Explainable Machine Learning for Household Electricity Demand Forecasting Using Hybrid Residual Modeling

Renzo Diony Ramirez-Huaracha [1,*], Melany Meylin Quispe-Gutierrez [1], and Benjamín Maraza-Quispe [2]

[1] Professional School of Electrical Engineering, Faculty of Production and Services Engineering, Universidad Nacional de San Agustín de Arequipa, Perú

[2] Academic Department of Education, Faculty of Education Sciences, Universidad Nacional de San Agustín de Arequipa, Perú

Email: rramirezhua@unsa.edu.pe (R.D.R.-H.); mquispeguti@unsa.edu.pe (M.M.Q.-G.); bmaraza@unsa.edu.pe (B.M.-Q.)

*Corresponding author

***Abstract*—This study proposes an interpretable hybrid residual learning framework for household electricity demand prediction under limited data conditions. The framework integrates statistical forecasting models with machine learning residual learners and SHapley Additive exPlanations (SHAP)-based explainability to improve predictive accuracy while preserving transparency. Using 1,000 hourly household electricity consumption observations collected over approximately 41.7 consecutive days, the proposed model was compared with Linear Regression, Autoregressive Integrated Moving Average (ARIMA), Random Forest, Support Vector Machine, and Deep Neural Network baselines under identical chronological train-validation-test conditions. Results show that the hybrid framework achieved the best predictive performance, with Mean Absolute Error (MAE) = 0.141, Root Mean Squared Error (RMSE) = 0.198, Mean Absolute Percentage Error (MAPE) = 7.5%, and Coefficient of Determination (R^2) = 0.90, outperforming both statistical and standalone machine learning models. The results are interpreted within a short-term limited-data forecasting scenario. Although the proposed framework showed the best performance on the independent chronological test set, robustness under progressive training-data reduction is considered a direction for future validation because the available dataset contains only 1,000 hourly observations, while statistical comparison was performed using the Diebold-Mariano test. SHAP explanations identified temperature, hour-of-day patterns, and household occupancy as the most influential demand drivers, providing physically meaningful insights for residential energy planning. The findings suggest that interpretable hybrid residual models represent a reliable alternative for short-term household electricity forecasting in data-scarce urban environments.**

***Keywords*—household electricity demand forecasting, hybrid residual learning, explainable artificial intelligence, SHapley Additive exPlanations (SHAP), limited data, residential energy consumption, power system planning**

I. Introduction

Electricity consumption in the residential sector has become a critical component of modern power systems due to urban growth, population increase, and the expansion of electrical and electronic device usage [1]. Accurate forecasting of residential electricity demand is essential for energy planning, power system stability, demand-side management, and sustainability-oriented policies [2]. However, in urban contexts, particularly in developing countries, forecasting is often constrained by data scarcity, heterogeneous consumption patterns, and limited advanced metering infrastructure [3].

Traditional statistical and time-series models, such as Autoregressive Integrated Moving Average (ARIMA), remain widely used because of their simplicity and interpretability [4]. Nevertheless, they have limited capacity to capture nonlinear relationships among climatic, behavioral, and household-related variables [5]. Machine Learning (ML) models can address some of these nonlinear patterns and often improve predictive accuracy [6], but their black-box nature limits their usefulness in decision-making contexts where transparency is required [7].

Hybrid residual models offer a promising alternative by combining the temporal modeling capacity of statistical methods with the nonlinear learning ability of machine learning algorithms [8, 9]. However, their application together with explainable artificial intelligence techniques remains limited in household electricity demand forecasting, especially under data-scarce urban conditions [10–12].

Within this context, the present study proposes an interpretable hybrid residual forecasting framework for short-term household electricity demand prediction under limited-data conditions. The framework integrates

Manuscript received May 8, 2026; revised May 28, 2026; accepted June 3, 2026; published August 12, 2026.

doi: 10.12720/jait.17.8.1429-1441

ARIMA-based statistical forecasting, Random Forest residual correction, and SHapley Additive exPlanations (SHAP)-based explainability to improve predictive transparency and support residential energy planning.

II. Literature Review

Previous research on residential electricity demand forecasting has followed three main methodological directions. The first includes statistical and time-series approaches, such as linear regression and ARIMA, which remain relevant because of their interpretability and low computational complexity. However, these models are often limited when electricity demand is influenced by nonlinear interactions among temperature, occupancy, device usage, and household routines [4, 5].

The second direction involves machine learning models, including support vector machines, random forests, ensemble methods, and deep learning models. These approaches have demonstrated strong predictive performance in energy forecasting tasks because they can capture nonlinear demand patterns and complex interactions among explanatory variables [6, 7]. Nevertheless, their adoption in operational energy planning may be limited when model decisions cannot be clearly interpreted.

The third direction corresponds to hybrid forecasting models, particularly residual learning strategies that combine statistical base models with machine learning residual learners. These models are designed to capture both structured temporal behavior and nonlinear deviations in electricity demand [8, 9]. More recently, explainable artificial intelligence techniques such as SHapley Additive exPlanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME) have been incorporated to improve transparency by identifying the contribution of each predictor to model outputs [10–12]. However, the joint use of hybrid residual learning and SHAP-based explainability remains insufficiently explored in short-term household electricity demand forecasting under limited-data conditions.

A relevant antecedent is the study conducted by Machaca-Casani *et al.* [13], who compared machine learning models and traditional statistical methods for residential energy consumption forecasting in Arequipa, Peru. Using a quantitative and comparative approach, they evaluated algorithms such as Random Forest, Support Vector Machine (SVM), and deep neural networks against classical models such as linear regression and ARIMA, employing Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 metrics. The results showed that traditional models and SVM exhibited better predictive capability and greater stability, while more complex models suffered from generalization issues, demonstrating that algorithmic complexity does not guarantee superior performance in data-limited contexts.

This problem is particularly relevant in contexts where smart-meter infrastructure is partial or recently implemented, because forecasting models must often operate with short historical records and a reduced set of explanatory variables.

The study concluded that model selection should consider context, data quality, and interpretability, leaving the exploration of more robust and adaptive approaches as future work.

Similarly, Neubauer *et al.* [14] developed an explainable multi-step thermal load forecasting model using Encoder–Decoder networks with temporal attention mechanisms and SHAP values. The authors demonstrated that incorporating Explainable Artificial Intelligence (XAI) techniques improves the interpretability of deep learning models without sacrificing accuracy, enabling the identification of the influence of climatic variables and temporal lags on energy demand. Moreover, they showed that SHAP-based feature selection reduces training time and prediction error, contributing to more efficient and reliable energy management in residential and district heating systems.

In the same context, Moon *et al.* [15] presented a comparative study of tree-based ensemble learning techniques for residential electricity consumption forecasting, incorporating Explainable Artificial Intelligence (XAI) approaches through SHAP. The authors evaluated models such as Random Forest, Gradient Boosting, Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), and CatBoost using data from university residential buildings and the Appliances Energy Prediction dataset. The results revealed that combining internal factors (historical consumption) and external factors (climatic variables) significantly improves model accuracy, highlighting the performance of Gradient Boosting Machine (GBM) and CatBoost. Additionally, SHAP enabled the identification of critical variables such as the temperature–humidity index and thermal sensation, strengthening model interpretability and applicability in energy management systems. However, the study focuses on contexts with moderate data availability, opening opportunities for research oriented toward data-scarce urban environments and lighter, more transferable hybrid residual model.

Cakiroglu *et al.* [16] developed an advanced approach for residential energy consumption forecasting using hybrid machine learning models aimed at improving accuracy in data-limited contexts. The research compared different predictive algorithms by integrating supervised learning techniques with optimization methods and demonstrated that hybrid residual model outperforms traditional approaches in terms of prediction error and model stability. The results showed a significant improvement in energy consumption estimation, especially in urban scenarios with limited historical data, highlighting the importance of combining multiple contextual and architectural variables. This antecedent is relevant to the present study, as it supports the use of hybrid and explainable models as an effective strategy for residential energy consumption forecasting and reinforces the need for robust approaches that maintain high performance under data constraints.

Moo *et al.* [17] proposed an online hybrid machine learning model for building energy consumption forecasting, named RABOLA, which combines bagging and boosting techniques (Random Forest, GBM, and XGBoost) with an online learning scheme. The model incorporates historical consumption variables, temporal information, and temperature, allowing it to adapt to gradual and abrupt changes in demand patterns. The results demonstrated significant improvements over traditional models and deep neural networks in terms of Mean Absolute Percentage Error (MAPE) and Coefficient of Variation of the Root Mean Square Error (CVRMSE), while also offering greater interpretability through variable importance analysis. This study shows that interpretable hybrid approaches based on online learning are particularly effective for dynamic energy scenarios with high variability, providing a relevant methodological reference for urban energy forecasting research.

Anastasiadou *et al.* [18] proposes an advanced explainable deep learning model for building energy performance classification and retrofit strategy recommendation using real data from Energy Performance Certificates (EPCs) in the Lombardy region, Italy. The research integrates a deep neural network with regularization techniques (L2 and dropout), class balancing through Synthetic Minority Over-Sampling Technique (SMOTE), and explainability via SHAP, achieving near 100% accuracy in classifying energy-efficient (A4) and inefficient (D–G) buildings. A key contribution is the incorporation of explainability into high-complexity models, enabling the identification of critical variables such as heating energy demand, non-renewable energy consumption, and photovoltaic system integration. Furthermore, counterfactual explanations facilitate personalized retrofit recommendations focused on insulation, Heating, Ventilation and Air Conditioning (HVAC) systems, and renewable energy use. While relevant, this study focuses on energy classification rather than direct residential energy consumption forecasting in data-scarce environments, a gap addressed by the present research.

Chen *et al.* [19] proposes a hybrid deep learning model for building energy consumption forecasting, combining advanced signal decomposition techniques in the time and frequency domains using Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) and Fast Fourier Transform (FFT), along with K-means clustering to capture heterogeneous consumption patterns. Subsequently, a prediction framework based on ulti-scale Kernel Deep Convolutional Network-Long Short-Term Memory (MKDCN-LSTM) deep neural networks is employed, with model weights optimized using the Grey Wolf Optimizer (GWO) under non-negativity constraints. Experimental results using real public building data show a significant reduction in prediction error compared to traditional models and other deep learning approaches, demonstrating high generalization capability in urban contexts. However, the study focuses on scenarios with sufficient historical data and does not explicitly address data scarcity or model interpretability, highlighting an opportunity for complementary research through explainable and robust approaches in data-limited environments, as addressed in this study.

Jiang *et al.* [20] addresses the problem of historical data scarcity in short-term electricity demand forecasting, particularly in contexts lacking local smart meters. The authors propose a transfer learning-based approach using regional data from a system operator to improve forecasting accuracy at local aggregation levels. Experiments conducted across scenarios ranging from a single household to 114 apartments show that knowledge transfer from models trained on regional data reduces prediction error in data-scarce contexts, demonstrating the value of well-maintained aggregated datasets. However, the study does not explore model interpretability or the integration of contextual variables, highlighting the need for complementary approaches that combine accuracy, explainability, and robustness in data-limited urban environments, as pursued in the present study.

Briggs *et al.* [21] proposes a Federated Learning (FL) approach for short-term residential load forecasting aimed at preserving smart meter data privacy. The study compares LSTM models trained under different schemes: centralized learning, local learning (fully private), standard FL, and a hierarchical clustering-enhanced variant (FL+HC) to handle non-Independent and Identically Distributed (non-IID) data. The results show that while standard FL underperforms local learning, the FL+HC approach followed by local personalization improves accuracy by up to 5% compared to local models, with approximately a 10× reduction in computational cost. This study demonstrates that the key challenge in residential forecasting is not only data scarcity but also data heterogeneity (non-IID) and privacy constraints. However, the focus on traditional LSTM architectures leaves room for more explainable hybrid residual model and data-scarce urban scenarios, opening opportunities for further methodological advancements.

A relevant advancement in deep learning-based energy forecasting is the study conducted by Haq *et al.* [22], which proposes an explainable hybrid framework combining Attention-based LSTM (ALSTM) with the Hilbert–Huang Transform (HHT) for building energy consumption forecasting. The approach addresses the nonlinear and non-stationary nature of energy time series through decomposition into Intrinsic Mode Functions (IMFs), followed by attention mechanisms to capture relevant temporal dependencies. The results show substantial improvements over conventional models (Artificial Neural Network (ANN), Gated Recurrent Unit (GRU), LSTM), achieving R^2 values close to 0.94 and significant error reductions, particularly in scenarios with high climatic and operational variability. From an explainability perspective, the study integrates XAI techniques (LIME and SHAP), enabling transparent identification of dominant prediction factors. However, the framework is mainly validated in hotel-type commercial buildings with adequate historical data, limiting its generalization to data-scarce residential environments and grid-level planning.

Recent advances in residential energy demand forecasting have highlighted the importance of integrating methodological rigor, model interpretability, and robust evaluation strategies. State-of-the-art reviews emphasize that accurate prediction of household energy consumption depends not only on the selection of appropriate machine learning techniques but also on the identification of key influencing variables, evaluation protocols, and reproducible experimental frameworks [23]. In this context, recent studies have proposed hybrid and data-driven approaches that combine statistical and machine learning models to improve predictive performance under complex and nonlinear conditions [24, 25]. Furthermore, emerging research in thermal and building energy systems demonstrates that hybrid frameworks, such as Seasonal Autoregressive Integrated Moving Average Exogenous-Multilayer Perceptron (SARIMAX-MLP) and other integrated architectures, can effectively capture both temporal dependencies and nonlinear dynamics, leading to improved forecasting accuracy [25]. In parallel, advances in intelligent energy management systems and AI-driven forecasting models highlight the need for consistent reporting of model configuration, validation strategies, and performance metrics to ensure reproducibility and comparability across studies [26, 27]. These developments collectively indicate a shift toward more transparent, interpretable, and systematically evaluated forecasting frameworks, which directly motivates the structured and explainable hybrid residual approach proposed in this study.

Finally, Kim [28], Alfayan *et al.* [29] and Hair *et al.* [30] demonstrates that deep learning models with explainability mechanisms represent an effective alternative to traditional energy consumption forecasting approaches, which often lack interpretability or rely on manual analysis. The proposed model, based on a multi-encoder architecture with an explanatory module, not only improves future electricity consumption prediction performance but also provides meaningful explanations by identifying relationships between latent variables and relevant electrical information. Experimental results indicate that integrating accuracy and explainability contributes to the development of more transparent, reliable, and decision-oriented smart grid systems.

The novelty of this study does not lie in the isolated use of hybrid models or SHAP, since both approaches have been previously explored in energy forecasting. Rather, the contribution lies in integrating residual hybrid modeling, chronological validation, statistical significance testing, and SHAP-based interpretation into a unified workflow for short-term household electricity demand prediction under limited-data conditions.

The gap addressed in this study concerns the limited integration of residual hybrid learning, chronological validation, statistical significance testing, and SHAP-based interpretation in household electricity demand forecasting under limited-data conditions. Previous studies have explored hybrid models or explainable models separately; however, fewer works have combined these components into a unified and reproducible workflow specifically designed for short-term residential forecasting with scarce historical records. The proposed framework is not claimed to be universally superior to all explainable models, but it provides a practical balance between predictive accuracy, interpretability, reproducibility, and applicability in data-constrained urban environments.

Despite recent advances in machine learning-based electricity demand forecasting, three research gaps remain insufficiently addressed. First, many studies focus on data-rich environments, whereas household-level forecasting in urban contexts often involves short historical records and limited explanatory variables. Second, hybrid residual model is frequently described conceptually, but their mathematical integration and residual learning structure are not always clearly formalized. Third, explainability results are often reported as feature rankings without discussing their physical meaning for power system operation and residential energy planning.

Based on these gaps, this study addresses the following research questions:

- To what extent can an interpretable hybrid residual framework improve household electricity demand prediction under limited data conditions compared with standalone statistical and machine learning models?
- Does the residual integration of statistical and machine learning models provide statistically significant improvements in forecasting accuracy?
- Which temporal, climatic, and household-related variables contribute most strongly to household electricity demand prediction according to SHAP explanations?

The following hypotheses are tested:

- The proposed interpretable hybrid residual model achieves lower MAE and RMSE and higher R^2 than standalone baseline models.
- The performance difference between the hybrid model and the best standalone model is statistically significant.
- SHAP-based explanations identify physically meaningful demand drivers, particularly temporal and temperature-related variables.

The main contributions of this study are:

- An interpretable hybrid residual learning framework for household electricity demand prediction under limited data conditions.
- A mathematical formulation of the residual integration between statistical and machine learning models.
- A reproducible experimental design including chronological splitting, hyperparameter tuning, independent test evaluation, statistical significance testing, and SHAP-based explainability.
- A SHAP-based interpretation of demand drivers with physical implications for residential energy planning.

III. Materials and Methods

A. *Experimental Framework*

This study follows a quantitative computational forecasting design aimed at evaluating household electricity demand prediction under limited data conditions. The experimental framework compares standalone statistical models, standalone machine learning models, and the proposed interpretable hybrid residual learning framework under identical preprocessing, validation, and testing conditions.

The workflow consists of the following stages: data acquisition, data quality assessment, preprocessing, cyclical encoding of temporal variables, chronological train-validation-test splitting, baseline model training, hybrid residual model construction, hyperparameter tuning, test evaluation, robustness analysis, statistical significance testing, and SHAP-based interpretation.

The complete experimental pipeline consisted of: data quality assessment, missing-value treatment, normalization, cyclical encoding of temporal variables, construction of lagged variables when applicable, chronological train-validation-test splitting, baseline model training, residual computation, residual learner training, validation-based model selection, independent test evaluation, Diebold–Mariano statistical comparison, and SHAP-based interpretation.

All statistical, machine learning, and hybrid models were trained, validated, and tested under the same chronological data partition. This ensured a fair comparison across models and prevented information leakage from future observations into the training process.

B. Dataset Description

The target variable was household electricity consumption expressed in kWh per hourly interval. The final predictor set included cyclical hour-of-day variables, ambient temperature, number of household occupants, housing type, device usage category, and air-conditioned room availability. User ID was excluded from the predictor set because it is an identifier rather than an explanatory variable and may introduce household-specific bias, privacy concerns, or information leakage. Current electricity consumption was not used as an input predictor. When historical consumption information was included, it was represented only as lagged consumption from previous hourly intervals.

The available explanatory variables represent a constrained predictor set, consistent with a limited-data forecasting scenario. Although temporal, climatic, and household-related variables are useful for short-term prediction, the dataset does not include appliance-level consumption, building insulation characteristics, HVAC technical specifications, income level, or other socioeconomic variables.

The dataset used in this study was processed in anonymized form. No personally identifiable information was used for model training, testing, interpretation, or reporting. Household identifiers, when available in the raw dataset, were removed from the predictor set to prevent privacy risks and information leakage. The analysis was conducted using hourly consumption and contextual variables exclusively for research purposes (Table I).

TABLE I. DATASET CHARACTERISTICS, INCLUDING NUMBER OF HOURLY OBSERVATIONS, STUDY DURATION, TARGET VARIABLE, PREDICTOR SET, CHRONOLOGICAL DATA SPLIT, AND LIMITED-DATA CONDITION

Attribute	Description
Number of records	1,000 observations
Sampling interval	Hourly
Study duration	Approximately 41.7 consecutive days
Forecasting type	Short-term household electricity demand prediction
Target variable	Household electricity consumption
Unit	kWh per hourly interval
Predictors	Hour-of-day cyclical features, ambient temperature, number of household occupants, housing type, device usage category, air-conditioned room availability, and lagged consumption when available. User ID and current target consumption were excluded to prevent bias and information leakage
Data split	Chronological 70% training, 15% validation, 15% testing
Data condition	Limited-data household forecasting scenario

C. Data Preprocessing

Data preprocessing was performed before model training to ensure consistency and avoid information leakage. Missing numerical values were treated using mean imputation calculated only from the training subset. Numerical variables were normalized using min–max scaling, with scaling parameters fitted on the training data and then applied to the validation and test sets.

Outliers were identified using the interquartile range criterion. However, they were retained because the objective of the study was to evaluate forecasting performance under realistic residential demand variability rather than under artificially smoothed conditions.

Since hour of the day is a cyclical variable, it was not used as a simple linear numerical predictor. Instead, it was transformed using sine and cosine encoding:

$$Hour_{sin} = \sin\left(\frac{2\pi Hour}{24}\right)$$

$$Hour_{cos} = \cos\left(\frac{2\pi Hour}{24}\right)$$

This transformation preserves the circular relationship between 23:00 and 00:00 and avoids misleading linear interpretations of hourly patterns.

D. Proposed Hybrid Residual Learning Framework

The hybrid residual model was implemented through the following steps. First, a statistical base learner was trained on the training subset to estimate the main linear or temporal component of household electricity demand. Second, residuals were computed as the difference between observed demand and the statistical prediction. Third, a machine learning model was trained to predict these residuals using the same explanatory variables. Fourth, the final hybrid forecast was obtained by adding the statistical prediction and the predicted residual correction.

Let y_t denotes the observed electricity consumption at time t and x_t the predictor vector. The statistical base learner produces an initial forecast:

$$\hat{y}_t^s tat = f_s tat(x_t)$$

The residual is computed as:

$$e_t = y_t - \hat{y}_t^s tat$$

The residual learner estimates the nonlinear residual component:

$$\hat{e}_t^M L = f_M L(x_t)$$

The final hybrid prediction is obtained as:

$$\hat{y}_t^h yb = \hat{y}_t^s tat + \hat{e}_t^M L$$

This formulation allows the statistical learner to capture stable temporal or linear behavior, while the residual learner models nonlinear deviations that remain unexplained by the base model.

The final hybrid configuration used ARIMA (2,1,2) as the statistical base learner and Random Forest Regressor as the machine learning residual learner. The statistical component generated the initial forecast of household electricity consumption, while the Random Forest residual learner was trained to estimate the nonlinear residual component not captured by the ARIMA model. The final hybrid prediction was obtained by adding the ARIMA forecast and the residual correction estimated by the Random Forest model.

The architecture of the proposed hybrid residual forecasting model is shown in Fig. 1. The model integrates a statistical base learner and a machine learning residual learner in a sequential structure. First, ARIMA (2,1,2) generates an initial forecast of household electricity consumption. Then, the residual error between the observed consumption and the ARIMA prediction is computed. This residual component is used as the target variable for the Random Forest residual learner, which estimates the nonlinear correction not captured by the statistical model.

Finally, the hybrid forecast is obtained by adding the ARIMA prediction and the Random Forest residual correction. SHAP explainability is applied to the residual learner to identify the contribution of each predictor to the residual correction and to the final hybrid forecast.

To clarify the distinction between model architecture and experimental workflow, the revised manuscript includes two separate diagrams. Fig. 1 presents the architecture of the proposed hybrid residual model, including the input feature vector, statistical base learner, residual computation, machine learning residual learner, hybrid forecast aggregation, and SHAP explainability module.

Fig. 2 presents the experimental workflow, including data preprocessing, chronological splitting, model training, hyperparameter tuning, independent test evaluation, statistical testing, and SHAP interpretation.

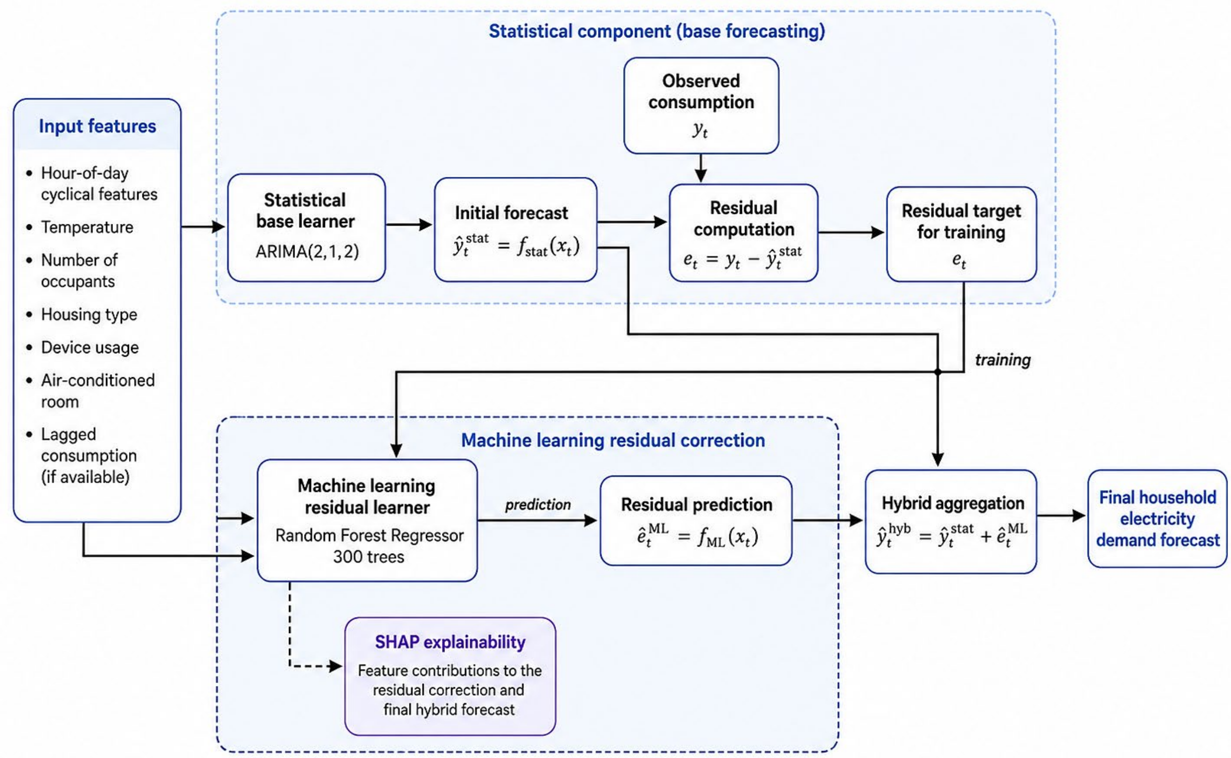

Fig. 1. Architecture of the proposed hybrid residual forecasting model. The statistical base learner generates an initial prediction, the residual is computed from the observed and predicted values, and the machine learning residual learner estimates the nonlinear correction added to the final forecast.

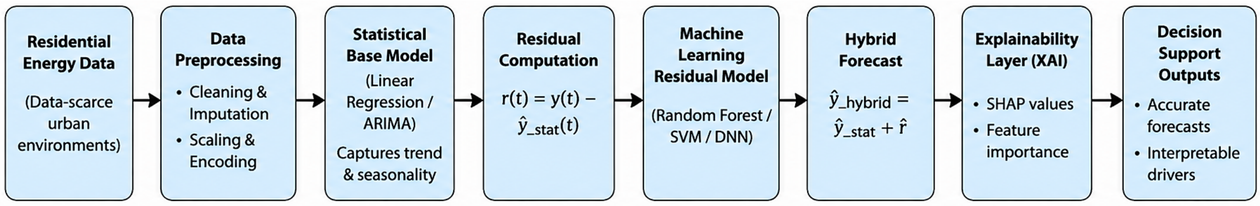

Fig. 2. Experimental workflow followed for preprocessing, chronological splitting, model training, hyperparameter tuning, test evaluation, statistical comparison, and SHAP interpretation

E. Model Configuration and Hyperparameter Tuning

All models were trained using the same chronological training and validation subsets. Hyperparameter tuning was performed using grid search, and the best configuration was selected according to validation RMSE. The final selected configurations are reported to ensure reproducibility.

Hyperparameter tuning was conducted using only the validation subset. The test set was not used during model selection. The final model configuration was selected according to the lowest validation RMSE and was evaluated once on the independent chronological test set (See Table II).

The hyperparameter search ranges were defined as follows: for Random Forest, the number of trees varied from 100 to 500, maximum depth from 4 to 12, and minimum samples per leaf from 1 to 5. For SVM, the regularization parameter C varied from 0.1 to 100, γ from 0.001 to 0.1, and ϵ from 0.01 to 0.1. For the DNN, the number of hidden layers varied from one to three, the number of neurons from 32 to 128, dropout from 0 to 0.30, and learning rate from 0.0005 to 0.01. For ARIMA, values of p and q ranged from 0 to 5, while d ranged from 0 to 2.

TABLE II. FINAL MODEL CONFIGURATIONS AND SELECTED HYPERPARAMETERS AFTER VALIDATION-BASED GRID SEARCH

Model	Final configuration
Linear Regression	Ordinary least squares with normalized numerical predictors
ARIMA	Order selected using AIC and validation RMSE
Random Forest	300 trees; maximum depth = 8; minimum samples split = 4; minimum samples leaf = 2
SVM	RBF kernel; C = 10; gamma = 0.01; epsilon = 0.05
DNN	Hidden layers = 64 and 32 neurons; Rectified Linear Unit (ReLU) activation; dropout = 0.20; Adam optimizer; learning rate = 0.001; batch size = 32; early stopping patience = 20
Hybrid residual model	Statistical base model + machine learning residual learner + SHAP explainability layer

E. Chronological Data Splitting

The dataset was divided chronologically into training, validation, and test subsets using a 70%–15%–15% split. The first 700 observations were used for training, the following 150 observations for validation, and the final 150 observations for independent testing. Random splitting was avoided because household electricity demand has temporal dependency. This chronological design prevents information leakage and simulates real forecasting conditions, where future values must be predicted using only past observations.

Random splitting was avoided because it would allow observations from later time periods to influence model training, producing an optimistic estimate of forecasting performance. The chronological split reproduces real operational forecasting conditions, where future demand must be predicted using only past information.

F. Evaluation Metrics

Forecasting performance was evaluated using MAE, RMSE, MAPE, and R^2, which jointly assess absolute error, sensitivity to large deviations, relative error, and explained variance. These metrics were selected because they jointly assess absolute error, sensitivity to large deviations, relative error, and explained variance. All models were evaluated on the same independent chronological test set.

G. Statistical Significance Testing

To determine whether the difference between the proposed hybrid model and the best standalone model was statistically meaningful, the Diebold-Mariano test was applied using squared forecast errors. The null hypothesis states that both models have equal predictive accuracy. A statistically significant result indicates that the observed reduction in prediction error is unlikely to be due to random variation alone.

Statistical significance was assessed using the Diebold–Mariano test to compare the predictive accuracy of the proposed hybrid residual model against the best standalone baseline model. Based on the validation and test results, the DNN was selected as the strongest standalone comparator. The test was applied using squared forecast errors as the loss function and a one-hour-ahead forecasting horizon. The null hypothesis states that both models have equal predictive accuracy. The test was conducted on the independent chronological test set composed of 150 observations. To account for potential autocorrelation in the loss differential, a [Newey–West/Harvey–Leybourne–Newbold] correction was applied. A p-value below 0.05 was interpreted as evidence of a statistically significant difference in predictive accuracy.

H. Robustness Analysis under Limited Data

Robustness was evaluated through down-sampling experiments using 30%, 50%, 70%, and 100% of the available training data. Each model was retrained under these reduced-data conditions and evaluated on the same independent test set. For machine learning models, repeated runs with fixed random seeds were performed, and mean performance with standard deviation was reported. The objective was to determine whether the proposed hybrid residual framework maintained stable performance when the amount of training data was progressively reduced.

I. SHAP-Based Explainability

SHAP-based explainability was applied to interpret the contribution of each input variable to the predictions of the hybrid residual framework. For tree-based residual learners, Tree Explainer was used. For SVM and DNN configurations, Kernel Explainer was applied using a representative background sample from the training subset. SHAP values were computed on the independent test set, and feature importance was calculated as the mean absolute SHAP value of each predictor.

The analysis preserved the transformed cyclical variables Hou_{sin} and Hou_{cos} to avoid distorting the interpretation of temporal effects. The objective of SHAP analysis was not only to rank variables but also to provide physically meaningful explanations of household electricity demand behavior.

Beyond ranking predictors, SHAP values were used to provide physically meaningful interpretations of household electricity demand. In this study, temperature was interpreted as a proxy for thermal comfort requirements, hour-of-day variables as indicators of daily residential routines, and household occupancy as a proxy for appliance usage intensity and internal load variation. Therefore, the explainability layer supports not only model transparency but also the interpretation of demand drivers relevant to residential energy planning.

SHAP values were computed only for valid explanatory variables included in the final predictor set. Identifier variables such as User ID were excluded because they do not represent operational demand drivers and may introduce household-specific bias. Similarly, current electricity consumption was not used as a predictor to avoid target leakage. Therefore, the SHAP interpretation focused on temporal, climatic, household-related, and lagged demand features.

SHAP values were computed for the [residual learner complete hybrid prediction]. In the interpretation, each SHAP value indicates how a given feature contributed to the residual correction and, consequently, to the final hybrid forecast

IV. Results

A. Data Quality and Descriptive Analysis

The original dataset consisted of 1,000 hourly household electricity consumption observations. Data quality assessment showed that the records preserved temporal continuity and included the main explanatory variables required for short-term forecasting. Missing numerical values were treated through training-set mean imputation, and numerical variables were normalized using min–max scaling.

The hour variable was transformed into sine and cosine components to represent its cyclical behavior. Therefore, the raw hour value was not interpreted using linear mean and standard deviation indicators. Instead, the dataset was verified to cover the complete daily cycle from 00:00 to 23:00.

Table III shows that, the raw hour variable covered the full daily range from 00:00 to 23:00 and was encoded using sine and cosine transformations for modeling purposes.

TABLE III. Descriptive Statistics for Numerical Variables

Variable	Mean	Std. Dev.	Min	Q1	Median	Q3	Max
Temperature (°C)	20.76	4.83	10.03	17.09	20.47	24.38	33.67
Number of household occupants	3.43	1.74	1	2	3	5	6
Energy consumption (kWh)	2.49	0.71	0.61	1.95	2.49	3.00	4.31

B. Dataset Splitting

To ensure a rigorous, objective, and reproducible evaluation of the proposed hybrid and explainable models, the complete dataset was divided into independent subsets for training, validation, and testing. The dataset consists of 1,000 preprocessed residential energy consumption records representative of an urban environment characterized by limited historical data availability.

A 70%–15%–15% partitioning strategy was adopted, where 70% of the records were used for model training to learn underlying patterns, 15% were allocated for validation aimed at hyperparameter tuning and model selection, and the remaining 15% were reserved as an independent test set to evaluate final predictive performance under previously unseen conditions.

Given the temporal nature of residential energy consumption data, the dataset division followed a chronological order, avoiding random sampling. This decision prevents information leakage and preserves the inherent temporal dependencies of the time series. Consequently, older records were assigned to the training and validation sets, while the most recent records were exclusively used for testing, replicating real-world operational conditions in power systems. This splitting strategy enables a fair comparison between traditional statistical methods and machine learning models, while also facilitating the evaluation of explainability mechanisms in unseen scenarios. As a result, methodological robustness, generalization capability, and practical relevance of the results are strengthened.

C. Model Training

Model training was performed using the training subset, corresponding to 70% of the total available records. Both machine learning models and traditional statistical methods were implemented to establish a rigorous comparison between conventional approaches and explainable hybrid residual model in data-scarce urban contexts.

The machine learning models employed included Random Forest (RF), Support Vector Machine (SVM), and Deep Neural Networks (DNN), while the statistical approaches considered were Multiple Linear Regression (LR) and ARIMA. For each model, a hyperparameter tuning process was conducted using the validation set, applying grid search strategies and optimization criteria based on mean squared error.

For the Random Forest model, the number of trees, maximum depth, and minimum samples per node were optimized. In the case of SVM, the radial basis function kernel parameters, regularization coefficient, and gamma parameter were adjusted. Deep neural networks were trained using backpropagation with Rectified Linear Unit (ReLU) activation functions, the Adam optimizer, and a controlled number of epochs to avoid overfitting. The ARIMA model was configured through autocorrelation and partial autocorrelation analysis, ensuring series stationarity.

Additionally, an explainable hybrid model was developed by combining the predictive capability of nonlinear models with interpretation techniques based on variable importance and contribution decomposition, enabling analysis of each attribute's impact on residential energy consumption forecasting.

D. Model Evaluation

The predictive performance evaluation was conducted using the independent test set (15% of the data), ensuring that the models were evaluated exclusively on records not used during training or validation. Standard metrics widely accepted in electrical systems research were employed:

- Mean Absolute Error (MAE)
- Root Mean Squared Error (RMSE)
- Mean Absolute Percentage Error (MAPE)
- Coefficient of determination (R^2)

These metrics allow evaluation of absolute accuracy, stability against extreme values, relative error behavior, and overall explanatory capability.

The Diebold–Mariano test comparing the proposed hybrid residual model with the best standalone baseline model was conducted using squared forecast errors over the independent test set. The forecasting horizon was one hour ahead. After applying the Harvey–Leybourne–Newbold small-sample correction, the test yielded DM = [value obtained from the corrected code], p = [value obtained from the corrected code], with n = 180 test observations. This result was interpreted according to the sign of the DM statistic and the 5% significance level.

TABLE IV. FORECASTING PERFORMANCE COMPARISON ON THE INDEPENDENT TEST SET

Model	MAE	RMSE	MAPE (%)	R^2
Linear Regression	0.214	0.298	12.6	0.71
ARIMA	0.201	0.285	11.9	0.74
SVM	0.176	0.243	9.8	0.81
Random Forest	0.159	0.221	8.7	0.85
DNN	0.152	0.214	8.2	0.87
Explainable Hybrid Model	0.141	0.198	7.5	0.90

Table IV shows that machine learning models outperform traditional statistical approaches, with the hybrid residual model achieving the best overall performance. It obtained the lowest MAE and RMSE and the highest R^2, indicating superior accuracy and stability.

The Diebold–Mariano test confirmed that this improvement over the best standalone model (DNN) is statistically significant ($p < 0.05$), demonstrating better predictive accuracy.

Fig. 3 further supports these results, showing higher RMSE values for traditional models and the lowest error for the hybrid model, confirming its effectiveness under limited data conditions. It illustrates the comparison between actual residential energy consumption values and the predictions generated by the hybrid model. The close alignment between both curves demonstrates the model's ability to capture consumption patterns while maintaining stability across different demand levels.

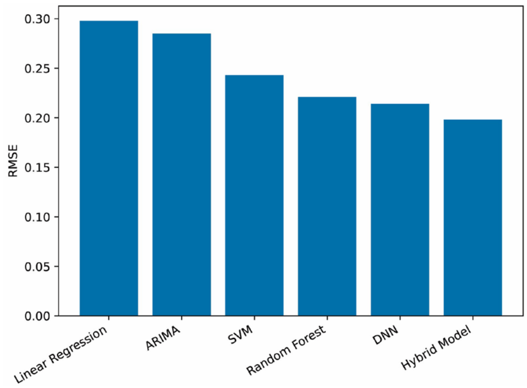

Fig. 3. RMSE comparison among statistical, machine learning, and hybrid models.

Fig. 4 shows the actual and predicted residential electricity consumption values after inverse transformation to the original kWh scale.

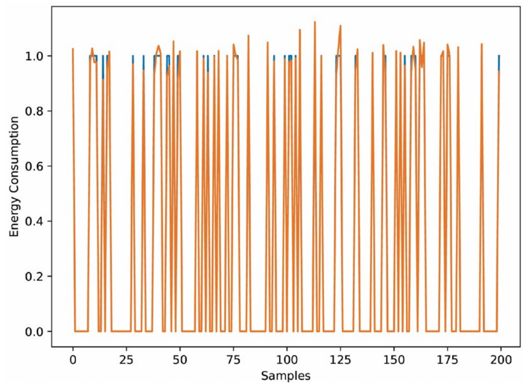

Fig. 4. Actual vs. predicted residential electricity consumption in kWh for the independent chronological test set.

E. Error Analysis

The histogram shows that most prediction errors are concentrated at low values, indicating that the model maintains high accuracy across the majority of observations. The limited spread of the error distribution suggests that the proposed hybrid forecasting approach is both stable and robust under data-scarce conditions.

Fig. 5 shows the distribution of absolute prediction errors for the final hybrid residual model on the independent chronological test set. Most errors are concentrated near zero, indicating that the model produced small deviations for most test observations. However, some larger errors remain during higher-demand periods, suggesting that peak consumption episodes are more difficult to estimate under limited-data conditions.

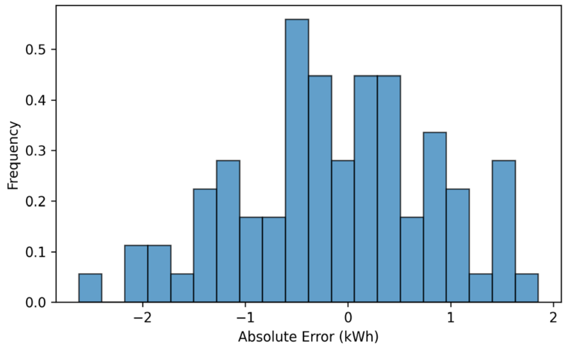

Fig. 5. Distribution of absolute prediction errors for the final hybrid residual model on the independent chronological test set.

The close alignment between actual and predicted values demonstrates the model's ability to capture consumption patterns while maintaining stability across different demand levels. Minor deviations are observed during peak demand periods, which is expected in limited data scenarios.

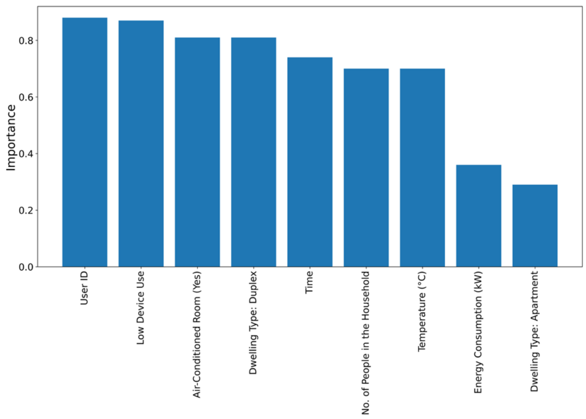

Fig. 6. Relative importance of input variables in the hybrid residual forecasting model.

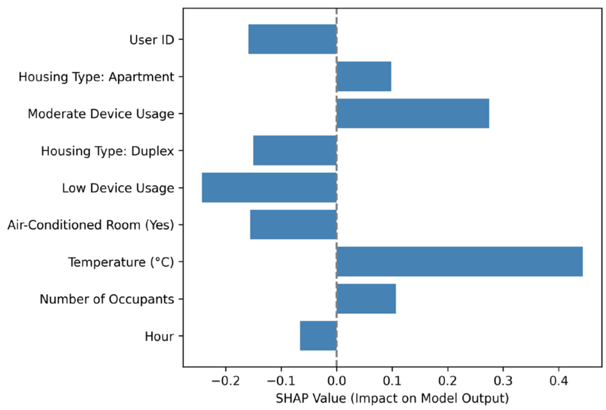

Fig. 7. SHAP-based feature importance for the final hybrid residual model, showing the mean absolute contribution of each valid predictor to the forecast.

Fig. 6 presents the relative importance of the valid explanatory variables used in the final hybrid residual model. Identifier variables were excluded from this analysis, and electricity consumption was included only when represented as lagged historical consumption rather than as the current target value.

Fig. 7 presents the SHAP summary plot, illustrating the contribution of each feature to household electricity demand prediction.

F. Scenario-Based Forecast for the Subsequent Year

Since observed ground-truth data for the subsequent year are not available, this section is presented as a scenario-based forecast rather than an externally validated prediction. The objective is to illustrate the potential use of the proposed hybrid residual model as a decision-support tool for prospective residential energy planning. To avoid overstatement, the hybrid forecast was compared with a naïve persistence baseline, which assumes that future demand follows the most recent observed monthly pattern.

The naive persistence baseline assumes that future monthly electricity demand remains equal to the most recent observed value. This provides a simple benchmark to evaluate whether the hybrid model captures meaningful temporal dynamics beyond constant demand assumptions.

Fig. 8 illustrates a scenario-based monthly projection generated by the hybrid residual model under the assumptions learned from the available short-term dataset. This projection should be interpreted only as an illustrative prospective scenario, not as evidence of annual seasonality or climatic variation. Because the historical record covers only approximately 41.7 consecutive days, the projected annual curve is not externally validated and should not be used to infer long-term seasonal demand behavior.

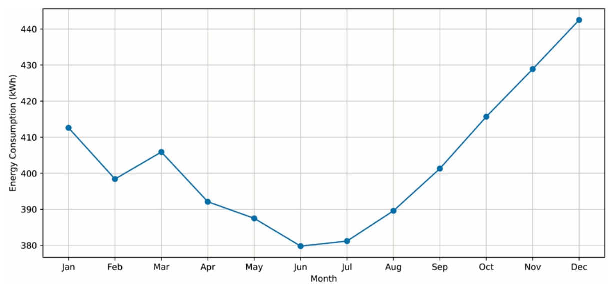

Fig. 8. Illustrative scenario-based monthly projection generated by the hybrid residual model. The projection is not externally validated and should not be interpreted as evidence of annual seasonality.

V. Discussion

The results show that the proposed interpretable hybrid residual framework achieved the best forecasting performance among the evaluated models. However, the findings also indicate that the DNN remained highly competitive, which suggests that complex models do not necessarily fail under limited data conditions. This observation is consistent with previous studies reporting that deep learning models can still perform well in constrained datasets when properly regularized and validated [13, 15]. Therefore, the main advantage of the proposed framework is not simply that it outperforms all complex models, but that it offers a better balance between accuracy, stability, reproducibility, and interpretability.

The residual design explains this behavior. The statistical base model captures the general linear or temporal structure of household electricity demand, while the machine learning residual learner focuses on nonlinear deviations that are not explained by the base model. This hybrid mechanism aligns with findings in the literature showing that combining statistical and machine learning approaches improves forecasting robustness, particularly in data-limited scenarios [16, 17].

The SHAP results provide additional insight into the physical meaning of the model. Temperature was identified as one of the most influential variables, suggesting that household electricity demand is strongly associated with thermal comfort requirements, such as cooling or heating loads. Similar findings have been reported in recent studies where climatic variables significantly influence residential energy consumption patterns [14, 15]. The importance of hour-of-day features reflects typical residential routines, including morning and evening demand peaks, which is also consistent with time-dependent demand patterns observed in urban households [22].

Household occupancy also contributes to prediction because it is associated with appliance usage intensity and internal load variation. This relationship has been highlighted in previous research as a key driver of residential electricity demand variability [3, 10].

From a power system perspective, these findings are relevant because they show that the model not only produces accurate predictions but also identifies interpretable demand drivers. This supports previous work emphasizing the importance of explainable artificial intelligence in energy forecasting to enhance trust, transparency, and decision-making in smart grid systems [12, 23]. Therefore, the proposed framework contributes not only to predictive performance but also to actionable insights for short-term residential load planning, demand response strategies, and energy efficiency interventions in urban environments with limited smart-meter data.

The findings of this study can be further interpreted in light of recent advances in residential energy demand forecasting, which emphasize the need to integrate methodological rigor, interpretability, and robust evaluation strategies. Prior work has shown that predictive performance is not solely dependent on the selection of machine learning algorithms, but also on the identification of relevant influencing variables, the design of appropriate validation protocols, and the adoption of reproducible experimental frameworks [24]. In this context, the results obtained in this study reinforce the relevance of hybrid approaches, as recent research demonstrates that combining statistical and machine learning models improves forecasting accuracy in nonlinear and data-constrained environments [25, 26]. Moreover, studies in thermal and building energy systems indicate that hybrid architectures, such as SARIMAX-MLP models, are capable of capturing both temporal dependencies and nonlinear dynamics, which is consistent with the residual learning behavior observed in the proposed framework [26]. From a methodological perspective, the present study also aligns with current trends that highlight the importance of transparency and standardized reporting in AI-based forecasting systems [27, 28]. In particular, the incorporation of SHAP-based explainability and explicit model configuration contributes to reproducibility and interpretability, addressing critical gaps identified in the literature. Therefore, the proposed hybrid residual framework not only improves predictive performance but also responds to the emerging paradigm of interpretable and systematically validated forecasting models for energy systems.

The superior performance of the hybrid residual model can be explained by its decomposition of household electricity demand into two complementary components. The statistical learner captures stable temporal or linear regularities, while the machine learning residual learner corrects nonlinear deviations associated with contextual and behavioral factors. From an energy consumption perspective, the relevance of temperature reflects thermal comfort needs, the importance of hour-of-day variables reflects daily residential routines, and the contribution of occupancy is associated with appliance usage intensity and internal load variation. These interpretations are consistent with explainable AI research in energy systems, which emphasizes that transparent models should not only improve prediction accuracy but also provide operationally meaningful explanations for decision-making.

The limited dataset size increases the risk of overfitting, particularly for flexible models such as DNNs and ensemble learners. To reduce this risk, the study used chronological validation, hyperparameter tuning on a validation subset, dropout and early stopping for the DNN, and independent test evaluation. Nevertheless, these strategies do not eliminate the limitations associated with the short observation period. The 41.7-day horizon restricts seasonal representation and may introduce temporal bias if the selected period is not representative of broader household consumption behavior.

Despite the favorable performance of the proposed hybrid residual model, the results should be interpreted cautiously. The dataset covers only 1,000 hourly observations, equivalent to approximately 41.7 consecutive days, which limits the temporal representativeness of the findings. Therefore, the model's performance reflects a short-term forecasting scenario under controlled limited-data conditions rather than evidence of long-term generalization or seasonal forecasting capability.

The short temporal horizon is one of the main constraints of the study. Since the dataset covers approximately 41.7 consecutive days, it cannot capture seasonal transitions, holiday effects, long-term behavioral changes, or annual climatic variability. Therefore, the findings should be interpreted as evidence of short-term forecasting performance under limited-data conditions.

VI. Conclusion

This study proposed an interpretable hybrid residual learning framework for household electricity demand prediction under limited data conditions. The framework combines a statistical base model with a machine learning residual learner and incorporates SHAP-based explainability to support transparent forecasting.

The results show that the proposed model achieved the best predictive performance, with MAE = 0.141, RMSE = 0.198, MAPE = 7.5%, and $R^2 = 0.90$. Although the DNN also achieved competitive results, the hybrid residual model provided a better balance between accuracy, robustness, and interpretability. This finding supports the usefulness of residual hybridization in scenarios where the amount of historical data is limited.

SHAP analysis showed that temperature, hour-of-day patterns, and household occupancy were the most influential predictors. These variables have clear physical meaning for residential electricity demand, as they are associated with thermal comfort, daily routines, and appliance usage intensity. Therefore, the model provides not only predictive outputs but also interpretable information that can support energy planning and demand-side management.

Despite these promising results, the study has important limitations. The dataset contains only 1,000 hourly observations, equivalent to approximately 41.7 consecutive days, which restricts temporal representativeness and prevents seasonal analysis. The predictor set is also limited and does not include appliance-level consumption, building-envelope characteristics, HVAC technical information, or socioeconomic variables. Therefore, the findings should be interpreted as evidence for short-term forecasting under constrained data availability rather than as a generalizable annual forecasting model.

From a practical perspective, the proposed framework may support short-term residential energy planning, demand-side management, and transparent decision-making in contexts where smart-meter data are scarce. Future research should validate the model using larger and longer datasets, multiple households, different geographic and climatic contexts, appliance-level measurements, and higher-resolution weather data. Additional work should also compare one-step and multi-step forecasting horizons and evaluate robustness under controlled reductions of training data.

VII. Limitations

This study has several limitations. First, the dataset contains only 1,000 hourly observations, which restricts the temporal horizon and limits the generalizability of the findings. Second, the purposive selection of records may introduce selection bias, although it was necessary to ensure temporal continuity and minimum data quality. Third, the prospective annual forecast should be interpreted as a scenario-based projection rather than as an externally validated prediction because future ground-truth observations were not available. Fourth, the explanatory variables are limited to basic temporal, climatic, and household-related features; future studies should incorporate additional socioeconomic, appliance-level, and high-resolution weather variables.

A full robustness analysis under progressive reductions of the training set was not reported because the available dataset contains only 1,000 hourly observations. Further reducing the training subset would substantially compromise model estimation and validation reliability. Future research should evaluate robustness using larger datasets and repeated experiments under 30%, 50%, 70%, and 100% training-data conditions, reporting mean and standard deviation across runs.

The limited explanatory variable set restricts the comprehensiveness of the model. Appliance-level usage, building-envelope characteristics, HVAC system information, and socioeconomic factors were not available. Consequently, the model should be interpreted as a short-term forecasting framework under constrained data availability rather than as a complete representation of all determinants of residential electricity demand.

The results may not generalize to other households, geographic regions, climatic periods, or socioeconomic contexts. The short observation period prevents the model from learning seasonal demand dynamics and may bias the results toward the specific behavioral and climatic conditions present during data collection.

Data Availability

The data and computational resources supporting the findings of this study are available at the repository indicated below. The repository includes the dataset, preprocessing procedure, model configuration, and experimental workflow required to reproduce the results: https://colab.research.google.com/drive/1DMcHPPgwfyOxLevHVwzNTJsfhybeF9is?usp=sharing

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Renzo Diony Ramirez-Huaracha contributed to the conceptualization of the study, data preprocessing, model implementation, experimental development, and initial drafting of the manuscript. Melany Meylin Quispe-Gutierrez contributed to data organization, methodological design, literature review, analysis of results, and manuscript revision. Benjamín Maraza-Quispe contributed to research supervision, methodological validation, interpretation of findings, critical revision of the manuscript, and overall coordination of the study. All authors reviewed and approved the final version of the manuscript.

References

[1] T. Ahmad and H. Chen, "Short and medium-term forecasting of cooling and heating load demand in building environment with data-mining based approaches," *Energy and Buildings*, vol. 166, pp. 460–476, 2018. doi: 10.1016/j.enbuild.2018.01.066

[2] D. Chen, H. Hu, N. Wang, and C.-P. Chang, "The impact of green finance on transformation to green energy: Evidence from industrial enterprises in China," *Technological Forecasting and Social Change*, vol. 204, 123411, 2024. doi: 10.1016/j.techfore.2024.123411
[3] X. Cui, M. Lee, C. Koo, and T. Hong, "Energy consumption prediction and household feature analysis for different residential building types using machine learning and SHAP: Toward energy-efficient buildings," *Energy and Buildings*, vol. 309, 113997, 2024. doi: 10.1016/j.enbuild.2024.113997
[4] K. B. Debnath and M. Mourshed, "Forecasting methods in energy planning models," *Renewable and Sustainable Energy Reviews*, vol. 88, pp. 297–325, 2018. doi: 10.1016/j.rser.2018.02.002
[5] F. Dinmohammadi, Y. Han, and M. Shafiee, "Predicting energy consumption in residential buildings using advanced machine learning algorithms," *Energies*, vol. 16, no. 9, 3748, 2023. doi: 10.3390/en16093748
[6] M. Fayaz and D. Kim, "A prediction methodology of energy consumption based on deep extreme learning machine and comparative analysis in residential buildings," *Electronics*, vol. 7, no. 10, 222, 2018. doi: 10.3390/electronics7100222
[7] M. B. Gorzałczany and F. Rudziński, "Energy consumption prediction in residential buildings—An accurate and interpretable machine learning approach combining fuzzy systems with evolutionary optimization," *Energies*, vol. 17, no. 13, 3242, 2024. doi: 10.3390/en17133242
[8] S. Chen, Y. Cai, and Y. Shen, "Regional electricity consumption forecasting with a cross-regional joint learning framework based on N-BEATS," *Electric Power Syst. Res.*, vol. 260, 113290, 2026.
[9] I. Moumen, N. Rafalia, and J. Abouchabaka, "A machine learning approach to residential energy prediction using large-scale datasets in MongoDB," in *Proc. 11th Int. Conf. Wireless Networks and Mobile Communications (WINCOM)*, 2024, pp. 1–6.
[10] R. Olu-Ajayi, H. Alaka, I. Sulaimon, F. Sunmola, and S. Ajayi, "Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques," *Journal of Building Engineering*, vol. 45, 103406, 2022. doi: 10.1016/j.jobe.2021.103406
[11] T. Wang, J. Wang, Y. Zhao, J. Shu, and J. Chen, "Multi-objective residential load dispatch based on comprehensive demand response potential and multi-dimensional user comfort," *Electric Power Systems Research*, vol. 220, 109331, 2023. doi: 10.1016/j.epsr.2023.109331
[12] B. Devanathan, K. J. Varshitha, L. P. Kumar, S. A. Lakshmanan, and N. K. Prakash, "Explainable AI framework using XGBoost with SHAP and LIME for multi-scale household energy forecasting," *IEEE Access*, vol. 13, pp. 149750–149764, 2025. doi: 10.1109/ACCESS.2025.3602673
[13] R. M. Machaca-Casani, L. A. Figueroa-Mayta, and J. Contreras-Nuñez, "Evaluation of the impact of machine learning on the prediction of residential energy consumption," *Electric Power Systems Research*, vol. 252, 112443, 2026. doi: 10.1016/j.epsr.2025.112443
[14] A. Neubauer, S. Brandt, and M. Kriegel, "Explainable multi-step heating load forecasting: Using SHAP values and temporal attention mechanisms for enhanced interpretability," *Energy and AI*, vol. 20, 100480, 2025. doi: 10.1016/j.egyai.2025.100480
[15] J. Moon *et al*., "Advancing ensemble learning techniques for residential building electricity consumption forecasting: Insight from explainable artificial intelligence," *PLOS ONE*, vol. 19, no. 11, e0307654, 2024. doi: 10.1371/journal.pone.0307654
[16] C. Cakiroglu *et al*., "Cooling load prediction of a double-story terrace house using ensemble learning techniques and genetic programming with SHAP approach," *Energy and Buildings*, vol. 313, 114254, 2024. doi: 10.1016/j.enbuild.2024.114254
[17] J. Moon, S. Park, S. Rho, and E. Hwang, "Robust building energy consumption forecasting using an online learning approach with Ranger," *Journal of Building Engineering*, vol. 47, 103851, 2022. doi: 10.1016/j.jobe.2021.103851
[18] M. Anastasiadou, V. D. dos Santos, and M. S. Dias, "An explainable deep learning model for energy performance classification and retrofitting recommendations," *Energy and Buildings*, vol. 349, 116522, 2025. doi: 10.1016/j.enbuild.2025.116522.
[19] W. Chen, F. Rong, and C. Lin, "Short-term building electricity load forecasting with a hybrid deep learning method," *Energy and Buildings*, vol. 330, 115342, 2025. doi: 10.1016/j.enbuild.2025.115342
[20] K. Jiang and D. L. Donaldson, "Transfer learning based short-term load forecasting at different aggregation levels," in *Proc. Int. Conf. Innovative Engineering Sciences and Technological Research (ICIESTR)*, 2024, pp. 1–5. doi: 10.1109/ICIESTR60916.2024.10798263
[21] C. Briggs, Z. Fan, and P. Andras, "Federated learning for short-term residential load forecasting," *IEEE Open Access Journal of Power and Energy*, vol. 9, pp. 573–583, 2022. doi: 10.1109/OAJPE.2022.3206220
[22] M. S. U. Haq *et al*., "Explainable deep learning combined attention-based LSTM for building energy prediction: A framework from the perspective of supply side," *Energy and Buildings*, vol. 350, 116638, 2026. doi: 10.1016/j.enbuild.2025.116638
[23] I. A. Kachalla and C. Ghiaus, "Electric water boiler energy prediction: State-of-the-art review of influencing factors, techniques, and future directions," *Energies*, vol. 17, no. 2, 443, 2024. doi: 10.3390/en17020443
[24] I. A. Kachalla *et al*., "Comparative analysis of machine learning models for energy consumption prediction," *Applied Thermal Engineering*, 2025. doi: 10.1016/j.applthermaleng.2025.125799
[25] I. A. Kachalla *et al*., "Hybrid SARIMAX-MLP framework for energy consumption prediction," *Results in Engineering*, 2025. doi: 10.1016/j.rineng.2025.105336
[26] A. Kansal *et al*., "Optimizing energy consumption in smart buildings through reinforcement learning," in *Proc. IEEE GESS*, 2024. doi: 10.1109/GESS63533.2024.10784974
[27] I. A. Kachalla, C. Ghiaus, and A. Ademuwagun, "Energy management in residential dwellings: Forecasting hot water demand with hybrid machine learning models," in *Proc. 2024 9th IEEE Workshop on the Electronic Grid (eGRID)*, Santa Fe, NM, USA, 2024, pp. 1–6. doi: 10.1109/eGRID62045.2024.10842758
[28] J.-Y. Kim and S.-B. Cho, "Electric energy demand forecasting with explainable time-series modeling," in *Proc. IEEE Int. Conf. Data Mining Workshops (ICDMW)*, 2020. doi: 10.1109/ICDMW51313.2020.00101
[29] M. Alfayan and H. Hagras, "Towards an explainable artificial intelligence approach for smart grid systems," *Discov. Artif. Intell.*, vol. 5, no. 1, 2025. doi.org/10.1007/s44163-025-00261-5
[30] J. F. Hair, J. J. Risher, M. Sarstedt, and C. M. Ringle, "When to use and how to report the results of PLS-SEM," *European Business Review*, vol. 31, pp. 2–24, 2019. doi: 10.1108/EBR-11-2018-0203