Vol. 2, No. 1, March 2025

IJAIS International Journal of Artificial Intelligence and Science

Publsihed By Asosiasi Doktor Sistem Informasi Indonesia

Personalized Energy Optimization in Smart Homes Using Adaptive Machine Learning Models: A Feature-Driven Approach

Oyeniran, M.[1]., Adekunle, J. D[2]., Sule, H.S[3]., Folorunso, O[4]., Alagbe, S.A[5]., Anifowoshe, T. J[6]., Robbert C, O[7]., Ebonyem, B. N[8]., Ideh, E. G[9] ., Oyelakin, S. O[10]., Ogu, C. K.[11]

[1,2,3,4,5,9]Federal University of Agriculture, Abeokuta, Ogun State, Nigeria

[6]Fisher of Men Technology Academy, Lagos, Nigeria

[7]Department of Managements Information Systems, Topdel Engineering Limited, Lagos, Nigeria

[8]Lagos State Co-operative College, Lagos, Nigeria

[10]Department of Mass Communication, Bayero University, Kano, Nigeria

[11]Medipolis GmbH, Otto-Schott-Straße, Jena, Germany

Email: oyeniranmatthew@gmail.com

Received: 5 January 2025

Revised: 17 February 2025

Accepted: 23 March 2025

Published: 30 March 2025

Corresponding Author:

Author Name*:

Oyeniran, M.

Email*:

oyeniranmatthew@gmail.com

DOI: 10.63158/IJAIS.v2i1.19

Abstract. The increase in demand for efficient energy smart homes has necessitates the personalized optimization strategies to have a reduction in energy consumption while maintaining user comfort. This research develops a Personalized Energy Optimization System using adaptive machine learning models to analyze household energy patterns and predict consumption in real time. Leveraging the Appliances Energy Prediction Dataset from the UCI repository, we applied supervised learning algorithms such as Gradient Boosting, XGBoost, CatBoost, LightGBM, and Random Forest to identify key factors influencing energy use, including occupancy patterns, appliance usage, and environmental conditions. Through feature engineering, normalization, and one-hot encoding, we enhanced model performance and interpretability. Among the evaluated models, LightGBM achieved the highest accuracy (R^2: 0.999573, RMSE: 0.013526), outperforming others in predicting energy consumption. The findings offer data-driven insights for dynamic energy management, optimizing household efficiency, and promoting sustainability.

Keywords: Personalized energy optimization, Smart homes, Machine learning, LightGBM, Energy consumption prediction, Regression analysis

IJAIS International Journal of Artificial Intelligence and Science
Publsihed By Asosiasi Doktor Sistem Informasi Indonesia

1. INTRODUCTION

The swift evolution of digital technologies has catalyzed the rise of smart homes, integrating a diverse array of Internet of Things (IoT) devices to automate, streamline, and personalize household management. These devices—ranging from smart thermostats [1] and advanced lighting systems [2] to sophisticated energy monitoring tools [3]—empower homeowners to oversee and optimize their energy consumption through mobile apps or voice-activated platforms, offering unprecedented convenience and control.

Recent projections indicate the global smart home market, valued at $84.50 billion in 2024, is expected to surge to $116.4 billion by 2029 [4]. This anticipated growth is largely driven by the escalating demand for energy-efficient (EE) technologies and the exponential spread of connected devices. As homeowners increasingly prioritize comfort, security, and sustainability, energy efficiency is no longer a luxury—it has become a foundational aspect of modern smart living.

The advantages of EE technologies in smart homes are well documented. These systems contribute to reduced operational costs, alleviate stress on power grids, enhance indoor air quality, and significantly lower carbon emissions [5]–[8]. From a holistic perspective, they not only support environmental sustainability but also deliver tangible financial and health benefits.

However, several challenges persist in achieving optimal energy efficiency. Many households continue to lack awareness of their consumption patterns, which leads to energy waste and higher utility bills [5], [9], [10]. More critically, many existing energy management solutions are built on generalized models that fail to reflect individual user preferences, behaviors, or unique household dynamics [11]. In addition, the integration of various smart technologies—each operating under different standards and protocols—can introduce complexities that hinder cohesive energy optimization strategies [12].

To confront these issues, this study proposes a novel, personalized energy optimization system driven by adaptive machine learning models. By analyzing granular energy usage data and tailoring predictions to individual household behavior, the system aims to

significantly improve energy efficiency, reduce consumption, and enhance user satisfaction in smart home environments.

2. LITERATURE REVIEW

As smart home technologies continue to evolve, efficient energy management has become an essential objective. The integration of real-time data from Internet of Things (IoT) devices with predictive modelling capabilities has allowed for smarter, more adaptive energy usage strategies in modern households. Machine learning (ML) approaches—including supervised, unsupervised, deep learning, and reinforcement learning—have emerged as promising solutions for managing the complexity and variability of residential energy consumption in real-world conditions.

2.1. Regression and Predictive Algorithms in Energy Management

Regression-based approaches have long played a pivotal role in forecasting energy consumption due to their interpretability and effectiveness. These models analyze historical data to identify patterns, enabling energy management systems (EMS) to schedule high-energy tasks during off-peak hours and reduce peak load strain. For instance, Fischer demonstrated the value of decision trees in this context, achieving a Root Mean Square Error (RMSE) of 0.25 on normalized datasets [13]. Decision trees are particularly useful for their capacity to model nonlinear relationships, making them ideal for capturing the interplay of occupancy, weather, and appliance usage variables [13], [14], [15].

Furthermore, Model Predictive Control (MPC) techniques have been successfully employed to dynamically optimize Heating, Ventilation, and Air Conditioning (HVAC) systems. Taheri et al. applied MPC algorithms that adjusted HVAC settings based on real-time environmental inputs like temperature and humidity, resulting in a 15% reduction in energy consumption [16]–[18]. While linear regression remains adequate for scenarios with straightforward consumption trends, nonlinear techniques such as polynomial regression and decision trees are better suited for complex residential environments. Arshad et al. emphasized that nonlinear models, including decision trees and random forests, outperform simpler linear methods in multivariable contexts where interactions are intricate and dependencies are nonlinear [19].

2.2. Ensemble Learning Techniques for Enhanced Prediction

Ensemble learning enhances prediction reliability by combining multiple models to exploit their individual strengths. These approaches are particularly effective in managing the heterogeneous and nonlinear nature of household energy data. Techniques like Random Forest, when paired with feature selection algorithms, have significantly improved energy forecasting accuracy. Gupta et al. demonstrated this by applying recursive feature elimination to isolate high-impact variables—such as appliance usage frequency, weekdays, and outdoor temperature—yielding a 10% increase in accuracy [20], [21].

Similarly, hybrid approaches integrating Artificial Neural Networks (ANNs) with Random Forest models have shown strong performance in balancing grid reliance with renewable energy inputs. Shafique et al. modeled complex interactions between solar energy production, user demand, and battery storage, facilitating more sustainable energy use [22]–[24]. These ensemble techniques are critical for optimizing smart home systems in environments where dynamic energy allocation is necessary to balance environmental sustainability with user comfort.

2.3. Reinforcement Learning for Adaptive Energy Control

Reinforcement Learning (RL) has emerged as a transformative technique for real-time energy optimization, particularly in smart homes where consumption patterns are highly variable. RL algorithms adjust their behavior based on environmental feedback, learning optimal strategies over time. In HVAC systems, Huang et al. demonstrated that RL could reduce energy consumption by 20% while maintaining user comfort through adaptive learning of temperature preferences [25], [26].

Baker et al. extended this approach by integrating RL with supervised classification models like Support Vector Machines (SVMs) and K-Nearest Neighbors (KNN). Their hybrid framework first categorized energy usage patterns, enabling the RL model to tailor its strategies to specific household profiles and appliance interactions [27]–[29]. This fusion of classification and reinforcement techniques offers a powerful method for personalizing energy control in complex, multi-device environments, enhancing both energy savings and responsiveness.

IJAIS International Journal of Artificial Intelligence and Science
Published By Asosiasi Doktor Sistem Informasi Indonesia

2.4. Feature Engineering and Data Integration

Feature engineering is a foundational aspect of improving model accuracy in energy forecasting. Including domain-specific attributes such as temporal markers (e.g., hour of the day, day of the week, and seasonal cycles) has been shown to significantly enhance prediction performance. Li et al. observed a 30% improvement in accuracy when incorporating time-based features into their models [14], [30]. Han and Qiu similarly noted that energy usage spikes and efficiency windows are closely tied to daily activity cycles, which can be captured through temporal feature inclusion [30].

The integration of IoT sensor data—covering parameters like temperature, lighting, and occupancy—provides granular insights that improve real-time energy modelling. Tabassum et al. employed IoT-derived data with Support Vector Regression (SVR) and ANNs, resulting in high alignment between predicted and actual consumption patterns. This responsiveness allows EMS to adapt dynamically to changing conditions, user behaviour, and environmental variables, fostering smarter energy usage [31].

Table 1. Overview of Existing Studies on Smart Energy

Author(s)	Model Used	Outcome Achieved	Metrics
Fischer [13]	Decision Trees	Optimized energy scheduling; reduced peak demand	RMSE = 0.25
Luthra et al. [17]	Model Predictive Control (MPC)	Improved HVAC efficiency; 15% reduction in energy	Accuracy, Cost Savings
Huang et al. [25]	Reinforcement Learning	20% energy efficiency improvement in HVAC	Energy Consumption Reduction
Shafique et al. [22]	ANN + Random Forest	Integrated renewable sources; minimized grid reliance	Enhanced Efficiency
Gupta et al. [20]	Random Forest & Feature Selection	10% accuracy improvement with relevant features	Accuracy, Feature Impact
Baker et al. [27]	SVM, KNN + Reinforcement Learning	Enhanced prediction accuracy with hybrid approach	Classification Accuracy

In synthesizing the advancements discussed across regression models, ensemble learning, reinforcement learning, and feature engineering, a key insight emerges: while each technique contributes significantly to energy optimization in smart homes, their isolated implementation often fails to capture the personalized nuances of household energy behavior. A truly adaptive energy optimization system requires a cohesive framework that merges these approaches. This necessitates further research into developing integrated machine learning architectures that are not only scalable and efficient but also deeply personalized to user-specific patterns—thereby maximizing energy savings, comfort, and environmental impact simultaneously.

3. METHODS

This study focused on developing a personalized energy optimization system in smart homes using supervised machine learning techniques. The adopted methods involve series of steps, starting from data collection to the implementation and evaluation of the machine learning models used.

3.1. Data Collection and Sources

The dataset used in this research is the Appliances Energy Prediction Dataset from the UCI Machine Learning Repository, which contains data from smart homes that monitor energy consumption over time. The dataset contains 19,735 instances recorded at 10-minute intervals over approximately four and a half months. It includes multiple features related to indoor and outdoor environmental conditions, such as temperature, humidity, and atmospheric pressure, as well as appliance-level energy consumption data. This dataset is well-suited for energy optimization research due to its granular time-stamped data and multiple influencing factors (Table 2).

Table 2. Feature Description Table for the Appliance Energy Prediction Dataset

Feature	Description	Quantitative / Qualitative	Missing Values
Date and Time	The timestamp of each energy reading	Quantitative	No
Appliances	Total energy consumption (watts)	Quantitative	No
Lights	Energy consumption for lighting (watts)	Quantitative	No

Feature	Description	Quantitative / Qualitative	Missing Values
T1-T9	Indoor temperature readings from various rooms (°C)	Quantitative	No
RH_1-RH_9	Indoor relative humidity readings from various rooms (%)	Quantitative	No
T_out	Outdoor temperature (°C)	Quantitative	No
Press_mm_hg	Atmospheric pressure (mmHg)	Quantitative	No
RH_out	Outdoor relative humidity (%)	Quantitative	No
Windspeed	Wind speed (m/s)	Quantitative	No
Visibility	Outdoor visibility (meters)	Quantitative	No
Tdewpoint	Dew point temperature (°C)	Quantitative	No
rv1, rv2	Random variables (unmodeled factors or noise)	Quantitative	No

3.2. Data Preprocessing

Data preprocessing plays a pivotal role in the development of reliable and accurate machine learning models, especially within the energy optimization domain of smart homes, where data integrity directly affects model effectiveness. Machine learning algorithms, particularly those based on tree and gradient methods, are sensitive to the scale and structure of input features. As such, appropriate preprocessing steps such as normalization and standardization are essential [32]. In energy datasets, parameters like temperature, humidity, wind speed, and energy usage are measured across different scales, potentially introducing model bias. Normalization (scaling values between 0 and 1) or standardization (rescaling to zero mean and unit variance) addresses this issue effectively [14]. In this study, key features such as "Appliances", "Lights", and indoor temperature readings were standardized to ensure consistent input ranges and accelerate model convergence during training.

Beyond numerical scaling, categorical data such as "Day of the Week" was converted into machine-readable formats using one-hot encoding. This transformation is crucial for algorithms like Random Forest and Gradient Boosting that are unable to interpret raw categorical inputs [33], [34]. Additionally, encoding temporal variables (e.g., hours, weekdays, seasons) enhances the model's sensitivity to daily and seasonal consumption cycles, enabling better performance in predicting time-dependent energy patterns [35]. To evaluate the integrity of the dataset, an outlier analysis was conducted. As shown in

IJAIS International Journal of Artificial Intelligence and Science
Published By Asosiasi Doktor Sistem Informasi Indonesia

Table 3, although certain features exhibited a non-negligible number of outliers (e.g., 10.83% for Appliances and 22.72% for Lights), their influence on the overall mean values was minimal. Consequently, these values were retained, as their exclusion would remove useful variance from the dataset that could benefit model learning.

Table 3. Outlier Detection Summary for Appliances Energy Prediction

Feature	Outliers Count	Outlier Ratio	Outliers Mean	With Mean	Without Mean
Appliances	2138	0.1083	348.61	97.69	67.21
Lights	4483	0.2272	16.74	3.80	0.00
Indoor Temp.	515	0.0261	21.67	21.69	21.69
Relative Humidity	146	0.0074	49.82	40.26	40.19
T_out	436	0.0221	22.51	7.41	7.07
Press_mm_hg	219	0.0111	733.17	755.52	755.77
RH_out	239	0.0121	32.90	79.75	80.32
Windspeed	214	0.0108	11.63	4.04	3.96
Visibility	2522	0.1278	58.18	38.33	35.42
Tdewpoint	10	0.0005	15.24	3.77	3.75
rv1, rv2	0	0.0000	-	24.99	24.99

3.3. Tools, Equipment, Software, and Library or Package

To train the algorithm and to facilitate data collection, cleaning, analysis, and visualization. This study relied on Jupyter Notebook, a popular integrated development environments (IDEs) and data analytics platforms. This environment not only streamlined the coding and testing processes but also facilitated impactful data visualization, which was critical for illustrating our findings. We leveraged a variety of Python libraries, each serving a specialized role in the project such as; Pandas, NumPy, Scikit-learn, Matplotlib and Seaborn. These tools offer enhanced aesthetics and built-in functions for training a complex algorithm, creative static and interactive visualizations.

3.4. The Model

This energy prediction research evaluated several regression models to find the best fit for accurately forecasting energy consumption in smart homes. These models include Gradient Boosting, XGBoost, CatBoost, and LightGBM.

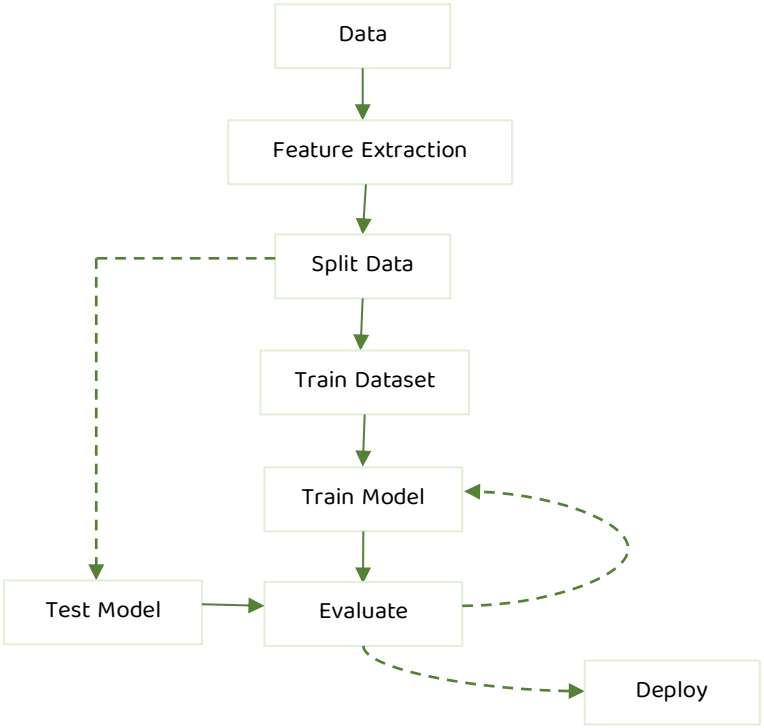

Figure 1. Proposed Model

Gradient Boosting is said to be a powerful model for capturing complex relationships but is known to be computationally intensive [36]. This algorithm seeks to minimize a loss of function $L\left(y,\underline{y}\right)$ by adding new models $f_m(x)$ that point in the direction of the negative gradient of the loss function (Equation. 1).

$$y_m = y_{m-1} - \alpha \nabla_{\hat{y}} L(y, y_{m-1}) \quad (1)$$

where α is the learning rate, y_m is the updated prediction, and $\nabla_y L$ is the gradient of the loss to predictions.

On the other hand, XGBoost (Extreme Gradient Boosting) is known to be an optimized version of Gradient Boosting, designed to be faster and more efficient [34]. It uses advanced regularization (L1 and L2) techniques to prevent overfitting and incorporates parallel processing and other optimizations, making it one of the preferred choices for structured data problems (Equation. 2). XGBoost adds a regularization term to the loss function to penalize model complexity and uses a weighted quantile sketch to handle sparsity in data.

$$L(\theta) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \quad (2)$$

Where $l(y_i, \hat{y}_i)$ is the loss function and $\Omega(f_k)$ represent the regularization function of the k-th features.

Similarly, CatBoost (Categorical Boosting) is a gradient-boosting algorithm specifically designed to handle categorical features more effectively (Equation 3). Incorporating ordered boosting mitigates the effect of target leakage, improving the accuracy and generalization of unseen data [35]. It also minimizes a loss function through gradient boosting, when uniquely encoding categorical features through combinations that optimize the categorical splits. where the categorical features are processed using the ordered boosting approach.

$$L=\sum_{i=1}^{n}\left(y_i - \sum_{k=1}^{K} f_k(x_i)\right)^2 \quad (3)$$

Where $f_k(x_i)$ represent the prediction from the k-th base learner and K is the total number of base learners combined in the model.

The LightGBM (Light Gradient Boosting Machine) is an efficient gradient-boosting framework that uses histogram-based algorithms and a leaf-wise tree growth approach [36]. It can handle large datasets with minimal memory usage and performs better than other boosting models in speed and accuracy, especially for high-dimensional data. LightGBM's leaf-wise growth strategy involves splitting the leaf with the largest reduction in loss, which can result in faster convergence (Equation 4).

$$L=\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2 +\alpha\left|\left|w\right|\right|_1 + \beta\left|\left|w\right|\right|_2^2 \quad (4)$$

where α and β are regularization terms for L1 and L2 penalties.

3.5. Performance Metrics

For this research, the model's performance was evaluated using five performance indicators such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R^2), which offer valuable insights into how well the model captures trends, manages errors, and explains the variability in the data. The mean squared error (MSE) in

the context of energy prediction helps to capture how well the model performs in minimizing large errors, as it penalizes larger deviations between predicted and actual values more heavily. This indicator is mathematically represented as shown in Equation 5.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 \quad (5)$$

Also, the root mean squared error (RMSE) provides a measure of prediction accuracy in the same units of energy usage (e.g., kilowatts or kilowatt-hours). as the target variable, helping ensure that the model's forecasted energy consumption aligns closely with actual usage patterns in smart homes– mathematically represented as shown in Equation 6.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \quad (6)$$

R-square (R^2), or the coefficient of determination, explains the proportion of variance in the energy consumption data that is captured by the model's predictors. It is calculated as shown in Equation 7.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \underline{y}_i\right)^2} \quad (7)$$

In addition to the performance indication, we use the Shapiro-Wilk test to assess whether the residuals from the model's predictions follow a normal distribution, which is a common assumption in predictive modeling [37]. The test is calculated as shown in Equation 8.

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \quad (8)$$

Where x_i are ordered residual sample values and a_i are constants.

4. RESULTS AND DISCUSSION

4.1. Performance Evaluation

The correlation result presented in Figure 2 indicates that there is a strong positive (0.79) relationship between the Outdoor Temperature and Tdewpoint. Similarly, Indoor Temperature also shows a moderate positive correlation with Outdoor Temperature

IJAIS International Journal of Artificial Intelligence and Science
Published By Asosiasi Doktor Sistem Informasi Indonesia

(0.68) and Tdewpoint (0.57), suggesting that outdoor conditions influence indoor temperatures. The relative humidity has a moderate positive correlation with Tdewpoint (0.64) and outdoor temperature (0.34), reflecting that warmer temperatures increase humidity levels. However, outdoor temperature shows a moderate negative correlation with RH_out (-0.57), indicating that higher outdoor temperatures are associated with lower outdoor humidity. Pressure has weak negative correlations with outdoor temperature (-0.14) and relative humidity (-0.29), suggesting that higher pressure is linked to cooler and drier conditions. Lights and Visibility show weak correlations with most variables, though Lights have a slight positive correlation with windspeed (0.06) and RH_out (0.06).

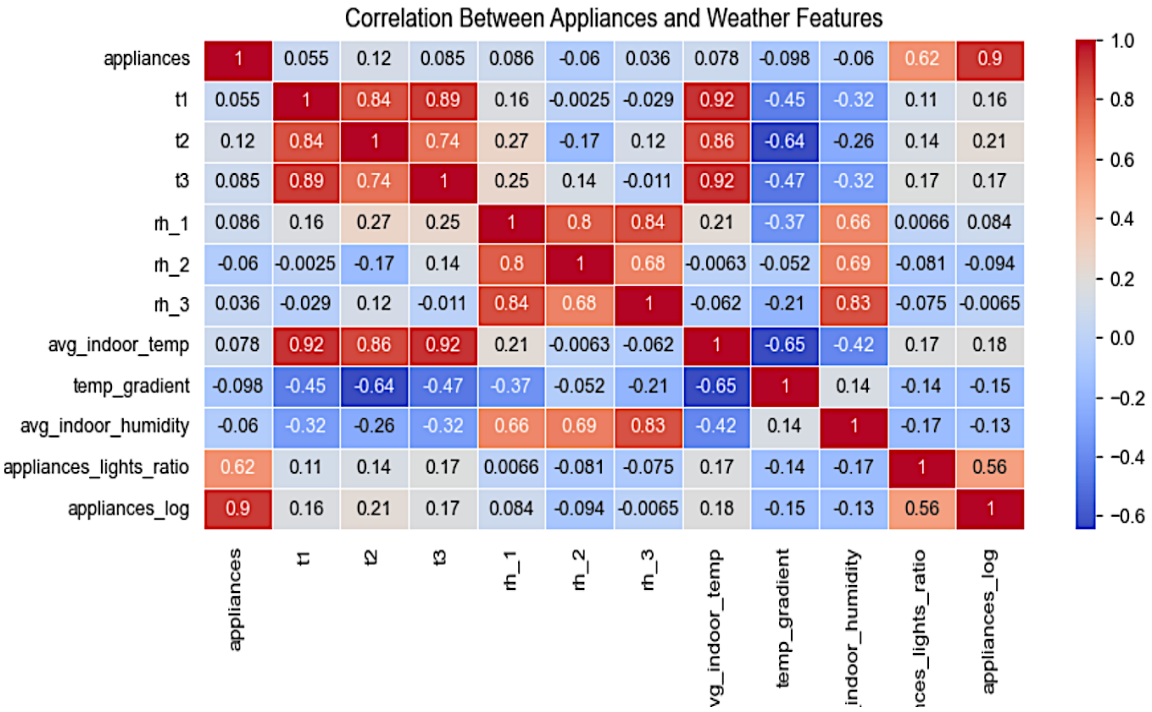

Figure 2. Relationship between the different energy consumption variables

The hourly energy consumption (figure 3) indicates a peak consumption period in the morning and evening. Conversely, energy demand drops during midday and late at night.

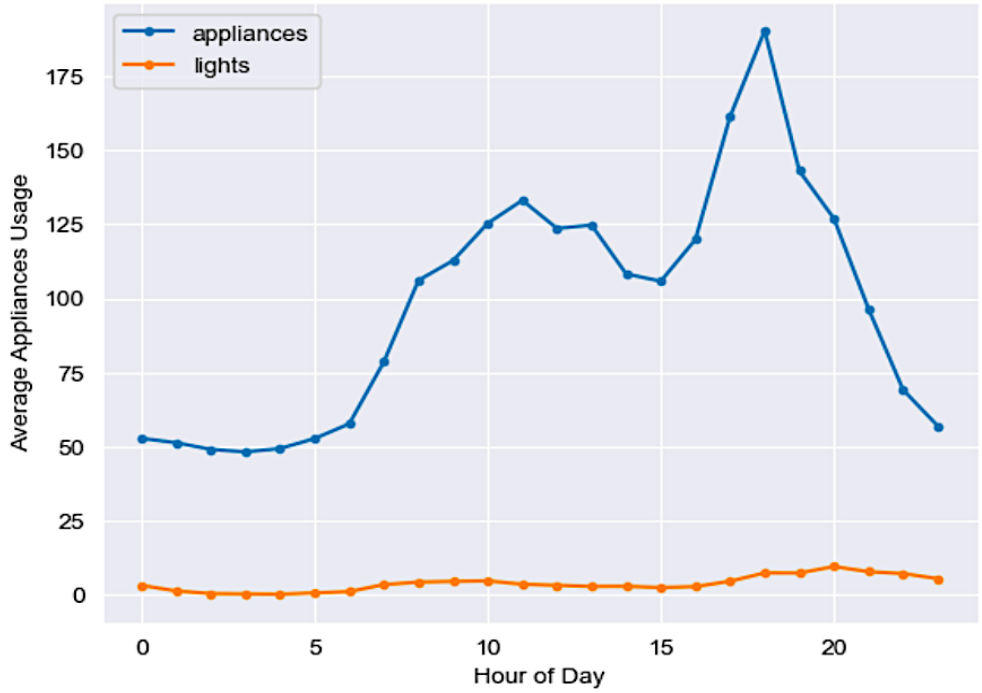

Figure 3. The energy consumption trends over the day and night

The daily energy consumption for appliances and lights depict variations in energy usage across the week (Figure 4). Monday and Saturday show the highest consumption, while energy usage stabilizes midweek. Additionally, the contribution of lighting remains relatively low compared to appliances.

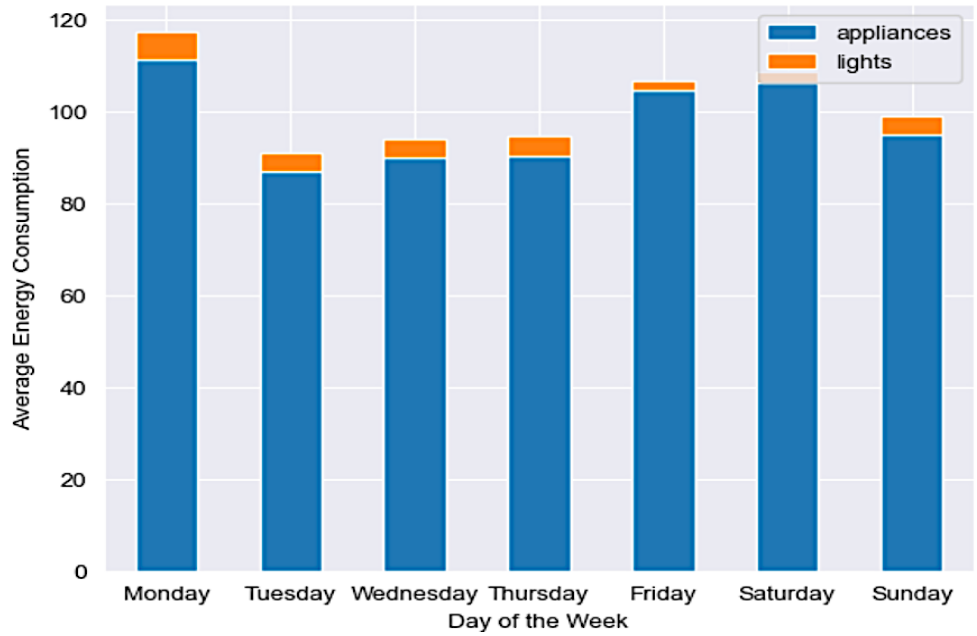

Figure 4. Energy consumption over the days within the week

IJAIS International Journal of Artificial Intelligence and Science
Published By Asosiasi Doktor Sistem Informasi Indonesia

The descriptive analysis of energy consumption by appliances reveals considerable variability. The average usage is 97.69 watts (median: 60 watts; std: 102.52, IQR: 10 – 1080) indicates a wide range of consumption levels. In contrast, lighting follows a different pattern, with a much lower average consumption of 3.80 watts (Median: 0.00; std: 7.94). Other environmental factors, such as indoor temperature and relative humidity, also show notable trends. The average indoor temperature is 21.67°C (std: 1.61°C), indicating a controlled environment with minor fluctuations. Relative humidity averages 40.26% (IQR: 27.02% - 63.36%), demonstrating moderate variability. Outdoor temperature (T_out) displays greater dispersion, with a mean of 7.41°C (std: 5.32°C; Min: -5.00°C; Max: 26.00°C), reflecting seasonal and daily fluctuations.

Table 4. Descriptive Statistics of Features

Feature	Count	Mean	Median	STD	IQR (Max-Min)
Appliances	19735	97.69	60.00	102.52	10.00 – 1080.00
Lights	19735	3.80	0.00	7.94	0.00 - 70.00
Indoor Temp.	19735	21.67	21.60	1.61	16.79 – 26.26
Relative Humidity	19735	40.26	39.66	3.98	27.02 – 63.36
T_out	19735	7.41	6.92	5.32	-5.00 – 26.00
Press_mm_hg	19735	755.52	756.10	7.40	729.30 – 772.30
RH_out	19735	79.75	83.67	14.90	24.00 – 100.00
Windspeed	19735	4.04	3.67	2.45	0.00 – 14.00
Visibility	19735	38.33	40.00	11.79	1.00 – 66.00
Tdewpoint	19735	3.76	3.43	4.19	-6.60 – 15.50
rv1, rv2	19735	24.99	24.90	14.50	0.01 – 50.00

The normality test (using the Shapiro-Wilk statistic) results indicate that all variables show significant deviations from normality ($p < 0.05$).

Table 5. Shapiro-Wilk Test Results for Features

Feature	W Statistic	P – value	Decision
Appliances	0.576415	1.839371e-111	Significant
Lights	0.542653	1.658712e-113	Significant
Indoor Temp.	0.991404	2.147862e-32	Significant
Relative Humidity	0.983720	5.157294e-42	Significant

IJAIS International Journal of Artificial Intelligence and Science
Publsihed By Asosiasi Doktor Sistem Informasi Indonesia

Feature	W Statistic	P – value	Decision
T_out	0.982098	1.478741e-43	Significant
Press_mm_hg	0.986587	5.865653e-39	Significant
RH_out	0.92095	1.166351e-71	Significant
Windspeed	0.929943	6.690935e-69	Significant
Visibility	0.922067	3.93784e-71	Significant
Tdewpoint	0.992333	8.253211e-31	Significant
rv1, rv2	0.954011	1.909266e-60	Significant

The predictive modeling results demonstrate the effectiveness of different machine learning algorithms in estimating appliance energy consumption. Among the models tested, LightGBM(MSE: 0.000183, RMSE: 0.013526, R^2: 0.999573) achieved the highest accuracy, making it the most reliable model. Also, random forest (MSE: 0.006176, R^2: 0.985577), followed closely by XGBoost (MSE: 0.006872, R^2: 0.983951) also performed well. Gradient Boosting (MSE: 0.008214, R^2: 0.980818) exhibited slightly lower performance, with an, while AdaBoost (MSE: 0.039015, R^2: 0.908890) lagged.

Table 6. Model Performance Metrics

Model	MSE	RMSE	R2 Score
Random Forest	0.006167	0.078590	0.985577
Gradient Boosting	0.008214	0.090633	0.980818
AdaBoost	0.039015	0.197523	0.908890
XGBoost	0.006872	0.082900	0.983951
LightGBM	0.000183	0.013526	0.999573

The KDE plot provides a smooth estimate of the distribution of residuals for different models. Ideally, residuals should be symmetrically distributed around zero, indicating that the models do not have systematic bias. In the KDE plot, it is observed that LightGBM exhibits a sharp peak at zero, suggesting it has the smallest residual variance. Other models, such as Random Forest and XGBoost, also show residuals centered around zero but with broader distributions, indicating more variability in prediction errors.

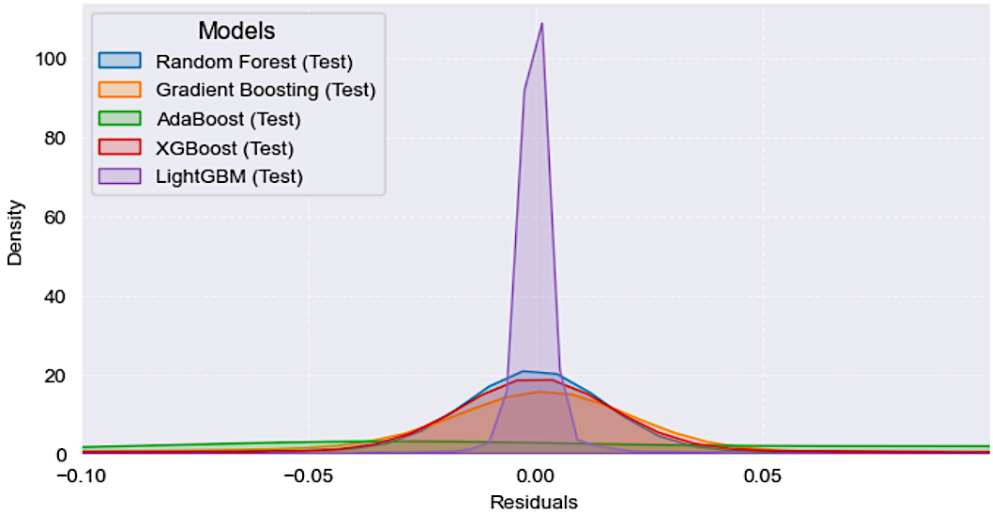

Figure 5. KDE Plot of Test Residuals

The boxplot representation provides additional insights into the spread and presence of outliers in the residuals. AdaBoost appears to have a wider spread of residuals compared to other models, as indicated by the interquartile range (IQR). LightGBM, on the other hand, shows a more concentrated distribution with fewer extreme outliers.

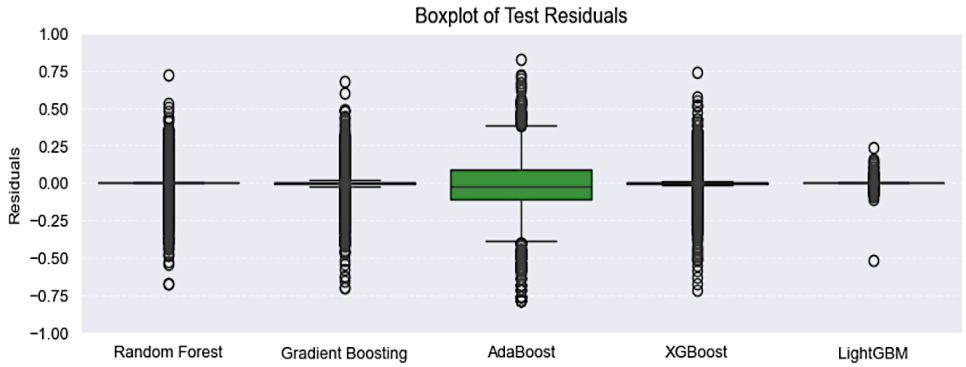

Figure 6. Outliers of Test Residuals

4.2. Discussion

The results of this study demonstrate the effectiveness of machine learning models—particularly LightGBM—in forecasting energy consumption in smart homes with high precision. Among all models evaluated, LightGBM achieved the best performance metrics (MSE: 0.000183, RMSE: 0.013526, R^2: 0.999573), indicating exceptional accuracy and minimal prediction error. This superior performance underscores its suitability for real-time

energy management applications, where precise load forecasting is critical to optimizing energy use and minimizing costs.

These findings align with established literature emphasizing the value of ensemble learning techniques in energy optimization. For example, Dalal et al. [38] proposed a hybrid framework combining transfer learning and LightGBM, which achieved a Mean Absolute Percentage Error (MAPE) of 2.83 and a GINI index of 0.97. Similar to the methodology adopted in this research, their approach emphasized the importance of feature extraction and preprocessing in enhancing model performance—validating the applicability of LightGBM to dynamic and high-resolution smart home datasets.

The descriptive statistical analysis (Table 4) and correlation results (Figure 2) highlight significant relationships among environmental features. Notably, outdoor temperature and dew point exhibited a strong positive correlation (0.79), while indoor temperature also showed moderate positive relationships with outdoor conditions (0.68 with outdoor temperature, 0.57 with dew point). These environmental dependencies suggest that indoor energy demand is significantly influenced by external weather conditions, reinforcing the need for context-aware predictive systems like LightGBM.

Hourly and weekly energy usage patterns (Figures 3 and 4) further emphasize the temporal variation in energy consumption, with distinct peaks during mornings and evenings. These patterns validate the importance of time-series features such as hour-of-day and day-of-week in improving forecasting accuracy. The KDE and residual boxplots (Figures 5 and 6) revealed that LightGBM not only achieved minimal prediction error but also had the most symmetrical and narrowly distributed residuals, indicating stable performance with minimal bias or outliers.

In the broader context, integrating such predictive models with smart automation systems holds transformative potential. As shown in recent studies, smart thermostats and lighting controls can yield energy savings of up to 20%, especially when coupled with adaptive forecasting tools. When integrated with models like LightGBM, these systems can dynamically adjust device settings—like temperature, lighting, and appliance schedules—based on real-time predictions and environmental feedback, thereby maximizing both comfort and energy efficiency.

IJAIS International Journal of Artificial Intelligence and Science
Publsihed By Asosiasi Doktor Sistem Informasi Indonesia

Moreover, AI-driven platforms like BrainBox AI's ARIA system exemplify the commercial viability of such approaches. Their automated HVAC optimization platform, which monitors variables like humidity and ventilation in real time, reported energy savings of up to 25% alongside reductions in carbon emissions. This reinforces the significance of combining advanced machine learning algorithms with real-time control systems for sustainable, data-driven smart home energy optimization.

Overall, the high accuracy and robustness of the LightGBM model, combined with its compatibility with high-dimensional and categorical data, make it a compelling choice for developing next-generation energy management solutions. Future work should focus on integrating user-specific behavioral data, renewable energy sources, and cost-based constraints to build fully personalized and economically viable energy ecosystems for smart homes.

5. CONCLUSION

This study's comparative analysis of several machine learning algorithms—namely Random Forest, Gradient Boosting, AdaBoost, XGBoost, and LightGBM—provides a comprehensive understanding of their predictive capabilities and residual behaviors in the context of smart home energy consumption forecasting. While all models demonstrated strong overall performance in capturing patterns within the dataset, residual diagnostics, particularly through Q-Q plots, revealed notable deviations from normality across all models. These deviations, particularly in the distribution tails, suggest the presence of underlying limitations such as model bias, sensitivity to outliers, or potential overfitting. Although XGBoost and Gradient Boosting models exhibited relatively better conformity to normality within their central distribution ranges, LightGBM and Random Forest models displayed more pronounced departures—especially in the tails—indicating variance instability and skewness in prediction errors. Despite LightGBM's exceptional statistical accuracy (R^2: 0.999573), its residual pattern implies that further refinement may be necessary to ensure consistent generalization across diverse household settings. These findings underscore the importance of complementing high-performance metrics with robust error analysis for more trustworthy model deployment.

IJAIS International Journal of Artificial Intelligence and Science
Publsihed By Asosiasi Doktor Sistem Informasi Indonesia

To improve residual normality and enhance predictive stability, future implementations should consider applying data transformations such as logarithmic or Box-Cox scaling. These techniques can help stabilize variance and improve the assumptions underlying model predictions. Additionally, more extensive hyperparameter optimization and the incorporation of regularization strategies may mitigate overfitting risks. Exploring alternative loss functions tailored to specific energy forecasting objectives can also contribute to greater model resilience and accuracy. Ultimately, while the current models offer strong foundations for smart home energy optimization, continuous refinement through advanced preprocessing, algorithmic tuning, and real-world integration will be key to achieving more personalized, adaptive, and energy-efficient solutions.

REFERENCES

[1] J. Lu, T. Sookoor, V. Srinivasan, G. Gao, B. Holben, J. Stankovic, E. Field, and K. Whitehouse, "The smart thermostat," *SenSys 2010 - Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems*, pp. 211–224, 2010.

[2] M. Soheilian, G. Fischl, and M. Aries, "Smart lighting application for energy saving and user well-being in the residential environment," *Sustainability*, vol. 13, no. 11, p. 6198, 2021.

[3] H. R. Khan, M. Kazmi, Lubaba, M. H. B. Khalid, U. Alam, K. Arshad, K. Assaleh, and S. A. Qazi, "A low-cost energy monitoring system with universal compatibility and real-time visualization for enhanced accessibility and power savings," *Sustainability*, vol. 16, no. 10, p. 4137, 2024.

[4] Markets and Markets, "Smart Home Market Report," 2024.

[5] R. Bray, R. Ford, M. Morris, J. Hardy, and L. Gooding, "The co-benefits and risks of smart local energy systems: A systematic review," *Energy Research & Social Science*, vol. 115, p. 103608, 2024.

[6] A. Oshilalu, M. Kolawole, and O. Taiwo, "Innovative solar energy integration for efficient grid electricity management and advanced electronics applications," *International Journal of Science and Research Archive*, vol. 13, pp. 2931–2950, 2024.

[7] B. Lin and Z. Li, "Is more use of electricity leading to less carbon emission growth? An analysis with a panel threshold model," *Energy Policy*, vol. 137, p. 111121, 2020.

[8] H. Altın, "The impact of energy efficiency and renewable energy consumption on carbon emissions in G7 countries," *International Journal of Sustainable Engineering*, vol. 17, no. 1, pp. 134–142, 2024.

[9] T. Olatunde, A. Okwandu, and D. Akande, "Reviewing the impact of energy-efficient appliances on household consumption," *International Journal of Science and Technology Research Archive*, vol. 6, pp. 001–011, 2024.

[10] A. N. A. Baidoo, J. A. Danquah, E. K. Nunoo, A. Opoku, A. Asiedu, and B. Owusu, "Households' energy conservation and efficiency awareness practices in the Cape Coast Metropolis of Ghana," *Discover Sustainability*, vol. 5, no. 2, 2024.

[11] M. Sinha, E. Chacko, P. Makhija, and S. Pramanik, "Energy-Efficient smart cities with green Internet of things," in *Green Technological Innovation for Sustainable Smart Societies: Post Pandemic Era*, pp. 345–361, 2021.

[12] X. Zhou, J. Xiong, T. Hong, D. Zhao, and Y. Zhang, "MATNilm: Multi-appliance-task non-intrusive load monitoring with limited labeled data," *IEEE Transactions on Industrial Informatics*, vol. 20, no. 3, pp. 3177–3187, 2022.

[13] D. Fischer, "Forecasting energy with decision trees," 2018.

[14] L. Li, A. J. Blomberg, J. Lawrence, W. J. Réquia, Y. Wei, M. Liu, and P. Koutrakis, "A spatiotemporal ensemble model to predict gross beta particulate radioactivity across the contiguous United States," *Environment International*, vol. 156, p. 106643, 2021.

[15] Y. Himeur, M. Elnour, F. Fadli, N. Meskin, I. Petri, Y. Rezgui, and A. Amira, "Next-generation energy systems for sustainable smart cities: Roles of transfer learning," *Sustainable Cities and Society*, vol. 85, p. 104059, 2022.

[16] S. Taheri, P. Hosseini, and A. Razban, "Model predictive control of heating, ventilation, and air conditioning (HVAC) systems: A state-of-the-art review," *Journal of Building Engineering*, vol. 60, p. 105067, 2022.

[17] R. Godina, E. M. Rodrigues, E. Pouresmaeil, J. C. Matias, and J. P. Catalão, "Model predictive control home energy management and optimization strategy with demand response," *Applied Sciences*, vol. 8, no. 3, p. 408, 2018.

[18] V. A. Freire, L. V. R. De Arruda, C. Bordons, and J. J. Márquez, "Optimal demand response management of a residential microgrid using model predictive control," *IEEE Access*, vol. 8, pp. 228264–228276, 2020.

[19] M. A. Arshad, S. Shahriar, and K. Anjum, "The power of simplicity: Why simple linear models outperform complex machine learning techniques—case of breast cancer diagnosis," *arXiv preprint arXiv:2306.02449*, 2023.

[20] R. Vishraj, S. Gupta, and S. Singh, "Evaluation of feature selection methods utilizing random forest and logistic regression for lung tissue categorization using HRCT images," *Expert Systems*, vol. 40, no. 8, p. e13320, 2023.

[21] X. Wen, Y. Xie, L. Jiang, Y. Li, and T. Ge, "On the interpretability of machine learning methods in crash frequency modeling and crash modification factor development," *Accident Analysis & Prevention*, vol. 168, p. 106617, 2022.

[22] M. Shafique, Y. Gong, H. Zhao, C. Han, L. Jing, and P. Yang, "A hybrid deep reinforcement learning ensemble optimization model for heat load energy-saving prediction," *Journal of Building Engineering*, vol. 58, p. 105031, 2022.

[23] J. Sun, M. Gong, Y. Zhao, C. Han, L. Jing, and P. Yang, "A hybrid deep reinforcement learning ensemble optimization model for heat load energy-saving prediction," *Journal of Building Engineering*, vol. 58, p. 105031, 2022.

[24] A. B. Çolak, A. Shafiq, and T. N. Sindhu, "Modeling of Darcy–Forchheimer bioconvective Powell Eyring nanofluid with artificial neural network," *Chinese Journal of Physics*, vol. 77, pp. 2435–2453, 2022.

[25] L. Y. Huang, L. Y. Yao, and J. C. Teo, "HVAC control based on reinforcement learning and fuzzy reasoning," *Journal of Building Engineering*, 2025.

[26] X. Zhang, X. Wang, H. Zhang, Y. Ma, S. Chen, C. Wang, and X. Xiao, "Hybrid model-free control based on deep reinforcement learning: An energy-efficient operation strategy for HVAC systems," *Journal of Building Engineering*, vol. 96, p. 110410, 2024.

[27] R. Baker, M. Tabassum, S. Zen, K. B. Perumal, and V. Raj, "Review of artificial intelligence techniques used in IoT networks," *International Journal of Engineering Systems Modelling and Simulation*, vol. 15, no. 4, pp. 189–198, 2024.

[28] J. Xiong, T. Hong, D. Zhao, and Y. Zhang, "MATNilm: Multi-appliance-task non-intrusive load monitoring with limited labeled data," *IEEE Transactions on Industrial Informatics*, vol. 20, no. 3, pp. 3177–3187, 2023.

[29] P. Clerici Maestosi, "Harmonizing urban innovation: Exploring the nexus between smart cities and positive energy districts," *Energies*, vol. 17, no. 14, 2024.

[30] H. Qiu, J. Zhang, L. Yang, K. Han, X. Yang, and Z. Gao, "Spatial–temporal multi-task learning for short-term passenger inflow and outflow prediction on holidays in urban rail transit systems," *Transportation*, pp. 1–30, 2025.

[31] M. Tabassum, K. B. Zen, S. Perumal, and V. Raj, "Review of artificial intelligence techniques used in IoT networks," *International Journal of Engineering Systems Modelling and Simulation*, vol. 15, no. 4, pp. 189–198, 2024.

[32] H. A. Abdul-Ghani and D. Konstantas, "A comprehensive study of security and privacy guidelines, threats, and countermeasures: An IoT perspective," *Journal of Sensor and Actuator Networks*, vol. 8, no. 2, p. 22, 2019.

[33] M. K. Dahouda and I. Joe, "A deep-learned embedding technique for categorical features encoding," *IEEE Access*, vol. 9, pp. 114381–114391, 2021.

[34] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *KDD '16 Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, 2016.

[35] Y. Zhang, Z. Zhao, and J. Zheng, "CatBoost: A new approach for estimating daily reference crop evapotranspiration in arid and semi-arid regions of Northern China," *Journal of Hydrology*, vol. 588, p. 125087, 2020.

[36] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T. Liu, "LightGBM: A highly efficient gradient boosting decision tree," 2017.

[37] J. D. Adekunle, M. I. Oyeniran, H. S. Sule, T. T. Akinpelu, E. J. Ayanlowo, C. K. Ogu, and C. O. Robert, "Let's boost house price predictions: A machine learning approach for Norwich," *Journal of Advances in Artificial Intelligence*, vol. 3, no. 1, pp. 1–18, 2025.

[38] S. Dalal, U. K. Lilhore, B. Seth, M. Radulescu, and S. Hamrioui, "A hybrid model for short-term energy load prediction based on transfer learning with LightGBM for smart grids in smart energy systems," *Journal of Urban Technology*, pp. 1–27, 2024.