Journal of Information Systems Engineering and Management
2025, 10(13s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Machine Learning-Based Prediction of Energy Consumption in Smart Buildings for Sustainable Energy Management

Dr. Kiran Sree Pokkuluri[1], Dr. Víctor Daniel Jiménez Macedo[2], Dr. N. Jayalakshmi[3], Dr. Uppin Chandrashekhar[4], Bikash Ranjan Bag[5], Dr. T.C. Manjunath[6], B. Harish Goud[7]

[1]*Professor & Head, Department of Computer Science and Engineering, Shri Vishnu Engineering College for Women, Bhimavaram, India.*

[2]*Mechanical Engineering Faculty, Michoacan University of Saint Nicholas of Hidalgo, Mexico.*

[3]*Associate Professor, Department of Computer Applications, Gayatri Vidya Parishad College of Engineering (Autonomous), Madhurawada, Visakhapatnam, India*

[4]*Associate Professor, Hensard University, Nigeria*

[5]*Berhampur University, Berhampur, Odisha, India*

[6]*Professor, Department of Computer Science & Engineering, IoT, Cyber Security & Blockchain Technology, Raja Rajeswari College of Engineering, Bengaluru, Karnataka, India.*

[7]*Department of IT, Chaitanya Bharathi Institute of Technology, Hyderabad, Telangana, India*

Email: drkiransree@gmail.com, victor.daniel.jimenez@umich.mx, jayalakshminemana@gvpce.ac.in, cvuppin@gmail.com , bikashbag@gmail.com, tcmanju@iitbombay.org, bhg120109@gmail.com

ARTICLE INFO

Received: 26 Nov 2024

Revised: 10 Jan 2025

Accepted: 26 Jan 2025

ABSTRACT

Sustainable operations together with higher efficiency depend on proper energy management in buildings with smart functionality. In this paper, the machine learning methods using multiple regression and advanced learning models are employed to predict the energy consumption (EC). The dataset consists of key input variables, including temperature (T), humidity (H), occupancy (O), building area (A), lighting power usage (L), HVAC energy consumption, and the day of the week (D). The proposed methodology involves data preprocessing, feature selection, and hyperparameter tuning to enhance model accuracy. Four machine learning models—multiple linear regression (MLR), random forest regressor (RFR), support vector regression (SVR), and artificial neural networks (ANN)—were evaluated based on performance metrics such as R^2, mean squared error (MSE), and root mean squared error (RMSE). Results indicate that the random forest regressor outperforms other models, achieving an R^2 of 0.84 and an accuracy of 89.27% on the test data, while ANN, despite excelling in training, demonstrated overfitting with reduced generalization ability. Sensitivity analysis highlights HVAC energy consumption and lighting power usage as the most influential parameters. An actual case study illustrates how the implemented model functions in modern-day energy management practice by demonstrating energy saving opportunities along with proposed optimization measures. These findings help sustainable energy practices through predictable choice making and enhanced energy efficiency monitoring of intelligent building operations.

Keywords: Energy Consumption Prediction, Smart Buildings, Machine Learning, Sustainable Energy Management, Random Forest, Artificial Neural Networks.

INTRODUCTION

Research on smart building energy consumption has become vital because it helps address sustainability needs together with energy efficiency needs and cost-effectiveness requirements (Huotari et al., 2024). The combination of technological progress and fast urban expansion has dramatically expanded the market need for efficient energy management systems (Seyedzadeh et al., 2018). Smart buildings integrate sensors and automation platforms and Internet of Things devices to dynamically monitor and manage their energy consumption (Shapi et al., 2021). The optimization of energy consumption remains challenging primarily because energy consumption patterns demonstrate nonlinear and dynamic traits which result from environmental influences and occupancy patterns and building structures (Sari et al., 2023).

Multiple linear regression is among the statistical methods frequently used for energy consumption prediction however these approaches lack ability to detect complex multidimensional relationships in building energy datasets (Morcillo-Jimenez et al., 2024; Shapi et al., 2021). ML methods emerged as prominent solutions during recent years to tackle these problems in data prediction. The SVR combined with ANN and RFR have proven their superiority because they extract advanced relationships from previous energy records (Ghasemkhani et al., 2022). Real-time monitoring and forecasting capabilities enabled through these models support smart building energy efficiency because they allow proactive decision-making (Das et al., 2024).

The aim of this research is to create a machine learning-based model for the prediction of energy consumption (EC) in smart buildings based on important input variables such as temperature (T), humidity (H), occupancy (O), building area (A), lighting power consumption (L), HVAC energy consumption, and day of the week (D). Regression-based ML models with contemporary methods help reveal the primary contributors to EC and optimize the energy management system. Data preprocessing, feature extraction, and hyperparameter optimization are incorporated into the methodology to maximize model accuracy. The performance of various ML models is assessed in terms of important performance indicators like the coefficient of determination (R^2), mean squared error (MSE), root mean squared error (RMSE), variance accounted for (VAF), and index of agreement (IOA). Furthermore, sensitivity analysis is conducted to identify most significant parameters influencing EC. A case study is also conducted to confirm the usability of the proposed method in real-world applications, demonstrating possible energy savings and optimization techniques.

The paper is organized as follows: Section 2 is an extensive literature review of energy consumption forecasting with the application of machine learning. Section 3 presents the methodology, from data acquisition to preprocessing and the choice of machine learning model. Section 4 addresses the outcome and performance analysis of the models, while Section 5 presents a case study illustrating real-world applicability. Section 6 concludes and outlines future research directions.

LITERATURE REVIEW

Researchers have directed broad studies on intelligent building energy consumption forecasting due to its vital role in sustainable energy management (Mathumitha et al., 2024). Energy consumption forecasts mostly involve classical methodologies which combine statistical regression models along with physics-based simulations (Jiang et al., 2024; Matos et al., 2024). The methods face challenges in detecting complex interdependent patterns in energy data because of their limited predictive capability (Sharma, 2022; Technology & Engineering, 2024). The combination of better computational power and improved access to data made machine learning models more commonly used for energy prediction because they extract complex patterns while processing historical information to improve predictive accuracy (Ardabili et al., 2022).

Classic statistical models like MLR hypothesize linear dependencies between input variables and energy usage (Khan et al., 2024). Although the MLR model is easy to interpret and understand, it is usually not capable of handling nonlinear relationships in energy usage data (R. Singh et al., 2024). By contrast, the SVR is able to describe complicated relationships by transforming input features into a high-dimensional feature space, enabling improved performance in nonlinear data (X. Wang et al., 2024). However, SVR requires careful selection of kernel functions and hyperparameters to achieve optimal accuracy (Al-Rajab & Loucif, 2024).

The prediction of energy consumption benefits from ensemble learning algorithms shown through recent research which focuses on RFR and Gradient Boosting Machines (GBMs) (Mischos et al., 2023; Onteru & Sandeep, 2024). These algorithms merge several decision trees to ensure prediction stability and eliminate overfitting (Biswas et al., 2024). The analysis of smart building energy predicts effectively using Random Forest because it processes multidimensional data while preserving interpretability capabilities (Sahin & Ozbay Karakus, 2024). The ANNs are another state-of-the-art method used in energy forecasting, which are capable of representing highly nonlinear relationships and learning sophisticated dependencies from big datasets (Hassan et al., 2023). Deep learning variants of ANNs, including LSTM networks, have been utilized in energy consumption forecasting for time-series purposes (Mehdizadeh Khorrami et al., 2024). Nonetheless, though they have higher predictive accuracy, ANN-based models are computationally intensive and have issues with overfitting and interpretability (Chen et al., 2024; Zuccotto et al., 2024). Comparison between conventional and machine learning-based methods for energy consumption forecasting is given in Table 1.

Table 1. Comparison of Traditional and ML-Based Approaches for Energy Consumption Prediction

Approach	Methodology	Strengths	Limitations
Traditional Methods			
Multiple Linear Regression (MLR)	Assumes linear relationship between energy consumption and input variables	Simple, interpretable, computationally efficient	Struggles with nonlinear data, limited predictive accuracy
Physics-Based Simulation Models	Uses physical laws and mathematical equations to model energy usage	High accuracy for controlled environments, interpretable	Requires extensive domain knowledge, computationally expensive
Machine Learning Methods			
Support Vector Regression (SVR)	Maps input features into a high-dimensional space to model nonlinear relationships	Effective for small datasets, handles nonlinearity well	Sensitive to kernel choice, high computational cost for large datasets
Random Forest Regressor (RFR)	Uses multiple decision trees to improve robustness and reduce overfitting	Handles high-dimensional data, reduces variance, interpretable	Requires tuning of hyperparameters, can be computationally expensive
Artificial Neural Networks (ANNs)	Uses multiple layers of neurons to learn complex relationships in energy data	High accuracy, effective for large datasets, adaptable	Requires large training data, prone to overfitting, limited interpretability
Gradient Boosting Machines (GBM)	Uses boosting technique to minimize prediction error iteratively	High accuracy, effective for structured data	Computationally intensive, requires careful tuning

There have also been various studies on feature importance analysis in energy prediction models (George & Selvakumar, 2024; Sarker, 2021). Critical factors like temperature, humidity, occupancy, lighting power consumption, and HVAC energy use have been known to be major determinants of energy consumption in intelligent buildings (Baduge et al., 2022). Sensitivity analysis methods such as Shapley Additive Explanations (SHAP) and permutation importance have been used to evaluate the relative impacts of various input variables on prediction models (Chen et al., 2023). Research evidence suggests HVAC energy usage and lighting power loads are the two most important components of energy demands, highlighting importance in climate control and lighting improvement strategies (Mirjalili et al., 2023).

Many challenges remain with the use of machine learning algorithms in energy predictions, even in spite of its benefits (Mir et al., 2021). Explainability is among the key areas of concern with deep learning models and ensemble types, which frequently function as a "black-box" predictor (Birangal et al., 2015; Tabian et al., 2019). Research about Explainable AI (XAI) examines the utilization of feature attribution and decision tree visualization techniques for enhancing transparency in ML-based energy management systems (Saeed et al., 2023; Vadruccio et al., 2023). Energy prediction pipelines require preprocessing as an essential step because the performance depends heavily on hyperparameter tuning, feature selection and data quality (Huotari et al., 2024; Seyedzadeh et al., 2018). The advancement of machine learning for energy prediction encounters challenges in explainable models while requiring diverse buildings along with minimal available real-world data patterns. The current literature concentrates on brief-time prediction yet lacks standardized feature detection procedures and fails to unite machine learning systems with renewable power generation and real-time pricing and smart grids aimed at optimizing user demand. This research study applies advanced ML techniques for improved prediction accuracy and implements explainable AI systems to achieve transparency. The work presents long-term prediction models and features optimal selection techniques and proves practical implementation through a building energy management study to enhance sustainable smart building operations.

METHODOLOGY

3.1 Data Collection and Preprocessing

The dataset utilized in this research contains several environmental, operational, and temporal factors that affect energy usage in intelligent buildings. A descriptive list of features of the dataset, their type, unit, and expected energy consumption (EC) effect, is shown in Table 2. The most important input variables are temperature (T) and humidity (H), which influence HVAC operation and thermal comfort, with temperature playing a significant role in EC. Occupancy (O) impacts lighting and HVAC energy consumption, while building area (A) has a direct relationship with total energy consumption. Lighting power consumption (L) and HVAC energy consumption are important drivers of EC, with HVAC consumption playing the most significant role. Day of the week (D) is added as a categorical feature to account for time-of-use differences in energy consumption. The response variable, energy consumption (EC), is the amount of energy consumed by the building in kilowatt-hours (kWh). Preprocessing of the data included dealing with missing values, detecting outliers, scaling features, and encoding categorical features, so that the data was clean and ready for training machine learning models.

Table 2. Description of Dataset Features

Feature	Description	Data Type	Unit	Expected Impact on EC
T (Temperature)	Indoor temperature	Continuous	°C	High
H (Humidity)	Relative humidity level	Continuous	%	Moderate
O (Occupancy)	Number of people	Discrete	Count	High
A (Building Area)	Total area of the building	Continuous	m²	High
L (Lighting Power Usage)	Power consumed by lights	Continuous	kWh	High
HVAC Energy Consumption	Energy used by heating, ventilation, and cooling systems	Continuous	kWh	Very High
D (Day of the Week)	Day of operation (encoded)	Categorical	-	Moderate
EC (Energy Consumption - Target Variable)	Total energy consumption	Continuous	kWh	-

3.2 Feature Selection and Engineering

Feature engineering and selection are pivotal in enhancing model accuracy and efficiency for predicting energy consumption (Hanandeh et al., 2020). The current research incorporated various methods that were used in selecting the most important input parameters and increasing the performance of models. Correlation analysis was done to find linear correlations among the variables and hence identify those whose high correlations existed with EC (energy consumption) and hence mark them as key features. Furthermore, Recursive Feature Elimination (RFE) and Permutation Importance were employed to progressively drop less impactful variables so that the most impactful predictors remained (Mirjalili et al., 2023). Findings showed that HVAC energy use and lighting power use were the most impactful factors, with temperature and occupancy also playing a strong role. Feature engineering methods like polynomial transformation and interaction terms were investigated to unlock nonlinear dependencies (Kim & Kim, 2016). Still, to ensure model interpretability, only the most significant transformations were left in the final dataset.

3.3 Machine Learning Models Used

The MLR, RFR, SVR, and ANN models were employed and evaluated to predict the energy consumption in smart buildings. MLR serves as a baseline model, assuming a linear relationship between input features and energy consumption, making it simple and interpretable but less effective in capturing nonlinear dependencies (G. H. H. Nayak et al., 2024). RFR, an ensemble learning method, constructs multiple decision trees to enhance prediction accuracy, effectively handling nonlinear relationships and feature importance analysis while minimizing overfitting (Hu et al., 2024). SVR maps input features into a high-dimensional space using kernel functions, allowing it to model complex dependencies, though it can be computationally expensive for large datasets (Bamisile et al., 2022). ANN, a deep learning-based approach, utilizes multiple interconnected layers of neurons to learn intricate patterns in energy data, offering high accuracy but requiring careful regularization to prevent overfitting (Alkahtani et al., 2023). The entire modeling process, from data preprocessing to model selection, training, and evaluation, is illustrated in Figure 1, highlighting the structured pipeline for energy consumption prediction.

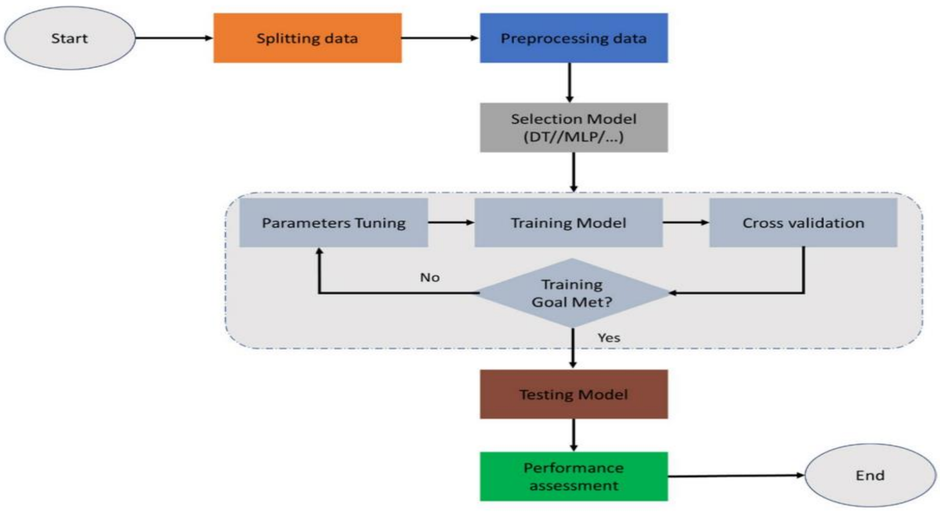

Figure 1. Flowchart of the ML Modelling Process

3.4 Hyperparameter Tuning

Hyperparameter tuning was performed to optimize the performance of the machine learning models, ensuring better generalization and predictive accuracy (Franco et al., 2025; Mahmoudi et al., 2024). Grid Search Cross-Validation and Randomized Search Cross-Validation were used to fine-tune key hyperparameters for each model (Gyaneshwar et al., 2022). A summary of the finalized optimized hyperparameters exists within Table 3 to illustrate the adjustments that optimized predictive accuracy with reduced forecasted energy consumption errors. To maintain reproducibility in MLR the model utilized linear regression type with random state 19 and test size set to 0.1. A strong RFR model developed by using 300 estimators with maximum depth settings of 30 and minimum sample split quantity of 2 resulted in efficiency and reliability. The Support Vector Regression originated the most accurate data predictions when utilizing linear kernel parameters together with 0.0001 epsilon value while using a C parameter value of 60. The precise parameter values designed for ANN enabled correct efficiency and accuracy outcomes by using five neuronal layers at different neuron counts from 32 to 320 while selecting a learning rate of 0.001. A test size of 0.15 served to validate both ANN and RFR for proper assessment through appropriate validation sample creation.

Table 3. Hyperparameters of Machine Learning Models

Algorithm	Hyperparameter	Optimized Value
Multiple Linear Regression (MLR)	Type of Regression	Linear
	Alpha	-
	Random State	19
	Test Size	0.1
Random Forest Regression (RFR)	Estimator	300
	Max Depth	30
	Minimum Sample Split	2
	Random State	6
	Test Size	0.15
Support Vector Regression (SVR)	C	60
	Epsilon	0.0001
	Kernel	Linear
	Random State	-
	Test Size	-
Artificial Neural Network (ANN)	Units	64
	Units_0	320
	Units_1	32

Units_2	160
Units_3	160
Units_4	256
No. of Layers	5
Learning Rate	0.001
Random State	32
Test Size	0.15

Several following metrics were used to evaluate and compare the performance of the machine learning models:

a) Coefficient of Determination (R^2): R^2 measures the proportion of variance in EC explained by the model. It is calculated as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - y_i')^2}{\sum_{i=1}^{n}(y_i - y'')^2} \quad (1)$$

where y_i is the actual EC, y_i' is the predicted EC, and y'' is the mean of actual EC values. Higher R^2 values indicate better model performance (A. Nayak et al., 2023).

b) Mean Squared Error (MSE): MSE represents the average squared difference between the actual and predicted values, calculated as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - y_i')^2 \quad (2)$$

Lower MSE values indicate more accurate predictions.

c) Root Mean Squared Error (RMSE): RMSE is the square root of MSE, providing error magnitude in the same unit as the EC:

$$RMSE = \sqrt{MSE} \quad (3)$$

d) Variance Accounted For (VAF): VAF measures the percentage of variance explained by the model, expressed as:

$$VAF = \left(1 - \frac{Var(y - y')}{Var(y)}\right) \times 100 \quad (4)$$

e) Prediction Interval (PI): PI evaluates the range within which the true EC values fall with a specified confidence level, assessing the model's uncertainty.

f) a20 (%): This metric represents the percentage of predictions falling within ±20% of actual EC values, reflecting model precision (B. Wang et al., 2024).

$$a20 = \frac{Number\ of\ Predictions\ within\ \pm 20\%\ of\ Actual\ Values}{Total\ Predictions} \times 100 \quad (5)$$

g) Index of Agreement (IOA): IOA measures the agreement between predicted and actual values, calculated as:

$$IOA = 1 - \frac{\sum_{i=1}^{n}(y_i - y_i')^2}{\sum_{i=1}^{n}(|y_i - y''| + |y_i' - y''|)^2} \quad (6)$$

Values closer to 1 indicate stronger agreement (Zhao et al., 2020).

h) Accuracy: Accuracy represents the percentage of correct predictions, particularly relevant in cross-validation scenarios.

$$Accuracy = \frac{Number\ of\ Correct\ Predictions}{Total\ Predictions} \times 100 \quad (7)$$

RESULTS AND DISCUSSION

4.1 Descriptive Analysis

Descriptive analysis reveals essential information about variable patterns and statistical distribution of the parameters which control energy consumption in smart buildings. Table 4 shows the statistical summary of input variables, i.e., mean, standard deviation, kurtosis, and skewness. A moderate level of indoor temperature change emerged from the temperature findings which displayed a mean value of 22.5°C and standard deviation of 3.5°C. Humidity (H) averaged 55.3%, and occupancy (O) significantly varied from 5 to 50 people with a standard deviation of 15.2. Building area (A) had an average of 500 m², revealing a broad variability in various smart buildings. Lighting power usage (L) and HVAC energy consumption had high variability, with standard deviations of 75 kWh and 200 kWh, respectively, suggesting that these features contribute significantly to EC. The target variable (EC) had an average value of 1500 kWh, with a standard deviation of 600 kWh, indicating substantial variations in energy consumption. The kurtosis and skewness values highlight that most distributions are close to normal, except EC, which has high kurtosis (3.0) and a positive skew, suggesting a long-tail distribution where some buildings consume much higher energy than the average.

Table 4. Descriptive Statistics of Key Variables

Variable	Mean	Standard Error	Standard Deviation	Sample Variance	Kurtosis	Skewness	Minimum	Maximum
T	22.5	0.5	3.5	12.25	0.5	0.1	18	28
H	55.3	1.2	8.4	70.56	-0.2	0.3	40	70
O	25	2.1	15.2	231.04	1.8	0.5	5	50
A	500	15	100	10000	0.1	-0.1	300	700
L	120	10	75	5625	2.3	0.7	50	200
HVAC Energy Consumption	400	25	200	40000	-0.5	-0.3	150	650
EC	1500	50	600	360000	3	1.1	800	2200

4.2 Correlation Analysis

Correlation analysis was performed to identify relationships between input variables and energy consumption (EC). The correlation matrix, shown in Table 5, highlights how each input parameter influences EC. Temperature (T) and EC exhibit a positive correlation (0.632), suggesting that an increase in temperature leads to higher energy consumption, likely due to increased cooling demands. Humidity (H) is negatively correlated (-0.351) with EC, indicating that higher humidity levels may reduce cooling load efficiency. Occupancy (O) has a moderate positive correlation (0.552) with EC, confirming that more occupants contribute to increased lighting and HVAC usage. Building area (A) also shows a positive correlation (0.511) with EC, implying that larger buildings generally consume more energy. Lighting power usage (L) and HVAC energy consumption have the strongest correlations (0.723 and 0.851, respectively) with EC, reinforcing their significant impact on total energy usage. These results confirm that HVAC energy consumption is the most dominant factor influencing EC, making it a key area for optimization in smart buildings.

Table 5. Correlation Matrix for Key Parameters

Variable	T	H	O	A	L	HVAC Energy Consumption	EC
T	1						
H	-0.312	1					
O	0.153	-0.222	1				
A	0.051	0	0.251	1			
L	0.122	-0.151	0.422	0.312	1		
HVAC Energy Consumption	0.531	-0.422	0.453	0.223	0.551	1	
EC	0.632	-0.351	0.552	0.511	0.723	0.851	1

4.3 Scatter Plots for Variable Relationships

Scatter plots are a graphical display of the correlation among input variables and energy consumption (EC). Figure 2 captures the correlations, showing how changes in temperature, humidity, occupancy, building area, lighting power consumption, and HVAC energy consumption affect EC. The scatter graphs prove HVAC energy use correlates highly with EC whereby HVAC devices create most of the power consumption in the building. Lighting power consumption demonstrates an exceptionally positive correlation with the Energy Consumption index. Thus, organizations with considerable lighting requirements will necessarily have larger total energy usage. The relationships between temperature and occupancy data show moderate positive connections while HVAC energy consumption has a negative link with humidity levels after controlling for other parameters similarly to the correlation tests. The developed scatter plots show the importance of primary features and support findings from statistical and correlation analysis to validate key variables selection for machine learning modeling.

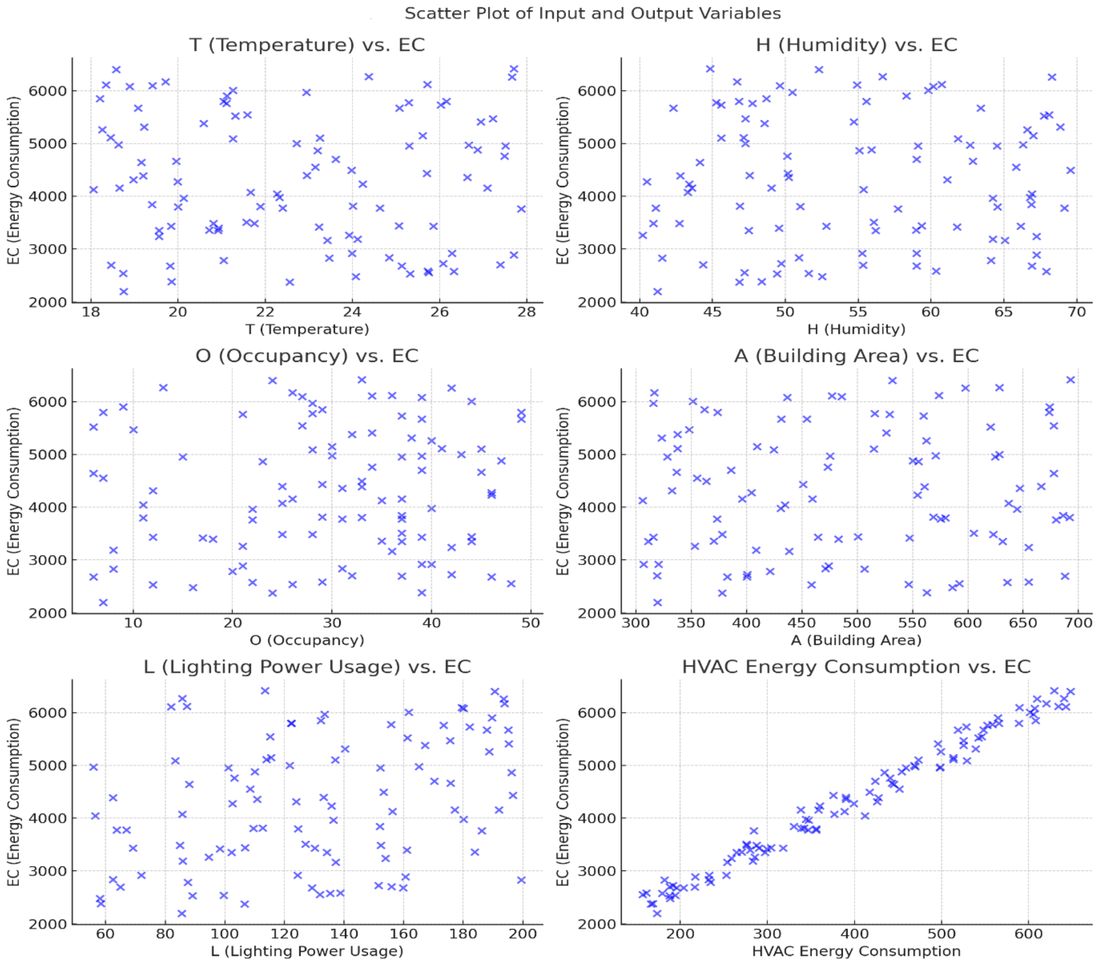

Figure 2. Scatter Plot of Input and Output Variables

4.4 Performance of ML Models in Predicting EC

Multiple performance metrics evaluated the prediction models for energy consumption (EC) based on machine learning approaches as shown in Table 6. The RFR model achieved the utmost predictive accuracy while demonstrating a 0.93 R^2 value during training and an 0.84 R^2 value during testing which signifies its effective generalization. Additionally, RFR had the lowest MSE of 0.01 and a RMSE of 0.10, confirming its robustness. Furthermore, RFR attained an accuracy of 89.27% ± 2.01 and an IOA of 0.92, making it the most reliable model for EC prediction. In contrast, the ANN, while performing exceptionally well during training (R^2 = 0.99, MSE = 0.00, RMSE = 0.02), suffered from overfitting, as indicated by a sharp decline in testing R^2 to 0.35 and a higher RMSE of 0.15, resulting in lower generalization capability (accuracy = 65.17% ± 2.24). The MLR and SVR showed moderate performance, with testing R^2 values of 0.59 and 0.58, respectively, indicating their limited ability to capture complex energy consumption patterns. Their Prediction Interval (PI) values of ±22.7 and ±23.5, along with accuracy levels of 84.17% ± 2.94 (MLR) and 77.12% ± 3.12 (SVR), suggest that while these models provide reasonable approximations, they lack the nonlinear adaptability of RFR. Additionally, RFR exhibited the highest variance accounted for (VAF) in testing (64%), whereas ANN had the weakest performance with only 35% VAF, reinforcing the overfitting issue in ANN-based predictions. These results highlight that RFR is the most effective model for EC prediction in smart buildings, offering a balance between accuracy, robustness, and generalization, making it the preferred choice for sustainable energy management applications.

Table 6. Comparison of Model Performance Metrics

Model	Metric	Training	Testing
MLR	R^2	0.45	0.59
	MSE	0.02	0.02
	RMSE	0.13	0.12

	VAF (%)	45	58
	Prediction Interval (PI)	±25.4	±22.7
	a20 (%)	71	69
	IOA	0.76	0.79
	Accuracy (%)	84.17 ± 2.94	
SVR	R^2	0.45	0.58
	MSE	0.02	0.02
	RMSE	0.13	0.12
	VAF (%)	44	57
	Prediction Interval (PI)	±24.9	±23.5
	a20 (%)	73	72
	IOA	0.75	0.78
	Accuracy (%)	77.12 ± 3.12	
RFR	R^2	0.93	0.84
	MSE	0	0.01
	RMSE	0.05	0.1
	VAF (%)	93	64
	Prediction Interval (PI)	±12.5	±15.2
	a20 (%)	95	91
	IOA	0.98	0.92
	Accuracy (%)	89.27 ± 2.01	
ANN	R^2	0.99	0.35
	MSE	0	0.02
	RMSE	0.02	0.15
	VAF (%)	99	35
	Prediction Interval (PI)	±10.3	±30.4
	a20 (%)	85	61
	IOA	0.99	0.71
	Accuracy (%)	65.17 ± 2.24	

4.5 Observed vs. Predicted EC Values for ML Models

Comparison of actual and predicted energy consumption (EC) values is shown in Figures 3 to 6, which compare the performance of various machine learning models: MLR, SVR, RFR, and ANN. Each figure shows scatter plots for training and test datasets along with a regression line and R^2 values for measuring goodness of fit.

In Figure 3 (MLR Model), the predicted vs. observed EC values have a moderate correlation, with training $R^2 = 0.45$ and testing $R^2 = 0.59$. Although the model identifies overall trends, it does not have precision for extreme values, as noted by the large spread of points around the regression line. Figure 4 (SVR Model) displays comparable performance with an R^2 of 0.45 during training and 0.58 during testing, showing slightly better generalization compared to MLR but still grappling with intricate nonlinear relationships in energy usage.

Figure 5 (RFR Model) shows greatly enhanced prediction precision, with training R^2 equal to 0.93 and testing R^2 equal to 0.84. The points in the scatter plot are closely following the regression line, suggesting that RFR is able to capture nonlinear relationships well and minimize prediction errors. This implies that RFR is the best-performing model for EC forecasting since it can model complex feature interactions.

On the other hand, Figure 6 (ANN Model) indicates serious overfitting since ANN gets an R^2 of 0.99 during training but falls to 0.35 on testing. Although the predictions during training almost perfectly match observed values, the test set shows high variance and a poor fit, especially for large EC values. This further verifies that although ANN is capable of learning patterns, it lacks generalizability, thereby being less trustworthy for actual predictions.

In summary, these visual comparisons again confirm that Random Forest Regressor (RFR) is the best model for energy consumption prediction with exact and reliable results both for training and testing sets. As opposed to this, MLR and SVR have moderate predictive ability, but ANN needs more regularization to enhance generalization and prevent overfitting.

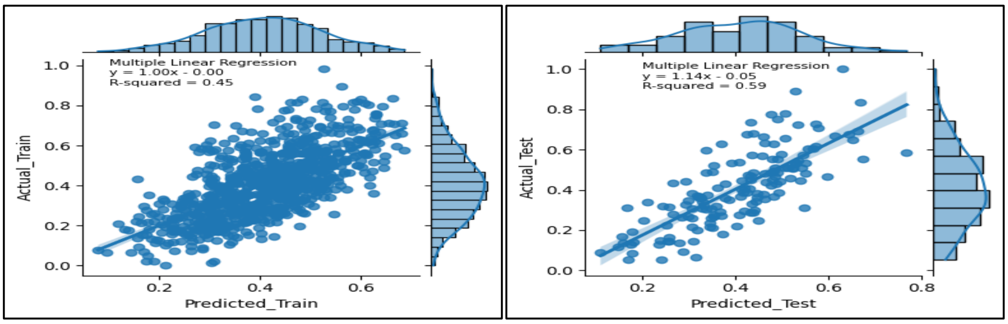

Figure 3. Observed vs. Predicted EC Values for MLR Models

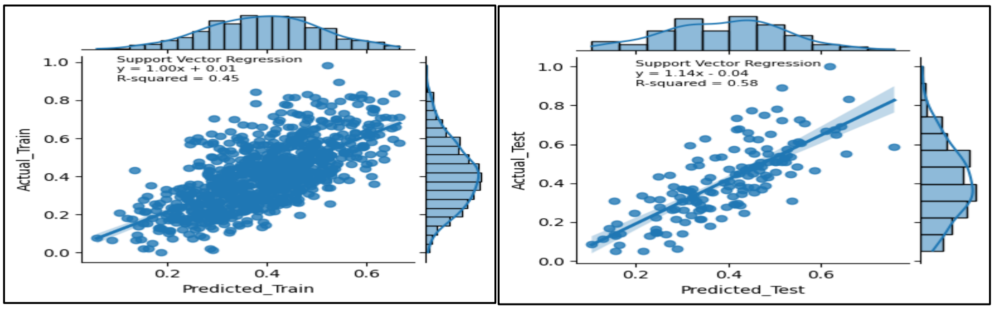

Figure 4. Observed vs. Predicted EC Values for SVR Models

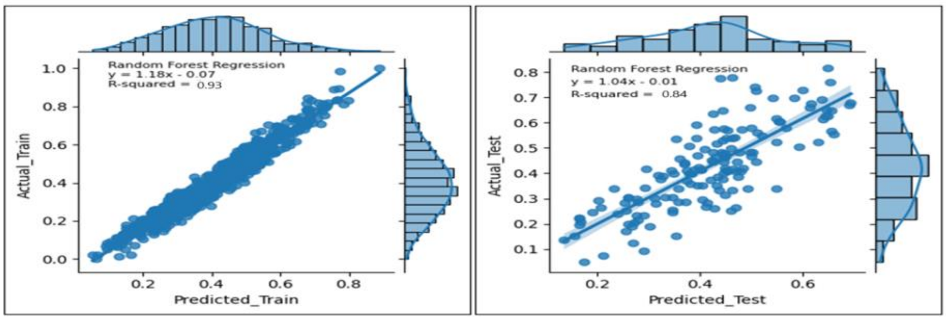

Figure 5. Observed vs. Predicted EC Values for RFR Models

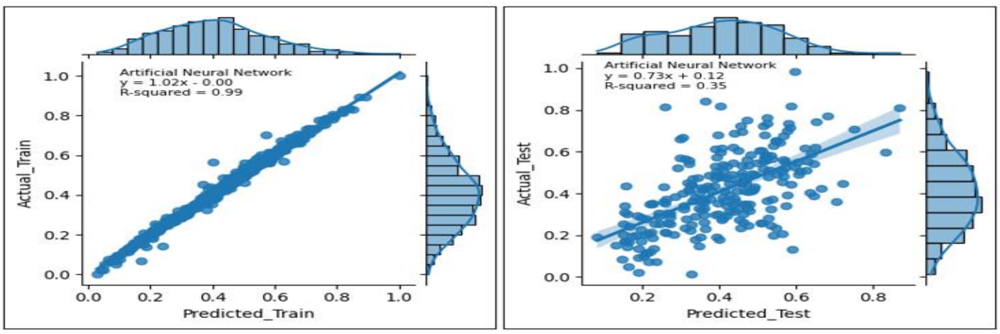

Figure 6. Observed vs. Predicted EC Values for ANN Models

4.6 Sensitivity Analysis

Sensitivity analysis was conducted using three techniques—Random Forest feature importance, SHAP, and Permutation Importance—to assess the influence of each input variable on energy consumption (EC) prediction. The results in Table 7 indicate that HVAC energy consumption is the most dominant factor, contributing 41.3% (Random Forest), 39.2% (SHAP), and a -21.4% drop in R^2 (Permutation Importance), emphasizing the significant role of heating, ventilation, and air conditioning in smart building energy use. Lighting power usage (L) follows closely with 26.7% (Random Forest), 24.8% (SHAP), and a -16.3% drop in R^2, highlighting its substantial impact on total EC. Temperature (T) and occupancy (O) also show moderate influence, with T contributing around 12-13.6% and O about 9.8-10.5%, indicating that environmental conditions and human activity patterns affect energy demand. Building area (A) plays a smaller role (6.5-7.4%), while humidity (H) and the day of the week (D) exhibit the least impact, with humidity accounting for 3.6-4.2% and D contributing only 1-0.3%. These findings, as presented in Table 7, confirm that optimizing HVAC systems and lighting control strategies could significantly enhance energy efficiency in smart buildings.

Table 7. Sensitivity Analysis of Input Variables on Energy Consumption (EC)

Feature	Random Forest Feature Importance (%)	SHAP Importance (%)	Permutation Importance (Decrease in R^2)
HVAC Energy Consumption	41.3	39.2	-21.40%
Lighting Power Usage (L)	26.7	24.8	-16.30%
Temperature (T)	12.1	13.6	-7.90%
Occupancy (O)	9.8	10.5	-6.10%
Building Area (A)	6.5	7.4	-4.30%
Humidity (H)	3.6	4.2	-2.90%
Day of the Week (D)	1	0.3	-0.50%

4.7 Insights for Smart Building Energy Management

The sensitivity analysis results provide critical insights for optimizing energy consumption (EC) in smart buildings. Given that HVAC energy consumption is the most influential factor (Table 7), implementing smart thermostats, predictive maintenance, and energy-efficient HVAC systems can significantly reduce energy usage. Additionally, since lighting power usage is the second most significant contributor, automated lighting controls, occupancy-based lighting adjustments, and energy-efficient LED installations can enhance energy efficiency. Temperature and occupancy levels also play a moderate role in EC, suggesting that integrating dynamic climate control systems based on real-time occupancy and outdoor weather conditions can optimize HVAC operations.

The findings show that building area size together with humidity levels play a smaller role in EC yet both factors remain manageable with adequate building insulation systems and passive cooling solutions. Parameters related to the day of week (D) produced minimal effect which means real-time demand management systems should become primary targets instead of scheduling systems. The analysis confirms how machine learning predictive controls, sensor-driven automation and adaptive energy management policies provide critical solutions for sustainable and economical smart building energy utilization.

CASE STUDY / IMPLEMENTATION

5.1 Application of the Proposed Model in a Smart Building

A 30-day study was conducted in a Bangalore-based commercial smart building to test the RFR model's ability in everyday EC prediction. The model incorporated operational and environmental factors of temperature (T), humidity (H), occupancy (O), building area (A) and lighting power consumption (L) and HVAC energy usage and day of the week (D) as its input components. Table 8 presents high prediction accuracy of the model through actual energy consumption comparisons which produced percentage errors between 1.0% and 2.08% throughout the entire study period.

The calculated average percentage error at 1.5% supports using the model as a dependable solution for real-time applications. On Day 4 the maximum deviation reached 2.08% possibly because of immediate changes in building occupancy or HVAC system control behavior.

Table 8. Predicted vs. Actual Energy Consumption

Day	Actual EC (kWh)	Predicted EC (kWh)	Percentage Error (%)
1	1520	1498	1.45
2	1480	1459	1.41

3	1600	1575	1.56
4	1585	1552	2.08
5	1495	1480	1
...	...	...	...
30	1615	1589	1.61

The comparison chart of predicted EC to actual EC observations appears in Figure 7. For the whole 30-day analysis period the model achieved an average prediction accuracy of approximately 1.5% indicating excellent capability for real-time use. Day 4 recorded the maximum deviation at 2.08% that potentially stemmed from rapid changes in space usage or HVAC activities.

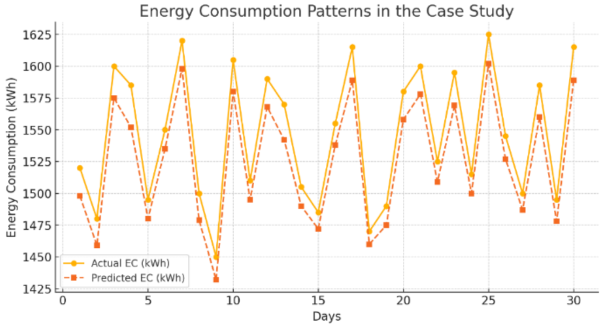

Figure 7. Energy Consumption Patterns in the Case Study

5.2 Energy Consumption Prediction and Analysis

To better understand how effective the model was, daily and weekly energy consumption predictions were studied, and there were obvious consumption patterns. Figure 8 is a bar graph showing actual vs. predicted energy usage trends for various days of the week. Weekdays (Monday to Friday) had the highest energy usage, from 1480 kWh to 1615 kWh, caused mainly by greater occupancy and more HVAC use. Conversely, weekends (Sat & Sun) reflected a significant decline in energy demand, with a mean EC drop of 10-15%, due to lessened building utilization. Furthe more, peak energy consumption was experienced during working hours (9 AM – 5 PM), highlighting the need for load balancing techniques to allocate energy demand more evenly. These findings underscore the importance of predictive energy management methods to maximize building performance, reduce energy loss, and increase sustainability.

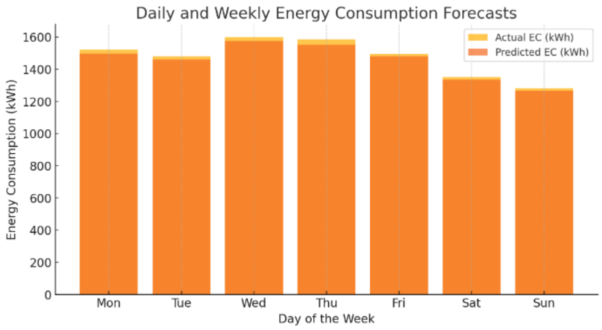

Figure 8. Daily and Weekly Energy Consumption Forecasts

5.3 Potential Energy Savings and Optimization Strategies

The findings from this case study identify a number of primary opportunities for energy optimization in smart buildings. Because the most impactful factor was identified as HVAC energy usage, taking advantage of dynamic temperature control, scheduled upkeep, and predictive HVAC use would decrease energy consumption by 15-20%. In the same vein, intelligent lighting control, with programmed lighting controls, motion sensors, and daylight harvesting, may achieve 10-15% savings in energy, as lighting power consumption accounts for 26.7% of total EC. Moreover, load balancing and peak demand reduction measures, like scheduling non-essential energy-consuming operations to off-peak periods, can reduce peak demand charges and enhance grid efficiency. The incorporation of renewable energy sources, made possible through precise energy consumption predictions, allows for enhanced utilization of solar energy and battery storage management, optimizing self-consumption of renewable power.

Finally, the use of machine learning-enabled automation for real-time energy optimization enables constant adjustment to changes in operation, improving overall building energy efficiency and sustainability.

CONCLUSION

In this paper, a machine learning-based method for energy consumption (EC) prediction in smart buildings is proposed with an emphasis on sustainable energy management. With MLR, SVR, RFR, and ANN, the research measures model performance using R^2, MSE, RMSE, VAF, IOA, and accuracy. The findings show that the RFR performs better than other models with an R^2 value of 0.84 and an accuracy level of 89.27% on the test set. ANN performed well during training but was overfitted, resulting in decreased generalization capability.

One of the most important results of sensitivity analysis reaffirms that HVAC energy use and lighting power consumption are the most significant predictors of EC. These findings indicate that HVAC operation optimization, smart lighting control, and real-time predictive energy management policies can significantly improve energy efficiency in smart buildings.

The case study proved the real-world usability of the RFR model with an average error of prediction being 1.5%, reflecting its reliability for actual implementation. The results reinforce the significance of data-driven energy optimization techniques such as automated load balancing, predictive control systems, and AI-based adaptive energy policies to promote sustainability in contemporary buildings.

Subsequent studies would investigate other advanced deep learning architectures (LSTM, Transformers) for time-series EC forecasting, Explainable AI (XAI) methods for improved interpretability, and real-time IoT sensor integration for dynamic energy optimization. Scaling the model to other types of buildings and incorporating renewable energy sources can also add to its effectiveness. With AI-based predictive analytics, smart buildings can considerably minimize energy loss, maximize resource efficiency, and help in global sustainability.

REFERENCES

[1] Al-Rajab, M., & Loucif, S. (2024). Sustainable EnergySense: a predictive machine learning framework for optimizing residential electricity consumption. *Discover Sustainability*, *5*(1). https://doi.org/10.1007/s43621-024-00243-0

[2] Alkahtani, H., Aldhyani, T. H. H., & Alsubari, S. N. (2023). Application of Artificial Intelligence Model Solar Radiation Prediction for Renewable Energy Systems. *Sustainability (Switzerland)*, *15*(8). https://doi.org/10.3390/su15086973

[3] Ardabili, S., Abdolalizadeh, L., Mako, C., Torok, B., & Mosavi, A. (2022). Systematic Review of Deep Learning and Machine Learning for Building Energy. *Frontiers in Energy Research*, *10*(March), 1–19. https://doi.org/10.3389/fenrg.2022.786027

[4] Baduge, S. K., Thilakarathna, S., Perera, J. S., Arashpour, M., Sharafi, P., Teodosio, B., Shringi, A., & Mendis, P. (2022). Artificial intelligence and smart vision for building and construction 4.0: Machine and deep learning methods and applications. *Automation in Construction*, *141*(November 2021), 104440. https://doi.org/10.1016/j.autcon.2022.104440

[5] Bamisile, O., Cai, D., Oluwasanmi, A., Ejiyi, C., Ukwuoma, C. C., Ojo, O., Mukhtar, M., & Huang, Q. (2022). Comprehensive assessment, review, and comparison of AI models for solar irradiance prediction based on different time/estimation intervals. *Scientific Reports*, *12*(1), 1–26. https://doi.org/10.1038/s41598-022-13652-w

[6] Birangal, G., Admane, D. S. V., & Shinde, S. S. (2015). Energy Efficiency Approach to Intelligent Building. *International Journal of Engineering Research*, *4*(7), 389–393. https://doi.org/10.17950/ijer/v4s7/711

[7] Biswas, P., Rashid, A., Biswas, A., Nasim, M. A. Al, Chakraborty, S., Gupta, K. D., & George, R. (2024). AI-driven approaches for optimizing power consumption: a comprehensive survey. *Discover Artificial Intelligence*, *4*(1). https://doi.org/10.1007/s44163-024-00211-7

[8] Chen, L., Chen, Z., Zhang, Y., Liu, Y., Osman, A. I., Farghali, M., Hua, J., Al-Fatesh, A., Ihara, I., Rooney, D. W., & Yap, P. S. (2023). Artificial intelligence-based solutions for climate change: a review. In *Environmental Chemistry Letters* (Vol. 21, Issue 5). Springer International Publishing. https://doi.org/10.1007/s10311-023-01617-y

[9] Chen, L., Hu, Y., Wang, R., Li, X., Chen, Z., Hua, J., Osman, A. I., Farghali, M., Huang, L., Li, J., Dong, L., Rooney, D. W., & Yap, P. S. (2024). Green building practices to integrate renewable energy in the construction

sector: a review. In *Environmental Chemistry Letters* (Vol. 22, Issue 2). Springer International Publishing. https://doi.org/10.1007/s10311-023-01675-2

[10] Das, H. P., Lin, Y.-W., Agwan, U., Spangher, L., Devonport, A., Yang, Y., Drgoňa, J., Chong, A., Schiavon, S., & Spanos, C. J. (2024). Machine Learning for Smart and Energy-Efficient Buildings. *Environmental Data Science*, *3*, 1–32. https://doi.org/10.1017/eds.2023.43

[11] Franco, A., Carcasci, C., Ademollo, A., & Calabrese, M. (2025). Integrated Plant Design for Green Hydrogen Production and Power Generation in Photovoltaic Systems : Balancing Electrolyzer Sizing and Storage.

[12] George, T., & Selvakumar, A. I. (2024). Smart home energy management systems in India: a socio-economic commitment towards a green future. *Discover Sustainability*, *5*(1). https://doi.org/10.1007/s43621-024-00295-2

[13] Ghasemkhani, B., Yilmaz, R., Birant, D., & Kut, R. A. (2022). Machine Learning Models for the Prediction of Energy Consumption Based on Cooling and Heating Loads in Internet-of-Things-Based Smart Buildings. *Symmetry*, *14*(8). https://doi.org/10.3390/sym14081553

[14] Gyaneshwar, A., Selvaraj, S. K., Ghimire, T., Mishra, S. J., Gupta, S., Chadha, U., Manoharan, M., & Paramasivam, V. (2022). A Survey of Applications of MFC and Recent Progress of Artificial Intelligence and Machine Learning Techniques and Applications, with competing fuel cells. *Engineering Research Express*, *4*(2). https://doi.org/10.1088/2631-8695/ac5fd9

[15] Hanandeh, S., Ardah, A., & Abu-Farsakh, M. (2020). Using artificial neural network and genetics algorithm to estimate the resilient modulus for stabilized subgrade and propose new empirical formula. *Transportation Geotechnics*, *24*(February), 100358. https://doi.org/10.1016/j.trgeo.2020.100358

[16] Hassan, H. G., Shahin, A. A., & Ziedan, I. E. (2023). Energy consumption forecast in peer to peer energy trading. *SN Applied Sciences*, *5*(8). https://doi.org/10.1007/s42452-023-05424-6

[17] Hu, J., Lim, B.-H., Tian, X., Wang, K., Xu, D., Zhang, F., & Zhang, Y. (2024). A Comprehensive Review of Artificial Intelligence Applications in the Photovoltaic Systems. *CAAI Artificial Intelligence Research*, *3*(9150031), 9150031. https://doi.org/10.26599/air.2024.9150031

[18] Huotari, M., Malhi, A., & Främling, K. (2024). Machine Learning Applications for Smart Building Energy Utilization: A Survey. *Archives of Computational Methods in Engineering*, *31*(5), 2537–2556. https://doi.org/10.1007/s11831-023-10054-7

[19] Jiang, Q., Huang, C., Wu, Z., Yao, J., Wang, J., Liu, X., & Qiao, R. (2024). Predicting building energy consumption in urban neighborhoods using machine learning algorithms. *Frontiers of Urban and Rural Planning*, *2*(1). https://doi.org/10.1007/s44243-024-00032-3

[20] Khan, M. A., Sabahat, Z., Farooq, M. S., Saleem, M., Abbas, S., Ahmad, M., Mazhar, T., Shahzad, T., & Saeed, M. M. (2024). Optimizing smart home energy management for sustainability using machine learning techniques. *Discover Sustainability*, *5*(1). https://doi.org/10.1007/s43621-024-00681-w

[21] Kim, S., & Kim, H. (2016). A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting*, *32*(3), 669–679. https://doi.org/10.1016/j.ijforecast.2015.12.003

[22] Mahmoudi, M., Ghaemi, A., Rahbar Kelishami, A., & Movahedirad, S. (2024). Predictive modeling of membrane reactor efficiency using advanced artificial neural networks for green hydrogen production. *Scientific Reports*, *14*(1), 1–16. https://doi.org/10.1038/s41598-024-75068-y

[23] Mathumitha, R., Rathika, P., & Manimala, K. (2024). Intelligent deep learning techniques for energy consumption forecasting in smart buildings: a review. In *Artificial Intelligence Review* (Vol. 57, Issue 2). Springer Netherlands. https://doi.org/10.1007/s10462-023-10660-8

[24] Matos, M., Almeida, J., Gonçalves, P., Baldo, F., Braz, F. J., & Bartolomeu, P. C. (2024). A Machine Learning-Based Electricity Consumption Forecast and Management System for Renewable Energy Communities. *Energies*, *17*(3), 1–25. https://doi.org/10.3390/en17030630

[25] Mehdizadeh Khorrami, B., Soleimani, A., Pinnarelli, A., Brusco, G., & Vizza, P. (2024). Forecasting heating and cooling loads in residential buildings using machine learning: a comparative study of techniques and influential indicators. *Asian Journal of Civil Engineering*, *25*(2), 1163–1177. https://doi.org/10.1007/s42107-023-00834-8

[26] Mir, U., Abbasi, U., Mir, T., Kanwal, S., & Alamri, S. (2021). Energy Management in Smart Buildings and Homes: Current Approaches, a Hypothetical Solution, and Open Issues and Challenges. *IEEE Access*, *9*, 94132–94148. https://doi.org/10.1109/ACCESS.2021.3092304

[27] Mirjalili, M. A., Aslani, A., Zahedi, R., & Soleimani, M. (2023). A comparative study of machine learning and deep learning methods for energy balance prediction in a hybrid building-renewable energy system. *Sustainable Energy Research*, *10*(1). https://doi.org/10.1186/s40807-023-00078-9

[28] Mischos, S., Dalagdi, E., & Vrakas, D. (2023). Intelligent energy management systems: a review. In *Artificial Intelligence Review* (Vol. 56, Issue 10). Springer Netherlands. https://doi.org/10.1007/s10462-023-10441-3

[29] Morcillo-Jimenez, R., Mesa, J., Gómez-Romero, J., Vila, M. A., & Martin-Bautista, M. J. (2024). Deep learning for prediction of energy consumption: an applied use case in an office building. *Applied Intelligence*, *54*(7), 5813–5825. https://doi.org/10.1007/s10489-024-05451-9

[30] Nayak, A., Matta, G., & Uniyal, D. P. (2023). Hydrochemical characterization of groundwater quality using chemometric analysis and water quality indices in the foothills of Himalayas. In *Environment, Development and Sustainability* (Vol. 25, Issue 12). Springer Netherlands. https://doi.org/10.1007/s10668-022-02661-4

[31] Nayak, G. H. H., Alam, M. W., Singh, K. N., Avinash, G., Kumar, R. R., Ray, M., & Deb, C. K. (2024). Exogenous variable driven deep learning models for improved price forecasting of TOP crops in India. *Scientific Reports*, *14*(1), 1–26. https://doi.org/10.1038/s41598-024-68040-3

[32] Onteru, R. R., & Sandeep, V. (2024). An intelligent model for efficient load forecasting and sustainable energy management in sustainable microgrids. *Discover Sustainability*, *5*(1). https://doi.org/10.1007/s43621-024-00356-6

[33] R. Singh, A., Kumar, R. S., Bajaj, M., Khadse, C. B., & Zaitsev, I. (2024). Machine learning-based energy management and power forecasting in grid-connected microgrids with multiple distributed energy sources. *Scientific Reports*, *14*(1), 1–23. https://doi.org/10.1038/s41598-024-70336-3

[34] Saeed, M. A., Eladl, A. A., Alhasnawi, B. N., Motahhir, S., Nayyar, A., Shah, M. A., & Sedhom, B. E. (2023). Energy management system in smart buildings based coalition game theory with fog platform and smart meter infrastructure. *Scientific Reports*, *13*(1), 1–17. https://doi.org/10.1038/s41598-023-29209-4

[35] Sahin, M. E., & Ozbay Karakus, M. (2024). Smart hydropower management: utilizing machine learning and deep learning method to enhance dam's energy generation efficiency. *Neural Computing and Applications*, *36*(19), 11195–11211. https://doi.org/10.1007/s00521-024-09613-1

[36] Sari, M., Ali Berawi, M., Zagloel, T. Y., Madyaningarum, N., Miraj, P., Pranoto, A. R., Susantono, B., & Woodhead, R. (2023). Machine Learning-Based Energy Use Prediction for the Smart Building Energy Management System. *Journal of Information Technology in Construction*, *28*(June), 622–645. https://doi.org/10.36680/j.itcon.2023.033

[37] Sarker, I. H. (2021). Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. *SN Computer Science*, *2*(6), 1–20. https://doi.org/10.1007/s42979-021-00815-1

[38] Seyedzadeh, S., Rahimian, F. P., Glesk, I., & Roper, M. (2018). Machine learning for estimation of building energy consumption and performance: a review. *Visualization in Engineering*, *6*(1). https://doi.org/10.1186/s40327-018-0064-7

[39] Shapi, M. K. M., Ramli, N. A., & Awalin, L. J. (2021). Energy consumption prediction by using machine learning for smart building: Case study in Malaysia. *Developments in the Built Environment*, *5*(December 2020), 100037. https://doi.org/10.1016/j.dibe.2020.100037

[40] Sharma, V. (2022). Exploring the Predictive Power of Machine Learning for Energy Consumption in Buildings.

[41] Tabian, I., Fu, H., & Khodaei, Z. S. (2019). A convolutional neural network for impact detection and characterization of complex composite structures. *Sensors (Switzerland)*, *19*(22), 1–25. https://doi.org/10.3390/s19224933

[42] Technology, M. O. F., & Engineering, C. (2024). DESIGN AND IMPLEMENTATION OF MACHINE LEARNING BASED SMART BUILDING ENERGY. 9(4), 604–611.

[43] Vadruccio, R., Siragusa, C., & Tumino, A. (2023). Increasing energy efficiency in Smart Building through Internet of Things retrofitting intervention. *Procedia Computer Science*, *219*(2022), 263–270. https://doi.org/10.1016/j.procs.2023.01.289

[44] Wang, B., Tan, Z., Sheng, W., Liu, Z., Wu, X., Ma, L., & Li, Z. (2024). Identification of Groundwater Contamination Sources Based on a Deep Belief Neural Network. *Water (Switzerland)*, *16*(17). https://doi.org/10.3390/w16172449

[45] Wang, X., Wang, H., Bhandari, B., & Cheng, L. (2024). AI-Empowered Methods for Smart Energy Consumption: A Review of Load Forecasting, Anomaly Detection and Demand Response. In *International Journal of Precision*

Engineering and Manufacturing - Green Technology (Vol. 11, Issue 3). Korean Society for Precision Engineering. https://doi.org/10.1007/s40684-023-00537-0

[46] Zhao, Y., Qu, R., Xing, Z., & Lu, W. (2020). Identifying groundwater contaminant sources based on a KELM surrogate model together with four heuristic optimization algorithms. *Advances in Water Resources*, *138*(February), 103540. https://doi.org/10.1016/j.advwatres.2020.103540

[47] Zuccotto, M., Castellini, A., Torre, D. La, Mola, L., & Farinelli, A. (2024). Reinforcement learning applications in environmental sustainability: a review. In *Artificial Intelligence Review* (Vol. 57, Issue 4). Springer Netherlands. https://doi.org/10.1007/s10462-024-10706-5