Energy & Buildings 349 (2025) 116589

Contents lists available at ScienceDirect

Energy & Buildings

journal homepage: www.elsevier.com/locate/enbuild

Enhancing residential load forecasting accuracy through dynamic feature selection and ensemble machine learning models: A real-world scenario in Southern Finland

Nabi Taheri [a,*], Lauri Karttunen [b], Sami Jouttijärvi [b], Antonio Piazzi [c], Mauro Tucci [a], Kati Miettunen [b]

[a] Department of Energy, Systems, Territory and Construction Engineering, University of Pisa, Via Carlo Francesco Gabba 22, Pisa, 56122, PI, Italy
[b] Department of Mechanical and Materials Engineering, University of Turku, Vesilinnantie 5, Turku, 20500, Turku, Finland
[c] Department of Data Science, i-EM S.r.l, Lampredi 45, Livorno, 57121, LI, Italy

ARTICLE INFO

Keywords:
Load forecasting
Residential electricity consumption
Dynamic feature selection
SHAP
Ensemble learning
Machine learning

ABSTRACT

This study aims to improve the accuracy of residential electricity consumption forecasts, a key component of modern energy management systems, especially in settings with high demand variability. The novelty of this work lies in combining dynamic feature selection–adapting to recent data patterns–with a stacking-based ensemble model that leverages multiple predictors. A comprehensive hourly dataset from 200 residential buildings in southwestern Finland was collected throughout 2023, including meteorological variables and real-time electricity prices. The forecasting task was defined as a day-ahead (24-hour ahead) prediction. The year was divided into cold (October–March) and warm (April–September) periods. Four machine learning models were employed per season, including XGBoost, Random Forest, Voting, and a Stacking ensemble with Ridge regression as the meta-learner. The stacking model achieved the best performance in 163 buildings in the cold period and 178 buildings in the warm period. Feature importance was assessed using SHAP values, comparing static and dynamic feature selection strategies. The dynamic approach reduced average prediction error from 11.85 % to 9.31 % in cold months and from 11.67 % to 9.14 % in warm months, outperforming the static method in over 96 % of buildings. These findings underscore the effectiveness of adaptive feature selection and ensemble learning in capturing seasonal and behavioral dynamics in residential electricity usage.

1. Introduction

Short-term load forecasting–typically referring to predictions from a few minutes up to a few days ahead–plays a vital role in ensuring the reliable and cost-effective operation of modern power systems. As the share of renewable energy sources grows and electricity pricing becomes more dynamic, the ability to accurately predict demand becomes increasingly important. A reliable forecast allows system operators to schedule generation more efficiently, reduce dependence on costly reserve power, and maintain grid stability. At the residential level, accurate forecasting enables more effective demand response, supports the integration of distributed energy resources, and contributes to overall improvements in energy efficiency. In this study, short-term load forecasting specifically refers to day-ahead prediction, where hourly consumption values are used to forecast the next 24 hours.

While much progress has been made in forecasting aggregated or industrial loads, residential electricity consumption remains significantly more challenging to predict due to its highly individualized, irregular, and behavior-driven nature. Unlike commercial buildings, where usage patterns are often systematic and equipment-driven, residential consumption is influenced by diverse factors such as occupant routines, weather sensitivity, economic behavior, and the presence of smart devices or automation systems [1,2]. Furthermore, load profiles in residential settings can change not only from season to season but also from one day to the next, making traditional fixed-model approaches less effective. To address these challenges, this study proposes a novel forecasting framework that integrates dynamic feature selection, adaptively tailored to recent behavioral and environmental patterns, with a stacking ensemble learning strategy. This novel integration enables the model to remain responsive to changing conditions, significantly enhancing forecasting accuracy and robustness across diverse residential contexts.

To navigate this complexity, researchers have increasingly turned to machine learning techniques that can capture nonlinear relationships and hidden patterns in the data [3,4]. However, the success of these

---

* Corresponding author.
E-mail address: nabi.taheri@phd.unipi.it (N. Taheri).

https://doi.org/10.1016/j.enbuild.2025.116589
Received 10 July 2025; Received in revised form 8 October 2025; Accepted 14 October 2025

models depends heavily on the relevance of the input features used. Without careful feature selection, even the most advanced algorithms may struggle with noisy data, risk overfitting, or produce results that are difficult to understand or explain. As such, the integration of interpretable learning methods with adaptive feature selection has become an emerging direction for improving forecasting performance in real-world residential applications [5]. In parallel, adaptive rule-based mechanisms have also been proposed to fine-tune forecasting models based on feedback from real-time operational conditions, combining fuzzy logic and neural networks to dynamically correct prediction errors [6].

Recent advancements in machine learning have introduced a variety of approaches to tackle the complexities of residential load forecasting. Traditional methods such as autoregressive integrated moving average (ARIMA) models remain popular for their simplicity but often struggle to capture the non-linear and stochastic nature of household energy use [2,7]. In response, more flexible techniques such as support vector regression (SVR) and Gaussian process regression (GPR) have been adopted for their ability to model complex relationships, although their performance heavily depends on kernel selection and hyperparameter tuning [8–10]. However, ensemble methods have emerged as particularly effective for residential load forecasting due to their robustness against noise and ability to combine multiple weak learners into a stronger predictive model. Random forests (RF), for instance, have been widely used for their inherent feature selection capability and resistance to overfitting, making them suitable for high-dimensional smart meter data [11]. Gradient boosting frameworks such as XGBoost and LightGBM further improve accuracy by iteratively correcting errors from previous models, often outperforming standalone algorithms in short-term load forecasting tasks [12,13]. Recent studies have also explored stacked generalization, where meta-learners combine predictions from diverse base models to enhance overall performance [14]. Meanwhile, explainable Artificial Intelligence (XAI) techniques like SHAP (Shapley Additive Explanations) are increasingly applied to interpret ensemble model decisions, providing insights into feature importance and consumption patterns [15]. A hybrid deep ensemble approach with hyperparameters optimised via Adaptive Wind-Driven Optimization (AWDO) has been shown to outperform conventional and state-of-the-art models in residential short-term load forecasting tasks [16]. Ultimately, while no single method universally dominates, ensemble techniques remain a cornerstone in residential load forecasting due to their adaptability and consistent performance across diverse datasets [17].

Feature selection plays a crucial role in improving the accuracy and efficiency of short-term load forecasting, especially in residential contexts where energy consumption is highly variable and nonlinear. Several recent studies have demonstrated that identifying a compact and relevant subset of input variables can significantly reduce model complexity and improve generalization [18–20]. For instance, integrating feature selection with machine learning models such as gradient boosting or random forests has been shown to enhance performance across different seasons and household profiles [18,21]. In particular, wrapper and filter methods like mutual information, recursive feature elimination, and permutation importance have been effective in identifying key predictors while minimizing redundancy [19,21,22]. Hybrid frameworks that combine feature selection with deep learning architectures have also emerged, using tools such as Pearson correlation, SHAP values, or tree-based importance rankings to guide the model input structure [23–25]. These techniques not only improve accuracy but also make the models more interpretable and adaptable to context-specific patterns. In comparative experiments, models with feature selection consistently have outperformed their full-feature counterparts, showing improved accuracy metrics such as lower RMSE and MAPE across diverse test conditions [25–28]. Taken together, these findings highlight that effective feature selection is not just a preprocessing step but a foundational component of modern short-term load forecasting systems.

Most reviewed studies above on feature selection adopt a fixed set of features determined from a specific time window and apply the same set throughout the entire forecasting period [21,24,25]. However, the short-term correlation degree and stability of the load variables have been shown to be inconsistent, significantly impacting the accuracy of the final prediction model [25]. While this approach may be sufficient in relatively stable domains, such as industrial electricity consumption or renewable energy generation from wind or solar, it is often inadequate for residential load forecasting. Electricity consumption in residential buildings is influenced by a variety of dynamic factors, including changing weather conditions, daylight hours, occupancy shifts, and behavioral patterns. These factors can fluctuate not only between seasons but even from week to week or day to day, making static feature sets less effective and potentially detrimental to prediction accuracy. This is particularly relevant in high-latitude regions such as Finland, where seasonal variations in daylight, temperature, and weather patterns are more extreme, further amplifying the need for adaptive feature selection strategies. It is worth noting that the Nordic conditions significantly impact the production of photovoltaic (PV) energy, as well as allow different PV installations compared to lower latitudes; in fact, there is very active research on optimizing the production of Nordic PV to match the consumption of electricity and increase its profitability [29–33]. The increase of solar energy in private households in Nordics motivates developing smart home energy management systems, where predicting consumption patterns is one key issue and motivation for this work.

To address the limitations arising from high seasonal variations, a more flexible and adaptive feature selection process is introduced in this study, one that can respond to changes in the relevance of input variables over time. The dataset is first introduced, followed by the application of SHAP as the baseline method for feature importance analysis. By applying SHAP to specific case datasets, it is demonstrated that the importance of input features is highly sensitive to time, further reinforcing the need for a dynamic approach. Several machine learning models are then evaluated, including Random Forest, XGBoost, and ensemble techniques such as Voting and Stacking. The most effective model is selected based on comparative performance across different feature configurations. A dynamic feature selection mechanism is subsequently developed to supply this model with an updated and context-aware set of inputs at each forecasting step. The effectiveness of this dynamic strategy is benchmarked against static feature selection and against using the full feature set without any selection.

2. Description of the dataset

This study focuses on analyzing the electricity consumption patterns of 200 residential buildings located in southern Finland. The anonymous dataset was provided by the local distribution system operator, Vakka Suomen Voima (VSV). The data includes hourly electricity consumption values for each building. Fig. 1 presents the hourly load profiles for the period from November 20, to November 26, 2023, across six sample buildings.

As can be observed, the electricity consumption patterns generally do not follow a regular trend. Among these, Building B appears to exhibit relatively consistent behavior over several days, showing a systematic daily pattern compared to the others. In contrast, the other five buildings exhibit highly irregular load patterns, making accurate prediction significantly more challenging. These observations highlight the inherent variability in electricity consumption among different buildings, which may stem from diverse operational schedules, occupancy rates, or external influences such as holidays and weather conditions during this period. In addition to electric energy usage, hourly weather data were also collected, including temperature (minimum, maximum, and average), Global Horizontal Irradiance (GHI), and humidity levels. Due to the significant temperature variations that can occur throughout the day and night in Nordic countries, all three temperature indicators were included in the dataset. While wind speed may have some influence on residential electricity use, it was not included in this study because such data are generally unavailable at the building level, especially in urban

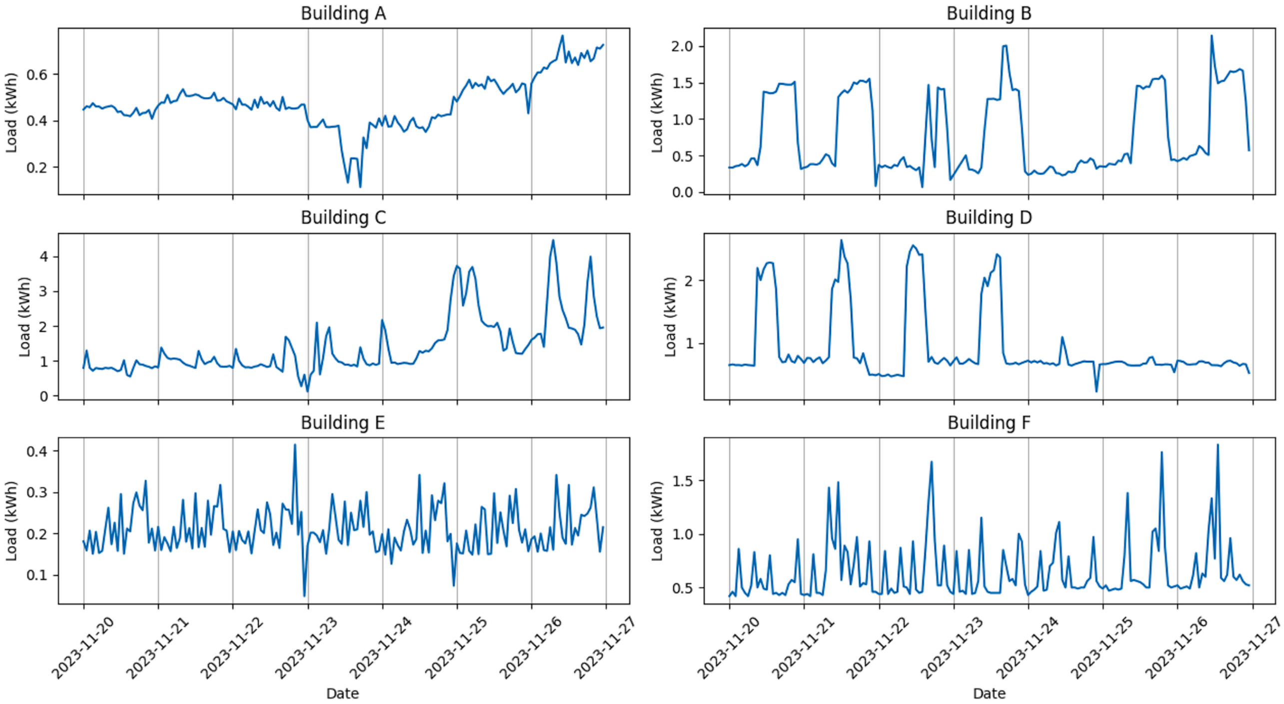

Fig. 1. Building electricity consumption – November 20, to November 26, 2023.

environments where most residential buildings are located. It is important to note that no information is available about the occupants or the type of occupancy (e.g., owner-occupied, long-term rental, or short-term rental) for these buildings. This absence of building-level metadata imposes certain limitations. Without information such as occupancy type, contract scheme, type of building structure, and insulation or household size, it is more difficult to interpret consumption patterns in relation to behavioral or contractual drivers. Moreover, the lack of such contextual data may limit the external generalization of the results to other populations of buildings. However, their effects are to some extent captured by features such as the hour of day, day of week, and electricity price. Since the exact locations of the residential buildings were unavailable, weather data from the city of Turku, which is the largest city in southwestern Finland and is located at most 100 km from the studied buildings, was used. The weather information was retrieved from the official website of the Finnish Meteorological Institute [34]. Furthermore, hourly electricity prices are included in the dataset, based on Finland's dynamic pricing system [35]. In Finland, one of the commonly used electricity pricing schemes is the spot price contract, where the electricity price varies hourly based on the real-time electricity market.

As stated, the primary goal of this study is to forecast the electric energy demand of residential buildings using machine learning (ML) techniques. For this purpose, input features include temperature, GHI, humidity, electricity price, hour of the day, and day of the week. The target variable is the building's electricity consumption. All data is structured on an hourly time resolution, and the prediction is made for the next day, hour by hour. It is important to note that many of the buildings may have a fixed price electricity contract. However, since detailed contract information for each building is not available in the dataset, no segmentation based on contract type has been performed. It is reasonable to assume that buildings using the spot pricing scheme will be more directly affected by changes in electricity prices, while those on fixed-rate or time-of-use contracts may show electricity consumption patterns influenced more by the time of day than by price fluctuations. In this study, Finnish public holidays are treated similarly to Sundays in the analysis, as per the national calendar. Fig. 2 illustrates the relationship between the input features and the target variable.

3. Methodology

This section begins with an introduction to the SHAP analysis method, detailing its application to the available dataset. The machine learning models employed in the study are then presented. Lastly, the implementation of the dynamic feature selection approach is described, along with a comparison to a static feature selection method.

3.1. Feature importance analysis

Forecasting electricity consumption in residential buildings involves several sources of uncertainty. Unlike industrial or commercial environments, residential electricity consumption is highly dependent on the daily routines and personal habits of individuals, making it inherently more variable and difficult to predict. For instance, some residents may prioritize maintaining indoor temperature, while others may focus more on natural lighting. Certain users follow predictable routines involving energy-intensive activities such as showering, doing laundry, or dishwashing. Although direct data on specific occupant actions, such as switching lights on or opening windows, were not available, these behavioral effects are indirectly captured through temporal features such as hour of the day, day of the week, and holidays, which reflect typical routines like returning home, turning on lights, or going to sleep. In humid regions, especially during summer, humidity can become a significant driver of electricity consumption due to air conditioning or dehumidification systems. However, in Nordic countries, dehumidification is relatively uncommon; interestingly, humidifying indoor air during the extremely dry winter months is more typical.

Moreover, the day of the week may impact consumption patterns, as residents with flexible or remote work arrangements may spend certain weekdays at home, and weekends are expected to have a different electricity consumption pattern than other days. Additionally, some buildings may be used for short-term rentals, introducing irregular consumption patterns. Beyond the direct electricity demand for heating or cooling, occupants' behavior is often shaped by weather conditions – for instance, in summer, people tend to spend more time outdoors on sunny days, while rainy or humid weather may lead them to stay indoors

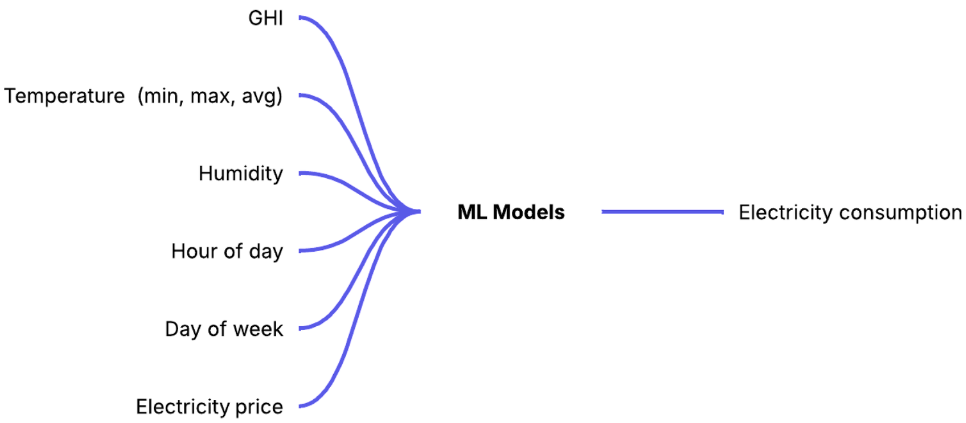

Fig. 2. Structure of Features and Target Variable.

and use more electrical appliances. These diverse patterns suggest that selecting appropriate input features is critical to achieving accurate predictions. Including all available features may not only increase the computational burden but also lead to poorer predictions due to noise from irrelevant variables. Therefore, analyzing feature importance on a per-building basis is essential. Understanding which features most influence electricity consumption helps tailor predictive models more effectively and improve accuracy. To identify the most relevant input variables, the SHAP (SHapley Additive exPlanations) method is applied–a widely used interpretability approach grounded in game theory. SHAP assigns each feature an importance value that represents its contribution to a specific prediction. SHAP values are based on Shapley values from cooperative game theory. The Shapley value for a feature i is given by:

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|! \cdot (|F| - |S| - 1)!}{|F|!} \left[f_{S \cup \{i\}}(x) - f_S(x) \right] \tag{1}$$

where F is the set of all features, S is a subset of features not containing i, $f_S(x)$ is the model trained with subset S, evaluated at input x, and ϕ_i represents the marginal contribution of feature i. Since computing exact Shapley values is computationally expensive, especially for large feature sets, SHAP uses efficient approximations [36]. In this study, the XGBoost method is employed for the computation of SHAP values. SHAP helps identify the most influential features for each building, facilitates the elimination of less relevant variables to reduce overfitting, and enhances the transparency of model predictions.

As mentioned in the data description section, this study analyzes electricity consumption across 200 residential buildings. The aim of this stage is to demonstrate that electricity consumption patterns are not governed by a single fixed set of features across all buildings. This analysis has a significant impact on the accuracy of electricity consumption forecasting. Therefore, it is crucial to personalize feature selection as much as possible for each building and to base predictions on the features that are most relevant in each case. To this end, as mentioned, the SHAP method is applied to analyze the importance of features. In this step, 8 selected features are analyzed for all 200 buildings to identify their individual impact on consumption behavior. The results of these analyses will be shown in the results section.

3.2. Overview of machine learning models

To develop a reliable residential electricity consumption forecasting framework, both individual and ensemble machine learning models are considered. Two widely used base models–Extreme Gradient Boosting (XGBoost) and Random Forest (RF)–are selected for their strong performance in time-series prediction tasks [1,37]. Additionally, ensemble techniques such as stacking and voting are introduced to combine model outputs and enhance predictive accuracy by leveraging the strengths of multiple algorithms.

Extreme Gradient Boosting (XGBoost)

XGBoost is a powerful ensemble learning method based on gradient boosting, designed to improve model performance through the sequential training of decision trees, where each new tree corrects the errors of its predecessors. It is known for its high efficiency, scalability, and ability to handle missing values and overfitting through regularization. In the context of load forecasting, XGBoost has been widely adopted due to its robustness and accuracy in capturing nonlinear relationships in time-series data [1].

Random Forest (RF)

RF is a learning algorithm that builds multiple decision trees during training and outputs the average prediction (in regression tasks) of the individual trees. It reduces overfitting by combining the results of many weak learners, offering robust performance even with noisy or complex datasets. Due to its ability to model non-linear relationships and handle high-dimensional data, Random Forest is frequently used in electric energy forecasting tasks. In residential load forecasting, RF has been shown to deliver strong baseline performance, particularly in cases where weather-related and temporal features influence consumption patterns [2].

Voting Ensemble Method

The voting ensemble method is a straightforward yet powerful technique that combines the predictions of multiple base models to improve overall performance. In regression tasks, such as load forecasting, soft voting (averaging predicted values) is typically used. This approach helps reduce the variance and bias of individual models by aggregating their outputs, leading to more stable and accurate predictions. In the context of electric energy forecasting, the voting ensemble method has shown promising results, especially when combining models that capture different aspects of the data [3]. For example, combining tree-based models like XGBoost and Random Forest allows the ensemble to leverage the strengths of both–XGBoost's gradient-boosted precision and Random Forest's robustness against noise and overfitting.

Stacking Ensemble Method

Stacking, or stacked generalization, is a more advanced ensemble technique that combines the predictions of multiple base models through a meta-model, which learns how to best combine the base predictions to improve overall accuracy. Unlike voting, which uses simple averaging or majority rules, stacking trains a second-level model–often a linear regressor or another machine learning algorithm–on the outputs of the base models to make the final prediction. This method is particularly useful when the base models capture different patterns or features in the data. In residential electric energy forecasting, stacking can effectively integrate diverse modeling approaches, such as decision tree ensembles (e.g., XGBoost and Random Forest), to capture complex consumption behaviors influenced by both environmental and temporal factors [1,4,14].

In stacked generalization, the choice of the meta-learner plays a crucial role in effectively combining the outputs of base models. Among the most common approaches, SVR has been widely used for nonlinear regression tasks. As explained [38], SVR fits a regression function within an ϵ-insensitive margin, penalizing only deviations larger than ϵ, and employs kernel functions to capture complex nonlinear relationships. Despite its strong generalization ability, SVR requires solving a quadratic optimization problem, which can become computationally demanding in practice. Another widely adopted option is LightGBM, a gradient boosting framework optimized for efficiency and scalability. As noted in [39], LightGBM is a highly efficient tree-based method that reduces training time and memory usage while maintaining state-of-the-art predictive accuracy. Owing to its regularization mechanisms and ability to capture nonlinear dependencies, LightGBM has become a popular choice for regression and forecasting tasks in recent years. Finally, Ridge Regression offers a simpler yet effective alternative. As described by [40], ridge regression introduces an L_2 penalty that shrinks coefficient estimates toward zero, mitigating instability caused by multicollinearity and reducing variance at the cost of a small bias. This bias–variance trade-off often leads to improved predictive accuracy, especially in scenarios with correlated features or limited sample sizes. Moreover, ridge regression has a closed-form solution that only requires solving a linear system, making it computationally lightweight compared to SVR or tree-based methods.

In this study, three meta-learners (SVR, LightGBM, and ridge regression) are evaluated as candidates for the stacking framework. As shown in Section 4.4, their forecasting accuracy is very similar in this case, where each sliding window includes only a small number of features and samples. Therefore, following the recommendation in [5,40] and considering its simplicity, stability, and efficiency, Ridge Regression was ultimately selected as the default meta-learner in the framework.

To assess the performance of the forecasting models, two commonly used evaluation metrics were employed: the Normalized Root Mean Squared Error (NRMSE), and the Mean Absolute Error (MAE). Using multiple complementary metrics provides a more comprehensive evaluation of model accuracy and robustness across different buildings. The NRMSE provides a scale-independent measure of prediction error, which is particularly useful when comparing results across buildings with varying consumption ranges. It is defined as:

$$\mathrm{NRMSE} = \frac{\mathrm{RMSE}}{\mathrm{Range}} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_{\mathrm{true},i} - y_{\mathrm{pred},i})^2}}{y_{\max} - y_{\min}} \tag{2}$$

where n denotes the number of samples, $y_{\mathrm{true},i}$ is the observed consumption at time i, $y_{\mathrm{pred},i}$ is the predicted consumption, and Range = $y_{\max} - y_{\min}$ is the range of the target variable. A lower NRMSE indicates better accuracy [41]. The MAE is a widely used error measure that captures the average magnitude of the prediction errors without considering their direction. Unlike RMSE, it is less sensitive to large outliers. The MAE is defined as:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_{\mathrm{true},i} - y_{\mathrm{pred},i}\right| \tag{3}$$

A smaller MAE indicates higher predictive accuracy. MAE is commonly used in energy forecasting studies due to its interpretability in the original measurement units (e.g., kWh) [42]. By combining NRMSE and MAE, the forecasting performance can be assessed from complementary perspectives, namely relative scale-independent accuracy, average absolute error in the original units, and explanatory power of the model.

3.3. Proposed method

As previously discussed, SHAP-based feature importance analysis provides valuable insight into identifying the most influential variables for each individual building. These findings emphasize the importance of adopting an effective feature selection strategy. However, it is important to note that SHAP, like other similar methods, only ranks features based on their relative importance and impact on the model output. It does not indicate which feature combination will lead to the highest prediction accuracy. For this reason, it becomes necessary to determine the optimal feature set. According to the dataset description, eight features, including temperature (minimum, maximum, and average), GHI, humidity, electricity price, hour of day, and day of week, are considered as the set of features available for prediction. In this study, the best-performing feature set can include any subset of the eight available features, from a single feature to combinations involving two up to all eight features. Two general approaches can be adopted to address this challenge: a static approach and a dynamic approach.

The static approach relies on a fixed training dataset, such as one from the previous year or a specific period that represents a larger seasonal pattern. Based on this dataset, features are ranked by importance, and the best-performing combination is selected and applied consistently throughout the target forecasting period. But as will be shown in the results section, the influence of features on prediction accuracy varies over time. In other words, a set of features that significantly improved prediction performance during one period may not necessarily be the most effective across all time frames. The dynamic approach, which is the primary focus of this work, involves selecting the best feature set for each training window and applying it to the nearest future prediction window. This approach allows the forecasting process to adapt to changing environmental conditions and user behavior patterns. To better illustrate the concept, both approaches are described in detail.

The static approach is considered first. In this setting, one representative period is selected to determine the feature set, which is then applied to the entire forecasting horizon. Following this convention, SHAP analysis results from January were used to represent the colder months (October–March), and results from June were considered representative of the warmer months (April–September). It should be emphasized that January and June were chosen only as illustrative periods for the static method and are not intended to represent the entire cold or warm seasons. This simplification reflects the common practice in fixed-window feature selection approaches and serves here as a baseline for comparison with the proposed dynamic method. Although July is typically the warmest month in southern Finland, it coincides with the holiday season, leading to atypical consumption behavior; for this reason, June–the next warmest month–was selected as a more reliable representative of the warm period. Based on this configuration, the forecasting process proceeds as follows: the first week of the 6-month time period studied is used as the training period, followed by the next day as the test period. Features identified through SHAP are incrementally added to the feature set, one at a time, and a stacking model is employed to generate forecasts using each updated feature combination. Next, the sliding window is shifted forward by one day, and the process is repeated until the entire 6 months are covered. The same process is then carried out for the next 6-month period.

In the dynamic approach, a separate SHAP analysis is performed for each training window to ensure that the selected features are continuously updated in response to changing environmental conditions or evolving electricity consumption behavior. The process begins by using the first week of 2023 as the training data, and the following day is designated as the test day. This rolling procedure was applied across the full year of 2023 for all 200 buildings, so that each forecast was based on the most recent 7 days (168 hours) of data to predict the subsequent 24 hours. In this way, training and testing were repeated in a sequential manner throughout the year, ensuring that the evaluation covered a wide range of seasonal and behavioral variations. This design provides a robust validation of the proposed method, since performance was assessed across different buildings, time periods, and operating conditions rather than being limited to a single short interval. SHAP analysis is carried out based on this training window. Forecasting for the next day is then conducted using the stacking model, with features incrementally added according to their SHAP-derived importance

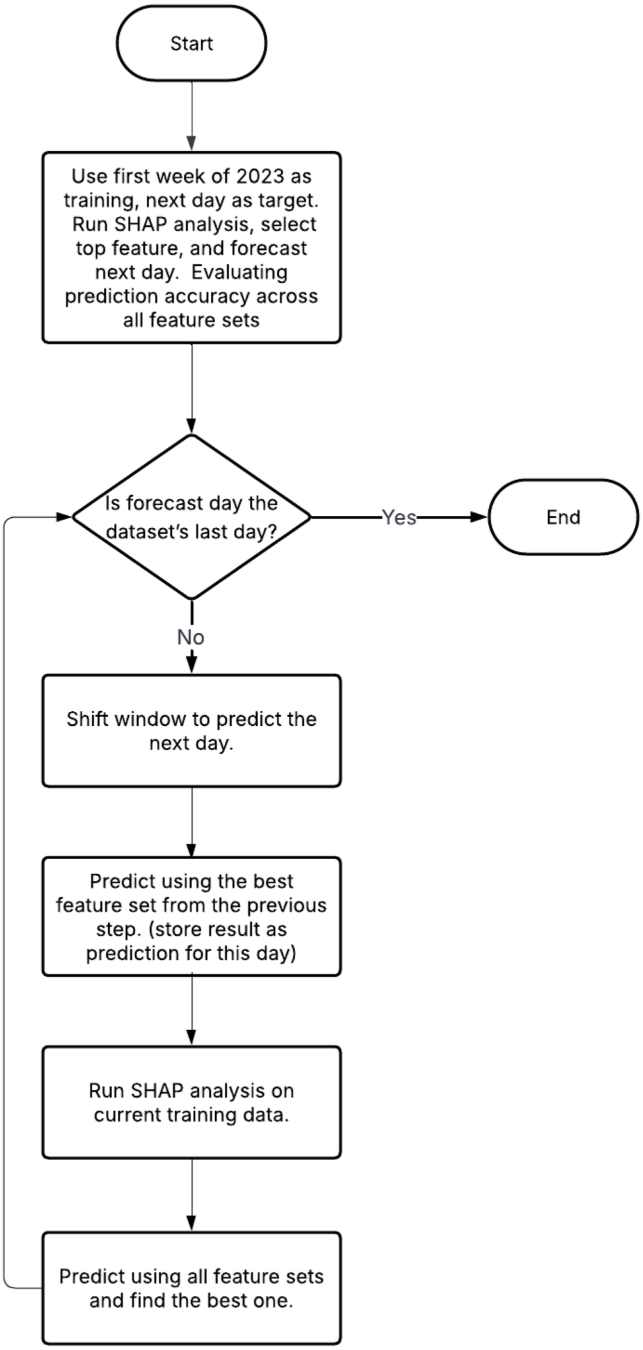

Fig. 3. Workflow of the dynamic feature selection and forecasting Process.

ranking. In each case, predictions are generated for all possible feature subsets. At this point, the actual electricity consumption data for the test day is unavailable, as the day has not yet occurred. Once the window is shifted forward by one day, the actual consumption data for the test day becomes available. At this point, the accuracy of all feature set combinations can be evaluated, and the best-performing configuration is identified. This winning configuration is then used to predict consumption for the next day. Naturally, for the first iteration (i.e., the first forecasted day), no prior performance result is available, and therefore, the prediction based on the most important SHAP-derived feature is assumed to represent the optimal feature set. Despite selecting a preferred configuration each day, the system continues to evaluate all possible combinations daily and stores their corresponding predictions. The sliding window advances by one day at a time, and after each new day, as the actual consumption values become available, the best configuration is determined and applied to the next prediction. This iterative process continues throughout the entire year, ensuring that the most recent and relevant training data is always used to determine the most appropriate feature set. To clarify the dynamic feature selection process, a summary of the steps involved in dynamic forecasting is illustrated in the flowchart shown in Fig. 3.

4. Results and discussion

This section first presents the results of SHAP analysis to show how changes in training data during different periods of the year affect the feature importance ranking. Then, with the results of this analysis, the best machine learning model is selected from the introduced methods, and finally, based on the static and dynamic methods introduced earlier, prediction models are developed, and their performance is compared.

4.1. SHAP-based feature importance evaluation

In this section, a SHAP-based feature importance analysis is conducted to explore how the influence of different input variables on electricity consumption changes across buildings and between cold and warm months. For this analysis, as mentioned in the previous section, January 2023 was selected as a representative of cold months, and June 2023 was chosen as a representative of warm months to serve as the reference dataset. Ideally, and as will be done in this study, the feature importance analysis should be performed every time a new prediction is made, allowing the model to be trained using the most relevant features for that specific time frame. However, at this stage of the study, the primary goal is to demonstrate the variability in feature importance across different buildings. The results of the feature importance analysis are presented in Tables 1 and 2. Since 196 distinct combinations of the eight analyzed features were observed, it is unfeasible to display all of them in the table. Therefore, the tables only include the number of buildings associated with the most essential feature and the top two most influential features. It should be noted that this number does not represent all theoretically possible combinations of the eight features, but only those that emerged in practice as the most influential in at least one building. This summary provides a clearer overview of which features most influence electricity consumption in the buildings studied in January and June.

Table 1
Feature combinations and their frequencies (January).

Top Feature	No.of Bldgs	Top Two Features	No.of Bldgs	Top Two Features	No.of Bldgs	Top Two Features	No.of Bldgs
hour	114	hour, temp_max	51	price, temp_max	11	temp (avg, min)	2
temp_max	56	hour, price	31	temp (avg, max)	7	humidity, temp_max	2
temp_avg	11	hour, temp_min	18	GHI, temp_max	5		
GHI	6	hour, temp_avg	17	hour, humidity	4		
day_of_week	6	temp (max, min)	16	price, temp_avg	4		
temp_min	4	GHI, hour	15	day_of_week, temp_max	3		
price	3	day_of_week, hour	12	GHI, day_of_week	2		

Table 2

Feature combinations and their frequencies (June).

Top Feature	No.of Bldgs	Top Two Features	No.of Bldgs	Top Two Features	No.of Bldgs	Top Two Features	No.of Bldgs
hour	96	GHI, hour	44	hour, temp_max	10	price, temp_max	4
GHI	27	hour, price	17	humidity, price	7	temp (avg, min)	3
price	22	hour, humidity	14	GHI, temp_max	7	humidity, temp_max	2
temp_avg	15	hour, temp_avg	13	temp (max, min)	7	GHI, temp_min	2
temp_max	15	GHI, temp_avg	12	price, temp_avg	5	GHI, day_of_week	2
day_of_week	10	hour, temp_min	11	temp (avg, max)	5	price, temp_min	2
temp_min	9	GHI, price	11	day_of_week, price	4	humidity, temp_avg	2
humidity	6	day_of_week, hour	11	GHI, humidity	4	day_of_week, temp_max	1

To clarify the interpretation of the results presented in these tables, it should be noted that, for example, the two-feature combination refers to the specific pair of features that would be selected based on SHAP analysis if exactly two out of the eight available features were to be used for prediction. Across the 200 buildings analyzed, a total of 196 distinct combinations actually appeared as top-ranked in the SHAP results. The main forecasting framework of this study, however, relies on incrementally adding features according to their SHAP ranking, leading to eight configurations per building, while the reported combination counts are provided only as a descriptive summary of the SHAP analysis. This analysis and the resulting feature rankings are determined based on the internal machine learning model used within the SHAP framework, which in this study is the XGBoost algorithm. As shown in the results presented in Table 1, the most influential feature across the buildings is the hour of the day, which was identified as the most important factor for 114 buildings. Another highly significant feature is temperature. Given that the analysis was conducted for the month of January, this result is expected, as colder weather typically leads to increased electricity consumption for heating purposes. Other features, such as electricity price, day of the week, and GHI, follow in terms of importance. Meanwhile, humidity appears to have had a minimal impact on electricity consumption during this period, and it is not chosen as a top feature in any building, which is reasonable considering the study was carried out in southern Finland during winter.

Looking at Table 2, which presents the results for the month of June, it is shown that temperature, which was the second most influential factor in January, has lost its dominant role. In fact, even in combinations of the top two features, temperature no longer holds a prominent position. This shift is expected given the seasonal conditions. June, July, and August in southern Finland typically bring milder weather and a reduced need for space heating, which naturally decreases the influence of temperature on electricity consumption. Interestingly, solar radiation gained more importance in June, which also made sense in the absence of strong thermal demand. In such conditions, the correlation between electric energy usage and sunlight availability becomes more prominent than the relationship with temperature. One notable observation across both tables is that Hour of Day consistently ranks as the most important feature in both January and June. This consistency suggests that most buildings follow relatively structured daily routines, making time of day a stable and highly predictive factor throughout the year. To provide a clearer illustration of the impact of individual features on building electricity consumption, a visual representation of the SHAP values is presented in Fig. 4. This visualization is shown for 6 randomly selected buildings.

To further explore the SHAP plots presented in Fig. 4 and analyze how electricity consumption responds to variations in input features, several buildings are examined in detail. For instance, in Building A during January, the maximum temperature appears at the top of the feature importance list, indicating that electricity consumption in this building is strongly influenced by this variable. Notably, the red data points–corresponding to higher temperature values–are concentrated on the left side of the plot, suggesting that increased temperatures are associated with lower electricity consumption. In contrast, blue points–indicating lower temperatures–are primarily located on the positive side, implying that colder days are linked to increased consumption, which aligns with expectations during the cold month of January. In the same building during June, humidity ranks third, suggesting that electricity usage is less influenced by this feature compared to electricity price and average temperature. However, a closer inspection reveals that the red points (high humidity values) tend to cluster on the positive SHAP side, indicating that electricity consumption tends to increase on more humid days. This might be due to residents spending more time indoors during rainy summer days, engaging in activities such as house cleaning or watching television, which could naturally raise electricity usage. Similarly, in Building E during June, focusing on the humidity feature shows that most blue points (low humidity values) are concentrated on the negative side of the plot. This indicates that lower humidity is associated with reduced electricity consumption. A plausible explanation is that on dry and sunny days, residents may prefer to spend more time outdoors, which in turn lowers their electricity use at home. It is important to note, however, that the impact of a single feature may not always be fully interpretable in isolation. In many cases, meaningful insights emerge only when a feature is considered alongside others, as interactions between variables often play a significant role in shaping consumption patterns.

Looking across these SHAP summary plots, one major pattern stands out: temperature-related features dominate electricity consumption predictions in January but lose much of their influence by June. This result is reasonable, considering the local climate. The winters in southern Finland are long and cold, requiring significant heating in homes; thus, temperature is a highly informative feature for predicting electricity usage. In contrast, summers in that region are generally mild and lack large-scale cooling. As a result, by June, behavioral patterns (like hour of day) and economic factors become more dominant drivers of electricity consumption, while temperature features drop in importance. What is fascinating, however, is the way this pattern is expressed differently across buildings. The SHAP plots do not show a universal ranking of features; instead, they reflect the individuality of each home – the structure, the occupants, their habits, and possibly even the appliances they use. This variation in electricity consumption behavior, both between different buildings and across months, highlights the importance of feature importance analysis in load forecasting. It confirms that a single, fixed set of input features cannot be applied uniformly to all buildings or throughout the year. A dynamic and flexible approach to feature selection is essential to accurately capture the unique characteristics of each situation. It should be noted that SHAP-based interpretability not only provides insights into the relative importance of features but can also support actionable decision-making. For example, grid operators may use the interpretability results to design targeted demand response programs, adjust tariff structures, or anticipate periods of high sensitivity to weather conditions. At the consumer level, SHAP analysis can highlight how electricity usage responds to factors such as price or temperature, thereby informing behavioral adjustments or the adoption of smart energy management practices. In this way, the interpretability offered by the proposed framework enhances not only model transparency but also its practical value for energy management and policy design.

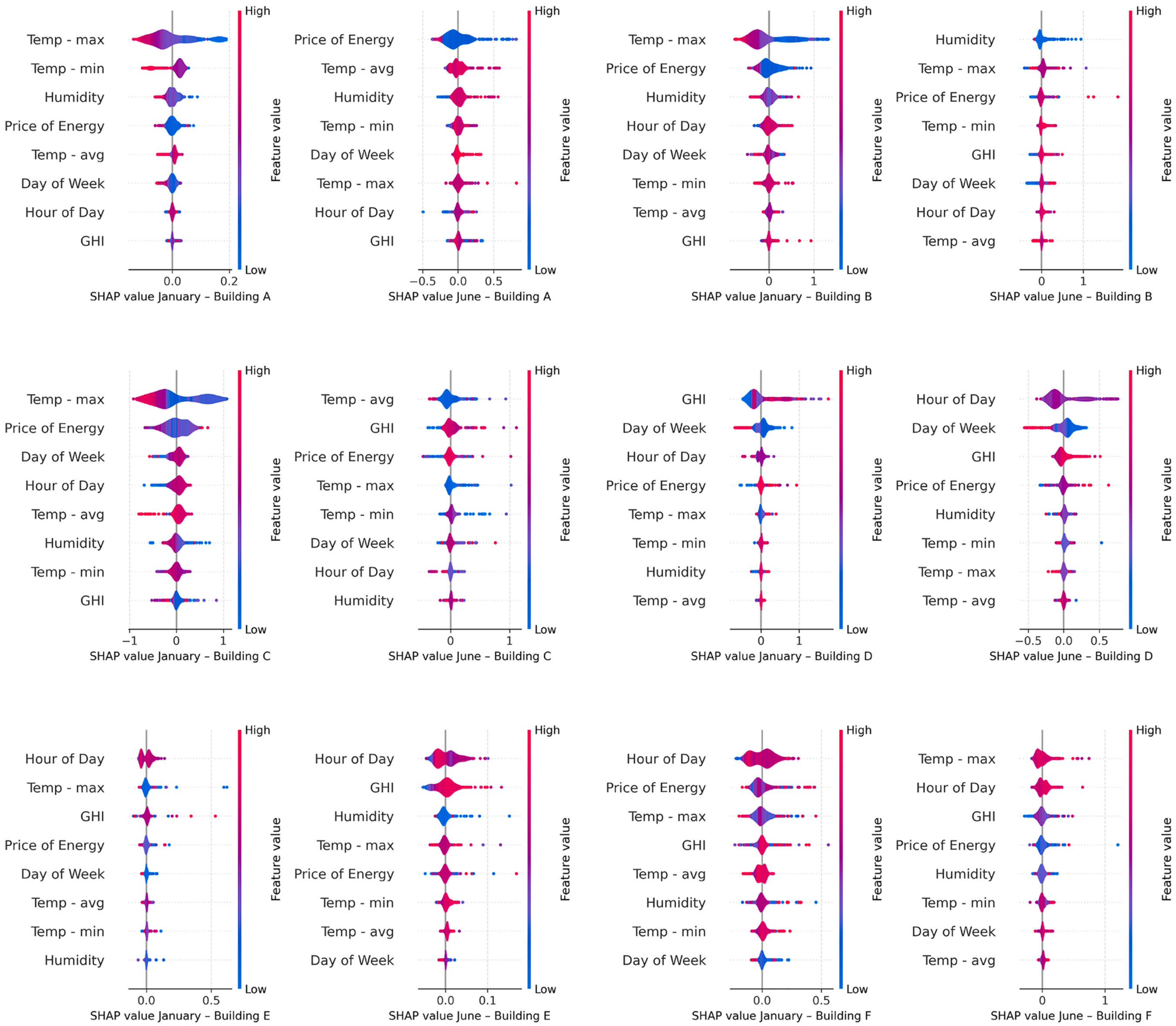

Fig. 4. SHAP summary plots for 6 randomly selected buildings.

4.2. Machine learning model selection

In this section, predictions are carried out using the feature importance analysis results and the machine learning methods introduced in the previous section. To evaluate the impact of incorporating different features based on their SHAP ranking, a step-by-step approach is applied. To simplify the process and reduce computational complexity, given the large number of buildings involved, the SHAP importance analysis conducted for January and June is used throughout the forecasting procedure to guide the incremental addition of features. January continues to serve as the representative month for the colder period (October to March), whereas June serves as the representative for the warmer period (April to September).

Initially, for all four models, only the most important feature (as identified by SHAP) is used as the input, and predictions are generated accordingly. Then, the second most important feature is added to the input set, and the prediction process is repeated. This iterative process continues until all eight features are included. All four machine learning models introduced earlier are applied in this experiment. The aim is to evaluate model performance under different feature combinations and identify the optimal set of features and the most suitable prediction model for each building. A sliding window approach is employed for the forecasting process. Initially, the first week of 2023 is used as the training dataset, and the following day is considered the test instance. Then, the entire window is shifted forward by one day, and the prediction is repeated. This procedure will continue until the entire year of 2023 is covered. Essentially, the model performs hourly forecasts for the next day, using a continuously updating training window. To present the results, Table 3 summarizes the best-performing machine learning model, based on the lowest NRMSE, across the entire test period for each of the 200 buildings.

Table 3
Number of buildings with lowest NRMSE per model.

Model	No. of Buildings (October to March)	No. of Buildings (April to September)
Stacking	163	178
Voting	2	1
RF	35	21
XGB	0	0

Table 4
Average NRMSE (%), and MAE (kWh), together with the best performance percentage of each method (October–March, using the stacking model).

Method	Avg. NRMSE	Avg. MAE (kWh)	Best Perf. (%)
All Features	11.85	0.49	0.00
Static Feature Selection	10.81	0.46	4.00
Dynamic Feature Selection	**9.31**	**0.40**	**96.00**

Table 5
Average NRMSE (%), and MAE (kWh), together with the best performance percentage of each method (April–September, using the stacking model).

Method	Avg. NRMSE	Avg. MAE (kWh)	Best Perf. (%)
All Features	11.67	0.38	0.00
Static Feature Selection	11.11	0.35	1.50
Dynamic Feature Selection	**9.14**	**0.32**	**98.50**

As shown in Table 3, the stacking method achieved the best performance across 163 buildings during the six coldest months of the year (October to March). During the warmer half of the year (April to September), stacking continued to outperform all other models, delivering the lowest NRMSE in 178 buildings. These results demonstrate the model's robustness and its consistent predictive strength across varying seasonal conditions. It is important to note that machine learning models can behave differently when faced with different datasets, forecasting horizons, or data arrangements. Therefore, it cannot be claimed that a single model will always deliver the best performance across all conditions and case studies. The goal of these analyses is to identify the method that performs best in most cases. However, when dealing with a specific case study, it is essential to tailor the approach and determine the most suitable method for that particular situation [43].

The findings highlight the superior effectiveness of ensemble-based approaches, particularly stacking. By integrating the outputs of multiple base learners, stacking leverages their complementary strengths, leading to enhanced predictive accuracy and greater model generalization. This advantage is especially critical in complex, heterogeneous environments such as residential load forecasting, where consumption patterns vary significantly across buildings and seasons. Given its dominant performance across both cold and warm periods, the stacking method has been selected as the primary modeling approach for all subsequent analyses presented in this study.

4.3. Results of dynamic and static methods

At this stage, the static and dynamic feature selection methods introduced in the methodology section are implemented. Forecasting results are presented for both approaches, along with a baseline scenario in which no feature selection is applied. This evaluation was conducted across all 200 buildings considered in the study. Tables 4 and 5, in conjunction with the comparative box plots in Fig. 5, offer a comprehensive and consistent validation of the effectiveness of the Dynamic Feature Selection method across seasonal conditions. Specifically, the average NRMSE values (expressed as percentages) are reported for each method. In addition, the tables include the percentage of buildings in which each approach achieved the best forecasting performance. These metrics provide a comprehensive comparison of the methods' effectiveness across different seasonal conditions and building profiles.

In Table 4, which refers to the performance during the colder months, the dynamic method demonstrates superiority across both evaluation metrics. It achieves the lowest average NRMSE (9.31 %) and MAE (0.40 kWh), indicating that it not only reduces the scale-normalized error but also provides more accurate predictions in the original units. This method outperforms all others in 96 % of the buildings. By contrast, the Static Selection approach yields a higher average NRMSE (10.81 %) and MAE (0.46 kWh) and is the best performer in just 4 % of the cases. The baseline model, which uses all features without any selection, performs worst, with the highest average errors (NRMSE 11.85 %, MAE 0.49 kWh).

Table 6
Selected features for each building based on January and February.

Building ID	Selected features (Jan)	Selected features (Feb)
A	temp_max, temp_min	temp_max, day_of_week, humidity, price
B	all features except GHI	all features except hour
C	all features except GHI and temp_min	temp_max, price, day_of_week
D	GHI, day_of_week	GHI
E	hour	hour
F	hour	hour

Table 5, which reports results for the warmer months, follows the same pattern. The dynamic method improves further, reducing the average NRMSE to 9.14 % and MAE to 0.32 kWh. It dominates the performance in 98.5 % of the buildings, confirming its robustness in different seasonal conditions. The static method yields slightly better results than the baseline, with an average NRMSE of 11.11 % and MAE of 0.35 kWh, and is the best method in only 1.5 % of the cases. The all-features baseline again records the poorest performance, with the highest NRMSE (11.67 %) and MAE (0.38 kWh), consistently failing to capture the variability in residential loads.

These quantitative insights from the tables are visually reinforced by the box plots in Fig. 5, which provide a clearer picture of the distribution and spread of errors. In both cold and warm seasons, the box plots for the dynamic method are not only centered at lower NRMSE values, but they also exhibit a tighter interquartile range (IQR), indicating more consistent and reliable performance across buildings. The dynamic approach also yields the lowest observed minimum error: 3.80 % in winter and 2.99 % in summer–values unmatched by the other strategies. The All Features method, in both periods, yields the highest mean errors and greater variability. This result confirms that including irrelevant or less informative features can introduce noise into the model, thereby degrading predictive accuracy. Static Selection, while an improvement over using all features indiscriminately, remains less effective than dynamically adjusting the input feature set. These findings collectively confirm that the dynamic feature selection strategy both improves forecasting accuracy on average and enhances reliability and robustness across both seasonal scenarios. By selecting features that are most relevant to each building and time period, the dynamic approach achieves substantial performance improvements, as clearly shown by the numerical results in Tables 4 and 5 and the visual distributions in Fig. 5. These results underscore the critical importance of context-aware, adaptive feature selection techniques in building-level electric energy forecasting, highlighting the significant advantages of the dynamic approach in terms of both stability and accuracy, regardless of seasonal variations.

To provide a more seasonally robust quantitative assessment, the representative month analysis is extended by including February alongside January for the colder period (October–March) and July alongside June for the warmer period (April–September). Tables 7 and 9 present the forecasting results for the same six buildings analyzed in the SHAP section. For each building, the tables indicate the error when all available features are used, the errors obtained with static SHAP-based subsets derived from the representative months (January/February or June/July), and the error achieved by the proposed dynamic method. NRMSE (%) and MAE (kWh) are reported together to provide a comprehensive view of predictive accuracy, reflecting both relative and absolute error magnitudes. Also, the exact static feature sets selected for each building are listed in Tables 6 and 8.

As shown in Tables 6–9, the results confirm that the dynamic feature selection method consistently provides the most accurate

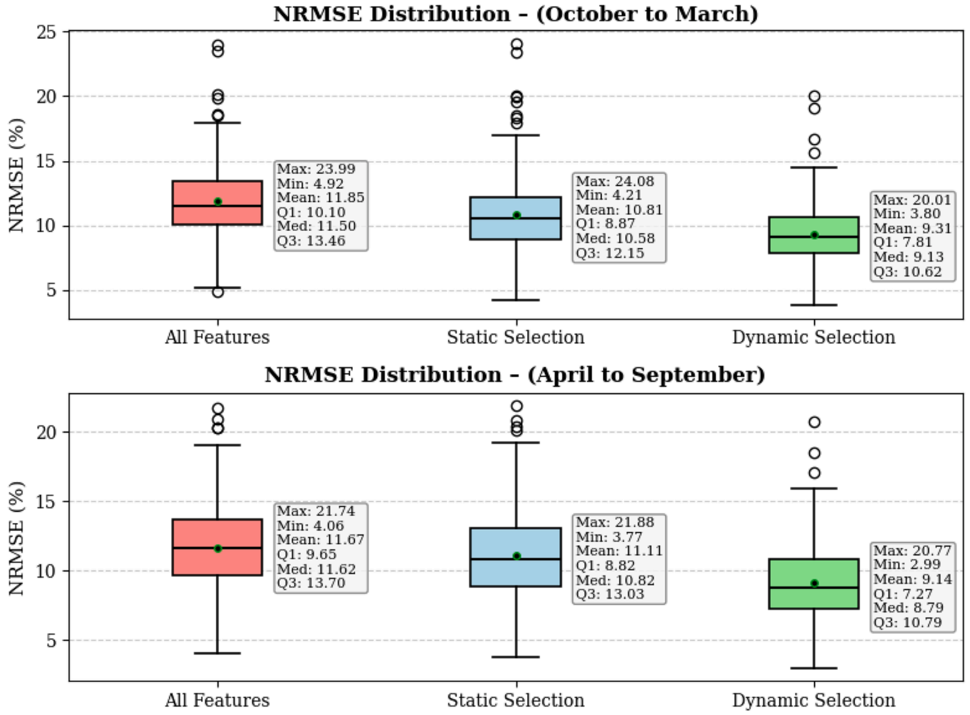

Fig. 5. Comparison of NRMSE distributions across feature selection strategies in cold and warm months.

Table 7
Comparison of NRMSE (%) and MAE (kWh) between different methods (October–March, using the stacking model).

Building (ID)	Evaluation Metric	All Features	Static January-based	Static February-based	Dynamic Feature selection
A	NRMSE(%)	6.79	7.36	6.14	**4.39**
	MAE (kWh)	0.08	0.08	0.08	**0.08**
B	NRMSE (%)	12.83	12.79	9.22	**5.86**
	MAE (kWh)	0.45	0.46	0.41	**0.31**
C	NRMSE (%)	16.77	16.93	14.76	**12.92**
	MAE (kWh)	0.52	0.53	0.45	**0.38**
D	NRMSE (%)	9.02	9.54	9.93	**6.90**
	MAE (kWh)	0.25	0.25	0.32	**0.20**
E	NRMSE (%)	6.09	4.92	4.92	**4.61**
	MAE (kWh)	0.05	0.04	0.04	**0.04**
F	NRMSE (%)	6.00	5.09	5.09	**4.85**
	MAE (kWh)	0.25	0.21	0.21	**0.20**

Table 8
Selected features for each building based on June and July.

Building ID	Selected features (Jun)	Selected features (Jul)
A	price	hour, price
B	humidity	hour, day_of_week
C	all features	all features except day_of_week
D	hour	hour
E	hour	hour
F	temp_max, hour, GHI, price, humidity	all features

forecasts across all buildings and seasonal conditions. Even after extending the static analysis to include February and July as additional representative months, the overall trend remains unchanged. Also, as presented in Tables 7 and 9, while incorporating these months modifies the selected feature sets and, in some cases, slightly improves the static model's performance–especially during the warmer period–the dynamic approach continues to outperform both static and full-feature configurations. The selected static feature sets for each case are detailed in Tables 6 and 8. In the warm season, the July-based static selection produces results that are relatively close to those of the dynamic model, suggesting that occupant behavior during the summer months aligns more closely with the feature patterns identified for July. Nonetheless, the dynamic method maintains a clear advantage by automatically identifying and updating the most relevant features for each time window. This adaptability allows it to adjust to evolving consumption behaviors and prevents the degradation of forecast accuracy that may occur when static features are derived from a time frame not fully representative of current conditions. It should be emphasized that the observed improvements are not merely the result of reducing the number of features. In the proposed framework, features are incrementally added in SHAP order, leading to up to eight combinations per building. The method does not inherently favor smaller subsets but instead identifies the most suitable feature configuration for each building and time period. Overall, while the static method offers noticeable benefits compared to no feature selection, the dynamic approach provides a more reliable and adaptive solution for forecasting in settings where electricity consumption is subject to change. To illustrate the alignment between the forecasted and actual load values, a sample of the hourly prediction results over

Table 9
Comparison of NRMSE (%) and MAE (kWh) between different methods (April–September, using the stacking model).

Building (ID)	Evaluation Metric	All Features	Static June-based	Static July-based	Dynamic Feature selection
A	NRMSE(%)	10.76	12.14	8.93	**8.19**
	MAE (kWh)	0.19	0.21	0.17	**0.15**
B	NRMSE(%)	19.04	19.11	16.92	**14.69**
	MAE (kWh)	0.88	0.95	0.88	**0.78**
C	NRMSE(%)	9.24	9.24	7.07	**6.80**
	MAE (kWh)	0.28	0.28	0.24	**0.22**
D	NRMSE(%)	10.26	9.00	9.00	**8.37**
	MAE (kWh)	0.26	0.25	0.25	**0.23**
E	NRMSE(%)	7.25	5.67	5.67	**5.66**
	MAE (kWh)	0.05	0.04	0.04	**0.04**
F	NRMSE(%)	8.09	8.58	8.09	**6.56**
	MAE (kWh)	0.20	0.20	0.20	**0.17**

several days is presented in Fig. 6. This visual comparison provides a clearer understanding of the model's performance and its ability to capture daily consumption patterns.

As illustrated in Fig. 6, the predictions generated by models using a fixed number of features often deviate from the actual consumption pattern during various hours of the day. In contrast, the output of the dynamic method moves in a consistent, aligned, and synchronized manner with the actual values. For instance, in Building D, the load profile is characterized by a well-defined daily rhythm with pronounced peaks during midday and evening hours. The dynamic model demonstrates its strength in tracking these patterns with both temporal precision and amplitude accuracy. Unlike static models, which either smooth out these peaks or misalign their timing, the dynamic method responds adaptively, capturing sharp transitions without overfitting. This result indicates a notable advantage in modeling structured yet fluctuating demand cycles.

Building C displays a subtler electric energy usage trend, marked by gradual variations over the course of the day. Here, the dynamic method outperforms by producing a prediction line that flows smoothly along the actual trajectory, avoiding the abrupt jumps frequently observed in static models. Such behavior suggests that the dynamic model is capable of selectively leveraging relevant features without overreacting to localized noise, a limitation often encountered when using fixed or redundant feature sets.

In Building B, the consumption behavior is irregular and less predictable, likely driven by variable occupant patterns or operational anomalies. Under these challenging conditions, the dynamic model still delivers a notably closer approximation to the actual data. It successfully tracks short-term fluctuations while maintaining overall trend fidelity–something that static models struggle with, either reacting too sluggishly or producing erratic outputs due to their lack of contextual flexibility.

In Building F, the forecasting errors are relatively large across all methods for the illustrated day, reflecting the inherent difficulty of predicting household electricity loads where occupant behavior may be irregular and non-repetitive. Sudden deviations caused by unpredictable activities lead to discrepancies that no model can fully capture. Nevertheless, the dynamic method is observed to follow the actual load profile more closely than both the static and full-feature approaches, particularly around the midday and evening peaks, highlighting its stronger adaptability under volatile consumption conditions.

Building A introduces a different type of challenge due to its relatively low consumption levels. In such settings, even small absolute deviations can result in disproportionately large relative errors. Nonetheless, the dynamic approach maintains close alignment with the ground truth, demonstrating that its feature selection mechanism remains robust even in low-signal environments, where overfitting or underfitting can easily occur.

Finally, in Building E, the load pattern is relatively stable across consecutive days, without pronounced fluctuations. As a result, the set of influential features identified dynamically is very similar to those selected by the static approach, which explains why the forecasting accuracy of the two methods is nearly identical (Table 7). In such cases, dynamic selection does not offer a strong advantage, although it still avoids the overgeneralization observed in the full-feature model. This behavior illustrates that the benefit of dynamic selection is less pronounced in buildings with stable and repetitive load patterns, which is expected since in these cases the dynamic method almost always selects the same features as the static method.

Collectively, these observations highlight the superior adaptability of the dynamic method. It consistently exhibits enhanced predictive accuracy and behavioral alignment with actual usage patterns across diverse building profiles, whether the demand is structured, erratic, gradual, or minimal. Such generalization capabilities make it a compelling approach for real-world deployment in electric energy forecasting systems. Another important observation is that the dynamic method performs well both in terms of value approximation and in accurately capturing the timing of consumption changes. The model does not merely predict numbers that are "close enough," but rather reconstructs the temporal sequence and rhythm of electric energy usage–an essential quality for time-sensitive applications such as load management, price response, or real-time control. It should be noted that although the main analysis focused on 24-hour ahead forecasting, the proposed framework is not restricted to this horizon. Since feature selection is performed dynamically within each sliding training window, the method can also be applied to multi-day horizons (e.g., 2–7 days ahead). In such cases, forecast accuracy may decrease as the horizon lengthens, but this can be partly mitigated by extending the length of the training window to capture longer-term dependencies.

Beyond accuracy evaluation, it is important to consider how the proposed framework could be integrated into operational forecasting systems. In practice, the dynamic feature selection step can be implemented as a pre-processing module that operates on each training window before the forecasting stage. This allows the method to be embedded into existing pipelines without requiring modifications to downstream prediction models, making it compatible with commonly used ensemble or deep learning approaches. The computational cost is modest: on a standard laptop (Intel Core i7, 16 GB RAM, Windows 11), generating one-day-ahead forecasts required, on average, about 11 seconds per building. Such efficiency suggests that the framework can be scaled to larger datasets and even deployed in near-real-time environments. Furthermore, its model-agnostic nature enables integration with hierarchical forecasting setups or aggregation schemes, offering flexibility for distribution system operators and energy management platforms. Having access to more accurate and up-to-date building-level load forecasts can directly support the design of advanced energy management systems in buildings where such implementations are feasible, leading to improved energy efficiency, increased economic profitability, and enhanced demand-side flexibility.

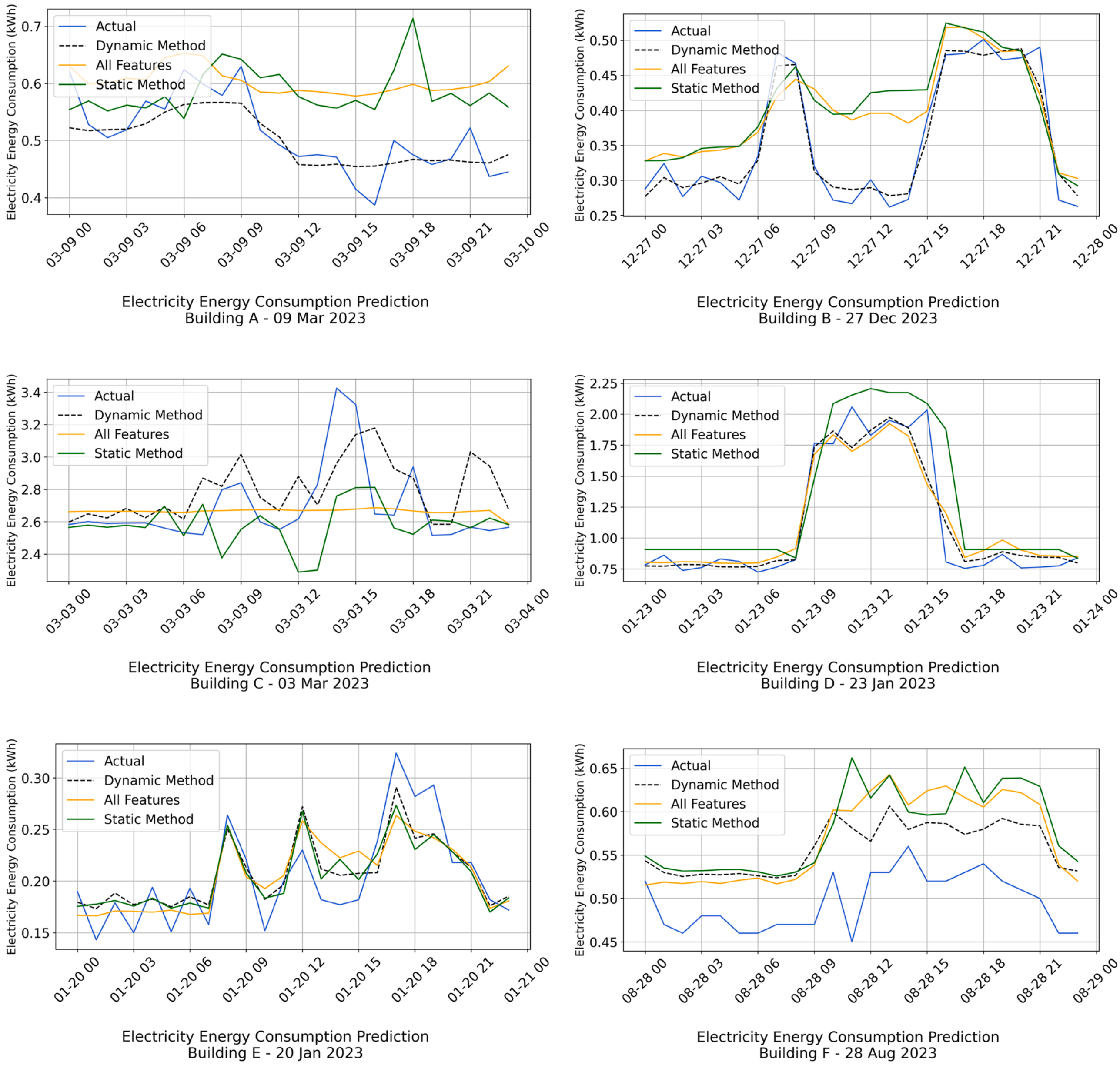

Fig. 6. Comparison of actual and predicted hourly electricity consumption for 6 sample buildings.

4.4. *Comparison of alternative meta-learners in the stacking framework*

As noted in Section 3.2, Ridge Regression was adopted as the meta-learner in this study. To examine the robustness of this choice, SVR and LightGBM were also evaluated on six representative buildings. Tables 10 and 11 present the results for both cold and warm seasons.

The outcomes indicate that the three meta-learners achieved almost identical forecasting accuracy across NRMSE and MAE. In several cases, Ridge delivered the lowest errors, for example, in Buildings B and E during the warm season and Building A in the cold season, while in other cases, SVR or LightGBM performed marginally better, such as Building D, where SVR slightly outperformed Ridge in both periods. LightGBM occasionally showed competitive accuracy (e.g., Building C in the cold season), but in some cases its performance dropped noticeably, as in Building D during the warm season, where the error increased compared to both Ridge and SVR. These variations confirm that no consistent superiority can be attributed to any single method. Overall, the differences remain small in absolute terms (typically within one or two tenths in NRMSE or MAE), and the ranking of the three methods fluctuates across buildings and seasons without a clear pattern. This suggests that the choice of meta-learner does not materially affect the final forecasting accuracy in this setting.

Table 10
Comparison of Ridge, SVR, and LightGBM as meta-learners on six representative buildings in cold seasons. (MAE reported in kWh).

Building (ID)	Ridge		SVR		LightGBM	
	NRMSE(%)	MAE	NRMSE(%)	MAE	NRMSE(%)	MAE
A	4.39	0.08	4.50	0.08	4.61	0.09
B	5.86	0.31	5.59	0.29	5.92	0.35
C	12.92	0.38	13.17	0.37	12.82	0.39
D	6.90	0.20	6.63	0.18	8.46	0.29
E	4.61	0.04	4.83	0.05	4.66	0.04
F	4.85	0.20	4.82	0.18	4.82	0.20

Given that each sliding window in this study contained only a few hundred samples and fewer than ten features, computational efficiency

Table 11
Comparison of Ridge, SVR, and LightGBM as meta-learners on six representative buildings in warm seasons. (MAE reported in kWh).

Building (ID)	Ridge		SVR		LightGBM	
	NRMSE(%)	MAE	NRMSE(%)	MAE	NRMSE(%)	MAE
A	8.19	0.15	8.15	0.14	8.37	0.15
B	14.69	0.78	15.56	0.64	15.16	0.86
C	6.80	0.22	6.70	0.20	6.71	0.22
D	8.37	0.23	8.19	0.21	9.54	0.29
E	5.66	0.04	5.97	0.05	5.72	0.04
F	6.56	0.17	6.50	0.15	6.66	0.17

becomes a key factor. Ridge involves only solving a linear system, while SVR requires quadratic optimization, and LightGBM relies on constructing decision trees. Considering the nearly identical accuracy observed and the substantially lower computational cost of Ridge, and consistent with prior recommendations in the literature [5,40], Ridge Regression was retained as the default meta-learner for its stability, simplicity, and efficiency.

5. Conclusion

This study focused on short-term forecasting of electricity consumption in residential buildings. The main innovation lies in combining a dynamic feature selection strategy with a stacking-based modeling approach, specifically tailored to capture the complexity of residential demand patterns. Leveraging a dataset from 200 residential buildings in southern Finland, the proposed approach integrates SHAP-based feature importance analysis with ensemble machine learning models, including XGBoost, Random Forest, and advanced ensemble strategies such as voting and stacking. Among these, the stacking model–employing Ridge regression as a meta-learner–consistently achieved the best performance across most evaluation scenarios. While the case study was conducted in Finland, where strong seasonal contrasts highlight the value of dynamic feature selection, the proposed framework is not limited to this context. The methodology can, in principle, be applied to other climates and grid structures, although the relative benefits may vary depending on local variability in load patterns and environmental conditions.

A comparison between static and dynamic feature selection strategies revealed that dynamically updating input features for each forecasting window leads to significantly better accuracy, particularly in environments characterized by seasonal and behavioral variability. For instance, during the six colder months of the year, the dynamic approach achieved an average NRMSE of 9.31 %, outperforming all other models in 96 % of the buildings. Similarly, in the warmer months, the proposed method maintained its effectiveness with an average NRMSE of 9.14 %, ranking as the best-performing model in 98.5 % of the buildings. Consistent improvements were also observed in terms of MAE, confirming that the method not only reduces absolute prediction errors but also enhances the overall accuracy of consumption forecasts. The results further highlighted that using fewer, highly relevant features can often outperform models trained on the full feature set. These findings underscore the importance of both interpretability and adaptability in predictive modeling for electric energy systems. Of course, it should be noted that the effectiveness of the dynamic method is more evident in buildings whose energy consumption patterns change over weeks compared to the static method. Moreover, the proposed framework is computationally efficient and can be seamlessly integrated into operational forecasting pipelines, providing grid operators and building managers with more reliable, context-aware forecasts to support energy management, improve efficiency, and enable more informed decision-making.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author(s) used ChatGPT by OpenAI to refine the language and improve clarity in parts of the manuscript. After using this tool, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

CRediT authorship contribution statement

Nabi Taheri: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Formal analysis, Data curation, Conceptualization; **Lauri Karttunen:** Writing – review & editing, Methodology, Conceptualization; **Sami Jouttijärvi:** Writing – review & editing, Supervision, Methodology; **Antonio Piazzi:** Writing – review & editing, Supervision, Methodology; **Mauro Tucci:** Writing – review & editing, Supervision, Methodology; **Kati Miettunen:** Writing – review & editing, Supervision, Funding acquisition.

Data availability

The meteorological data and electricity price data used in this study are publicly available and can be found in references [34,35], respectively. The electricity consumption data were provided by VSV Group and are not publicly available due to data privacy agreements.

Declaration of competing 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.

Acknowledgment

Nabi Taheri gratefully acknowledges the support from the University of Pisa, the University of Turku, and i-EM Company during the preparation of this work. This project was funded by the Strategic Research Council (SRC) established within the Research Council of Finland (RealSolar, Decision No. 358542). The work was supported by the Strategic Research Council within the Research Council of Finland, Decision No. 358542 (KM, SJ). Lauri Karttunen thanks the City of Salo and the University of Turku (project HEMS), and UTUGS.

References

[1] R.A. Abbasi, et al., Short-term load forecasting using XGBoost, in: Proceedings of the Workshops of the 33rd International Conference on Advanced Information Networking and Applications, the Workshops of the 33rd International Conference on Advanced Information Networking and Applications, Springer International Publishing, 2019. WAINA-2019.
[2] D. Chirag, et al, A review on time series forecasting techniques for building energy consumption, Renew. Sustain. Energy Rev. 74 (2017) 902–924.
[3] P.-P. Phyo, Y.-C. Byun, N. Park, Short-term energy forecasting using machine-learning-based ensemble voting regression, Symmetry 14 (1) (2022) 160.
[4] S. Luo, et al, Stacking integration algorithm based on CNN-BiLSTM-Attention with XGBoost for short-term electricity load forecasting, Energy Reports 12 (2024) 2676–2689.
[5] H. Huang, et al., Stacking and ridge regression-based spectral ensemble preprocessing method and its application in near-infrared spectral analysis, Talanta 276 (2024) 126242.
[6] T.S. Mahmoud, et al., Modelling self-optimised short-term load forecasting for medium voltage loads using tunning fuzzy systems and artificial neural networks, Energy Convers. Manage. 106 (2015) 1396–1408.
[7] A.V. Kychkin, G. C. Georgios, Feature and model selection for day-ahead electricity-load forecasting in residential buildings, Energy Build. 249 (2021) 111200.
[8] W. Cai, et al., Predicting the energy consumption in buildings using the optimized support vector regression model, Energy 273 (2023) 127188.
[9] A. Zeng, H. Ho, Y. Yu, Prediction of building electricity usage using gaussian process regression, J. Build. Eng. 28 (2020) 101054.
[10] K. Dab, et al., A compositional kernel based gaussian process approach to day-ahead residential load forecasting, Energy Build. 254 (2022) 111459.
[11] P. Koukaras, et al, Optimizing building short-term load forecasting: a comparative analysis of machine learning models, Energies 17 (6) (2024) 1450.
[12] K. Venkat, et al., An xgboost ensemble model for residential load forecasting, in: Proceedings of International Conference on Intelligent Computing, Information and Control Systems (ICICCS 2020), International Conference on Intelligent Computing, Information and Control Systems (ICICCS 2020) Singapore, Springer, 2021.

[13] S. Wu, et al., Research On The Daily Electricity Forecast Model Based On LightGBM, 2477, IOP Publishing, 2023.

[14] Y. He, et al, Residential net load interval prediction based on stacking ensemble learning, Energy 296 (2024) 131134.

[15] G. Mitrentsis, H. Lens, An interpretable probabilistic model for short-term solar power forecasting using natural gradient boosting, Appl. Energy 309 (2022) 118473.

[16] Ibrahim, I. Anwar, M.J. Hossain, Short-term multivariate time series load data forecasting at low-voltage level using optimised deep-ensemble learning-based models, Energy Convers. Manage. 296 (2023) 117663.

[17] Y. Yang, et al., A dynamic ensemble method for residential short-term load forecasting, Alexandria Eng. J. 63 (2023) 75–88.

[18] A. Yousaf, et al., An improved residential electricity load forecasting using a machine-learning-based feature selection approach and a proposed integration strategy, Sustainability, 13 (11), 2021 6199.

[19] A.M. Pirbazari, A. Chakravorty, C. Rong, Evaluating feature selection methods for short-term load forecasting, in: 2019 IEEE International Conference on Big Data and Smart Computing (BigComp), IEEE, 2019.

[20] I. Koprinska, M. Rana, V. G. Vassilios, Correlation and instance based feature selection for electricity load forecasting, Knowl. Based Syst. 82 (2015) 29–40.

[21] N. Huang, G. Lu, D. Xu, A permutation importance-based feature selection method for short-term electricity load forecasting using random forest, Energies 9 (10) (2016) 767.

[22] S. S. Subbiah, J. Chinnappan, Deep learning based short-term load forecasting with hybrid feature selection, Electr. Power Syst. Res. 210 (2022) 108065.

[23] X. Gao, et al., Short-term electricity load forecasting model based on EMD-GRU with feature selection, Energies, 12 (6), 2019 1140.

[24] X. Yang, et al., An optimized method for short-term load forecasting based on feature fusion and ConvLSTM-3D neural network, Front. Energy Res. 12 (2025) 1501963.

[25] J. Cui, et al., Advanced Short-Term Load Forecasting with XGBoost-RF Feature Selection and CNN-GRU, Processes, 12 (11), 2024 2466.

[26] X. Liu, A novel correlation feature self-assigned kolmogorov-Arnold networks for multi-energy load forecasting in integrated energy systems, Energy Convers. Manage. 325 (2025) 119388.

[27] A. Forootani, M. Rastegar, A. Sami, Short-term individual residential load forecasting using an enhanced machine learning-based approach based on a feature engineering framework: a comparative study with deep learning methods, Electr. Power Syst. Res. 210 (2022) 108119.

[28] W. Kim, Electricity load forecasting using advanced feature selection and optimal deep learning model for the variable refrigerant flow systems, Energy Reports 6 (2020) 2604–2618.

[29] S. Jouttijärvi, Techno-economic analysis on optimizing the value of photovoltaic electricity in a high-latitude location, Appl. Energy 361 (2024) 122924.

[30] J. Sami, Benefits of bifacial solar cells combined with low voltage power grids at high latitudes, Renew. Sustain. Energy Rev. 161 (2022) 112354.

[31] H. Gholami, Technical potential of solar energy in buildings across Norway: capacity and demand, Sol. Energy 278 (2024) 112758.

[32] S. Jouttijärvi, A comprehensive methodological workflow to maximize solar energy in low-voltage grids: a case study of vertical bifacial panels in Nordic conditions, Sol. Energy 262 (2023) 111819.

[33] S. Vinay, Determining the optimal azimuth for solar-ready buildings: Simulating for Maximising the Economic Value of Solar PV Installations in Lapland, Finland Renewable Energy, 2025.

[34] Weather Data Portal, Technical Report, Finnish Meteorological Institute. https://www.ilmatieteenlaitos.fi/.

[35] Finland'S dynamic pricing system, in: Price Data Portal. https://data.nordpoolgroup.com/auction/day-ahead/prices.

[36] S. M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, Adv. Neural Inf. Process. Syst. 30 (2017).

[37] R.A. Davis, M.S. Nielsen, Modeling of Time Series using Random Forests: Theoretical Developments, 2020 3644–3671.

[38] A.J. Smola, B. Schölkopf, A tutorial on support vector regression, Stat. Comput. 14 (2004) 199–222.

[39] G. Ke, Lightgbm: A Highly Efficient Gradient Boosting Decision Tree, Advances in Neural Information Processing Systems, 2017.

[40] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical clearning, 2009.

[41] N. Taheri, M. Tucci, Enhancing regional wind power forecasting through advanced machine-learning and feature-selection techniques, Energies 17 (21) (2024).

[42] K. Amasyali, N.M. El-Gohary, A review of data-driven building energy consumption prediction studies, Renew. Sustain. Energy Rev. 81 (2018) 1192–1205.

[43] N. Taheri, A. Piazzi, M. Tucci, Analyzing the impact of training data availability on machine learning models accuracy for regional photovoltaic production forecast, in: 2024 IEEE 8th Forum on Research and Technologies for Society and Industry Innovation (RTSI), IEEE, 2024.