



Dynamic Energy Demand Forecasting and Economic Optimization of Nepalese Industries through AI-Assisted Differential Equation Modeling

Avishek Shah¹, Prem Kumar Nayak², Suresh Kumar Sahani^{3*}

Faculty of Science, Technology, and Engineering, Rajarshi Janak University, Janakpurdham, Nepal^{1,2,3}

*Corresponding author: avishekshah@rju.edu.np | premnayak6511@gmail.com | sureshsahani@rju.edu.np

Received: 27 September 2026 | Revised: 11 October 2026 | Revised: 20 October 2026 | Published: 4 November 2026

Abstract

Purpose: This study develops a methodological framework integrating artificial intelligence (AI) and differential-equation models for dynamic industrial energy-demand forecasting and economic optimization in Nepal, considering production schedules, electricity conditions, tariffs, equipment characteristics, and data limitations.

Research Methodology: The framework models production and equipment dynamics using state-space ordinary differential equations, employs AI to predict short-term demand disturbances and residuals, and applies numerical integration, including the fourth-order Runge–Kutta (RK4) method, to simulate operational trajectories. Constrained receding-horizon optimization minimizes energy costs while satisfying production requirements and operational limits. Laplace-domain analysis supports interpretation of linearized system dynamics. Nepalese institutional energy sources provide contextual grounding.

Results: The framework demonstrates that industrial AI performance requires evaluation beyond forecasting accuracy, incorporating energy costs, production-normalized consumption, constraint violations, uncertainty calibration, and operational reliability. Synthetic demonstrations illustrate load profiles, forecasting comparisons, sensitivity assessments, and uncertainty analyses.

Conclusions: A staged validation approach is proposed, emphasizing chronological testing, transparent benchmarks, shadow-mode evaluation, operator supervision, and independent safety safeguards.

Limitations: The absence of plant-level operational data restricts the analysis to synthetic demonstrations, preventing empirical verification of energy savings, forecasting accuracy, or causal relationships.

Contributions: This study provides a reusable methodological framework integrating AI forecasting, dynamic system modeling, uncertainty assessment, and constrained economic optimization, establishing a foundation for future empirical applications in Nepalese industries.

Keywords: *AI-Assisted Forecasting, Constrained Optimization, Differential Equations, Industrial Energy Demand, Model Predictive Control*

How to Cite: Shah, A., Nayak, P. K., & Sahani, S. K. (2026). Dynamic Energy Demand Forecasting and Economic Optimization of Nepalese Industries through AI-Assisted Differential Equation Modeling. *Jurnal Teknik dan Informatika (JTI)*, 6(2), 103-129.

1. Introduction

Energy is an operating input, a cost exposure, and in many industrial processes a determinant of product quality and throughput. Factories may use electricity to operate motors, compressors, pumps, cooling systems, and electric process heating equipment, while thermal processes may also depend on biomass, coal, petroleum products, or other fuels. Industrial energy demand is inherently dynamic and varies according to production mix, shift patterns, equipment conditions, startup cycles, weather conditions, and maintenance activities. Forecasts can support procurement, scheduling, peak management, and energy-efficiency decisions, but forecast accuracy alone does not guarantee lower operating cost. An economically useful forecast must be translated into actions that remain feasible for the production process ([Venu & Namitha, 2026](#)).

Nepal's energy context makes this integration worth examining, but it also demands caution. The Nepal Electricity Authority publishes annual reports on the electricity system and its financial and operational activities [Nepal \(2025\)](#), while the Water and Energy Commission Secretariat publishes energy-consumption and supply reports, including provincial annexes ([Water & Energy, 2024](#)). These institutional sources are appropriate starting points for understanding national and provincial context. They do not, by themselves, provide a universal high-frequency dataset of factory-level consumption, tariffs, process states, and production outcomes. This data gap limits the validity of claims about how well a particular AI model would perform across Nepalese industries.

At the global level, the Energy Efficiency 2024 and 2025 analyses of the International Energy Agency (IEA) identify industry as a major area for energy-efficiency improvement and process optimization ([International, 2024](#); [International, 2025](#)). These global findings motivate attention to industrial energy management, but they should not be imported as Nepal-specific effect sizes. The cost and feasibility of interventions depend on local equipment, tariffs, industrial composition, supply reliability, and operational practice.

Nepal's industrial sector is heterogeneous. Small and medium enterprises may have limited metering infrastructure and rely on monthly electricity bills or manually recorded readings, whereas larger industrial plants may maintain process historians, power meters, programmable logic controllers, and shift-level production records. Industrial loads can include continuous processes, batch heating, seasonal production, refrigeration, compressed air, pumping, and machine-driven operations. A single model architecture may be reused across these settings, but its state equations, features, sampling period, and economic objective must be adapted to the process.

National energy statistics and electricity-sector annual reports can establish contextual trends and institutional responsibilities ([Nepal, 2025](#); [Water and Energy, 2024](#)). For a factory-level optimization study, however, the minimum required dataset should include timestamped energy consumption or power measurements, production quantities and product mix, shift schedules or operating modes, relevant weather or ambient conditions, equipment status, tariff categories, and operational event logs. Monthly energy bills alone can support coarse benchmarking but cannot identify sub-hourly peaks or the dynamic response to a control change.

Energy cost is not always proportional to total kilowatt-hours. Depending on the applicable tariff and contract, bills may include time-dependent energy prices, demand charges, power-factor provisions, or other components. The model must use the tariff that actually applies to the participating facility and period. This paper does not reproduce a current tariff schedule or infer one from an illustrative chart. Instead, it formulates cost generically so that the correct schedule can be inserted during implementation.

Production requirements constrain the timing and magnitude of energy use. A cement process, food-processing line, textile operation, metal workshop, or industrial dryer cannot necessarily move demand freely to a cheaper hour. A thermal batch may require a minimum soaking period, while a compressor must maintain pressure within a specified range, and a motor-driven process may be

constrained by throughput and mechanical operating limits. The optimization problem is therefore constrained scheduling, not unconstrained energy minimization.

Context for Industrial Energy Decisions in Nepal

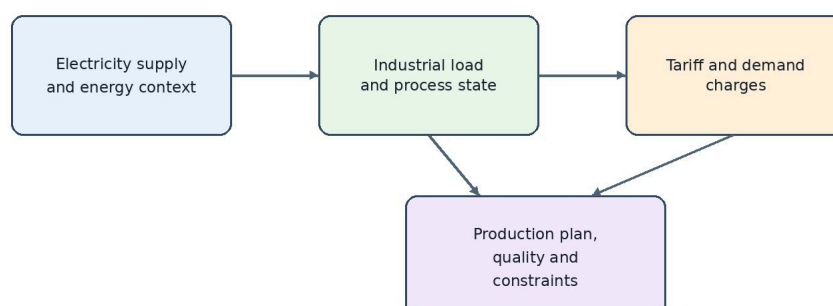


Figure 1. Contextual links for industrial energy decisions in Nepal

Figure 1 shows that the electricity supply and energy context influence the industrial load and process state, which, together with tariff and demand charges, must be reconciled with the production plan, quality requirements, and operating constraints. These interdependencies define the decision problem addressed in this paper.

AI methods can learn nonlinear associations from historical energy and process data. Differential-equation models provide a complementary description of dynamic behavior by explaining how states evolve, disturbances propagate, and processes respond over time. Hybrid approaches can therefore connect statistical forecasting with physical interpretability. The challenge is not merely to combine tools, but to allocate each tool a defensible role and test whether the combined system improves decisions against simpler alternatives.

The research gap has three dimensions. The first methodological consideration is that demand forecasting is often evaluated primarily as a prediction task, whereas industrial decision-making depends on state evolution and operational feasibility. The second contextual consideration is that public national energy statistics cannot substitute for high-frequency factory data, particularly given the diversity of industrial processes in Nepal. The third consideration concerns the risk of interpreting synthetic or retrospective simulations as evidence of actual energy savings. This study distinguishes contextual evidence, model derivation, synthetic demonstrations, and empirically validated findings to ensure transparency and avoid unsupported performance claims.

The objectives of this study are to formulate a dynamic state model for industrial energy demand, integrate an AI predictor for short-horizon demand forecasting or residual estimation, explain the complementary roles of Laplace-domain analysis and RK4 integration, define a constrained economic optimization objective, establish baseline models and validation metrics, and identify deployment requirements relevant to Nepalese industries.

This paper addresses four key research aspects, including the integration of AI-based forecasts with differential-equation state models for industrial energy demand, the complementary roles of Laplace analysis and numerical integration in prediction and optimization, the balance between economic objectives and production and equipment constraints, and the empirical evidence required for practical deployment in Nepalese factories. The study contributes a transparent methodological framework, an explicit evaluation protocol, synthetic illustrations, and a critical discussion of data requirements and implementation considerations.

The contribution is a reusable research framework rather than a claim that a new forecasting algorithm has been empirically proven superior. The framework specifies data requirements, state equations,

uncertainty handling, constraints, economic measures, and a staged deployment pathway. It is designed so that a future empirical study can insert plant data and report performance without changing the central logic.

2. Literature Review and Hypothesis/es Development

Classical load forecasting includes statistical time-series models, regression, exponential smoothing, and engineering-based models. [Bouktif, Fiaz, Ouni, and Serhani \(2018\)](#) describe a systematic forecasting workflow that includes exploratory analysis, time-series validation, and benchmark comparisons. [Talaei and Kaabouch \(2023\)](#) provide the foundational Box–Jenkins framework for identifying and estimating autoregressive integrated moving-average models. These approaches remain valuable when data are limited, trends are interpretable, or the forecasting horizon is short. Their limitations arise when operating regimes shift or nonlinear interactions dominate.

These sources support careful model comparison rather than a presumption that deeper networks are always preferable. General machine-learning foundations such as [Talaei and Kaabouch \(2023\)](#) also emphasize the importance of validation, regularization, and model complexity relative to the available data.

Reviews of AI-enabled energy consumption management discuss forecasting, anomaly detection, and demand response as connected functions ([Wang, Wang, Bhandari, & Cheng, 2024](#)). Adaptive forecasting research likewise emphasizes changing data distributions and the need to respond to nonstationarity ([Jaramillo, Pavón, & Jaramillo, 2024](#)). For industrial demand, this means that a model trained on stable production should be tested on starts, shutdowns, recipe changes, holidays, maintenance, and other operational regimes. Randomly splitting adjacent time points into training and testing sets can leak highly similar observations and overstate generalization.

Physics-informed and hybrid models attempt to balance data-driven flexibility with physically meaningful structure. Differential equations are widely used to represent dynamic systems, while state-space formulations support estimation and control. [Ljung \(1999\)](#) system-identification framework explains the need to select informative data and assess parameter uncertainty. [Sahani, Oruganti, and Kumar \(2025a\)](#) provide a rigorous introduction to feedback systems and dynamical models. These references establish the foundations for using measured transients to estimate dynamic behavior instead of treating demand as a sequence of unrelated values.

Model Predictive Control (MPC) repeatedly predicts future behavior, optimizes a finite-horizon action sequence, implements the first action, and then resolves the problem when new measurements arrive. [Michailidis, Michailidis, Minelli, Coban, and Kosmatopoulos \(2025\)](#) give a detailed treatment of MPC theory, computation, and design. [Huotari, Ritari, Vepsäläinen, and Tammi \(2020\)](#) provide a widely used treatment of predictive control. MPC's ability to represent constraints is especially relevant when production requirements, ramp rates, and equipment limits matter. Nevertheless, optimization quality depends on the prediction model, and solver feasibility does not by itself establish safe operation.

Numerical integration is essential when candidate trajectories are governed by differential equations. [Grant \(2022\)](#) describes Runge–Kutta methods and their accuracy properties. The classical fourth-order method is attractive for smooth, non-stiff ordinary differential equations because it offers a practical balance between implementation and accuracy. This method is not universally optimal because stiff systems with multiple states may require implicit numerical methods, while discontinuous switching and event-based constraints demand special consideration. Laplace transforms complement time-domain integration by exposing transfer functions and poles in linearized subsystems.

[Sahani, Oruganti, Kumar, and Karna \(2025\)](#) investigated the mechanical and operational behavior of steel manufacturing plants to improve process efficiency and performance. Mathematical modeling using Laplace transforms and the fourth-order Runge–Kutta method was applied to woolen knitting dynamics and plastic pipe performance ([Sahani, 2022](#); [Sahani et al., 2025](#)). AI-enhanced finite

element methods, IoT-based structural health monitoring, and machine learning techniques were explored for structural analysis and numerical error prediction ([Sahani et al., 2025](#); [Sahani, 2022](#); [Sahani, 2024a](#)). [Sahani \(2022\)](#); [Sahani \(2023a\)](#); [Sahani \(2023b\)](#) employed numerical optimization, Laplace transforms, and Chapman–Kolmogorov differential equations to analyze heat exchanger design and industrial plant reliability. Maintenance optimization, jaw crusher performance, and reliability-centered maintenance in cement manufacturing plants were examined in earlier studies ([Sahani et al., 2025](#); [Sahani, 2022](#); [Sahani, 2024b](#)). Furthermore, numerical methods were applied to mechanical vibration and heat conduction problems ([Sahani et al., 2025](#); [Sahani, Sah, & Sahani, 2023b](#)), while the reliability of plywood production facilities was assessed using Laplace transforms and the fourth-order Runge–Kutta method ([Sahani et al., 2025](#); [Sahani, Oruganti, & Kumar, 2025c](#)). Overall, these studies highlight the significance of mathematical modeling, numerical techniques, artificial intelligence, and reliability analysis in enhancing industrial efficiency, system performance, and maintenance optimization.

Energy management also requires a measurement and governance structure. ISO 50001:2018 sets requirements for energy-management systems, including systematic improvement and monitoring. The global efficiency reports of the IEA provide current policy and sector context ([International, 2024](#); [International, 2025](#)). Nepal's WECS and NEA publications provide local institutional grounding ([Nepal, 2025](#); [Water, & Energy, 2024](#)). Taken together, these sources suggest that forecasting, optimization, and measurement-and-verification should be treated as a connected cycle.

The broader forecasting literature also considers smart-grid load forecasting and integrated energy-system planning ([Biswal, Deb, Datta, Ustun, & Cali, 2024](#); [Krishnamurthy, Adewuyi, Ratshitanga, & Moodley, 2024](#)). The gap addressed here is the limited connection, in many forecasting demonstrations, between predictive error and downstream economic outcomes under operational constraints.

3. Research Methodology

3.1 Framework Overview and Data Pipeline

Figure 2 presents the overall architecture. Meter readings and process tags are first validated and aligned. A forecasting model estimates near-term demand, a disturbance, or a residual correction. The differential-equation model propagates the system state under candidate decisions. The optimizer evaluates economic cost and production constraints, and a safety supervisor or operator reviews the recommendation. During initial deployment, recommendations remain advisory, while independent safety systems retain full operational authority.

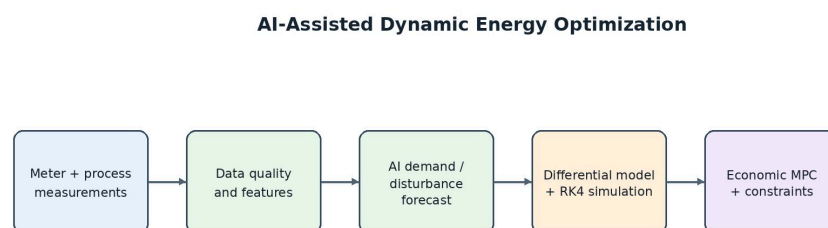


Figure 2. AI-assisted forecasting and economic optimization architecture

As shown in Figure 2, the framework proceeds sequentially from meter and process measurements, through data-quality checks and feature generation, AI-based demand or disturbance forecasting, and differential-model simulation with RK4, to economic model predictive control subject to constraints.

The data pipeline should preserve the original raw records and create a separate quality-controlled dataset. Time stamps must be synchronized, units standardized, duplicate samples detected, and gaps labelled. Long gaps should not be silently interpolated across shutdowns or process transitions. Feature generation should use only information available at the prediction time. Each sample should

carry operating-mode context where possible, because the same energy level can mean normal production in one regime and an anomaly in another.

Candidate predictors include persistence, linear regression, regularized regression, tree ensembles, and recurrent neural networks. The selection should depend on data volume, inference requirements, auditability, and the forecast horizon. For a plant with limited data, a simple baseline may be the most defensible choice. AI can forecast demand directly, estimate a disturbance, or learn the residual between a physical model and measured behavior. The residual approach should be bounded and monitored so it does not conceal a major model defect.

3.2 Data Quality and Measurement Architecture

An energy forecasting model can only be as reliable as the measurement chain that feeds it. Electrical power may be recorded by a smart meter, a panel meter, a programmable logic controller, or a building-management system, and these sources can differ in accuracy, aggregation interval, and clock synchronization. A study should document whether a reading is instantaneous power, interval energy, cumulative meter energy, or a calculated value. If cumulative readings are differenced to estimate interval energy, meter resets and rollover events must be detected before the values are used for training ([Chicco, 2021](#)).

Sampling frequency should be selected according to the dynamics and billing interval of interest. A monthly bill cannot reveal short-duration peaks, while high-frequency data may impose storage and processing costs without improving a slow supervisory controller. The chosen frequency should be justified by the shortest relevant dynamic, the demand-charge interval, and the control sampling period. Resampling must preserve energy: averaging instantaneous power and summing interval energy are different operations and should not be confused (([Dewangan, Abdelaziz, & Biswal, 2023](#)); ([Mathumitha, Rathika, & Manimala, 2024](#))).

Timestamp quality is particularly important when joining energy records to production logs. A shift log may use local time, a historian may store UTC, and a meter may have a clock offset. Daylight-saving transitions are less central in Nepal than in some jurisdictions, but clock drift, manual data entry, and inconsistent time zones can still misalign events. Before model fitting, a researcher should inspect lag relationships and verify several known production transitions against the raw records.

Sensor validation should include physical plausibility checks, calibration records, duplicate detection, and cross-sensor consistency. Missingness should be categorized as random dropout, planned shutdown, communication failure, or equipment outage whenever that information is available. The model should not impute across long shutdowns as if the process had continued normally. Where data quality is insufficient for control, the appropriate conclusion is that more instrumentation is required not that a more complex AI model will solve the problem.

3.3 Differential-Equation Model and Laplace Analysis

Let denote a vector of industrial process states, such as thermal temperature, stored energy, pressure, equipment condition, or an aggregate production state. Let $u(t)$ denote controllable inputs, $r(t)$ production requirements, and $d(t)$ disturbances such as ambient conditions or unplanned load. A general nonlinear state model is given in Formula 1, and energy demand is expressed in Formula 2, where θ is a parameter vector and represents measurement noise and omitted effects. Figure 3 shows the conceptual relationship among production, dynamic states, and demand.

$$\dot{x}(t) = f(x(t), u(t), r(t), d(t); \theta) \quad (1)$$

$$P(t) = g(x(t), u(t), r(t); \theta) + \varepsilon(t) \quad (2)$$

State-Space View of Industrial Energy Demand

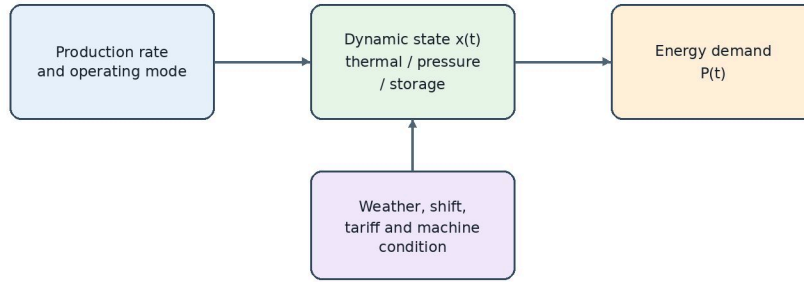


Figure 3. State-space view of dynamic industrial energy demand

Figure 3 indicates that the production rate and operating mode drive the dynamic state $x(t)$, which is also affected by weather, shift, tariff, and machine condition, and that the energy demand $P(t)$ is obtained from this state rather than directly from the inputs.

For a thermal subsystem, a lumped energy balance can be written as Formula 3, where C_{th} is the effective thermal capacitance, η is the conversion efficiency, $Q(t)$ is the commanded heat input, UA is the aggregate heat-loss conductance, $T_a(t)$ is the ambient temperature, and $Q_{load}(t)$ is the process heat extraction.

$$C_{th} \frac{dT(t)}{dt} = \eta Q(t) - UA[T(t) - T_a(t)] - Q_{load}(t) \quad (3)$$

Defining $a = UA/C_{th}$ and $b = \eta/C_{th}$ yields the first-order model in Formula 4, where $d(t)$ collects ambient and process disturbances. This model is illustrative, as a specific industrial plant may require multiple thermal states, nonlinear radiation effects, time delays, or material-flow dynamics to accurately represent its operational behavior.

$$\frac{dT(t)}{dt} = -aT(t) + bQ(t) + d(t), \quad a = \frac{UA}{C_{th}}, \quad b = \frac{\eta}{C_{th}} \quad (4)$$

Linearizing around an operating point and applying the Laplace transform under zero initial deviation gives Formula 5. The transfer function from heat input to temperature, , and the dominant time constant, G_{TQ} , follow as in Formula 6.

$$(s + a)\Delta T(s) = b\Delta Q(s) + \Delta D(s) \quad (5)$$

$$G_{TQ}(s) = \frac{\Delta T(s)}{\Delta Q(s)} = \frac{b}{s + a}, \quad \tau = \frac{1}{a} = \frac{C_{th}}{UA} \quad (6)$$

The pole at $s = -a$ describes the local response speed. This is useful for interpreting transients and estimating parameters, but a linear transfer function should not be assumed valid across all loads or operating modes.

RK4 is used to propagate candidate trajectories in the time domain. For $x' = f(t, x, u)$ with step size h , the four stage slopes are defined in Formula 7, Formula 8, Formula 9, Formula 10, where u_{mid} denotes the input at the midpoint of the step, and the state is updated by Formula 11.

$$k_1 = f(t_n, x_n, u_n) \quad (7)$$

$$k_2 = f\left(t_n + \frac{h}{2}, x_n + \frac{h}{2}k_1, u_{\text{mid}}\right) \quad (8)$$

$$k_3 = f\left(t_n + \frac{h}{2}, x_n + \frac{h}{2}k_2, u_{\text{mid}}\right) \quad (9)$$

$$k_4 = f(t_n + h, x_n + hk_3, u_{n+1}) \quad (10)$$

$$x_{n+1} = x_n + \frac{h}{6}(k_1 + 2k_2 + 2k_3 + k_4) \quad (11)$$

Under smoothness assumptions, RK4 has fourth-order global accuracy. Numerical convergence should be checked by reducing the step size and comparing trajectories and decisions. If the system is stiff, an implicit solver may be more appropriate.

Figure 4 illustrates the receding-horizon sequence, which involves defining the continuous model, estimating the current state and parameters, simulating candidate trajectories, optimizing the decision sequence, applying only the first control action, and repeating the process as new measurements become available. The distinction between forecasting and dynamic modeling is essential because the AI predictor estimates uncertain exogenous behavior, the differential equation maintains dynamic consistency, and the optimizer determines feasible operational actions within explicitly defined constraints.

Receding-Horizon Simulation and Optimization



Figure 4. Receding-horizon workflow with differential-equation simulation

Figure 4 shows the closed loop of the receding-horizon procedure: the state and parameters are estimated, candidate trajectories are simulated with RK4, cost is optimized under hard constraints, only the first action is applied, and the cycle is repeated after new measurements are received.

3.4 State Estimation and Observability

The state vector in a differential-equation model may include quantities that are not directly measured. For example, the effective thermal state of a product may differ from the measured air temperature,

while the true heat-loss coefficient may change slowly with fouling or door openings. A state estimator combines model predictions with measurements to infer these latent quantities. Depending on the model and noise assumptions, candidates include a Kalman filter, extended Kalman filter, unscented Kalman filter, or moving-horizon estimator. The choice should reflect the degree of nonlinearity and the available computational resources.

Observability determines whether the measured outputs contain enough information to infer the internal state. A model with multiple thermal masses may be mathematically elegant but poorly observable if only one temperature sensor is available. Additional states should therefore be introduced only when the data can support them or when reliable physical priors constrain their behavior. A more complex model may improve fit on training data while worsening parameter uncertainty and control reliability.

State estimates should include uncertainty, and the optimizer should respond conservatively when that uncertainty is large. If a sensor fails, the system may use a redundant measurement or model-based estimate for a limited time, but the permitted degraded mode must be defined in advance. The model should never convert an unobservable state into an apparently precise value merely because the optimization algorithm requires a number.

3.5 Economic Optimization Formulation

At decision time k , the optimizer selects an input sequence $(U_{k:k+H-1})$ over a horizon H . A general objective is given in Formula 12, subject to the constraints in Formula 13. Here, $C_{(k+j)}$ is the applicable energy price, P_j is the predicted purchased power, P_j is the sampling interval, E_j may represent a demand-peak or emissions term, λ_E , λ_R , λ_Q , and are weighting coefficients, L_Q penalizes deviation from production or quality requirements, and J_T is a terminal cost. The exact terms depend on the industrial process and applicable billing arrangement.

$$J = \sum_{j=1}^H \left[c_{k+j} P_j \Delta t + \lambda_E E_j + \lambda_R \|u_j - u_{j-1}\|^2 + \lambda_Q L_Q(x_j, r_j) \right] + J_T \quad (12)$$

$$u_{\min} \leq u_j \leq u_{\max}, \quad |u_j - u_{j-1}| \leq R_u \Delta t, \quad x_j \in X_{\text{safe}} \quad (13)$$

In Formula 13, and are the actuator bounds, is the maximum ramp rate, and is the approved safe operating set. The constraints include equipment limits, minimum and maximum production rates, ramp-rate bounds, product temperature or pressure windows, maintenance availability, and completion deadlines. Hard safety constraints should not be converted into finite penalties that can be violated to save money. Soft preferences, such as smoother actuator movement or modest deviation from a preferred production schedule can be represented as objective penalties. If the initial state makes the problem infeasible, the controller needs a separately defined recovery mode rather than silently relaxing a critical limit.

Tariff-aware optimization requires the actual tariff schedule and the plant's billing rules. Figure 5 illustrates how time-varying electricity prices can influence production scheduling decisions using a hypothetical relative tariff index rather than official tariff data from the Nepal Electricity Authority. A future Nepal-specific study must obtain and cite the applicable tariff notice or customer bill, including demand charges and power-factor rules where relevant. The economic objective should be computed from measured or predicted energy quantities using consistent units.

Production-normalized energy metrics are essential. If energy use falls because a plant produces less, the result is not necessarily an efficiency gain. The evaluation should report total energy, energy per

accepted unit, throughput, product quality, production downtime, and cost. Where the production mix changes, normalization should use a transparent model or compare matched production periods. Cost savings and energy savings must be reported separately because tariff shifts can change the bill without reducing consumption.

Table 1. Illustrative objective terms and constraint classes

Component	Representative Formulation	Type	Purpose
Energy charge	$\Sigma[c_j P_j \Delta t]$	Objective	Reflect applicable unit prices
Peak demand	$\max(P_j)$ or tariff-defined demand	Objective/constraint	Represent demand charge if applicable
Production tracking	$\Sigma[w_r (r_j - r_j^*)^2]$	Soft objective	Meet schedule and throughput targets
Equipment limits	$u_{min} \leq u_j \leq u_{max}$	Hard constraint	Respect actuator ratings
Ramp limit	$ u_j - u_{j-1} \leq R_u \Delta t$	Hard constraint	Avoid abrupt changes
Process safety	$x_j \in X_{safe}$	Hard constraint	Preserve approved operating envelope

Table 1 summarizes how each economic term and operational limit enters the formulation in Formula 12 and Formula 13, distinguishing objective terms and soft penalties from hard constraints that must not be traded against cost.

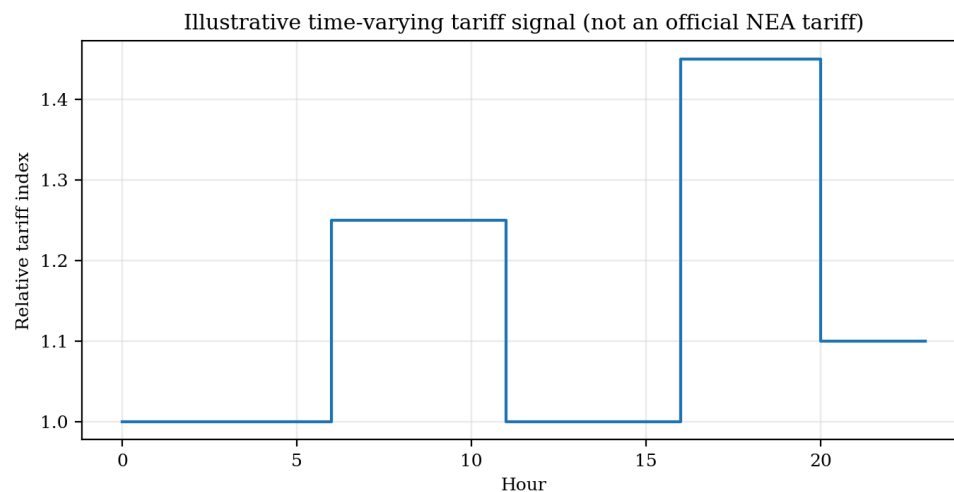


Figure 5. Illustrative time-varying tariff signal; not an official NEA tariff

Figure 5 depicts a stepwise relative tariff index over 24 hours, with a higher-price block in the morning and the highest block in the evening. It is used only to explain how a time-varying price signal may influence scheduling and is not an official NEA tariff.

3.6 Computational Complexity and Solver Selection

The computational burden of the proposed method depends on the number of states, prediction horizon, number of candidate control moves, nonlinearities, and the solver used for optimization. A low-order linear model with quadratic objective and linear constraints may lead to a quadratic program that can be solved efficiently. Nonlinear state equations, discrete operating modes, minimum run times, and nonconvex tariffs may require nonlinear or mixed-integer optimization. A method that works offline may still miss its deadline on a low-cost edge computer.

Move blocking can reduce the number of decision variables by allowing control changes only at selected points in the horizon. A shorter horizon, simpler model, or warm-started solver can also reduce computation, but each may sacrifice predictive performance. The engineering decision should be based on a measured trade-off between solve time and decision quality. Worst-case execution time is more important than average solve time for a controller that must produce a valid action by a fixed deadline.

Numerical integration and optimization errors should be considered together. If RK4 predicts the state with small truncation error but the optimizer uses an inaccurate linearization, the chosen action may still be poor. If the solver terminates early, the solution may be suboptimal or infeasible. A deployment report should state solver tolerances, termination conditions, numerical step, horizon, and what the controller does when no feasible solution is returned.

3.7 AI-Assisted Forecasting Design

Forecasting should align with the control decision interval. If optimization is updated every 15 minutes, a model that predicts only a daily total may not resolve intraday peaks. Features may include recent demand lags, shift and calendar indicators, production plan, equipment state, ambient conditions, and tariff periods. Each feature must be available at the time of inference, allowing planned production data to be incorporated only when genuinely known in advance, while excluding actual future production data to prevent data leakage.

The baseline hierarchy should start with persistence and a mode-specific average, then linear regression, a tree ensemble, and more complex deep-learning methods if justified. Recent forecasting reviews ([Baur et al., 2024](#); [Eddaoudi et al., 2024](#); [Jaramillo et al., 2024](#); [Tsegaye et al., 2024](#); [Wang et al., 2024](#)) support the importance of model selection, explainability, and nonstationarity. A model should not be selected from a single random split. Chronological train-validation-test blocks are more realistic, and tests should include unusual but operationally relevant regimes.

Prediction intervals or ensembles can represent uncertainty, but uncertainty must be calibrated. A nominal 90% interval should contain approximately 90% of held-out observations under the conditions for which it is claimed. Coverage should be reported by operating mode and near constraint boundaries. Changes in the process distribution may reduce prediction interval coverage, making drift monitoring and conservative fallback mechanisms essential components of the system rather than optional features.

Figure 6 provides a synthetic illustration of how forecasts should be compared on the same held-out series. Its curves are not measured evidence and do not imply that the AI-assisted forecast outperforms a baseline in real factories. A real study should report Mean Absolute Error (MAE), Root Mean Square Error (RMSE), bias, peak-event recall, interval coverage, and inference latency, with uncertainty assessed across multiple chronological test windows.

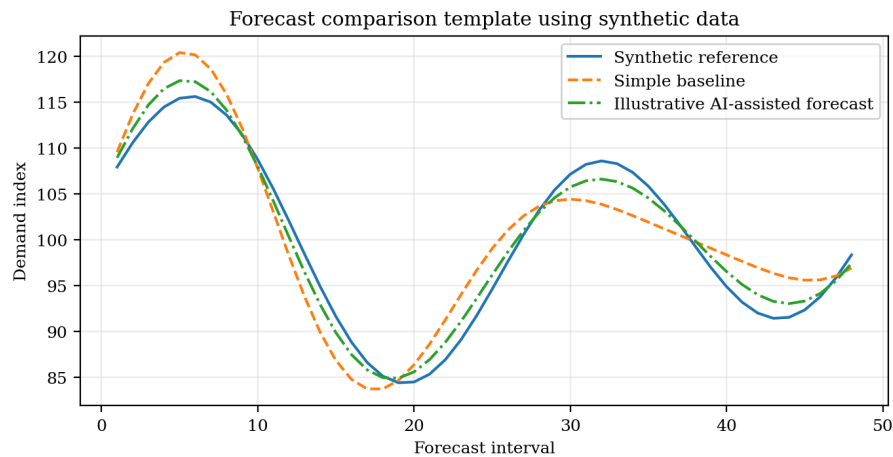


Figure 6. Forecast comparison template using synthetic data

Figure 6 compares a synthetic reference signal with a simple baseline and an illustrative AI-assisted forecast over 48 forecast intervals. Because all three series are synthetic, the figure serves only as a template for presenting held-out forecast comparisons.

3.8 Forecast Horizons and Decision Timescales

Energy forecasting operates at multiple timescales, and a single model should not be expected to serve every decision. A minutes-ahead forecast may support peak shaving, compressor sequencing, or thermal control. A day-ahead forecast can support production scheduling and energy procurement. A seasonal forecast can support capacity planning and maintenance. The explanatory variables and acceptable error differ across these horizons: immediate predictions rely heavily on recent state and machine status, while longer horizons depend more on production plans, weather, and business calendars. The proposed framework is principally a short-horizon supervisory controller, but its forecast module can be adapted to longer planning tasks.

The control horizon must be aligned with the physical time constants and the rate at which decisions can be changed. If a process responds slowly, frequent changes in the command may create unnecessary actuator movement without improving the temperature trajectory. If a process includes rapid cycling equipment, an overly slow update interval can miss peaks. A useful design begins by measuring the dominant time constants, then choosing a sampling period that captures relevant dynamics while leaving enough time for estimation and optimization. The RK4 integration step is a separate numerical choice and should be smaller than the decision interval when necessary.

A forecast can also be useful even when the optimizer does not directly control the equipment. For example, a day-ahead demand estimate may help schedule energy-intensive batches, while an intraday forecast can alert an operator to an impending peak. These advisory uses have lower automation risk and can generate evidence about forecast quality before control authority is considered. The deployment plan should distinguish advisory forecasting, decision support, and automatic actuation as separate maturity levels.

3.9 Experimental Design and Evaluation Metrics

A credible evaluation separates prediction performance from control performance. Prediction metrics include MAE, RMSE, mean absolute percentage error where denominators are meaningful, bias, and interval coverage. MAPE can become unstable when actual demand is near zero, so MAE or scaled errors may be preferable for shutdown-heavy data. Peak-focused metrics are useful when demand charges or equipment limits make high-load intervals disproportionately important.

Control metrics include total energy, cost under the applicable tariff, maximum demand, production-normalized energy, time within operating bounds, number and duration of violations, actuator movement, production throughput, quality yield, and solver timeout rate. Each metric evaluates a

distinct aspect of system performance. A controller that reduces operational costs but increases scrap rates may not deliver overall economic benefits, while a forecasting model that improves average prediction accuracy but fails to capture peak demand may be unsuitable for effective demand management.

The experimental design should compare at least three baselines: a current rule-based or manual policy, a forecasting-only benchmark, and a physics-based optimizer without AI corrections. The hybrid model should be tested against these baselines using identical initial conditions and disturbances in simulation. In field trials, matched shifts or production-normalized comparisons may be used, but confounders such as weather, product mix, maintenance, and operator decisions must be recorded.

Table 2. Recommended forecast and control evaluation metrics

Metric	Definition	Interpretation	Caution
MAE	Mean absolute forecast error	Average absolute miss	Does not emphasize peaks
RMSE	Square root of mean squared error	Penalizes large errors	Outlier-sensitive
Bias	Mean forecast minus actual	Systematic over/underforecast	Opposing errors can cancel
Interval coverage	Share of observations inside forecast interval	Uncertainty calibration	Must be checked by regime
Energy intensity	Energy / accepted output	Production-normalized efficiency	Requires consistent output basis
Constraint violations	Count/duration outside approved bounds	Operational feasibility	Sampling may miss short events

Table 2 lists the recommended metrics together with their interpretation and cautions, showing that no single error measure is sufficient and that forecast and control performance must be reported jointly.

Figure 7 is a synthetic weekly load pattern, included to illustrate shift and daily cycles that a forecasting pipeline may need to represent. Figure 8 depicts a conceptual feasible region, and Figure 9 shows a sensitivity-analysis template. Neither figure contains measured coefficients. In a real analysis, sensitivity values must be computed from fitted parameters or controlled perturbations and accompanied by confidence intervals. Figure 10 illustrates why point predictions are inadequate near an operating threshold, the tail of the predictive distribution may cross a constraint even when its mean does not.

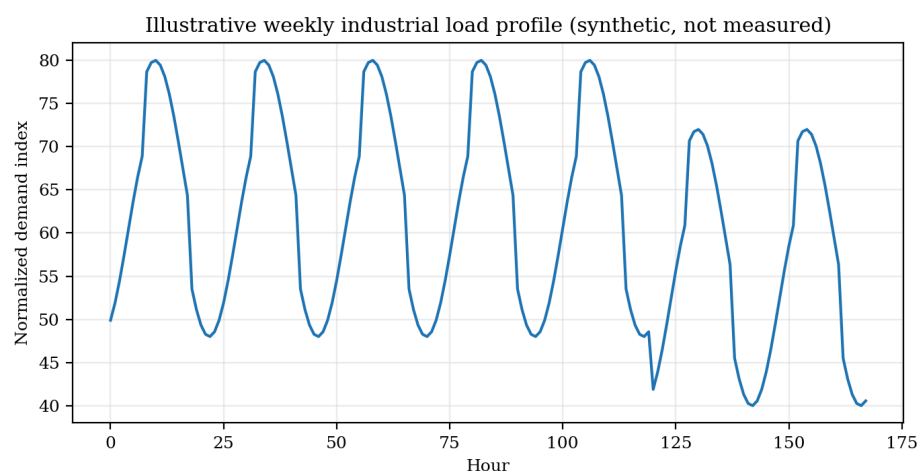


Figure 7. Illustrative weekly industrial load profile; synthetic, not measured

Figure 7 displays a synthetic 168-hour profile with five similar workday cycles followed by two lower-amplitude cycles, illustrating the shift and daily patterns that a forecasting pipeline must be able to represent.

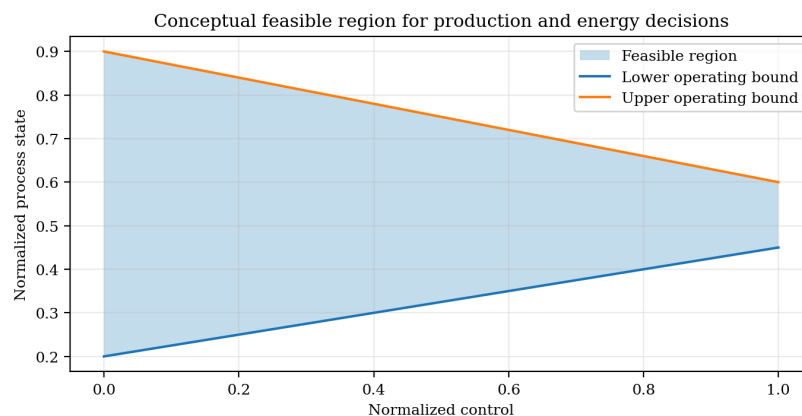


Figure 8. Conceptual feasible region for production and energy decisions

Figure 8 shows that the feasible region between the lower and upper operating bounds narrows as the normalized control increases, indicating that only a subset of control-state combinations is admissible.

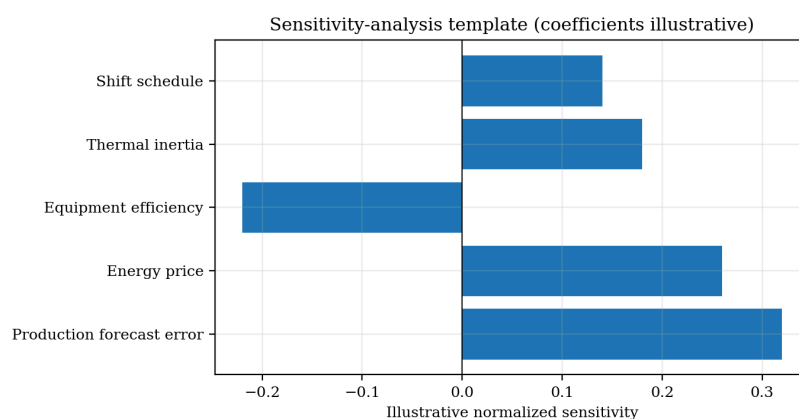


Figure 9. Sensitivity-analysis template; coefficients are illustrative

Figure 9 ranks illustrative normalized sensitivities, with production forecast error and energy price showing the largest positive values and equipment efficiency showing a negative value. These coefficients are illustrative and are not estimates for Nepalese industries.

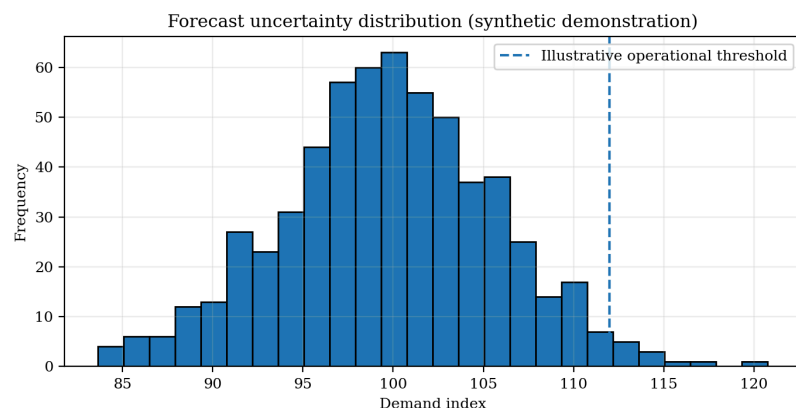


Figure 10. Forecast uncertainty distribution; synthetic demonstration

Figure 10 shows a synthetic distribution of demand predictions centered near an index of 100, with a small upper tail extending beyond the illustrative operational threshold. This demonstrates why the tail of the distribution, rather than only its mean, must be considered near a constraint.

3.10 Model Identification, Uncertainty, and Robustness

Parameter estimation should be treated as a scientific inference task rather than a one-time curve fit. For a linearized model, least-squares or prediction-error methods can estimate coefficients from observed inputs and outputs (Ljung, 1999). Yet an apparently good fit does not guarantee that individual parameters are identifiable. Thermal capacity and heat-loss conductance, for example, may be strongly correlated when the data cover only a narrow temperature range. Informative excitation, engineering priors, and confidence intervals help reveal this ambiguity.

Residual analysis should test whether errors are centered, temporally independent enough for the intended purpose, and free of systematic patterns across operating modes. If residuals correlate with throughput or ambient temperature, the model may be missing a relevant state or input. AI residual correction can improve local forecasts, but a residual model that compensates for a structural defect may fail under extrapolation. It is therefore important to compare physical-only, residual-corrected, and direct forecasting variants on the same held-out conditions.

Uncertainty comes from sensor error, parameter estimation, forecast error, unobserved disturbances, and structural model mismatch. These sources should be distinguished where possible. Parameter uncertainty can be propagated through simulations by sampling plausible parameter vectors, while forecast uncertainty can be assessed using scenario-based analysis and sensor uncertainty can be incorporated into the state estimation process. A risk-aware optimizer can then reserve a margin or enforce a probabilistic constraint, provided that the probability model is calibrated and the consequences of violation are understood.

Robustness tests should include a sustained forecast bias, a sudden load step, a temporary sensor outage, an efficiency reduction, a recipe change, and a solver timeout. The system should either continue within a verified envelope or return control to a known safe policy. Figure 10 is a simple visual reminder that the distribution tail, not only the mean matters near an operating threshold. The illustrated distribution is synthetic and must not be interpreted as a measured probability of overload.

3.11 Research Protocol for a Nepalese Industrial Pilot

A pilot should begin by selecting one process boundary with clear inputs and outputs. The boundary should include all energy carriers that materially contribute to the process and define how shared utilities are allocated. The study team should agree with plant management on operating constraints, production-quality measures, baseline periods, and criteria for stopping the pilot. These decisions should be documented before the model is tuned.

The baseline phase should collect data across representative shifts, production mixes, start-ups, shutdowns, and ambient conditions. It should also document maintenance events and abnormal periods. A short dataset dominated by one stable production mode is insufficient to validate behavior under recipe changes. If the observation period does not include an important regime, that regime should be marked as outside the validated envelope and handled conservatively.

The modeling phase should compare simple forecasts against candidate AI models, identify the differential-equation parameters, and test model residuals. The validation phase should evaluate forecasts on untouched chronological data and check RK4 convergence. The simulation phase should run stress scenarios, including demand surges, sensor faults, and infeasible initial conditions. Only after these phases pass should the optimizer run in shadow mode.

During shadow mode, every recommendation should be logged with input timestamps, model version, predicted trajectory, active constraints, objective terms, solve time, and the operator's decision where available. The review should examine whether the recommendation is physically sensible and whether

unmodeled production rules are repeatedly invoked. A pilot should start with a narrow operating envelope and conservative limits, with a clear manual override and rollback process.

The final evaluation should compare the pilot with a pre-specified baseline and report energy, cost, output, quality, peak demand, and safety indicators. Any change in production volume or product mix should be accounted for. The final report should include uncertainty intervals and disclose negative or inconclusive results. If the effect is smaller than measurement uncertainty, the correct conclusion is that the evidence is inconclusive rather than that the intervention has succeeded.

4. Results and Discussions

4.1 Methodological Demonstration Results

No plant-level interval data were supplied for this manuscript. Consequently, the results section does not claim measured prediction accuracy, cost reduction, energy savings, or industrial validation. Instead, it demonstrates the analysis that a future empirical study should conduct and identifies what can be concluded from the model structure alone.

Figure 7 shows an illustrative 168-hour demand profile with daily variation and a simplified workday pattern. This synthetic series is not a representation of measured Nepalese industrial demand. Its purpose is to show how operating schedules can create predictable structure while leaving residual variability that a model must forecast. The actual profile of a Nepalese factory should be obtained from interval meters and reconciled against shift logs and production quantities.

Figure 6 compares a synthetic reference signal, a simple baseline, and a notional AI-assisted forecast. The illustration demonstrates the format of a benchmark comparison but cannot establish which model is better. A defensible result would report the exact test window, number of observations, metrics by regime, prediction intervals, and inference time. It would also test whether better forecast accuracy changes the selected action or the final economic outcome.

Figure 4 and the differential equations establish a computational route from predicted disturbances to candidate trajectories. The model can be checked against an analytical first-order response in a simple case, then tested with smaller RK4 steps. If trajectory predictions change materially as the step is reduced, the integration configuration is not reliable enough for decision-making. Even converged numerical trajectories can be wrong if the underlying model is poorly identified.

Figure 5 demonstrates how an optimizer may respond to time-varying relative prices. It does not use an official tariff schedule. A plant-specific optimizer would incorporate actual electricity tariffs and production constraints, enabling flexible operations to be shifted away from high-cost periods while ensuring that required production completion times are maintained. Whether that strategy lowers total cost depends on thermal losses, start-up energy, production scheduling, and any demand charge. It cannot be inferred from the schematic alone.

Figure 8 illustrates that only a subset of control-state combinations may be feasible. Figure 9 suggests that production forecast error, price, equipment efficiency, thermal inertia, and shift schedule can be examined through sensitivity analysis. The displayed coefficients are illustrative and must not be reported as estimated Nepalese industrial sensitivities. In an empirical study, one-at-a-time perturbations, global sensitivity analysis, or Monte Carlo sampling should be performed over credible parameter ranges.

Figure 10 presents a synthetic distribution of demand predictions alongside an illustrative threshold, emphasizing the importance of accounting for uncertainty when evaluating the margin between a predicted operating state and a hard operational constraint. A risk-aware optimizer may maintain a reserve margin or use scenario constraints, but the margin must be derived from calibrated error and the consequences of a violation. A generic threshold from an illustrative figure has no operational authority.

Figure 11 outlines the staged validation pathway. The most important analytical finding of this methodological study is that credible deployment requires comprehensive evidence of data integrity, predictive validity, dynamic model adequacy, numerical convergence, optimization feasibility, and economic benefits. If any layer fails, a favorable metric elsewhere is insufficient to justify closed-loop use.

Staged Validation and Assurance Pathway

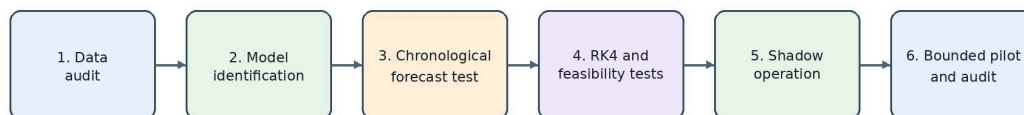


Figure 11. Staged deployment and assurance pathway

Figure 11 illustrates six sequential stages, including data auditing, model identification, chronological forecast testing, RK4 integration and feasibility testing, shadow operation, and a bounded pilot with auditing. Each stage must be successfully completed before proceeding to the next to ensure systematic validation and operational reliability.

4.2 Comparison with Alternative Architectures

The proposed framework integrates three methods with different strengths. AI is suited to learning nonlinear patterns and forecasting disturbances from historical data. Differential equations express the evolution of physical or operational states. Laplace analysis supports interpretation of linearized dynamics, while RK4 enables simulation of time-varying trajectories. Constrained optimization converts these predictions into decisions. Their combination is valuable only if the interfaces among components are validated.

Compared with a forecast-only approach, the hybrid framework evaluates what the prediction implies for a future state and an action. Compared with a physics-only model, AI can adapt to residual patterns and exogenous disturbances that are difficult to model from first principles. Compared with a purely learned control policy, an explicit model and hard constraints provide greater transparency. These potential advantages do not establish universal superiority, as variations in data quality, model mismatch, computational limitations, and process complexity may alter or even reverse the expected trade-offs.

Table 3. Comparison of forecasting and decision architectures

Architecture	Advantages	Principal Risk	Recommended Evaluation
Persistence / simple baseline	Low cost; transparent; strong short-horizon reference	Cannot represent changing regimes well	Chronological error and peak miss rate
AI-only forecast	Learns nonlinear patterns from tags	No inherent physical or constraint guarantee	MAE, RMSE, bias, interval calibration
Physics-only model	Interpretable state evolution	Parameter mismatch and omitted disturbances	Residual diagnostics and transient tests
AI + differential equation	Forecast flexibility plus dynamic consistency	Data, calibration, and integration complexity	Forecast plus state and constraint metrics
Hybrid constrained optimizer	Links prediction to economic decisions	Solver infeasibility and model risk	Cost, production-normalized energy, violations

Table 3 shows that each architecture offers distinct advantages but also carries a principal risk, and that the hybrid constrained optimizer requires the broadest evaluation, combining cost, production-normalized energy, and constraint-violation metrics.

The Nepal-specific implication is primarily methodological. Public institutional data provide a starting point for context, but factory-level adoption requires plant-specific metering and production data. Smaller enterprises may benefit first from energy baselining, anomaly detection, and advisory forecasts rather than full MPC. A low-order model, simple baseline, and transparent economic dashboard can be more maintainable than a high-capacity model that needs extensive labelled data and continuous connectivity.

Energy efficiency should be understood as a measurement-and-management process, not as a one-time algorithmic intervention. ISO 50001 emphasizes systematic monitoring and continual improvement. Accordingly, an AI model should be governed through versioning, performance audits, change management, and documented operational ownership. Savings claims should be normalized for production and evaluated over sufficient periods to distinguish intervention effects from ordinary variability.

4.3 Demand Peaks, Load Shifting, and Economic Interpretation

Peak demand is a distinct objective from total energy consumption. A plant may use the same number of kilowatt-hours but incur a higher bill if consumption is concentrated in a short interval that determines billed demand. Conversely, shifting flexible loads may lower peak demand while leaving energy use unchanged or even slightly increasing it due to additional starts or thermal losses. The economic objective must reflect the actual billing rules rather than assuming that minimizing the integral of power is equivalent to minimizing the bill.

For a demand-charge tariff, a simplified objective can include a term proportional to the maximum predicted demand over the billing interval. In practice, the billing demand may be calculated over a specific averaging window and may include ratchets or contract-specific conditions. The optimization should reproduce the actual rule. It should also recognize that the peak may be caused by multiple assets acting together, so plant-wide coordination may be more effective than optimizing a single machine in isolation.

Load shifting is constrained by production dependencies, as delaying a batch oven is feasible only when downstream processes can accommodate the adjustment, while compressors must maintain adequate header pressure and refrigeration systems must operate within safe temperature limits. These constraints are naturally represented in a state-space model and a constrained schedule. However, shifting a load into a lower-price period can create a new peak or increase the time that equipment remains warm and idle. The full trajectory and total cycle cost must be simulated.

Figure 5 deliberately uses a normalized illustrative price signal rather than an official tariff. This separation is important for academic integrity. An empirical study should attach the tariff source, effective dates, customer class, and applicable demand calculation. If tariff changes occur during the study period, the analysis should either normalize costs to a common tariff or clearly distinguish operational changes from price changes.

4.4 Multi-Objective Optimization and Stakeholder Trade-Offs

Industrial energy decisions often involve conflicting objectives: lower cost, lower energy consumption, lower emissions, high product quality, short cycle time, and limited equipment wear. A weighted sum can combine these terms, but its result depends on scaling and weights. If the cost term is measured in currency and the quality term in an arbitrary index, weights may dominate simply because the units differ. Each term should be normalized or assigned a weight with an explicit operational interpretation.

A Pareto analysis can show solutions for which no objective can be improved without worsening another. Rather than presenting one optimum as objectively best, a decision-support interface can expose a small set of feasible trade-offs. Production managers may prioritize delivery, maintenance engineers may prioritize equipment life, and finance staff may prioritize cost. The chosen solution should be documented with the weights, active constraints, and expected outcomes so that the decision can be audited.

The economic objective should include the cost of rejected production when credible data are available. A controller that reduces electricity expenditure but increases scrap can worsen total economic performance. Likewise, a strategy that minimizes immediate energy use may lengthen cycle time and reduce capacity. A fair comparison therefore includes output, quality, downtime, and maintenance indicators rather than treating energy cost as the sole performance measure.

Multi-objective optimization also helps explain why the same controller can produce different actions under different business conditions. High electricity prices may justify postponing a flexible production batch, whereas urgent delivery requirements may make such scheduling adjustments impractical. The mathematical model does not eliminate the need for business priorities. It makes the trade-offs explicit and allows stakeholders to inspect the implications of a decision.

4.5 Statistical Reporting and Measurement of Benefits

Energy outcomes vary naturally across days and shifts. A small difference between a baseline and a pilot can be due to production mix, ambient conditions, equipment cleanliness, operator behavior, or random variation. Statistical reporting should include the number of evaluated periods, the distribution of outcomes, uncertainty intervals, and the normalization method. Where enough repeated observations exist, paired comparisons across similar shifts may reduce variability, but the matching criteria must be stated before outcomes are inspected.

For a controlled intervention, the strongest design may be randomized or staggered introduction where operationally feasible. In many factories, randomization is impractical because the process must meet delivery commitments. In such cases, a matched-period comparison, interrupted time series, or difference-in-differences design may be considered, provided its assumptions are discussed. The method should control for changes in production volume, product type, weather, maintenance, and tariff whenever those factors materially affect energy use.

The energy baseline should be chosen with care. A baseline based only on the previous month may be misleading if the product mix or operating hours changed. A regression-based baseline can adjust for production and weather, but it must be fitted on a period representative of normal operation and checked for extrapolation. Baseline models should not be tuned after seeing the intervention outcome in a way that systematically favors the proposed controller.

The study should distinguish gross from net benefit. Gross benefit may include a reduction in purchased energy or billed peak. Net benefit subtracts instrumentation, integration, maintenance, training, software support, and downtime costs. Payback estimates should be accompanied by assumptions about energy prices, operating hours, model maintenance, and expected equipment life. In the absence of empirical measurements, a payback estimate should not be supplied as a result.

4.6 Interpretation of Synthetic Figures and Reproducibility Boundaries

The figures in this manuscript serve different evidentiary purposes, with architecture diagrams explaining the proposed information flow, state-space and optimization diagrams illustrating mathematical relationships, and synthetic plots demonstrating how evaluation results could be presented. They are not derived from a Nepalese factory historian. Captions explicitly state when a series is synthetic or conceptual so that visual clarity is not mistaken for empirical validation.

For a future empirical paper, each measured graph should identify the source dataset, measurement period, sampling interval, units, data-cleaning procedure, and relevant operating conditions. If the

graph is derived from a published report, the source should be cited in the caption and the underlying values should be traceable to the original table or dataset. If the data are generated by simulation, the governing equations, parameters, initial conditions, random seed, and code should be disclosed. This is especially important for uncertainty plots and sensitivity charts, where an attractive visualization can otherwise imply unjustified precision.

The Nepal Electricity Authority annual report [Nepal \(2025\)](#) and WECS publications [Water and Energy \(2024\)](#) are used here as contextual institutional sources rather than as evidence for a synthetic hourly demand curve. This distinction avoids fabricating a Nepal-specific industrial profile from national aggregates. A plant-level study should request or collect meter data directly and clearly state any access limitations. If no suitable dataset can be obtained, the work should remain framed as a methodological proposal or simulation study.

4.7 Extended Operational Scenarios

A start-up scenario tests whether the model represents a cold initial state and the associated warm-up energy. The energy required to reach operating temperature depends on effective thermal mass, heat losses, and the allowed ramp rate. A controller optimized only for steady-state operation may underestimate start-up demand and recommend an infeasible schedule. Start-up periods should therefore be represented explicitly in the training data and state model, or handled through a separate validated operating mode.

A production surge tests whether the model can respond when throughput rises faster than forecast. The optimizer may need to prioritize product requirements and equipment constraints over energy price. If the forecast error is persistent, the state estimator should update its estimate and the uncertainty margin should widen. A conservative policy may be appropriate until the model has re-established a reliable operating regime.

A sensor fault scenario tests whether the system detects impossible values, stale readings, or inconsistent measurements. A model should not continue issuing high-confidence recommendations from stale state estimates. Redundant sensors, plausibility checks, and a clear degraded mode can reduce this risk. Fault detection should not simply clip every unusual measurement, because an extreme value may be a genuine hazard signal.

A tariff-change scenario tests whether economic behavior is robust to a new price schedule. The optimizer should reload the approved tariff configuration, verify its effective date and units, and rerun a known test case. The cost outcome should be separated from physical energy use. This scenario is particularly important when the model is used for scheduling decisions over a billing period.

A maintenance or efficiency-loss scenario tests parameter drift. Fouling, insulation damage, leakage, and worn components may change the relationship between input energy and process state. Monitoring residuals can reveal that the existing model no longer predicts well, but diagnosis may require maintenance records or dedicated tests. The optimizer should not treat a persistent equipment fault merely as a reason to consume more energy.

4.8 Testable Hypotheses

The framework suggests several testable hypotheses for a future plant-level study. H_1 : an AI-assisted disturbance or residual forecast reduces out-of-sample prediction error relative to a simple baseline under operating regimes represented in training. H_2 : adding differential-equation state propagation improves predicted trajectory consistency relative to direct forecasting alone. H_3 : constrained optimization reduces tariff-adjusted cost or production-normalized energy relative to a documented baseline without degrading product quality or violating hard constraints. H_4 : uncertainty-aware margins reduce the frequency of near-limit excursions at the cost of some economic conservatism. These hypotheses should be tested rather than presented as established findings.

The null outcome is also meaningful. The AI model may fail to outperform persistence when the horizon is short and the process is stable. A more complex differential-equation model may not improve performance if sensors do not observe the relevant states. Optimization may have little economic value if production is inflexible or the tariff provides no incentive to shift demand. Reporting these outcomes would help identify where hybrid control is worthwhile and where simpler energy-management practices are preferable.

The hypotheses should be evaluated with pre-specified metrics and decision thresholds. The researcher should define a practically meaningful improvement, not merely a statistically detectable difference. If multiple models and metrics are tested, the report should disclose the comparison set and avoid highlighting only the best result. This disciplined approach supports credible conclusions for both academic readers and industrial decision-makers.

4.9 Practical Deployment for Nepalese Manufacturing

Implementation should begin with a specific use case rather than a nationwide model. Candidate starting points include compressed-air systems, refrigeration, batch heating, drying, pumping, and other processes where energy and operational data can be synchronized. The selected process should have a measurable baseline, identifiable controllable inputs, and production constraints that operators can describe. An advisory forecast may provide value before closed-loop optimization is justified.

Data infrastructure is often a more binding constraint than model selection. If only monthly bills exist, the first step may be interval metering. If energy data exist but production is not timestamped, output and shift records need to be integrated. If process states are unavailable, a low-order model may be too weak to support reliable control. Each missing data layer should be documented rather than hidden behind model complexity.

The software should run close enough to the process to meet latency and reliability requirements. A local industrial computer can reduce dependence on continuous connectivity, but it needs versioned models, local logs, access controls, backups, and a rollback mechanism. During commissioning, data acquisition should be read-only. The optimizer should not bypass existing control logic, burner management systems, electrical protection, pressure trips, or other safety functions.

Operator acceptance is central to deployment. The interface should show the expected demand trajectory, estimated cost effect, active constraints, uncertainty, and reasons for rejecting alternatives. Recommendations that are repeatedly overridden may indicate missing production constraints or poor model behavior. Operator overrides should be systematically recorded and reviewed rather than automatically used as training labels, as the reasons for rejecting system recommendations must first be identified and evaluated.

4.10 Model Governance and Lifecycle Management

A forecasting model is not static once deployed. Equipment wears, sensors are replaced, production recipes change, and operators adopt new practices. These changes can create data drift or concept drift: the distribution of inputs changes, or the relationship between the inputs and demand changes. Monitoring should therefore include feature distributions, residual bias, missingness, prediction interval coverage, and model performance by operating mode. A single aggregate accuracy score can conceal a failure in one important production regime.

Model updates should be versioned and tested offline before release. A new model should be compared against the incumbent using the same chronological evaluation protocol and a fixed acceptance checklist. During initial release, it can run in shadow mode and generate predictions without influencing equipment. Rollback should be straightforward, and the system should preserve enough logs to reconstruct why a recommendation was produced. These practices are consistent with the continual-improvement philosophy of energy management systems such as ISO 50001.

Data governance should identify who owns the measurements, who can alter model parameters, who approves a deployment, and who responds to a detected fault. Production data may be commercially sensitive. Access should follow least-privilege principles, and any data sharing for research should remove identifiers that are not necessary for analysis. A model's explanation should be based on recorded features and constraints, not on a post-hoc narrative that cannot be reproduced.

Operational monitoring should include the optimizer itself. The system should track solve time, infeasibility, active constraints, fallback activation, and the difference between predicted and realized outcomes. If the solver repeatedly reaches its time limit, the horizon or model complexity may be excessive for the available hardware. If a particular constraint is frequently active, that may indicate a genuine production bottleneck or an incorrectly specified model. Monitoring should trigger engineering review rather than automatic self-modification of safety-critical behavior.

4.11 Research Integrity, Reproducibility, and Validity

A reproducible study should report the data dictionary, sampling interval, time range, missing-data policy, preprocessing steps, chronological split, model architecture, hyperparameters, parameter estimates, solver settings, objective weights, and constraint definitions. Synthetic plots should be labelled in captions and in the surrounding discussion. If proprietary data cannot be shared, authors can provide anonymized aggregate data, a schema, and code that reproduces the analysis pipeline without disclosing sensitive operations.

Construct validity asks whether the variables represent the real objective. Power commands do not necessarily reflect actual purchased energy, while furnace air temperature may differ from product core temperature, and nominal production counts may not accurately represent the quantity of products meeting acceptable quality standards. Internal validity asks whether a measured change can be attributed to the intervention rather than weather, production mix, maintenance, or demand changes. External validity asks whether a model calibrated on one facility applies to another. These threats should be explicitly assessed.

Research integrity includes reporting unsuccessful experiments and negative findings. If a hybrid model does not outperform persistence, if uncertainty intervals are poorly calibrated, or if the optimizer frequently becomes infeasible, these are valuable findings. Selective reporting of favorable shifts can exaggerate benefits. A credible study should preserve all eligible test periods and document exclusions before interpreting results.

Table 4. Principal validity threats and required controls

Threat	Possible consequence	Diagnostic / mitigation	Reporting requirement
Data leakage	Overstated forecast accuracy	Chronological splits; feature audit	Split dates and feature availability
Regime shift	Poor performance after process change	Mode-specific evaluation and drift alerts	Errors by regime and time
Model mismatch	Incorrect predicted state trajectory	Residual tests and held-out transients	Model equations and parameter uncertainty
Numerical sensitivity	Different action at different step sizes	RK4 convergence study	Step size and solver settings
Confounding	False attribution of cost savings	Normalize production and match periods	Baseline and adjustment method
Unsafe infeasibility handling	Critical bound may be violated	Verified fallback and independent interlocks	Fault-injection results and stop rules

Table 4 maps each principal validity threat to its possible consequence, the required diagnostic or mitigation, and the information that must be reported, thereby linking research integrity to concrete reporting obligations.

4.12 Environmental and Energy-Security Implications

Economic optimization may also affect emissions and energy security, but these outcomes should be calculated transparently. The emissions associated with a unit of electricity depend on the generation mix and accounting boundary, while direct fuel combustion has different emission factors. A study should identify whether it uses location-based, market-based, or marginal emissions factors and cite the relevant authoritative source. It should not infer emissions reductions directly from a lower electricity bill.

In Nepal, the electricity supply context and hydropower resources make seasonal and temporal energy conditions relevant to planning, but the emissions implications of a specific industrial action depend on the actual system and fuel displaced. National energy reports can provide context ([Nepal, 2025](#); [Water, & Energy, 2024](#)) and the IEA discusses the broader role of efficiency in transitions ([International, 2024](#); [International, 2025](#)). A plant-level analysis should avoid claiming that every shifted kilowatt-hour has the same carbon intensity throughout the year.

Energy security can also be represented through resilience objectives, such as limiting production disruption during supply constraints or preserving critical loads. Such objectives may conflict with short-term tariff minimization. If an industrial site has backup generation, the optimization should include generator efficiency, fuel costs, start-up constraints, and maintenance implications. Backup equipment should not be dispatched solely from an AI recommendation without the existing protection and operating rules.

4.13 Implications for Engineering Education and Practice

The framework also demonstrates a valuable educational connection among differential equations, numerical analysis, machine learning, and operations research. Differential equations describe system dynamics, numerical methods approximate their solutions, AI identifies patterns that are difficult to represent analytically, and optimization determines feasible operational actions. Students and practitioners can explore these concepts using synthetic data before applying them to sensitive industrial records. However, simulation exercises should clearly distinguish theoretical model assumptions from empirical evidence.

For engineering applications, a fundamental consideration is selecting an appropriate level of model complexity based on available data and operational decision requirements. A mathematically sophisticated model does not necessarily provide greater practical value than a transparent baseline approach. Model effectiveness depends on well-calibrated predictions, clearly identified limitations, and the ability to support improved decisions within actual operational constraints. This consideration is particularly important in industrial environments with limited instrumentation, computational capacity, and specialized maintenance resources.

A responsible implementation strategy should follow a gradual and systematic approach, beginning with reliable measurements of energy consumption and production output, followed by baseline establishment and identification of major operating regimes. Subsequent stages should involve demand forecasting, prediction error evaluation, and testing constrained operational recommendations in shadow mode. Closed-loop control may be considered during later implementation stages rather than being treated as an immediate requirement. This phased approach enhances practical feasibility while maintaining operational safety, empirical validation, and evidence-based decision-making as central priorities.

4.14 Limitations and Threats to Validity

The primary limitation of this study is the absence of real industrial data, as synthetic demonstrations cannot establish empirical performance, energy savings, or generalizability to Nepalese industries. The proposed state-space model is simplified and may not capture complex nonlinear dynamics, operational delays, or hidden production constraints. The AI forecasting component remains conceptual, requiring plant-level data for model selection, training, validation, and uncertainty

calibration. Numerical integration and optimization also depend on accurate parameter estimation and cannot independently guarantee operational safety. Furthermore, variations in industrial processes, equipment conditions, energy sources, electricity tariffs, and data infrastructure limit external validity. Future empirical validation, facility-specific calibration, and independent safety mechanisms are therefore essential before practical implementation. Consequently, this study presents a methodological framework rather than a validated industrial energy management solution.

4.15 Future Research and Implementation Roadmap

Future research should prioritize collecting high-resolution industrial data integrating energy consumption, production variables, operational conditions, electricity tariffs, and equipment records while ensuring data privacy and security. Physical models should be calibrated and validated using observed operational dynamics, with parameter uncertainty and model adequacy systematically assessed. AI forecasting models should be evaluated through chronological data splitting, stress testing, uncertainty calibration, and cross-facility validation to establish predictive reliability and generalizability. Optimization strategies should initially undergo simulation and shadow-mode testing before controlled pilot implementation, supported by operator supervision, independent safety assessments, and reliable fallback mechanisms. Economic evaluations should compare energy consumption, operational costs, peak demand, production output, quality, and downtime against transparent baselines while accounting for uncertainty and changing operating conditions. Further studies should explore low-cost edge computing, interpretable decision-support systems, robust optimization, and adaptation to data distribution shifts. Particular attention should be given to affordable monitoring infrastructure, maintainable software, and workforce capabilities to ensure practical, scalable, and sustainable implementation in resource-constrained Nepalese industries.

Table 5. Staged validation and deployment gates

Stage	Required Evidence	Decision Gate	Fallback
1. Data audit	Timestamp, meter, missingness and unit checks	Dataset accepted for analysis	Repair or exclude documented periods
2. Model fit	Identifiability and residual diagnostics	Model adequate in target regime	Simplify or revise model
3. Forecast test	Chronological baselines and calibration	Predefined metrics met	Use baseline predictor
4. Simulation	RK4 convergence and feasible solutions	No unexplained constraint failures	Use verified conventional policy
5. Shadow mode	Operator review and recommendation log	Risks and missing constraints resolved	Remain advisory only
6. Bounded pilot	Normalized economics and safety review	Approved by process owner	Stop pilot and revert

Table 5 operationalizes the pathway in Figure 11 by specifying, for each stage, the evidence required, the decision gate, and the fallback action to be taken if the gate is not passed.

5. Conclusions

This study proposes a methodological framework for dynamic energy-demand forecasting and economic optimization in Nepalese industries by integrating artificial intelligence (AI) with differential-equation modeling. The framework combines AI-assisted forecasting for demand and disturbance estimation, state-space models for representing system dynamics, Laplace analysis for interpreting linearized responses, the fourth-order Runge–Kutta (RK4) method for simulating operational trajectories, and constrained receding-horizon optimization for economic decision-making. The findings emphasize that forecasting accuracy alone does not guarantee improvements in energy efficiency, operational costs, or production performance. Comprehensive evaluation must

consider production-normalized energy consumption, electricity tariffs, operational constraints, uncertainty calibration, numerical convergence, and system reliability. The framework also highlights the importance of reliable measurement systems, transparent baseline models, and facility-specific operational data. A structured implementation process involving data collection, model identification, numerical verification, shadow-mode testing, and controlled pilot deployment is recommended to ensure technical feasibility and operational safety.

The study concludes that AI-assisted differential-equation modeling provides a systematic approach to integrating data-driven forecasting with dynamic system representation and economically constrained optimization. Its principal contribution is the development of a transparent, adaptable, and auditable methodological framework that can support industrial energy management in Nepal. However, the reliance on synthetic demonstrations limits the ability to establish empirical improvements in forecasting accuracy, energy savings, or operational efficiency. Consequently, the proposed framework should be interpreted as a methodological foundation rather than evidence of demonstrated industrial performance. Future research should incorporate high-frequency plant-level operational data, compare conventional and AI-based forecasting methods, validate differential-equation models against observed system behavior, and evaluate economic outcomes under realistic operating conditions. Additional attention should be given to uncertainty management, independent safety mechanisms, and operator-supervised implementation. Empirical validation across different industrial facilities would further establish the framework's generalizability, practical applicability, and potential contribution to sustainable industrial energy management.

Acknowledgement

The authors gratefully acknowledge the Faculty of Science, Technology, and Engineering, Rajarshi Janak University, Janakpurdham, Nepal, for the academic support provided during the preparation of this study.

Author Contributions

AS contributed to the conceptualization, methodology development, formal analysis, and preparation of the original manuscript draft. PKN was responsible for the investigation, visualization, validation, and critical review and editing of the manuscript. SKS contributed to the conceptualization, methodology development, research supervision, and manuscript review and editing. All authors have read and approved the final version of the manuscript.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Baur, L., Ditschuneit, K., Schambach, M., Kaymakci, C., Wollmann, T., & Sauer, A. (2024). Explainability and interpretability in electric load forecasting using machine learning techniques – A review. *Energy and AI*, 16, Article 100358. <https://doi.org/10.1016/j.egyai.2024.100358>
- Biswal, B., Deb, S., Datta, S., Ustun, T. S., & Cali, U. (2024). Review on smart grid load forecasting for smart energy management using machine learning and deep learning techniques. *Energy Reports*, 12, 3654–3670. <https://doi.org/10.1016/j.egy.2024.09.056>
- Bouktif, S., Fiaz, A., Ouni, A., & Serhani, M. A. (2018). Optimal deep learning lstm model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches. *Energies*, 11(7), 1636. <https://doi.org/10.3390/en11071636>

- Chicco, G. (2021). Data consistency for data-driven smart energy assessment. *Frontiers in Big Data*, 4, Article 683682. <https://doi.org/10.3389/fdata.2021.683682>
- Dewangan, F., Abdelaziz, A. Y., & Biswal, M. (2023). Load forecasting models in smart grid using smart meter information: A review. *Energies*, 16(3). <https://doi.org/10.3390/en16031404>
- Eddaoudi, Z., Aarab, Z., El Ghazi, A., Boudmen, K., & Rahmani, M. D. (2024). A brief review of energy consumption forecasting using machine learning models. *Procedia Computer Science*, 236, 33–40. <https://doi.org/10.1016/j.procs.2024.05.001>
- Grant, Z. J. (2022). Perturbed Runge–Kutta methods for mixed precision applications. *Journal of Scientific Computing*, 92(1), 6. <https://doi.org/10.1007/s10915-022-01801-2>
- Huotari, J., Ritari, A., Vepsäläinen, J., & Tammi, K. (2020). Hybrid ship unit commitment with demand prediction and model predictive control. *Energies*, 13(18), 4748. <https://doi.org/10.3390/en13184748>
- International Energy Agency. (2024). Energy efficiency 2024. *IEA*. Retrieved from <https://www.iea.org/reports/energy-efficiency-2024>
- International Energy Agency. (2025). Energy efficiency 2025. *IEA*. Retrieved from <https://www.iea.org/reports/energy-efficiency-2025>
- Jaramillo, M., Pavón, W., & Jaramillo, L. (2024). Adaptive forecasting in energy consumption: A bibliometric analysis and review. *Data*, 9(1). <https://doi.org/10.3390/data9010013>
- Krishnamurthy, S., Adewuyi, O. B., Ratshitanga, M., & Moodley, P. (2024). Artificial intelligence-based forecasting models for integrated energy system management planning: An exploration of the prospects for South Africa. *Energy Conversion and Management: X*, 24, Article 100772. <https://doi.org/10.1016/j.ecmx.2024.100772>
- Ljung, L. (1999). *System identification: Theory for the user* (2nd ed.).
- Mathumitha, R., Rathika, P., & Manimala, K. (2024). Intelligent deep learning techniques for energy consumption forecasting in smart buildings: A review. *Artificial Intelligence Review*, 57, Article 35. <https://doi.org/10.1007/s10462-023-10660-8>
- Michailidis, P., Michailidis, I., Minelli, F., Coban, H. H., & Kosmatopoulos, E. (2025). Model predictive control for smart buildings: Applications and innovations in energy management. *Buildings*, 15(18), 3298. <https://doi.org/10.3390/buildings15183298>
- Nepal Electricity Authority. (2025). Annual report 2024/2025. Retrieved from <https://www.nea.org.np/index.php/en/detail/annual-report-20242025>
- Sahani, S. K. (2022). Evaluation of plastic pipe performance in industrial engineering utilizing Laplace transform and the fourth-order Runge–Kutta method. *Fuel Cells Bulletin*, 1–7. <https://doi.org/10.52710/fcb.210>
- Sahani, S. K. (2023a). Application of numerical methods in structural health monitoring using IoT sensors. *Journal of Electrical Systems*, 19(1), 194–207. <https://doi.org/10.52783/jes.8941>
- Sahani, S. K. (2023b). A finite difference scheme for solving partial differential equations in heat conduction problems in mechanical engineering. *Metallurgical and Materials Engineering*, 29(3), 69–81. <https://doi.org/10.63278/mme.v29i3.1798>
- Sahani, S. K. (2024a). AI-enhanced finite element method (FEM) for structural analysis. *Journal of Electrical Systems*, 20(1), 661–676. <https://doi.org/10.52783/jes.8946>
- Sahani, S. K. (2024b). A numerical study of second-order differential equations in vibration analysis of mechanical systems. *Membrane Technology*, (3). <https://doi.org/10.52710/mt.429>

- Sahani, S. K., Oruganti, S. K., & Kumar, K. S. (2025a). Advanced mathematical modeling of woolen knitting dynamics using Laplace transform and fourth-order Runge–Kutta method. *Journal of Mathematics*. <https://doi.org/10.1155/jom/4985087>
- Sahani, S. K., Oruganti, S. K., & Kumar, K. S. (2025c). Optimization of maintenance strategies for cement manufacturing plants. *SGS—Engineering & Sciences*, 1(1).
- Sahani, S. K., Oruganti, S. K., Kumar, K. S., & Karna, S. K. (2025). Reliability assessment of a plywood production facility utilizing Laplace transform and Runge–Kutta fourth-order differential equations: Overview of industrial plant. *Metallurgical and Materials Engineering*, 31(4), 417-423. <https://doi.org/10.63278/1453>
- Sahani, S. K., Sah, S. K., & Sahani, K. (2023b). Statistical reliability modeling of cement kilns and crushers. *Panamerican Mathematical Journal*, 33(4), 115-120. <https://doi.org/10.52783/pmj.v33.i4.4223>
- Talaei Khoei, T., & Kaabouch, N. (2023). Machine learning: Models, challenges, and research directions. *Future Internet*, 15(10), 332. <https://doi.org/10.3390/fi15100332>
- Tsegaye, S., Sanjeevikumar, P., Tjernberg, L. B., & Fante, K. A. (2024). Short-term energy forecasting using deep neural networks: Prospects and challenges. *The Journal of Engineering*, 2024(11). <https://doi.org/10.1049/tje2.70022>
- Venu, V., & Namitha, M. R. (2026). Applications of Auto-regressive Integrated Moving Average (ARIMA) Model in Agricultural Engineering: A Review. *Agricultural Reviews*, 47(4), 539-549. <https://doi.org/10.18805/ag>
- Wang, X., Wang, H., Bhandari, B., & Cheng, L. (2024). AI-empowered methods for smart energy consumption: A review of load forecasting, anomaly detection and demand response. *International Journal of Precision Engineering and Manufacturing-Green Technology*, 11(3), 963-993. <https://doi.org/10.1007/s40684-023-00537-0>
- Water and Energy Commission Secretariat. (2024). Energy sector synopsis report 2024 (FY 2079/80). *Government of Nepal*. Retrieved from <https://wecs.gov.np/content/56/neis/>