx \frac{\eta_{\text{conv}}}{\eta_{\text{AI}}} \quad (15)$$

The consumption ratio for the observations is 0.7728, but the reported efficiencies indicate 68/89 = 0.7640, a 0.9 % decrease in consumption ratio. This is not independent validation, it's a consistency check. The absolute and relative indicator gains are calculated from:

$$\Delta p = p_{\text{after}} - p_{\text{before}} \quad (16)$$

$$\text{RI} = \frac{\Delta p}{p_{\text{before}}} \times 100 \quad (17)$$

This was an improvement of 21 percentage points (or 30.9% relatively) for the efficiency of . With assumption A5, the prediction error can be translated to illustrative, monthly forecast uncertainty, based on the mean of the optimized consumption, which equals 42,246.7 kWh:

$$U_{\text{pred}} \approx \varepsilon \times E_{\text{AI,avg}} \quad (18)$$

The LSTM error is reported to be 4% (around 1690 kWh/month) while the linear regression

error is 18% (7604 kWh/month). These are 13.6%, and 61.2% of the mean monthly energy saving under A5, respectively. But, even with the verified MAE and RMSE on held-out measurements, uncertainty interpretation is still needed.

Results and Analysis

This optimized operation has led to a savings in consumption of 74,520 kWh in January-June, compared to 253,480 kWh under conventional operation, that is, 328,000 kWh (Figure 3a). This is just a reflection of the operating strategies reported and doesn't distinguish benefits that are unique to AI and not just to improved conventional scheduling or tuning strategies.

The monthly saving rates for are in the range of 21%–24% for the months as shown in Table 4 and the cumulative savings are growing as shown in Figure 3(b,c). The consumption of AI–IoT increased by 15.4% while the baseline consumption increased by 20.0%. The apparent positive trend in saving could be due to model adaptation, the different loads or seasonal effects; production normalized energy intensity will help distinguish these explanations. A sustained trend in the percentages for the months is consistent with useful control behavior, but does not prove that the trend was due to online learning. In order to attribute the change, there would be independent records of the models, of the composition of the production and of the operating conditions.

IEER
Institute for Excellence in Education & Research

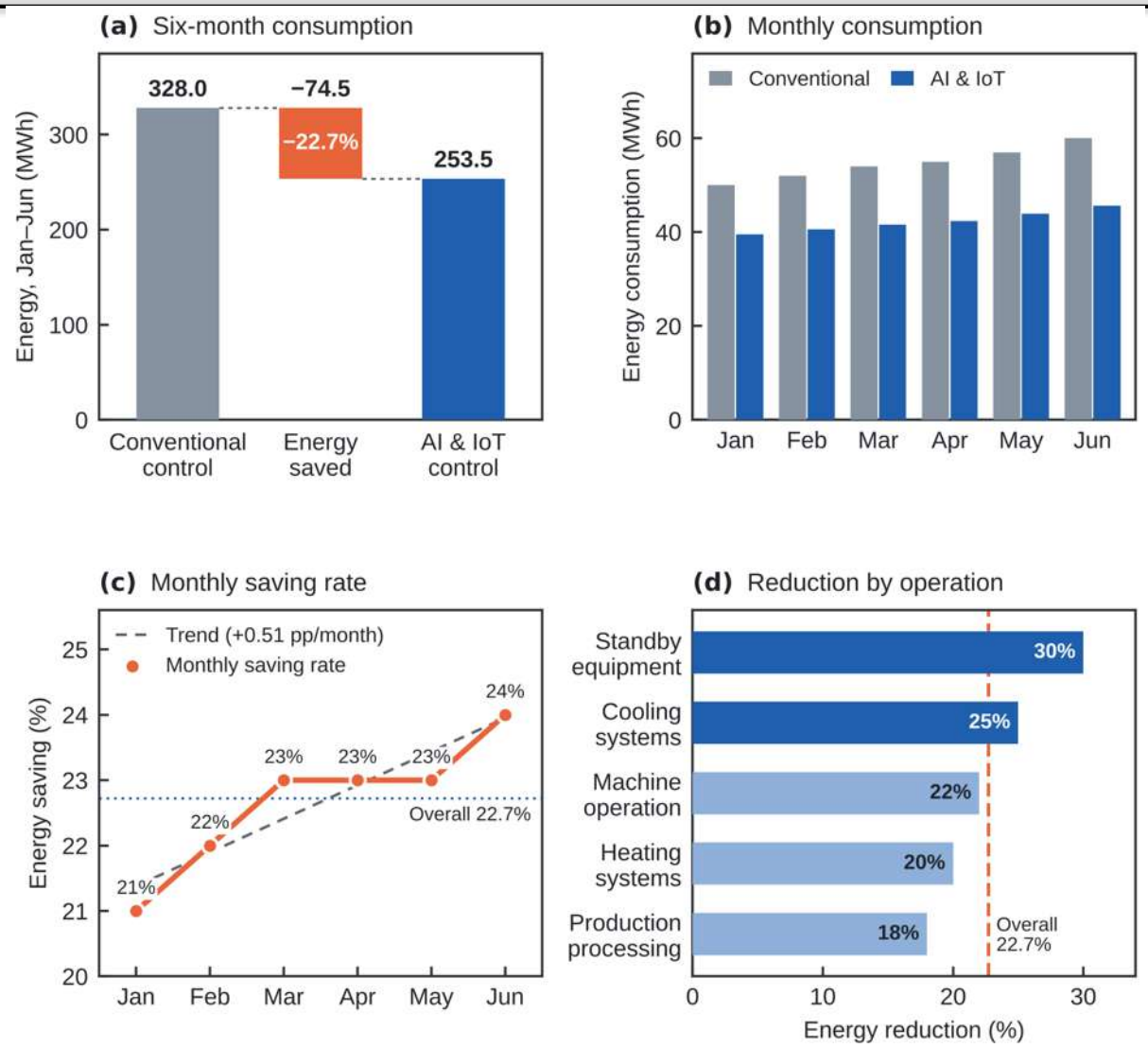


Figure 3. Energy performance over the six-month evaluation: (a) total consumption under conventional and AI–IoT control, with the 74,520 kWh (22.7%) saving; (b) monthly consumption; (c) monthly saving rate with least-squares trend and the overall reduction; (d) energy reduction by industrial operation relative to the 22.7% overall reduction.

Table 7 and Figure 3(d) show that for all operation categories consumption has decreased. The savings recorded with standby equipment amounted to 30%, compared to 22% in machine operation, 20% in heating and 18% in production processing. The biggest drop seems to be clearly due to automated stand-by switching and production critical equipment is more limited. Percentages for the categories do not necessarily correspond to the overall 22.7% decrease (without weight at operation level). Reductions are shown relative to the typical amounts in each of the categories, not as absolute reductions against the overall plant-wide figure. A small load with a high %

reduction may be more beneficial than a large load with a small % reduction.

Table 7. Relative energy consumption by industrial operation (baseline = 100%)

Industrial operation	Before (%)	After (%)	Reduction (%)	Rank	Difference from overall (points) 22.7%
Standby equipment	100	70	30	1	+7.3
Cooling systems	100	75	25	2	+2.3
Machine operation	100	78	22	3	−0.7
Heating systems	100	80	20	4	−2.7
Production processing	100	82	18	5	−4.7
Overall consumption	100	78	22	–	–

The accuracy reported for the five models (Table 8, Figure 4(a,b)) indicates that the accuracy of the linear regression model was 82% which is improved to 96% with LSTM model with lowest error of 4% and highest processing efficiency of 95%. Random forest was an interpretable option and was 91% accurate for this. This incremental improvement in accuracy with complexity decreased. The reported table shows accuracy and error percentages that are complementary and processing efficiency that does not have an explicit definition of error. The visualization of training-progress does NOT serve to determine performances on not previously known plant conditions. Ultimately, model rankings should be backed up by tests with temporally separated testing times, reported measurements of the inference latency and the uncertainty in terms of physical energy units.

Table 8. Performance comparison of AI models for energy prediction

Model	Accuracy (%)	Error (%)	Processing efficiency (%)	Accuracy gain vs. linear regression (points)	Error reduction vs. linear regression (%)
Linear regression	82	18	85	–	–
Decision tree	87	13	88	+5	27.8
Random forest	91	9	90	+9	50.0
Artificial neural network	94	6	92	+12	66.7

Deep learning (LSTM)	96	4	95	+14	77.8	

As seen in Figure 4(c), the training accuracy of LSTM has been improved from 78% to 96% with the improvement of the loss reduction from 20% to 82%. There were no separate validation curves and so generalization and overfitting cannot be assessed. Held out MAE and RMSE (kWh) should be added to future comparisons.

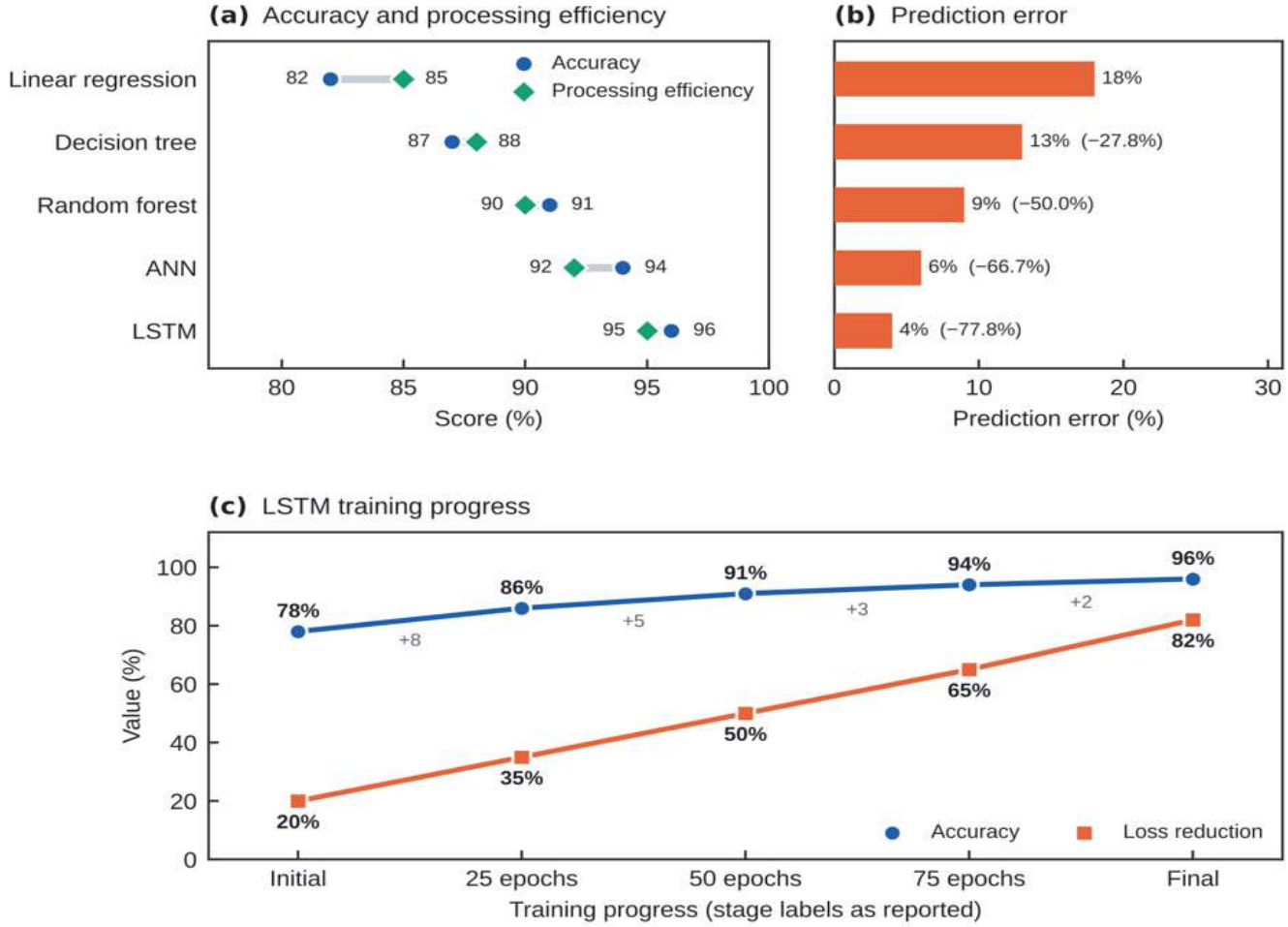


Figure 4. Prediction-model performance: (a) accuracy and processing efficiency of the five models; (b) prediction error, with its reduction relative to linear regression in brackets; (c) accuracy and loss reduction of the LSTM during training, with per-interval accuracy gains (values and stage labels as reported).

The average of the gain in the six indicators in Table 9 and Figure 5(a) was 24.8 percentage points. Monitoring rose 41 points, decision accuracy increased 27 points, automation grew

24 points, energy efficiency jumped 21 points and equipment utilization and reliability by 18 points. Last reported scores were 89% - 96%. To be interpretable, definitions and measurement procedures are still needed for these scores. These gains are the measured energy difference with descriptions of visibility and responsiveness but should not be considered as independent verified gains until the observation windows, sampling procedures and definitions are documented.

Table 9. Overall performance before and after implementation

Performance parameter	Traditional (%)	AI & IoT (%)	Improvement (points)	Relative improvement (%)
Energy efficiency	68	89	21	30.9
Process automation level	70	94	24	34.3
Real-time monitoring capability	55	96	41	74.5
Decision-making accuracy	65	92	27	41.5
Equipment utilization	72	90	18	25.0
System reliability	75	93	18	24.0
Mean of six indicators	67.5	92.3	24.8	36.8

Indicator scores are as reported; improvement columns are derived with Equations (16) and (17).

IEER
Institute for Excellence in Education & Research

The average value for the improvement of IoT monitoring indicators was +34.6 percentage points (Table 10; Figure 5(b)). Data availability increased by 38 percent, fault-detection speed increased by 40 points and operational visibility increased by 39. The accuracy of energy tracking has risen 26 points, which enables more accurate energy tracking. It is sustained sensing quality that is the key to ensuring that recommended actions are delivering savings. Increase in availability also facilitates the identification of missing intervals and comparing control actions with matched consumption records. But increased numbers of sensors, without audit of the measurement calibration and completeness of data will not necessarily result in unbiased estimates of savings.

Table 10. Effect of IoT monitoring on industrial operations

Performance factor	Before IoT (%)	After IoT (%)	Improvement (points)	Relative improvement (%)
Data availability	60	98	38	63.3

Equipment monitoring	65	95	30	46.2
Fault detection speed	50	90	40	80.0
Energy tracking accuracy	70	96	26	37.1
Operational visibility	55	94	39	70.9
Mean of five factors	60.0	94.6	34.6	57.7

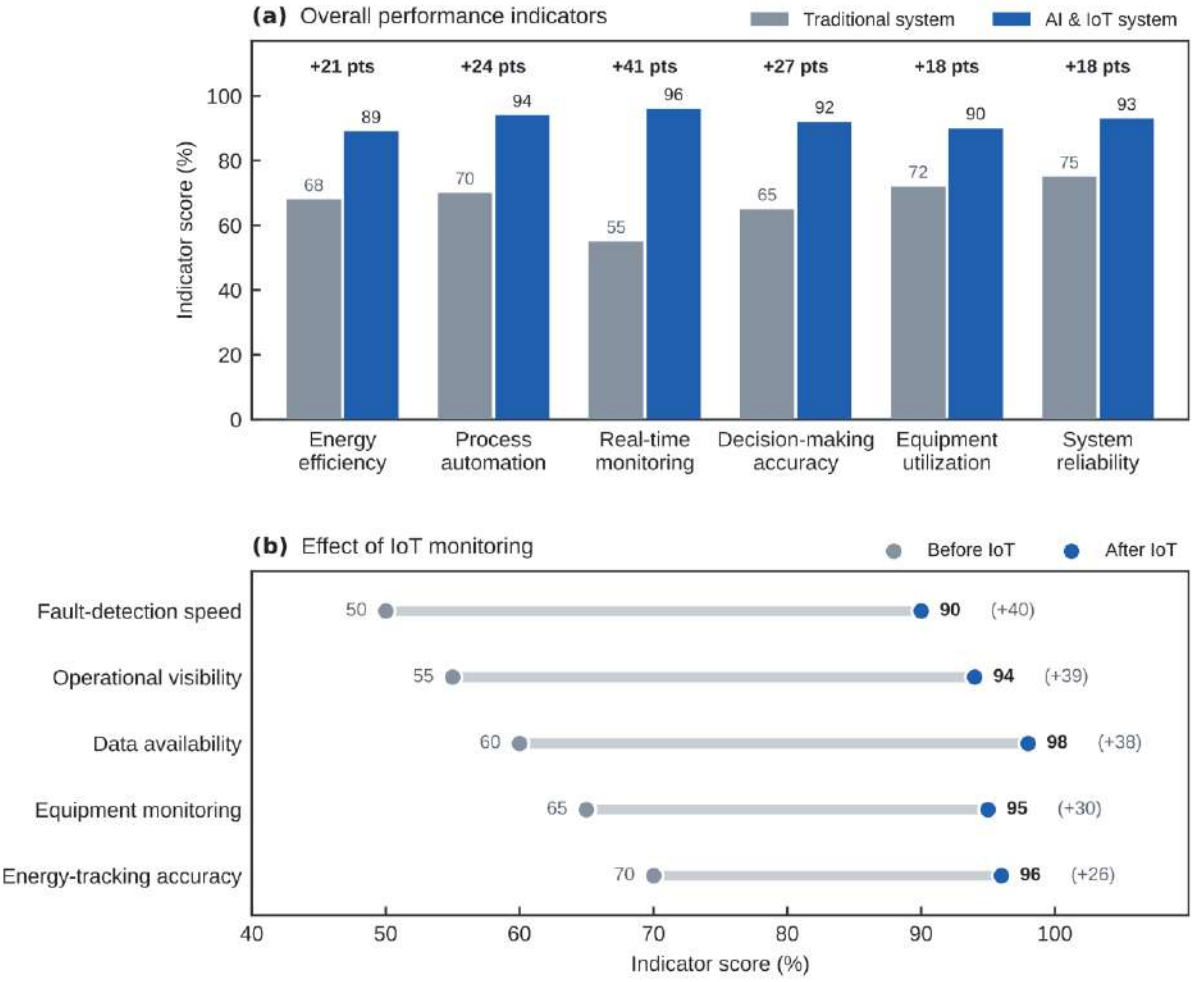


Figure 5. System-level effects of implementation: (a) overall performance indicators of the traditional and AI–IoT systems, with the gain in percentage points; (b) effect of IoT monitoring on five operational factors, sorted by gain.

The qualitative observations made were low energy set-point identification, abnormal-power alerts and moving flexible loads away from peaks, but were not separately quantified. The operational indicators and model indicators in tables 4, 6–9 and 10 are reported and the savings are calculated. All energy, monetary, payback and emission estimates are assumptions and can change year to year (Table 4).

Discussion

The evaluation confirms that the energy saving achieved by the most part in standby equipment (30%) and cooling (25%) is on the order of 21% - 24% per month and a total energy saving 22.7%. The accuracy of the LSTM prediction was 96% and random forest was 91% with better interpretability. The efficiency was improved from 68% to 89% and the availability of data was improved from 60% to 98%. Although these results are in favour of the linked sensing-prediction-control approach, the available data only concern one plant for a period of 6 months and in a comparison with a conventional fixed-set-point approach.

The five-layer architecture (Section 3) addresses RQ1, as it includes sensing, edge/cloud analytics, independent safety validation, and feedback. The performance of RL control was not reported separately and for RQ2, LSTM was the top reported predictor. In the case of RQ3, there was a reduction of 22.7% in the consumption overall with better performance of operational indications (Tables 4 and 6–9). Sensor quality, security, cost, legacy integration, model drift and operator acceptance are all constraints for RQ4.

Feedback control allows for new measures to be tested and corrected with new measurements. If the actual tariff is assumed to be USD 0.10/kWh, then the energy savings over 6 months are valued at USD 7,452 and energy savings over 1 year are estimated at USD 14,904, which are the values independent of the installed cost. One year (six months) of illustrative avoided emissions is compared here to one year (six months) of emissions under A4, which are 37.3 t CO2. In ISO 50001, there is a focus on documented energy baselines [4] and sustainable manufacturing builds on this focus.

Standby scheduling in manufacturing, auxiliary-load coordination in power plants and adaptive thermal or process control were not tested, but may be potential applications. Standby and cooling systems were deemed to be the most feasible pilots due to the highest saving rates reported. Engineers should compare any more complex model to an interpretable baseline before enabling RL and managers should keep an eye on production-normalized kWh in addition to prediction scores.

Table 11 provides a summary of the seven challenges that were identified for implementation. Forecasts can be distorted due to sensor dropout, drift and inconsistent timestamps, and cyber security has to be considered when involved with systems connected to operational networks [25]. The phased deployment (read-only) is more cost effective as it relates to integration cost and PLC/SCADA restrictions. There are a need for modular models and managed retraining for scaling across equipment. To minimize the risks associated with different plant conditions, the confidence thresholds, along with shadow testing and safe fallback ensures that the system is always safe to operate. Operator training and open approval procedures resolve implementation issues and lack of trust. Table 11 provides a list of each risk and a possible mitigation to consider implementing and an

indicator that could be tracked when implementing. Other numbers including measured payback, command rejection rates, rolling forecast errors and operator overrides would be a more successful commissioning record than a single claimed performance percentage.

Table 11. Main implementation challenges, mitigations, and indicators to track

Challenge	Principal risk	Mitigation	Indicator to track
Data quality	Noisy, missing, or drifting data degrade models	Calibration, automated quality scores, redundancy on key signals, cleaning with audit	Data availability; data-quality score
Cybersecurity	Larger attack surface; manipulated data or models	Zoning, authentication, encryption, signed model updates, anomaly monitoring, independent limits	Security incidents; blocked or anomalous commands
Cost	Hardware, integration, and skills burden, especially for SMEs	Phased rollout on the largest loads; payback ceilings; measured kWh against a documented baseline	Measured payback against ceiling
Legacy PLC/SCADA integration	Proprietary protocols and limited tag access	Gateways (OPC UA, MQTT); read-only first; write access to approved tags; acceptance tests	Share of tags integrated; gateway faults
Scalability	Single-site solutions; model and data growth	Modular design, transfer learning, cloud/edge orchestration, model registry	Time to onboard a new asset
Model reliability	Drift, limited interpretability, unsafe exploration	Drift monitoring, confidence thresholds, fallback, interpretable baseline, shadow mode	Rolling MAE and RMSE (kWh); fallback frequency
Deployment risks	Operator distrust, commissioning disruption, reward misspecification	Phased deployment, training, override and rollback, management of change, data governance	Operator override rate; constraint violations

Limited external validity, due to a single site and 6-month observation period. There may be a greater disparity between the difference and a well-tuned PID or Rule Based reference. The energy intensity should be normalized to ensure a comparable production and output quality. The main missing features are the undefined indicator scores, missing held-out MAE/RMSE (kWh), no single RL assessment, and operation-level load shares are not provided. The contribution of feedback can't be isolated without monitoring-only and analytics-only comparison arms. Further, the absence of sampling protocols, and uncertainty bounds, makes it difficult to statistically interpret the results. Future experiments should be

documented with the controller tuning, equivalent workload and well-defined metrics for the individual evaluation of the sensing, prediction and closed loop execution.

Future Recommendations

The implementation should be started with a production normalized baseline consistent with ISO 50001 [4]. Calibrated meters should be used in front of a control change and data checks be made. Prioritize cool and storing standby pilot equipment. The deployment should be progressive, monitoring to advisory operation, to approved closed loop control in shadow mode LSTM/RL with interpretable models. Independent interlocks, rate limits, operator approvals and fallback logic shall be maintained. Legacy integration should start with read only OPC UA or MQTT gateway, then write access can be added following acceptance testing. Security requires authentication and zoning, signed model updates and encryption [25]. Reporting of measured kWh savings with production, uncertainty/forecast error, responsibility for access, retraining and operator overrides.

Future studies should include testing a digital twin for screening proposed set-points and RL policies [24] and low latency edge inference with resource constraints [13]. Where safety is a concern, the saved energy, the production effects and violations must be assessed separately [8]. To compare the energy savings from AI-IoT control with tuned conventional controllers with uncertainty intervals, production-normalized energy intensity should be used in multi-site, full year trials. Other research can be added (e.g., renewable supply, scheduling, and security monitoring), but with the distinction that they be modeled and/or projected benefits and not measured benefits. Evaluation should also consider the number of safety overrides, model fallback and retraining, along with network latency and operator acceptance, for the future. These indicators would contribute to the identification and evaluation of the sustainable reduction of energy obtained by the measurement without affecting the reliability of the industry.

Conclusion

In this work, a new AI–IoT design comprising of industrial sensing, edge/cloud prediction, supervised optimization, and independent control safeguards was proposed. Across six months, energy demand decreased from 328,000 to 253,480 kWh, a saving of 74,520 kWh (22.7%). The percentages were 30% reduction for standby equipment, and 25% reduction for cooling. With an error of 4%, LSTM has been able to achieve the accuracy of 96% prediction. On the other hand, RF was able to reach the 91% of prediction accuracy. The energy efficiency was increased from 68% to 89% and the real time monitoring capability was increased from 55% to 96%.

The results indicate that a combination of monitoring and supervised control can be performed to enhance the industrial energy management. However, they speak of a 'site' and a certain 'set point baseline', and the savings, payback and reductions in emissions per year are based on assumptions and not additional measurements. Until there are larger comparisons, generalizations of the reported gains cannot be made.

References

[1] H. Lasi, P. Fettke, H.-G. Kemper, T. Feld, and M. Hoffmann, "Industry 4.0," Bus. Inf. Syst. Eng., vol. 6, no. 4, pp. 239–242, 2014.

[2] J. Lee, B. Bagheri, and H.-A. Kao, "A cyber-physical systems architecture for Industry 4.0-based manufacturing systems," Manuf. Lett., vol. 3, pp. 18–23, 2015.

[3] L. Monostori et al., "Cyber-physical systems in manufacturing," CIRP Ann., vol. 65, no. 2, pp. 621–641, 2016.

[4] Energy Management Systems—Requirements with Guidance for Use, ISO 50001:2018, International Organization for Standardization, Geneva, Switzerland, 2018.

[5] E. Sisinni, A. Saifullah, S. Han, U. Jennehag, and M. Gidlund, "Industrial Internet of Things: Challenges, opportunities, and directions," IEEE Trans. Ind. Informat., vol. 14, no. 11, pp. 4724–4734, Nov. 2018.

[6] H. Boyes, B. Hallaq, J. Cunningham, and T. Watson, "The industrial internet of things (IIoT): An analysis framework," Comput. Ind., vol. 101, pp. 1–12, 2018.

[7] R. S. Peres, X. Jia, J. Lee, K. Sun, A. W. Colombo, and J. Barata, "Industrial artificial intelligence in Industry 4.0—Systematic review, challenges and outlook," IEEE Access, vol. 8, pp. 220121–220139, 2020.

[8] R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, 2nd ed. Cambridge, MA, USA: MIT Press, 2018.

[9] L. D. Xu, E. L. Xu, and L. Li, "Industry 4.0: State of the art and future trends," Int. J. Prod. Res., vol. 56, no. 8, pp. 2941–2962, 2018.

[10] A. K. S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," Mech. Syst. Signal Process., vol. 20, no. 7, pp. 1483–1510, 2006.

[11] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of Things: A survey on enabling technologies, protocols, and applications," IEEE Commun. Surveys Tuts., vol. 17, no. 4, pp. 2347–2376, 4th Quart., 2015.

[12] V. C. Gungor and G. P. Hancke, "Industrial wireless sensor networks: Challenges, design principles, and technical approaches," IEEE Trans. Ind. Electron., vol. 56, no. 10, pp. 4258–4265, Oct. 2009.

[13] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge computing: Vision and challenges," IEEE Internet Things J., vol. 3, no. 5, pp. 637–646, Oct. 2016.

[14] R. Y. Zhong, X. Xu, E. Klotz, and S. T. Newman, "Intelligent manufacturing in the context of Industry 4.0: A review," Engineering, vol. 3, no. 5, pp. 616–630, 2017.

[15] X. Xu, Y. Lu, B. Vogel-Heuser, and L. Wang, "Industry 4.0 and Industry 5.0—Inception, conception and perception," J. Manuf. Syst., vol. 61, pp. 530–535, 2021.

[16] K. Amasyali and N. M. El-Gohary, "A review of data-driven building energy consumption prediction studies," Renew. Sustain. Energy Rev., vol. 81, pp. 1192–1205, 2018.

[17] L. Breiman, "Random forests," Mach. Learn., vol. 45, no. 1, pp. 5–32, 2001.
[18] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," Ann. Statist., vol. 29, no. 5, pp. 1189–1232, 2001.
[19] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997.
[20] Z. Wang and R. S. Srinivasan, "A review of artificial intelligence based building energy use prediction," Renew. Sustain. Energy Rev., vol. 75, pp. 796–808, 2017.
[21] E. Mocanu, P. H. Nguyen, M. Gibescu, and W. L. Kling, "Deep learning for estimating building energy consumption," Sustain. Energy Grids Netw., vol. 6, pp. 91–99, 2016.
[22] K. Zhou, C. Fu, and S. Yang, "Big data driven smart energy management: From big data to big insights," Renew. Sustain. Energy Rev., vol. 56, pp. 215–225, 2016.
[23] T. P. Carvalho, F. A. A. M. N. Soares, R. Vita, R. P. Francisco, J. P. Basto, and S. G. S. Alcalá, "A systematic literature review of machine learning methods applied to predictive maintenance," Comput. Ind. Eng., vol. 137, Art. no. 106024, 2019.
[24] F. Tao, Q. Qi, L. Wang, and A. Y. C. Nee, "Digital twins and cyber-physical systems toward smart manufacturing and Industry 4.0: Correlation and comparison," Engineering, vol. 5, no. 4, pp. 653–661, 2019.
[25] K. Stouffer et al., "Guide to operational technology (OT) security," Nat. Inst. Standards Technol., Gaithersburg, MD, USA, NIST Spec. Publ. 800-82 Rev. 3, 2023.