Journal of Computer Applications and Information Technology
Vol. 2, Issue. 3, Sep 2026, pp. 35~46
ISSN: 3107-6424, DOI: 10.32595/jcait/v2i3.2026.40

Adaptive Ensemble Learning for Short-Term Energy Consumption Prediction Using Smart Building Data

Ben Sujin[1], Misoon Rashid Ali Almarzouqi[2], Alsaadah Saif Mohammed ALabri[3]

[1]Lecturer, Computer Engineering Department, University of Technology and Applied Sciences, Nizwa, Oman. E-mail: bensujin.bennet@utas.edu.om
[2,3]Department of Information Technology, University of Technology and Applied Sciences, Shinas, Oman.

Article Info

Article History:

Received Jun 20, 2026
Revised Jul 19, 2026
Accepted Aug 27, 2026

Keywords:

Smart Buildings,
Energy Forecasting,
Ensemble Learning,
Temporal Leakage,
XGBoost,
Gradient Boosting,
Machine Learning,
Sensor Data

ABSTRACT

Accurate short-term energy forecasts can help schedule, monitor and make data-driven energy management decisions in smart environments. This paper proposes a leakage-aware adaptive ensemble framework that combines a persistence baseline, linear regression, histogram-based gradient boosting and XGBoost for short-horizon appliance-energy prediction. The study uses the publicly available UCI Appliances Energy Prediction dataset containing 19,735 ten-minute observations from a low-energy house. This approach constructs timestamp variables and lagged consumption features from only information available before the forecast instant. A chronological 70:15:15 train-validation-test split was used to prevent future information from entering model selection. Base-model contributions were determined from inverse validation MAE and normalized prior to generating the final forecast. On the held-out test period, the adaptive ensemble achieved MAE 28.99 Wh, RMSE 59.83 Wh, MAPE 27.12% and R^2 0.567. It outperformed the individual machine-learning models evaluated when considering RMSE and R^2, although the persistence baseline achieved lower MAE and MAPE. This result highlights that adaptive combination can improve overall variance-sensitive forecasting while also underlining the need for multiple metrics and leakage-safe evaluation.

Corresponding Author:

Ben Sujin,
Lecturer, Computer Engineering Department,
University of Technology and Applied Sciences, Nizwa, Oman.
E-mail: bensujin.bennet@utas.edu.om

1. INTRODUCTION

The problem of building energy forecasting is now a serious data-intensive challenge since sensor networks can record relevant information about the surrounding environment on a perpetual basis. First, a database survey of 431 machine-learning studies discovered a rapidly rising trend in building-energy prediction with open issues of multi-source learning, transfer, and real-time reproduction [1]. Second, a review of 146 ensemble-learning articles published in 2025 revealed

that heterogeneous ensembles are valuable, but there is a lack of understanding of combination strategies, algorithm choice, and generalization [2].

The UCI Appliances Energy Prediction data set is attractive for conceptual experiments since it combines information about appliances' energy consumption with indoor temperature and humidity, and the outdoor weather conditions measured every ten minutes [3]. Candanedo et al. proved that appliance energy use could be successfully predicted using two groups of variables: weather observations and wireless-sensor-network measurements [4].

While a large mean value for randomly shuffled time series suggests good predictive performance, this approach allows indirect use of future data for training or validation. For short-term forecasts, a temporal protocol is required, where the split of data into training, validation, and test sets follows a chronological order. Therefore, the current study controls for leakage by making it one of the core issues under investigation rather than an implementation detail.

2. LITERATURE SURVEY

Building-energy prediction has been explored using statistical, machine-learning and deep-learning approaches. A systematic review spanning 60 selected studies categorized past work according to building type, data components, temporal granularity, preprocessing routines, feature selection, model family and evaluation metrics [5]. The review stresses that no single model has universal value across building-energy forecasting scenarios.

Candanedo et al. constructed data-driven models predicting appliance energy in a low-energy house and noted that weather-station variables and temperature/humidity measurements improved prediction accuracy, with multiple indoor variables registering among the highest importance scores [4]. The result offers direct justification for examining the UCI dataset and combining environmental with historical energy measurements.

Ensemble learning methods have been studied specifically for short-term building-energy forecasting. Li et al. proposed a self-adaptive ensemble scheme utilizing a number of data-driven predictors and validated this approach on two real-building electricity datasets, using feature selection and cross-validation techniques to avoid overfitting [6]. The article highlights the value of model combination, though its ensemble construction is significantly more complex than the lightweight weighting protocol investigated in this study.

More recent works have continued to explore adaptive ensembles. A 2025 study employed a two-layer stacking ensemble with support-vector regression and random forest, which incorporated sliding-window inputs to account for changing building-consumption profiles [7]. Another 2025 Energy paper proposed evolutionary adaptive bagging, stacking and voting mechanisms and compared these with 15 predictors [8]. Overall, these works demonstrate that adaptive and heterogeneous ensemble learning techniques are already an active area of research.

Recent reviews further support this point. Wang et al. analyzed 82 ensemble learning studies, highlighting limitations including subjective algorithm selection, insufficient model diversity assessment, inappropriate combination rules, computational expense and lack of generalization validation [2]. Ji et al. also identified transfer learning between buildings and real-time forecasting as open directions [1]. Therefore, the contribution of this work is deliberately specific: it investigates the efficacy of transparent validation-error weighting within a chronological protocol and reports cases where the ensemble does or does not improve individual error measures.

3. DATASET DESCRIPTION

The experiment uses the UCI Appliances Energy Prediction dataset. The repository houses 19,735 instances and 29 attributes (including the target), ten-minute sampling and no missing values in the released dataset [3]. The target is appliance energy consumption in watt-hours. Inputs include indoor temperature and relative humidity measurements, outdoor weather variables, lighting consumption and timestamp information. Table 1 summarizes the main characteristics of the dataset and the evaluation protocol used in this study.

Table 1. Summary of the UCI Appliances Energy Prediction dataset and evaluation protocol

Property	Value
Dataset	UCI Appliances Energy Prediction
Observations	19,735
Sampling interval	10 minutes
Attributes	29 including target
Target	Appliances energy (Wh)
Missing values	None reported in dataset
Temporal span	Approx. 4.5 months
Evaluation	70% train / 15% validation / 15% test, chronological

3.1. Dataset Characteristics

The UCI Appliances Energy Prediction dataset is ideal for experimenting with short-term forecasting problems, as the data are collected at regular ten-minute intervals. The dataset contains appliance-energy measurements alongside environmental readings indoors and outdoors. Consequently, the prediction task can be approached from both a temporal and a sensor-based viewpoint. The regular sampling interval is particularly convenient for creating lag-based predictors. The value of the one-step historical feature is the immediately preceding observation, and larger lags represent progressively older values. Features of this kind allow the forecasting models to learn short-term temporal dependence without needing their own sequential architecture.

The dataset spans roughly four and a half months of observations from one energy-efficient house. Thus, the volume is sufficient for comparative experiments, but there is no guarantee that the produced model will generalize to other buildings, seasons, occupancy rates, or geographic locations. This limitation is discussed explicitly in the discussion and future-work sections.

3.2. Data Preparation

The data preparation procedure was kept as non-intrusive as possible with respect to the original temporal structure of the data. Timestamp information was retained and converted into temporal variables. Hour, day-of-week, and month were extracted for capturing repeating patterns in consumption. Hour and day-of-week were also converted into cyclic representations to prevent their endpoints from being artificially far.

Historical consumption features were generated from preceding observations. The selected lags represent multiple distances in time, making the models capable of capturing both immediate and more elaborate temporal dependencies. Rolling averages were only created after shifting the target sequence, preventing the current target point from being part of its feature representation. After feature construction, the observations were split into chronological training, validation, and test subsets. This step is critical for proper temporal validation, as the validation interval always

occurs after the training interval, and the test interval is even later and only available at the end of model development.

4. PROPOSED METHODOLOGY

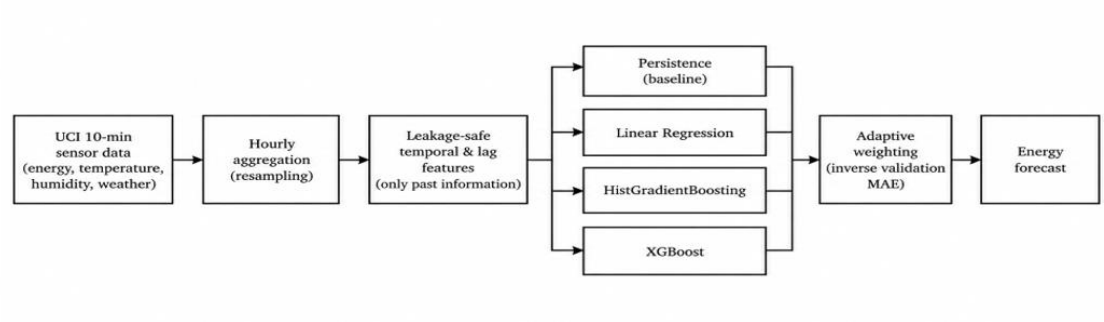


Figure 1. Leakage-aware adaptive ensemble architecture

As shown in Figure 1, the proposed framework initially aggregates the original 10-minute UCI sensor data into hourly values. The leakage-safe temporal and lag features are constructed using only the information available before the prediction time. Then, these features are independently processed by the persistence baseline, linear regression, HistGradientBoosting, and XGBoost models to generate their validation performance and corresponding adaptive weights based on the inverse validation MAE. Finally, the weighted model predictions are combined to obtain the short-term energy consumption forecast.

4.1. Temporal Feature Engineering

The timestamp is then split into hour, day-of-week and month features. Cyclical sine/cosine versions of hour and day-of-week are also included to make the encoding of periodicity non-discontinuous with respect to time. Historical target features are created using lags of 10, 20, 30, 60, 120, 180 and 360 minutes, along with rolling means, calculated after shifting the target by one observation, to include only historic values.

4.2. Base Predictors

Four predictors were used in the adaptive combination. The first one used the persistence forecast with the previous value of the appliance-energy pair as input. The second predictor was a standardized linear regression, which provided a low complexity baseline. The third method was Histogram-based Gradient Boosting, which was chosen for fast non-parametric tabular regression and belongs to the family of decision-tree ensemble methods [9]. Lastly, XGBoost regressor [10] with log transformed target was used as a regularized gradient-tree boosting method.

4.3. Adaptive Weighting

Validation MAE was calculated for each predictor separately. For each model, i, an inverse-error score was defined as the ratio of one over the model's validation MAE. Weights for the models were then selected so that all the weights are non-negative and sum up to one. The final test prediction was calculated as a weighted average of the base predictions. As you can see, the procedure is straightforward: no test observations were used during weights estimation.

4.4. Experimental Protocol

The chronological data sample is split into 70% training, 15% validation, and 15% testing subsets after lagging. The models are trained on the training subset. The weights of the model are

tuned using the validation subset, which is then used to compute the model's prediction on the unseen test subset. The data subsets are not randomly shuffled. A random seed of 42 is used for stochastic models.

4.5. Evaluation Metrics

MAE is the average absolute error, while RMSE is sensitive to extreme errors. MAPE represents relative error in terms of consumption, and R^2 is the proportion of variance in target explained by the model. Since MAPE is often unstable near zero, it should be interpreted in combination with the other three indicators.

4.6. Overall Processing Sequence

The entire forecasting process can be interpreted as a six-step sequence. First, original sensor observations are grouped according to their timestamps. Second, temporal variables and historical consumption features are generated. Third, the result is split chronologically into training, validation, and test periods. Fourth, each base predictor is trained independently with the training information. Fifth, validation predictions are used to calculate the relative contribution of each model. Finally, weighted predictions are combined and evaluated on the unseen test period.

This sequence isolates model development from final evaluation. In particular, the test observations are not used to decide the adaptive weights. This is critical because using test performance to select model contributions would turn the test set into part of the model-development process and thus degrade the final performance estimate.

4.7. Persistence Baseline

The persistence predictor is incorporated as an important reference because short-term energy consumption frequently exhibits local continuity. Rather than learning a complex mapping from multiple features, the persistence strategy simply uses the previous observed consumption as the forecast for the next prediction interval.

Although persistence is simple, it provides a meaningful benchmark. A machine-learning model should demonstrate useful predictive behaviour beyond simply repeating the last observation. In our current experiment, the persistence model performs especially well in MAE and MAPE. This suggests that immediate temporal continuity remains a key characteristic of the target series.

Its competitive performance also provides an important interpretation of the ensemble results. The proposed adaptive framework should not be seen as universally superior to persistence under every objective. Instead, the results suggest that model combination can improve variance-sensitive measures while a simple historical baseline proves highly competitive for pointwise absolute and relative error.

4.8. Linear Regression

Linear regression serves as a low-complexity learned predictor. The model assumes that the target can be approximated using a weighted combination of the available features. Standardization is used prior to model fitting to ensure that numerical differences among input variables do not disproportionately affect the coefficient estimation.

The role of linear regression within the ensemble is important because it provides a different modelling assumption from the tree-based predictors. Although boosting methods can capture nonlinear relationships and interactions, linear regression represents a simpler global

relationship. When combined with more flexible models, these differences can contribute to ensemble diversity.

4.9. Histogram-Based Gradient Boosting

Histogram-based gradient boosting is included as a computationally efficient nonlinear regression method. Rather than assuming a purely linear relationship, the algorithm constructs an ensemble of decision trees that progressively reduce prediction errors. The histogram representation enables continuous feature values to be grouped into bins during the tree construction, improving computational efficiency for the tabular dataset.

For the present forecasting problem, this model can capture relationships between lagged consumption, temporal variables, and environmental measurements that cannot be adequately represented by a linear model. Its feature-importance analysis is also used to examine which variables provide the greatest contribution to prediction performance.

4.10. XGBoost

XGBoost is another gradient-tree boosting approach [10] included for its regularization and strong performance on structured tabular data. Within the study, the target is log-transformed for the XGBoost model. The transformation provides a different modelling representation of the target distribution and allows the predictor to participate in the ensemble as a complementary model.

The purpose of including XGBoost is not to assume a priori that it will outperform every other predictor. Instead, its validation error determines its contribution to the final ensemble. This ensures the weighting procedure is data-dependent and avoids assigning equal importance to models with substantially different validation behaviours.

5. RESULTS

Table 2. Held-out chronological test results

Model	MAE (Wh)	RMSE (Wh)	MAPE (%)	R^2
Persistence	26.78	66.89	21.54	0.459
Linear Regression	27.89	60.51	25.60	0.557
HistGradientBoosting	34.35	63.93	33.98	0.506
XGBoost (log-target)	35.44	64.30	35.77	0.500
Adaptive Ensemble	28.99	59.83	27.12	0.567

In Table 2, the results of the individual models and the proposed adaptive ensemble are compared on the unseen test data. Ideally, the results should have a low value of MAE, RMSE, and MAPE, while having a high R^2 value. From the results, Adaptive Ensemble has the best prediction results with the lowest RMSE of 59.83 Wh and highest R^2 of 0.567 among the compared models. However, the Persistence model has a better MAE and MAPE than the Adaptive Ensemble. This result suggests that while the adaptive ensemble improves the overall fit, it does not improve all metrics over the simple baseline model.

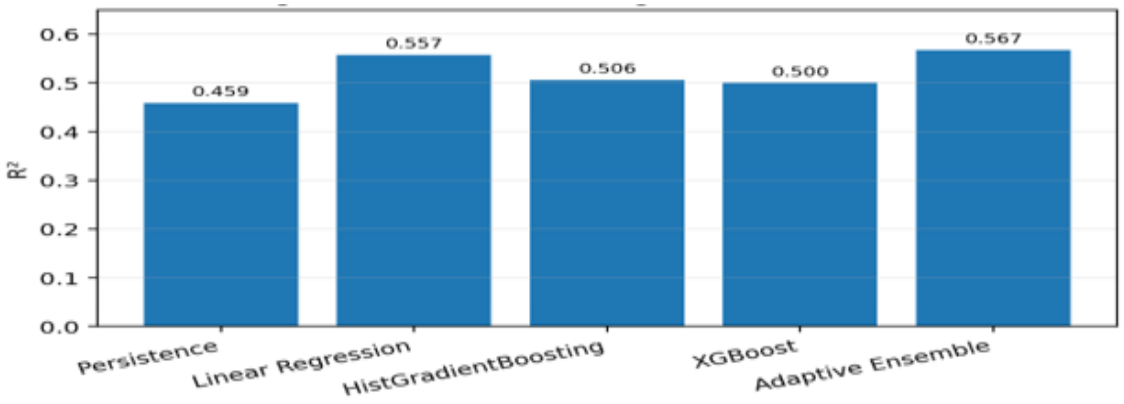


Figure 2. R^2 comparison on the held-out chronological test interval

Figure 2 highlights the comparison between the R^2 values of all the models. From the diagram, one can notice that the Adaptive Ensemble is superior to other prediction models with a percentage of 0.567, whereas the Linear Regression takes the second place with an R^2 equal to 0.557.

5.1. Comparative Analysis

The adaptive ensemble achieved MAE 28.99 Wh, RMSE 59.83 Wh, MAPE 27.12% and R^2 0.567. Its RMSE was lower than the persistence baseline (66.89 Wh), linear regression (60.51 Wh), HistGradientBoosting (63.93 Wh) and XGBoost (64.30 Wh). Its R^2 of 0.567 was also the best of the evaluated models. However, persistence produced the lowest MAE (26.78 Wh) and MAPE (21.54%). This is a significant point: the ensemble reduces substantially large errors and improves the explained variance, but it does not dominate all metrics.

The experiment shows that later-period appliance consumption is considerably more difficult to predict, highlighting the importance of temporal validation. The differences between MAE and RMSE provide valuable insights into error distribution. If a model exhibits a relatively low MAE and higher RMSE, it can be concluded that it rarely produces very large deviations. Nevertheless, the occasional outlier may significantly impact the RMSE. The persistence model demonstrates this tendency, which can be observed in the current experiment.

The adaptive ensemble, in turn, has a slightly higher MAE than persistence, but a notably lower RMSE. This finding suggests that the ensemble method does not necessarily reduce each prediction error. Instead, it appears to minimize large outliers, which ultimately improves the squared-error-based metric. Operationally, this characteristic can be highly desirable if extreme forecasting errors are relatively intolerable compared to numerous small errors.

The results therefore support a multi-objective interpretation of forecasting performance. Depending on the use case, an energy-management system may seek to minimize average error, large error occurrences, variance not explained, or relative error. The proposed ensemble provides its strongest advantage in terms of RMSE and R^2 rather than MAE and MAPE.

5.2. Adaptive Weights

Table 3. Validation errors and normalized inverse-MAE weights used by the ensemble

Base predictor	Validation MAE (Wh)	Adaptive weight
Persistence	25.70	0.252
Linear Regression	26.79	0.241
HistGradientBoosting	26.37	0.245
XGBoost (log-target)	24.74	0.261

As shown in Table 3, the weights are relatively balanced since the prediction errors of the four predictors are close to each other. This is desirable for heterogeneous ensemble because no one learner dominates the output. The results are consistent with a general observation in the literature that model-combination strategy is a significant research issue in building-energy ensemble prediction [2]. The validation errors of the four predictors are relatively close, yielding to a comparatively balanced distribution of the ensemble weights. Persistence receives a weight of 0.252, linear regression receives 0.241, HistGradientBoosting receives 0.245 and XGBoost receives 0.261.

The largest contribution is allocated to XGBoost as it achieves the lowest validation MAE among the four predictors. However, its contribution remains only slightly higher than the contributions of the other models to avoid the ensemble to become effectively equivalent to a single predictor.

The balanced weighting is an important characteristic of the proposed framework: if one model exhibited a substantially lower validation error than the others, the inverse-MAE mechanism would naturally enhance its contribution. Conversely, if a predictor underperformed on the validation interval, its contribution would be reduced. Therefore, the weighting mechanism provides with a simple way of adapting the final prediction according to the observed model reliability. The resulting ensemble can be interpreted as a validation-guided compromise among different forecasting behaviours. Persistence provides with strong short-term continuity, linear regression provides with a low-complexity global relationship and the two boosting models provide with nonlinear modelling capacity.

5.3. Overall Performance Interpretation

The held-out results demonstrate that the proposed adaptive ensemble provides a balanced prediction strategy, although its performance varies depending on the evaluation criterion. The ensemble obtains an MAE of 28.99 Wh and an RMSE of 59.83 Wh, with its R^2 reaching 0.567. These values indicate that the combined model captures a meaningful proportion of the variation in the appliance-energy signal during the unseen chronological test period.

The RMSE result is particularly relevant because it indicates that combining the predictors reduces the influence of larger prediction deviations compared with the individual models. The ensemble RMSE is lower than the RMSE obtained from persistence, linear regression, HistGradientBoosting, and XGBoost. This suggests that the combination process can reduce some of the larger errors produced by individual predictors.

The R^2 result provides a similar indication. The adaptive ensemble achieves the highest R^2 among the evaluated models. Therefore, although the ensemble does not achieve the lowest absolute error, it provides the strongest overall representation of target variance among the compared approaches.

6. FORECAST BEHAVIOUR AND DISCUSSION

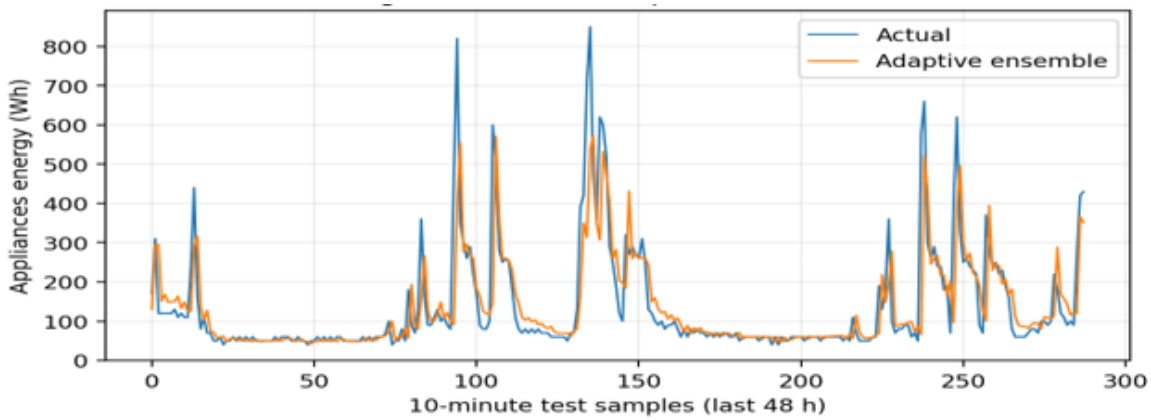


Figure 3. Actual and adaptive-ensemble predictions for the final 48 hours of the held-out test period

The held-out trace in Figure 3 demonstrates that the adaptive ensemble generally tracks the short-term evolution of the appliance-energy series while averaging out some sudden fluctuations. The aggregated metrics suggest that this averaging benefits RMSE and R^2 but also introduces some pointwise deviations that increase MAE to persistence. This behavior is to be expected when a model is optimized for the representation of overall variance rather than just the one-step persistence.

6.1. Feature-Importance Analysis

Table 4. Permutation importance of the HistGradientBoosting model on the held-out test set. Importance values represent the decrease in R^2 caused by permutation and are model- and dataset-specific

Rank	Feature	Permutation importance (R^2 decrease)
1	lag_1	1.269
2	lag_2	0.258
3	rollmean_12	0.019
4	rollmean_3	0.006
5	lag_12	0.004
6	RH_9	0.003
7	lag_6	0.003
8	RH_2	0.003
9	hour	0.003
10	RH_3	0.003

The preponderance of lag_1 in Table 4 suggests a strong short-term persistence in the appliance-energy signal. The second lag also contributes significantly and the rolling features and selected humidity variables add a smaller amount of information. This is consistent with previous work on the same data set which also showed that measured environmental variables and derived sensor information had significant predictive value [4].

6.2. Temporal Behaviour of the Predictions

The final period of the test forecast illustrates how the adaptive ensemble tracks on average the movement of the observed appliance-energy series. Since the ensemble includes multiple predictors, jumps characteristic of individual model outputs are smoothed when the predictors disagree.

This smoothing can be advantageous when isolated fluctuations are poorly predictable from the available sensor information, but it comes at the expense of a deviation from individual observations when the true series changes rapidly. This gives another reason why the ensemble could have lower RMSE but not the lowest MAE. The forecasting behaviour thus reflects a trade-off between responsiveness and stability. Persistence is very responsive to the latest observation because its prediction is directly taken from the previous value. The ensemble includes information from multiple learned relationships and therefore results in a more aggregated estimate. The final behaviour depends on the relative agreement between the four component models.

6.3. Relationship Between Historical and Environmental Information

The feature-importance results suggest that recent historical consumption is the major source of predictive information in the evaluated model. The first lag has significantly higher permutation importance than the other listed variables, followed by the second lag and rolling features.

This finding is compatible with the short-term forecasting nature of the experiment. If the prediction horizon is near the current observation, the most recent consumption value can have more direct information than slowly varying environmental measurements. However, the presence of humidity-related variables among the ranked features suggests that environmental information also affords additional predictive context.

The results therefore support a combined feature representation in which historical energy measurements provide the temporal foundation while environmental and temporal variables provide supplementary information. Removing either category without experimentation could unnecessarily limit the information available to the forecasting models.

6.4. Significance of Lag Features

The dominance of lag_1 highlights the value of immediately preceding time points. Its permutation importance value is substantially greater than the other features. Permuting this feature leads to an especially large reduction in the model's predictive performance, suggesting that the model was relying heavily on the most recent consumption value.

While the second lag adds meaningful information, its effect is substantially smaller. Rolling features summarize recent consumption, and so, together with the lag values, allow the model to capture both the most recent point and a broader local consumption level. This finding has practical implications for deployment. A forecasting system based on this framework must maintain reliable recent energy measurements because missing or delayed historical values could reduce prediction quality. It also suggests that improving the quality and availability of recent sensor observations may be as important as increasing model complexity.

7. DISCUSSION

Three observations stand out. Firstly, evaluating in chronological order produces significantly more conservative results than randomly splitting the data and should be the preferred

approach for deployment-oriented forecasting. Secondly, a lightweight ensemble can improve the RMSE and R^2 even if a simple persistence model continues to have a competitive MAE. Thirdly, the experiment illustrates that the novelty of an ensemble paper cannot solely rely on combining common algorithms – the value must come from the evaluation protocol, adaptive combination strategy, ablation or robustness analysis, and clearly stated limitations. The study has several limitations. The dataset represents a single low-energy house and covers only ≈ 4.5 months, so the results cannot establish generalization to offices, campuses or multiple buildings. Recent reviews highlight cross-building transfer and generalization as important open issues [1], [2]. Future work should therefore evaluate the same leakage-aware pipeline on a multi-building benchmark such as ASHRAE Great Energy Predictor III, add rolling-origin validation, investigate uncertainty intervals and compare adaptive weighting against stacking and optimized blending.

8. CONCLUSION

A leakage-aware adaptive ensemble framework for short-term appliance-energy forecasting was developed and evaluated on the UCI dataset. The actual chronological experiment resulted in R^2=0.567 and RMSE=59.83 for the proposed ensemble. The ensemble was improved RMSE and R^2 compared to the individual machine-learning predictors but did not achieve the lowest MAE or MAPE, and thus multiple metrics should be considered when evaluating a model. The work thereby provides a replicable baseline for further smart-building forecasting research rather than the claim for universality. The extension to multi-building data and use of a rolling validation is the main next step.

REFERENCES

[1] Ji, J., Yu, H., Wang, X., & Xu, X. (2025). Machine learning application in building energy consumption prediction: A comprehensive review. *Journal of Building Engineering*, *104*, 112295. https://doi.org/10.1016/j.jobe.2025.112295

[2] Wang, Z., Hong, Y., Huang, L., Zheng, M., Yuan, H., & Zeng, R. (2025). A comprehensive review and future research directions of ensemble learning models for predicting building energy consumption. *Energy and Buildings*, *335*, 115589. https://doi.org/10.1016/j.enbuild.2025.115589

[3] Candanedo, L. (2017). *Appliances Energy Prediction* [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C5VC8G

[4] Candanedo, L. M., Feldheim, V., & Deramaix, D. (2017). Data driven prediction models of energy use of appliances in a low-energy house. *Energy and Buildings*, *140*, 81-97. https://doi.org/10.1016/j.enbuild.2017.01.083

[5] Khalil, M., McGough, A. S., Pourmirza, Z., Pazhoohesh, M., & Walker, S. (2022). Machine learning, deep learning and statistical analysis for forecasting building energy consumption: A systematic review. *Engineering Applications of Artificial Intelligence*, *115*, 105287. https://doi.org/10.1016/j.engappai.2022.105287

[6] Li, K., Tian, J., Xue, W., & Tan, G. (2021). Short-term electricity consumption prediction for buildings using data-driven swarm intelligence based ensemble model. *Energy and Buildings*, *231*, 110558. https://doi.org/10.1016/j.enbuild.2020.110558

[7] Utama, P. H. K., Leksono, E., Haq, I. N., Indrapraja, R., Nanda, R. M., Friansa, K., Iskandar, R. F., & Pradipta, J. (2025). Adaptive ensemble learning for enhancing building energy consumption prediction: Insights from COVID-19 pandemic energy consumption dynamics.

Journal of Engineering and Technological Sciences, *57*(2), 161-174. https://doi.org/10.5614/j.eng.technol.sci.2025.57.2.2

[8] Neshat, M., Thilakaratne, M., El-Abd, M., Mirjalili, S., Gandomi, A. H., & Boland, J. (2025). Smart buildings energy consumption forecasting using adaptive evolutionary bagging extra tree learning models. *Energy*, *333*, 137130. https://doi.org/10.1016/j.energy.2025.137130

[9] Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5-32. https://doi.org/10.1023/A:1010933404324

[10] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785-794). Association for Computing Machinery. https://doi.org/10.1145/2939672.2939785