Proceedings of the 4th International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/2026.37476

Practical Applications of Machine Learning in Manufacturing: Virtual Power Plants and Coordinated Computing and Power Management

Andy Ren

Oaks Christian School, Westlake Village, USA
ARen@student.oakschristian.org

Abstract. This paper uses publicly available UCI electricity consumption data to examine how machine learning affects day-ahead load forecasting and energy storage dispatch. Hourly loads from 20 large electricity users are aggregated to simulate a virtual power plant. The experiments compare naive baselines with direct regression models and use the weekly load pattern to develop residual models that predict deviations from the load at the same hour of the previous week. The gradient boosting residual model achieves a mean absolute percentage error of 3.49%, a 13.2% reduction from the baseline of 4.02%, and reduces root mean squared error by 28.0%. The error in the period of peak falls from 1.69% to 1.22%, and the mean absolute error in daily peak load falls by 25.4%. By using a 15 MW/30 MWh storage system, assumed the rates are based on the time it uses, and a fixed peak-shaving target, peak-shaving not only effectively increases from 78.2% to 79.0% over the roughly 38-day test period but also the saves more electricity cost rise from CNY 511,000 to CNY 515,000. Weekly residual learning improves the accuracy forecasting with a larger error reductions during peak periods than the full test period. Better forecasts bring additional peak-shaving and cost benefits although the storage capacity and dispatch rules limit gains. This paper also shows that how the approach can potentially support coordinated computing and power management through shifts in task execution times; However, the performance in computing task scheduling might requires more validation.

Keywords: machine learning, virtual power plant, load forecasting, energy storage dispatch, coordinated computing and power management

1. Introduction

The International Energy Agency's 2025 report forecasts that global data center electricity consumption will more than double between 2024 and 2030, driven primarily by artificial intelligence [1]. Surging power consumption across data centers and industrial manufacturing accentuates the importance of active energy management. Because wind and solar output fluctuate with weather conditions, balancing supply and demand heavily relies on storage assets and flexible loads. Consequently, tailoring energy storage dispatch to real-time load shifts offers manufacturers an effective way to lower energy procurement expenses while bolstering grid reliability.

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.37476

Long short-term memory networks are frequently applied in load forecasting to extract time patterns from historical data, which assists in predicting short-term residential energy needs [2]. In contrast, demand response strategies rely more on reinforcement learning to shift daily power consumption based on electricity rates, battery status, and user habits [3]. Since the primary evaluation targets differ between these fields, prediction models prioritize accuracy metrics, whereas control systems must weigh operating expenses against physical system limits; consequently, enhancing forecast precision does not automatically lead to a matching improvement in overall demand-side control performance.

A virtual power plant combines data streams from diverse users and hardware, meaning forecast errors stem from data quality, signal stability, and control decisions; in data-driven energy management, researchers increasingly evaluate data collection, demand forecasting, and power dispatch within a single setup to observe how each stage affects full system performance [4]. Additionally, coordinating distributed energy assets introduces operational challenges such as variable generation output, uneven response speeds, and mismatched control intervals [5]. Ultimately, these physical constraints determine the actual value of predictions in routine operations, proving that higher forecast accuracy alone produces limited operational gains.

Coordinated computing and energy management regulates power demand primarily by leveraging spatial and temporal flexibility in task execution; for instance, carbon-aware computing systems schedule delayable workloads using next-day forecasts of grid carbon intensity and computing demand, which lowers daily operational emissions [6]. In addition, load-shaping strategies combine day-ahead planning with real-time job distribution, enabling systems to balance execution schedules and server locations within physical bounds [7]. However, unlike conventional battery storage, scheduling computational tasks must respect strict completion deadlines and service requirements, while shifting workloads across regions introduces added data transfer overhead.

The study combines the power that loads from different users in order to build a virtual power plant that is simulated which tests the load predictions of the next day made by basic methods. Then the test checks all predictions errors from all power needs, peak hour and from daily peak loads. It illustrates how the forecast errors affect peak-shaving performance and electricity buying costs according to the basic operating rules. Last but not least, it shows the needs and practical points in order to expand the setup to handle those computing tasks.

2. Research approach and application setting

The virtual power plant needs digital systems to work in order to gather the distributed power generation and battery storage and also the flexible electricity loads. The study mainly focuses on load collection and the scheduling for batteries by adding up demand from different users during multiple hours so it can predict the need of power for the next day. Simulation models basic power demand and battery movement to leave out power market bidding, grid line limitations and the direct control of generators.

The dataset is split into training and test parts by time order; direct regression models and weekly residual learning approaches are compared against two simple baseline methods that just copy past electricity use from the same hour of the previous week or day; prediction accuracy is checked by looking at overall forecast errors, peak-period errors, and daily peak-load errors; also, the energy dispatch test keeps battery settings and running rules fixed, which lets the study measure net changes in power cost savings, extra energy left over above the limit, and total peak-reduction results.

DOI: 10.54254/2755-2721/2026.37476

3. Models and methods

3.1. Data and the simulated virtual power plant

The test uses the ElectricityLoadDiagrams20112014 dataset from the UCI Machine Learning Repository , taking about 13 months of records from January 2011 to early February 2012 [8]. Out of the 160 users with full 2011 meter data, the 20 biggest electricity users are picked, as shown in Figure 1. Their hourly power use is added together to make a simulated virtual power plant, this combined load set includes 9,678 hourly numbers in a row with no missing parts, having an average of 81,449 kW and a top high of 132,744 kW. Since the dataset does not list actual business types, this group acts as a general stand-in for combined business and factory power needs rather than one specific industry.

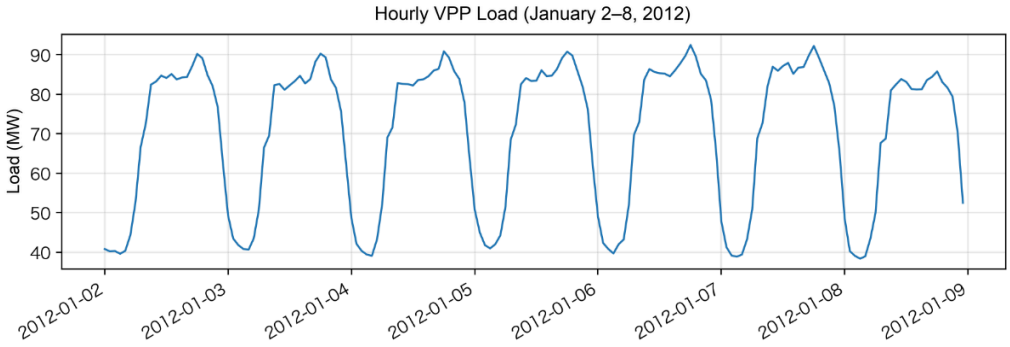


Figure 1. Hourly load of the VPP portfolio, January 2–8, 2012 (picture credit: original)

The training portion covers data from January 8 to December 31 in 2011, which adds up to 8,592 total hours, while early records from January 1 through January 7 are picked just to build older lag features. Next, the testing portion runs from January 1 into part of February 8 in 2012, giving a total of 918 hours, or roughly 38 days. Every time a next-day prediction is made, midnight works as the hard cutoff point for taking in data, and future hourly power needs are predicted for the next 24 hours. Historical load data and rolling averages strictly incorporate observations recorded prior to the midnight cutoff. For the final partial day, evaluation is confined exclusively to hours with valid observed measurements.

3.2. Baselines and direct regression models

The experimental evaluation incorporates two benchmark baselines alongside three direct regression models: multiple linear regression, random forest, and a two-hidden-layer multilayer perceptron (MLP) [9]. The input feature set comprises the hour of the day, day of the week, historical load values at 24-, 48-, and 168-hour lags, and the rolling 24-hour mean load. As detailed in Table 1, the weekly persistence baseline achieves a mean absolute percentage error (MAPE) of 4.02%, outperforming linear regression (4.11%), random forest (10.42%), and the MLP (5.00%). In terms of root mean squared error (RMSE), however, the MLP achieves the best performance at 4,753 kW, compared to 5,826 kW for the baseline.

DOI: 10.54254/2755-2721/2026.37476

Table 1. Test-period errors for baseline and direct regression models

Model	MAPE (%)	RMSE (kW)
Naive baseline: same hour last week	4.02	5,826
Naive baseline: same hour yesterday	5.05	7,769
Multiple linear regression	4.11	5,288
Random forest	10.42	11,879
Multilayer perceptron (64-32)	5.00	4,753

The strong performance of the weekly persistence baseline in terms of MAPE aligns with the distinct weekly seasonality inherent in the load profile. Starting with that pattern, a model can estimate how the current week differs from the previous one. The direct regression results provide a reference for assessing whether this change in the prediction target reduces error.

3.3. Residual learning based on weekly load patterns

The weekly residual is defined as $r = y - y_{-168}$, where y_t is the current hourly load and y_{t-168} is the load at the same hour of the previous week. The predicted residual is added to the previous week's baseline value to obtain the final load forecast. Inputs include one-hot encodings of hour of day and day of week, differences between historical loads and the corresponding hours of the previous week, and the difference between the mean of a completed 24-hour window and that of the corresponding window one week earlier. Linear regression, random forests, gradient boosting, and multilayer perceptrons are each used to fit the residuals. Gradient boosting combines weak learners by sequentially fitting the negative gradient of the loss function [10]. The weekly residual is the prediction target; gradient boosting is the algorithm used to fit that target.

In Table 2, the gradient boosting residual model has the lowest MAPE, at 3.49%, a 13.2% reduction from the weekly persistence baseline. RMSE falls from 5,826 kW to 4,192 kW, a reduction of 28.0%. The multilayer perceptron residual model achieves a MAPE of 3.53% and an RMSE of 3,688 kW. The two models therefore offer different advantages: gradient boosting has a lower average percentage error, while the multilayer perceptron has a lower RMSE. The gradient boosting residual model is used in the subsequent dispatch experiments.

Table 2. Test-period errors and peak-period accuracy of residual models

Model	MAPE (%)	RMSE (kW)	Peak MAPE (%)	Daily peak MAE (kW)
Weekly persistence baseline	4.02	5,826	1.69	2,112
Linear regression residual model	4.40	4,840	—	—
Random forest residual model	4.27	5,745	—	—
Gradient boosting residual model (selected)	3.49	4,192	1.22	1,576

DOI: 10.54254/2755-2721/2026.37476

Table 2. (continued)

Multilayer perceptron residual model	3.53	3,688	—	—

3.4. Storage dispatch simulation

The storage system has a rated power of 15 MW, an energy capacity of 30 MWh, and a round-trip efficiency of 90%. Its rated power is approximately 18% of the VPP's average load. Charging and discharging are scheduled a day ahead using the load forecast: charging occurs during hours in the lowest load quartile, and discharging occurs when the forecast exceeds the 82,000 kW target. System running is limited by physical bounds on power capacity, energy throughput, and battery charge levels; the cost test applies a time-of-use electricity rate, charging 1.00 CNY/kWh during peak hours, 0.70 CNY/kWh during middle hours, and 0.38 CNY/kWh during off-peak hours; four running setups are compared: a basic setup without any battery, a battery managed by simple persistence predictions, a battery guided by gradient boosting error forecasts, and a battery working under perfect future knowledge; peak-reduction efficiency is measured as the ratio of real peak energy cut to total energy above the target limit, while remaining load shows how much power stays above this target line. Electricity cost savings are calculated relative to the no-storage scenario and exclude storage investment, operation and maintenance, and degradation costs. The perfect-forecast scenario uses actual loads to prepare the schedule and serves as an ideal forecast reference under the same dispatch rules.

4. Discussion and Analysis

4.1. Forecast accuracy and errors during critical periods

The highest-load 10% of the test period comprises 92 hours. During those hours, the selected model has a MAPE of 1.22%, a 27.8% reduction from the weekly persistence baseline of 1.69%. Mean absolute error (MAE) in daily peak load falls from 2,112 kW to 1,576 kW, a reduction of 25.4%. Both reductions exceed the 13.2% improvement in overall MAPE, indicating that the forecast improves more during high-load periods. Peak periods are identified retrospectively from actual test-period loads and are used only to evaluate error. Figure 2 compares the forecasts with actual loads for January 16–18, 2012.

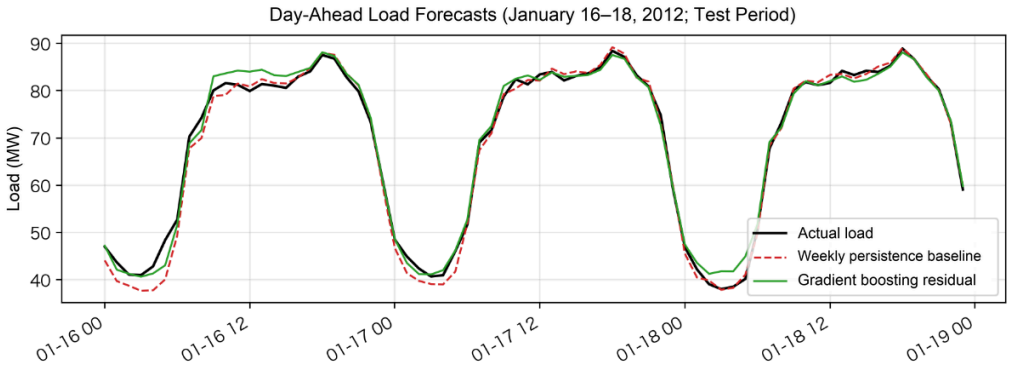


Figure 2. Day-ahead load forecasts, January 16–18, 2012 (picture credit : original)

DOI: 10.54254/2755-2721/2026.37476

4.2. Storage dispatch performance and electricity costs

As shown in Table 3, storage scheduled with the naive baseline forecast saves CNY 511,000 over approximately 38 days and reduces residual energy above the target from 1,115.2 MWh without storage to 242.9 MWh. Using the gradient boosting residual forecast raises savings to CNY 515,000, approximately CNY 4,000 more than the baseline. Peak-shaving effectiveness increases from 78.2% to 79.0%, while residual energy above the target falls by 8.7 MWh, or approximately 3.6%. Relative to the naive baseline, perfect forecasting would improve peak-shaving effectiveness by 1.3 percentage points. The chosen model reaches 0.8 percentage points of that total gain, which equals roughly 62% of the max improvement possible; these percentage numbers are worked out using the rounded values shown in the main table. Adding a battery creates most of the electricity cost savings, while better forecast accuracy mainly helps lower the extra power demand that goes over the target limit; this distinction shows that physical storage provides the main financial benefit, whereas improved predictions serve mostly to refine detailed operational limits. Looking at the example day in Figure 3, the battery charges up during low-use hours and drains power during heavy-use hours, which brings the final net load pretty close to the 82 MW target during several time slots.

Table 3. Power dispatch outcomes across roughly 38 days using an 82,000 kW target mark

Scenario	Cost savings (CNY 10,000)	Excess energy (MWh)	Peak shaving (%)	Mean daily peak (kW)
No storage	0.0	1,115.2	0.0	86,677
Baseline forecast + storage	51.1	242.9	78.2	83,765
Selected model + storage	51.5	234.2	79.0	83,666
Perfect forecast + storage (upper bound)	51.8	228.5	79.5	83,666

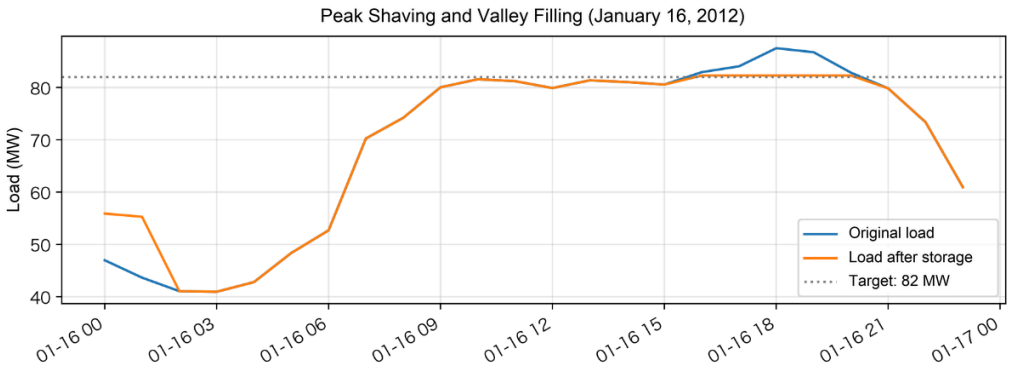


Figure 3. Peak shaving and valley filling on January 16, 2012 (picture credit: original)

4.3. Discussion of results

None of the direct regression models gets a smaller MAPE score than the basic weekly baseline, but the gradient boosting residual model manages to do so. For this group of power loads, tweaking the target prediction to rely on weekly usage shapes works better than just making a model more

DOI: 10.54254/2755-2721/2026.37476

complex. Still, model rankings change depending on which test measure is checked; the gradient boosting residual model gets the lowest MAPE, whereas the simple neural network residual model scores the lowest RMSE. Because of this, picking a model ought to match the real dispatch task you plan to run. On top of that, even when predictions are completely accurate, 228.5 MWh of extra power still sits above the target, which proves that battery size and daily running rules also limit overall peak shaving. While options like reinforcement learning could be applied to compare long-term gains across various dispatch choices, that kind of comparison goes beyond the main tests done in this work [3].

Battery storage moves power usage around by charging up and draining, whereas joint computing and power management shifts electricity demand by changing when and where computer tasks are processed. While the next-day forecasting method can help estimate total power needs under a planned task lineup, basic battery rules cannot be used directly for computer workloads. This happens because task start times, running lengths, and final deadlines decide how flexible the timing really is, while service quality goals and network data transfers set hard limits on moving tasks across locations. All of these extra requirements need to be built into a scheduling model first before anyone can test the true real-world gains of combining computer tasks with power management.

Since the tests only use one combined group of power loads along with a test timeframe of about 38 days, the findings mainly show how things work during winter months. Extra factors like weather conditions, factory schedules, and holiday breaks are left out of the input factors, which means their actual effects on prediction errors remain untested. In addition, battery power dispatch relies on fixed daily rules and assumed price rates, and the saved money figures skip equipment costs; because of this, the numbers do not reflect a real return on investment. Future studies can look at longer testing periods, compare seasonal changes across different types of power users, and check how battery size and schedule changes affect final operating results; also, testing these methods under varied weather conditions will help show how well the models work over time; in addition, bringing in more flexible load types can give a clearer picture of real-world energy storage benefits; overall, these extra tests will help build better guidelines for managing combined power and computing setups.

5. Conclusion

This study aggregates hourly loads from 20 users to simulate a virtual power plant, compares naive baselines, direct regression, and weekly residual learning, and uses the forecasts to schedule storage charging and discharging. Comparing forecast error, peak-shaving effectiveness, and electricity purchase costs leads to the following conclusions about machine learning in aggregate load management.

First, among the models tested, the gradient boosting residual model achieves the lowest MAPE, at 3.49%, while the multilayer perceptron residual model achieves the lowest RMSE, at 3,688 kW. Weekly residual learning improves forecasts for this portfolio, although the best model depends on the evaluation metric. Second, relative to the weekly persistence baseline, gradient boosting reduces overall MAPE by 13.2%, peak-period MAPE by 27.8%, and daily peak load MAE by 25.4%, with larger improvements during high-load periods. Third, under the same storage configuration and dispatch rules, peak-shaving effectiveness rises by 0.8 percentage points, residual energy above the target falls by 3.6%, and electricity purchase costs decrease by an additional CNY 4,000. Improved accuracy thus produces additional dispatch benefits. Fourth, storage capacity, operating rules, and data coverage limit the applicability of these results. Expanding this setup to combined computing and power management requires adding clear task completion deadlines and service quality rules; in

DOI: 10.54254/2755-2721/2026.37476

addition, computing jobs cannot be delayed freely like standard battery storage because finishing tasks on time directly impacts system performance; therefore, planning models must balance energy cost savings with strict computing service limits.

References

[1] International Energy Agency. (2025). *Energy and AI*. Pairs. IEA. [Online]. Available: https://www.iea.org/reports/energy-and-ai
[2] Kong, W., Dong, Z. Y., Jia, Y., et al. (2019). Short-term residential load forecasting based on LSTM recurrent neural network. *IEEE Transactions on Smart Grid, 10*(1), 841-851.
[3] Vázquez-Canteli, J. R., & Nagy, Z. (2019). Reinforcement learning for demand response: A review of algorithms and modeling techniques. *Applied Energy, 235*, 1072-1089.
[4] Ruan, G., Qiu, D., Sivaranjani, S., et al. (2014). Data-driven energy management of virtual power plants: A review. *Advances in Applied Energy, 14*, Art. no. 100170.
[5] Gao, H., Jin, T., Feng, C., et al. (2024). Review of virtual power plant operations: Resource coordination and multidimensional interaction. *Applied Energy, 357*, Art. no. 122284.
[6] Radovanović, A., Koningstein, R., Schneider, I., et al. (2023). Carbon-aware computing for datacenters. *IEEE Transactions on Power Systems, 38*(2), 1270–1280.
[7] Hall, S., Micheli, F., Belgioioso, G., et al. (2024). Carbon-aware computing for data centers with probabilistic performance guarantees. *arXiv* preprint, arXiv:2410.21510.
[8] Trindade, A. (2015). ElectricityLoadDiagrams20112014, UCI Machine Learning Repository. [Online]. Available: https://doi.org/10.24432/C58C86.
[9] Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.
[10] Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics, 29*(5), 1189–1232.