J. Electrical Systems 20-11s (2024): 4439-4453

[1]Shibna Hussain,
[2]Santosh Kumar Sharma[2]

Electricity Demand Forecasting For Residential Community: A Comparative Study of RF, KNN and DT Models

Abstract
Forecasting energy demand is crucial for residential communities in institutional buildings to optimize energy use and reduce costs. This study aims to develop a machine learning-based model for forecasting the energy demand of a staff quarter residential community in a campus of an institutional building in Kota, India. K-Nearest Neighbors, Decision Tree and Random Forest are selected for comparison because of their interpretability, computational efficiency and suitability for handling non-linear relationships in electrical demand forecasting. The model's performance has been evaluated using MAE and R2 performance metrics. The results show RF forecast outperforms with the lowest MAE of 0.3 compared to KNN (0.4) and DT (0.42) for optimizing energy management strategies within the site. The actual vs predicted load plots, error trend plots and error distribution plots help to understand model behaviour by means of dynamic prediction topologies. The integration of weather data analysis showcases seasonal patterns that affect energy demand. This work contributes to the growing field of smart grid optimization and sustainable energy management by offering a data-driven approach to forecasting energy use on-site for institutional residential settings which is a unique context as most existing studies focus on commercial or urban residential areas. A comparison with existing forecasting techniques shows that the compared model outperforms existing models in terms of accuracy thus providing a solution associated with peak load forecasting and opening the door to better smart grid performance.

Keywords: Load Forecasting, Machine Learning, Regression Models, Demand Side Management, Smart Grid Optimization, Sustainable Energy.

Nomenclature:

Acronyms		R^2	Coefficient of Determination
AI	Artificial intelligence	RF	Random forest
ANN	Artificial neural network	RMSE	Root Mean Square Error
CNN	Convolutional neural network	**Symbols**	
DNN	Deep neural networks	$l(t)$	Load at time t
DSM	Demand side management	$l(t+p)$	Predicted load
DT	Decision tree	$M(t)$	Temporal and calendar features.
GRU	Gated recurrent units	$N(t)$	Weather and external features
KNN	K Nearest Neighbor	n	Total number of predictions
LSTM	Long short-term memory	p	Forecast period
MAE	Mean absolute error	R^2	Coefficient of Determination
MAPE	Mean Absolute Percentage Error	y_i	Actual value
ML	Machine learning	$\hat{y}_i$	Predicted value
NMSE	Normalized Mean Square Error		

Highlights

1. To compare a machine learning-based model to predict energy demand for a residential community staff quarter in a campus of institutional buildings in Kota, India.

2. Time-series data management in ML using Python to predict future energy consumption trends.

3. The compared model outperforms existing accuracy forecasting techniques offering a better solution for peak load forecasting and enhancing smart grid performance.

1Department of Renewable Energy, Rajasthan Technical University Kota, India-324010

2Department of Electrical Engineering, Rajasthan Technical University Kota, India-324010

*Correspondence: rtushibna@gmail.com, (M): +91-77269284

1. Introduction

With significant implications on the environment the load profile has become more demanding and changeable due to the exponential development in the use of electrical devices. Reducing waste and improving grid stability depend on matching power demand with generation capacity. Power distribution shortage may be effectively managed with the use of accurate power load forecasts. To identify when a load increase will occur and to take appropriate action a load prediction is necessary.

Demand projections are also essential for propelling urbanization and industrialization [1], [2], [3]. Forecasting electrical loads is a crucial tool for efficiently managing power systems [4], [5], [6]. It enables utility providers to make informed decisions regarding the generation, transmission and distribution of energy [7], [8]. Load forecasting models can be classified based on the forecasting horizon with short-term models focused on timeframes ranging from hours to weeks medium-term models covering weekly to yearly periods and long-term models applicable to time windows exceeding a year [9], [10]. Even though there are many statistical models available for electrical load prediction achieving high estimate precision is still quite difficult [11]. Demand has been modeled and forecasted using a variety of statistical functions including partial linear models, autoregressive average models, linear regression and gray models are often employed [11]. This paper also examines how load forecasting has changed over time moving from mathematical to hybrid forecasting techniques that incorporate clever strategies for dealing with intricate and nonlinear problems [12]. Although effective for steady-state forecasting traditional approaches frequently struggle with the intricacy of variable loads and neglect the non-linear connections that are prevalent among contemporary energy systems. Because of this, there is a need for sophisticated data-driven forecasting techniques that are supported by artificial intelligence (AI) and machine learning (ML).

Using three years of real-time data from the Wavi substation in India [13] assessed machine learning-based electrical load prediction using regression learners obtaining 99.99% correlation between actual and anticipated loads using 14 regression methods The limits of statistical approaches particularly in dealing with long-term dependencies and nonlinear dynamics prompted researchers to investigate machine learning (ML) techniques for load forecasting in Table 1.

Table 1 Review based on different ML techniques

Authors	ML Techniques	Dataset	Forecasting Horizon	Building Type	Key Findings
Díaz et al. 2018 [14]	Random Forest (RF), Gradient Boosting	Electricity consumption data	1-day (Short-term)	Commercial	RF model showed superior performance with low MAE and high R^2 score.
Huang et al. 2019 [15]	LSTM (Long Short-Term Memory)	Energy demand data from grid	1-hour (Short-term)	Residential, Industrial	LSTM effectively captures temporal dependencies in energy demand.
Zhou et al. 2020 [16]	Hybrid (ANN + PSO)	Historical load data	7 days (Short-term)	Residential, Commercial	Hybrid ANN + PSO performed better than standalone ANN models.
Sahoo et al. 2020 [17]	K-Nearest Neighbor (KNN)	Daily energy consumption	24 hours (Short-term)	Residential, Commercial	KNN showed moderate performance with lower MAE but struggles with complex patterns.
Kumar et al. 2021 [18]	CNN (Convolutional Neural Network)	Smart meter data	1 day (Short-term)	Residential, Industrial	CNN applied successfully to time-series forecasting capturing spatial dependencies in smart meter data.
Wang et al. 2022 [19]	LSTM, GRU (Gated Recurrent Units)	Smart grid data, weather data	1 hour (Short-term)	Residential, Commercial	LSTM and GRU models outperformed traditional models, with GRU being more efficient for shorter horizons.
Li et al. 2023 [20]	Transformer Network, LSTM	Energy consumption data	12 hours (Short-term)	Residential, Commercial	Transformer-based models excelled in capturing long-term dependencies and achieving low error.
Zhao et al. 2023 [21]	Deep Neural Networks (DNN)	Multi-year load data	7 days (Short-term)	Residential, Commercial	Deep Neural Networks showed strong generalization, outperforming traditional forecasting methods.

Liu et al. 2024 [22]	Hybrid Model (RF + LSTM)	Historical load data, weather	24 hours (Short-term)	Residential, Commercial	Hybrid models combining RF and LSTM achieved the lowest MAE and high R^2 for both building types.

The utilization of machine learning algorithms has become increasingly prominent in the field of electrical load forecasting due to their proven capability to effectively model the complex, nonlinear relationships and intricate patterns inherent in energy consumption data [23]. In particular researchers have explored ML techniques for electrical load forecasting including traditional statistical methods like regression models and time series analysis as well as more advanced approaches such as decision trees, random forests and k-nearest neighbor algorithms.

It is motivated towards the crucial need to bridge the gap between energy demand and generation in real time. As the background of energy gets more complicated with the emergence of renewable energy sources accurate and adaptable load forecasting systems become critical.

This study intends to help construct a durable and intelligent grid system capable of fulfilling the needs of today's energy infrastructure with the aims to assist the advancement of a robust and intelligent grid system competent of fulfilling the demands of a modern energy infrastructure. The objectives of the study are:

1. To investigate power load forecasting through the use of artificial intelligence techniques like machine learning.
2. The goal of the study is to compare a model design for power load forecasting by doing experimental analyses on datasets using different machine learning methods.
3. The use of three algorithms and comparing by applying AI-based load forecasting to manage peak-load prediction complexity is a novel contribution which can improves the existing techniques and increases the smart grid's dependability by means of dynamic prediction topologies.

1.1 Novelty and Key Contributions

The novelty of work can be summarised with the following points:

1. Accuracy is improved as compared to traditional forecasting methods by using advanced machine learning algorithm, integration of weather data, dynamic prediction topologies to forecast energy demand in real-time.
2. A Smart grid optimisation and energy management has been provided for staff quarter residential community in campus is a unique context as most studies are done in commercial and urban residential areas within specific operational conditions.
3. The enhanced energy efficiency and reduced operational cost contribute to sustainable management of energy resources by providing data-driven solutions to peak load forecasting.

This research contributes to the ongoing advancements in the field of electrical load forecasting particularly by exploring the application of machine learning techniques to model the energy consumption patterns of staff quarters residential community in institutional buildings in the Kota region of India. The main contribution of this study is twofold. First, it compares the performance of the RF, KNN and DT models in projecting short-term energy load emphasizing their respective strengths and limitations. Second, the study presents an architectural upgrade by integrating these ML models into an internet-enabled gateway allowing for real-time load predictions. This design not only improves forecast accuracy but also allows service providers to respond immediately to demand changes lowering the probability of overloads and improving grid stability. Furthermore, the suggested technology is highly applicable across a variety of building types and geographic sizes making it ideal for large-scale implementation in smart cities and decentralized energy networks.

The paper is structured as follows. Section 2 describes the suggested technique including the ML algorithms used, data preparation and the architecture for real-time forecasting. It also covers the experimental setup which includes the characteristics of the dataset and model evaluation metrics. Section 3 analyses the outcomes of each ML model comparing MAE and R^2 scores to determine their usefulness. Finally, Section 4 summarizes major findings and suggests future research areas focusing on the potential of hybrid models and real-time adaptive forecasting for smart grid applications.

2. Materials and Methods

2.1 Site Selection

The framework employs a dataset derived from staff quarters in Rajasthan Technical University (RTU) Kota in Fig. 1 at latitude and longitude 25.143899°N, 75.808871°E with the statistics partitioned into training of 70% and testing of 30% subsets. The training and validation of the model utilizing three distinct regression methodologies was conducted. The real-time load forecasting system analyzes more than 15,000 data points of electricity generation from year 2024-2025 facilitating strong predictions informed by high-frequency measurements. The total staff quarters for teaching staffs are A type-01, type-12, C type-16 and D type-34. The A and B type quarters are single story and C&D types are double story quarters.

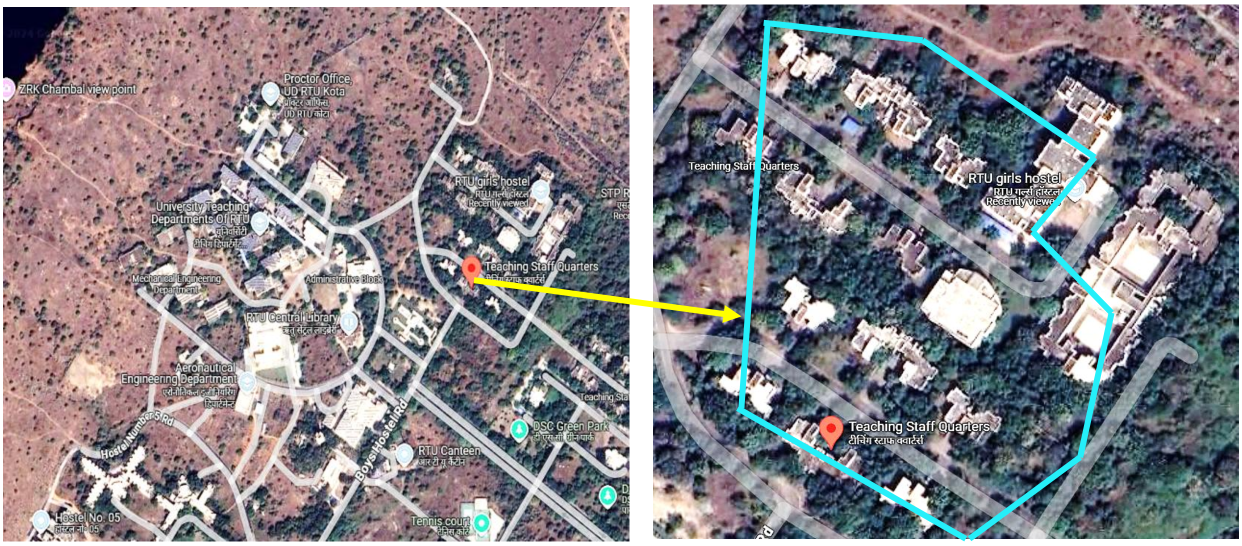

Fig. 1 Residential Community Teaching Staff Quarters at RTU

2.2 Mathematical and ML Modelling

2.2.1 Demand Side Management (DSM) using Machine Learning

Machine Learning (ML)-based regression algorithms are ideal for Demand Side Management (DSM) applications involving electrical load forecasting where the aim is to properly estimate future power demand. Accurate load forecasting enables utilities to optimize demand-side management by employing tactics like load shifting, peak shaving and demand response. Here are several ML-based regression approaches used for DSM in load forecasting, along with their advantages and typical applications.

In our suggested design the DSM system uses supervised machine learning to forecast and control demand. Three machine learning regression models are used in our methodology:

a. Random Forest (RF) Regression

The Random Forest method is a prominent ML tool for sorting and reversion problems. It is part of the ensemble learning approach which combines numerous decision trees produced during training to make a more accurate and reliable prediction. Random forests are designed to take use of trees' intrinsic variability by including randomization into the selection of both data points and parameters. Instead of selecting one of the generated trees the final forecast is made by combining them. A collection of random trees constructed across n_{tree} bootstrap samples is aggregated in this process r_{tree} of the training set $r^1 \ldots\ldots r^{n_{tree}}$. The sum of all the estimators that come from those trees, represented by $\hat{E}_1 \ldots . \hat{E}_{n_{tree}}$ is the final learning rule. The aggregation process involves calculating the average prediction value in order to create a forecast at a new point x as in Eq. (1):

$$\hat{E}_{RF}(x) = \frac{1}{n_{tree}} \sum_{s=1}^{n_{tree}} \hat{E}_s(x) \qquad (1)$$

b. K-Nearest Neighbor (KNN) Regression

KNN regression makes predictions by averaging the target values of the k closest neighbors of a given data point. They are commonly used for short-term forecasting in DSM due to its simplicity especially when the data has clusters or recurring patterns. Memory-intensive as it requires storing all data points, and computationally expensive for large datasets. The nearest historical point is , J_k is set of k nearest neighbor and m_i is the actual load value of the neighbor. The predicted load value is given by Eq. (2)

$$\hat{m}_{new} = \frac{1}{k} \sum_{i \epsilon J_k} m_i \qquad (2)$$

c. Decision Tree (DT) Regression

The decision tree approach in machine learning identifies patterns in data to anticipate events. The objective is to design an ideal decision tree with minimal generalization error. The Decision Tree divides the dataset into regions by splitting features at certain thresholds which helps in learning non-linear relationships. The decision tree is given a fresh input $D = (D_1, D_2, \ldots. D_n)$, it proceeds along the route that the feature splits decide until it reaches a leaf node $\bar{C}_{leaf}$. $\hat{C}$ is the target electrical load (kW). Dataset is $D = \{(D_1, C_1), (D_2, C_2), \ldots. (D_n, C_n), \}$. The forecasted load as in Eq. (3)

$$\hat{C} = \bar{C}_{leaf} \tag{3}$$

Useful for both short and long-duration load forecasting especially in complex and noisy data environments where understanding feature importance is beneficial. Simple to interpret works well on large datasets and is effective for non-linear relationships prone to overfitting especially with deep trees though cropping can mitigate this issue.

2.2.2 Mathematical Formulation of the Model

The $l(t)$ represent the load at time t. To forecast $\mathrm{l}(\mathrm{t}+\mathrm{p})$ where p is the forecast period.
The model could be defined as Eq. (4):

$$\hat{l}\,(t+p) = f\,(l(t), l(t-1) \ldots, l(t-v), M(t), N(t)) \tag{4}$$

$\hat{l}\,(t+p)$ is the predicted load. $\hat{l}\,(t), l(t-1) \ldots, l(t-v)$ are historical load values up to v previous timesteps. $M(t)$ includes temporal and calendar features. $N(t)$ includes weather and other external features.

2.2.3 Model Performance Metrics in DSM

The matrices NMSE, RMSE, MAE and MI are used for model visualisation. The assessment metrics are coefficient of determination R2 and mean absolute error (MAE) along with training time is taken into consideration [24]. As R^2 approaches 1 and MAE approaches 0 the model becomes more accurate.

a. Mean Absolute Error (MAE)

The average of the absolute discrepancies between projected and actual values is MAE [25]. The overall number of mistakes in a series of forecasts without taking into account their direction measured as in Eq (5)

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y_i}| \tag{5}$$

Where n is the total number of predictions, y_i is the actual value, $\hat{y_i}$ is the predicted value. A lower MAE indicates a model that is closely aligned with the actual values making it highly effective for assessing the overall prediction accuracy of models used in DSM applications. MAE is particularly useful because it is easy to interpret and isn't overly affected by outliers.

b. R^2 Score (Coefficient of Determination)

The predictability of dependent variable from independent variable is indicated by the R^2 score [26], also known as Coefficient of Determination or Goodness of fit as in Eq (6).

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y_i})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{6}$$

where $\bar{y}$ is the mean of the actual values. An R^2 score closer to 1 suggests a strong model fit where most of the variation in the load data is explained by the model. R^2 is a widely used metric in DSM forecasting as it provides insight into the model's explanatory power indicating whether the model captures both overall trends and specific variations in the demand.
Using these criteria allows to discover models that are not only accurate but also stable and trustworthy which is critical for DSM's load forecasting requirements [27].

2.3 Methodology

The forecast methodology is demonstrated in Fig. 1 where a machine learning-based novel architectural configuration to forecast the energy demand of an staff quarters site of an institutional building in Kota, India. The different phases are as follows: (1) online/offline dataset collecting; (2) data preparation process; (3) the mining of features; (4) most appropriate attribute selection; (5) AI model construction; (6) predicted value extraction and (7) outcome evaluation. After dataset collection comes data preprocessing to reject the spikes and replace the missing values introduced due to climatic conditions. Then follows extraction of features step which is lined up by characteristic selection which affects ML model in forecasting to load. The developed model is trained and tested to validate the performance and further results compared for forecasting.
The suggested design depicted in Fig.2 includes an internet-enabled gateway in the smart grid enabling for real-time data analysis and modeling in Python. This gateway configuration improves the system's potential for continuous load prediction and gives utility providers timely insights for better demand-side management. Each model's performance is evaluated based on two essential metrics: Mean Absolute Error (MAE) and R-squared (R^2) score which measure prediction accuracy and explanatory power. This technique seeks to produce high-accuracy flexible forecasting solutions that can be scaled for various building types and grid scenarios by embedding ML models in an internet-connected infrastructure.

2.4 Data Utilisation

2.4.1 Data Collection

Three distinct datasets are employed in study. The initial dataset pertains to the historical record of power demand of the staff quarters residential community at RTU, Kota has been obtained from the KDEL's previous electricity bills. The preceding dataset of weather conditions that contains outside temperature, temperature at dew point, wind speed average and last dataset contains the hourly energy prices that effect load pattern. All the three models were trained and tested with same dataset to get a fair comparison among the three.

Table 2 Selected dependent and independent parameters

Description—Type	Variable	
The estimated hourly load for the system (kW)	HL	Dataset 1
Previous year load at the same time (kW)	PY	
Last week load for the same time (kW)	LW	
Outside temperature (°C)	OT	Dataset 2
Dew point temperature (°C)	DPT	
Real humidity (%)	RH	
Average wind speed (km/h)	AWS	
Outside air pressure (kPa)	OAP	Dataset 3
KDEL hourly energy price (INR/kWh)	KDEP	

Fig.2 Comparative approach for load forecasting

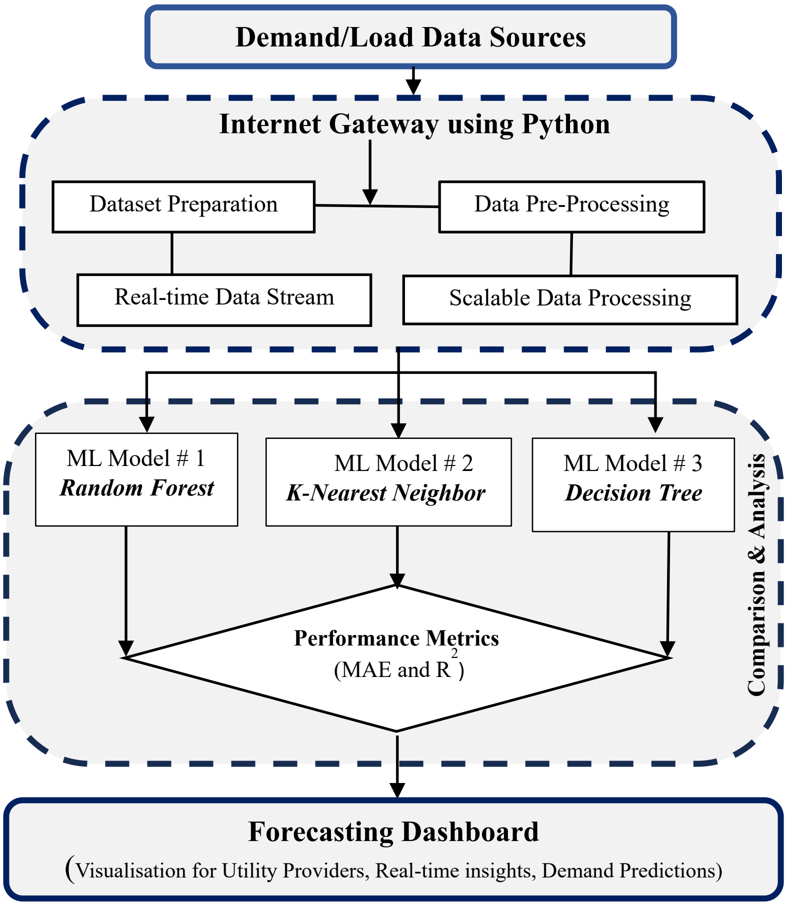

Table 2 provide the dependent and independent parameters. In this figure ML Model#1, Model#2 and Model#3 represent ML models analysed in paper. The datasets compiled table 2 parameters as shown in Fig. 3 actual load of previous year, electrical load of staff quarter collected from January 2024 till January 2025. This data is further trained and tested for forecasting.

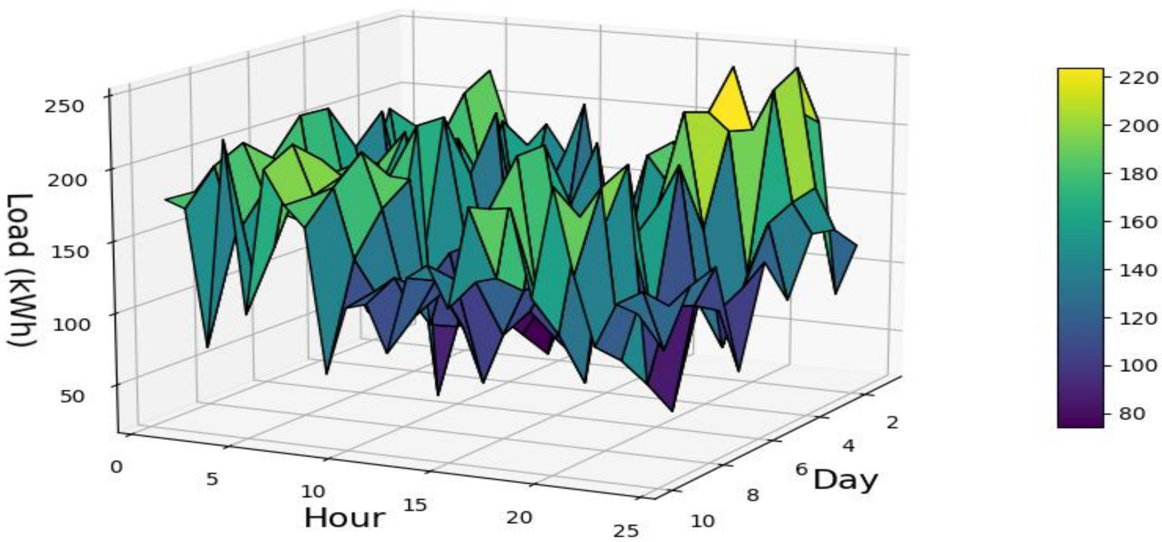

Fig.3 PY-HL-LW actual load 3D profile in day-hr from January 2024 till January 2025

2.4.2 Data Preprocessing and Forecasting Set-up

The preprocessing phase entails cleaning the raw data collected from the Staff Quarter electricity consumption and adding new characteristics such as lagged values which can give useful insights into the seasonal patterns of energy usage [28].

The detailed steps of data preprocessing for this load forecasting are as follows.

3.4.2.1 Data cleaning

Meteorological conditions and historical demand data are the two primary criteria on which load forecasting is based, together with temporal parameters such as hourly rate of the day, daily and days annually. The goal is to forecast the output parameters using input variables of electrical loads selected for their intrinsic association with fluctuations across time intermissions [29], [30]. Specifically, the time of day makes it less complicated to identify daily variations. The day of the week indicates weekly developments whereas the day of the year permits identify seasonal characteristics.

3.4.2.2 Regression

This supervised machine learning technique predicts continuous variables by connecting them and analysing how each one affects the others [31]. When analysing regression algorithm predictions, it is necessary to consider variance and bias measures. In this model random samples of 250 days hourly readings are analysed from dataset 1 and mapped with 50^{th} hour in such case all the variables will represent the relationship with the selected hour and produce the first reading of new dataset.

3*.4.2.3. Data integration and Correlational Analysis*

The data integration is done by merging data from several sources into a sole dataset [32]. This involves schematic integration which is the process of combining data from several sources and resolving data value disagreements caused by disparities between units of measurement, representation and other variables. Furthermore, duplicated data must be controlled.

3.4.2.4. Scaling

All datasets were scaled from minimum to maximum. To do this the training set was scaled. The test set's attributes were scaled for predicting with the same scaler. The resulting forecast was then inversely scaled back to the initial data representation.

3.4.2.5 Train and Test

The rigorous comparison of the three ML models included in the study is an important part of our approach. Mean Absolute Error (MAE) and the R-squared (R^2) score are two performance evaluation measures that offer an extensive overview of each model's accuracy and capacity to account for data variance. To imitate a real-world situation the dataset is split into training and testing sets with 30 % dedicated to testing and 70 % to training.

3. Results and Discussion

3.1 Weather data analysis

Weather data in Fig.4, Fig.5 and Fig.6 such as temperature, solar irradiance, wind speed and humidity substantially influence on power consumption.

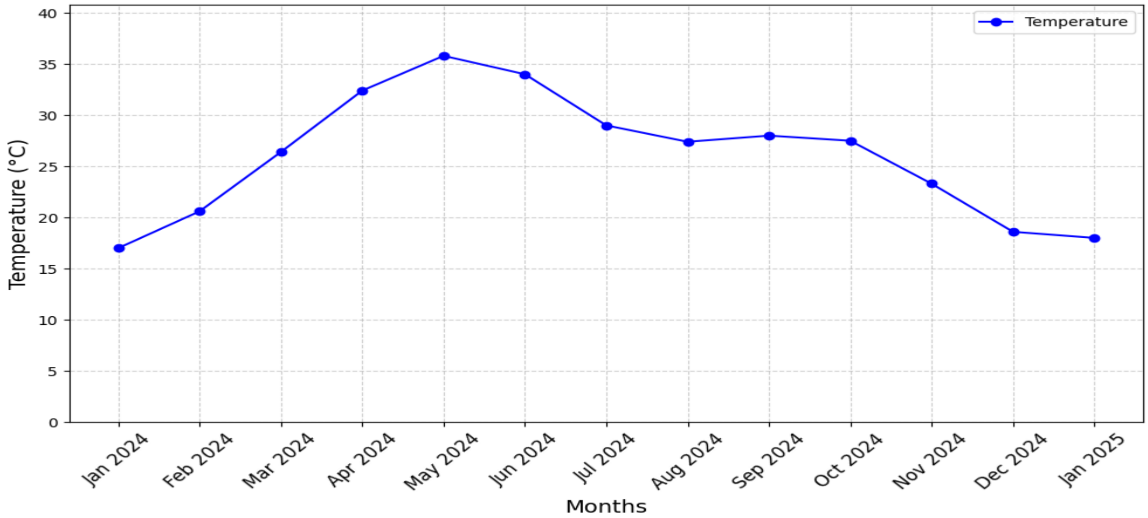

Fig.4 OT-Temperature profile

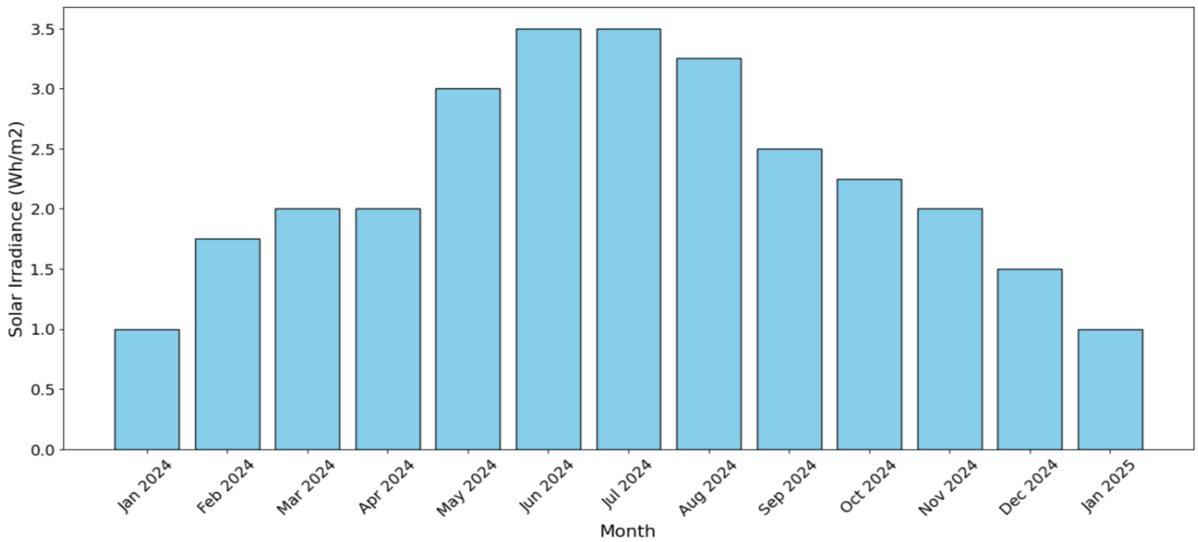

Fig.5 AWS-Average Wind Speed Pattern

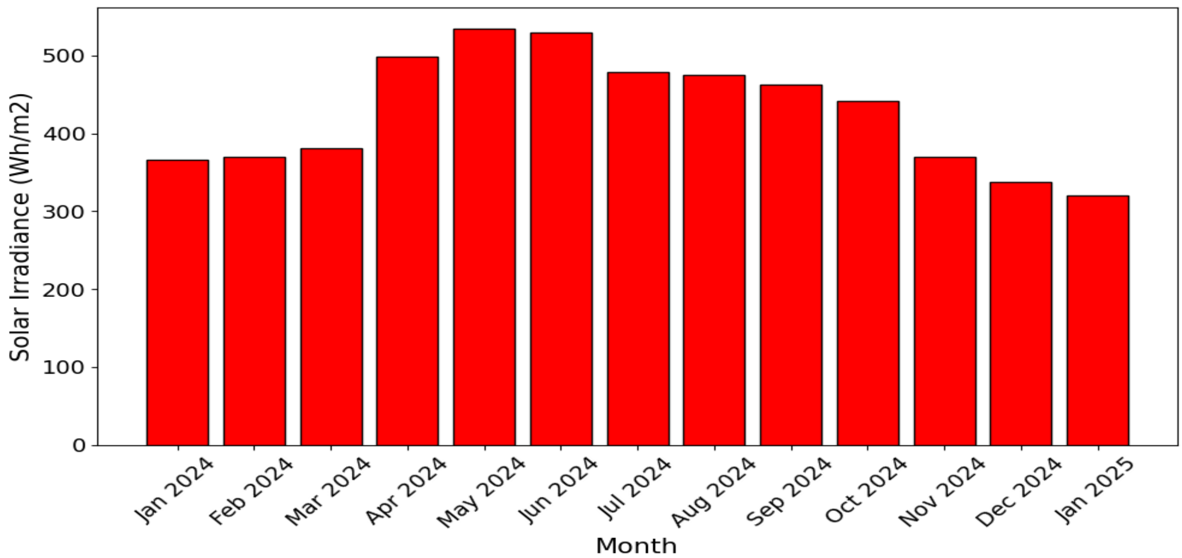

Fig.6 Solar Irradiance (Wh/m^2)

3.2 Model Validation and Comparison to Existing Techniques

In Table 3 RF indicates the best generalisation in maintaining the balance between training and testing. KNN performs reasonably well but struggles with complex patterns especially in commercial buildings as reflected by its lower testing accuracy. Decision Tree (DT) suffers from overfitting and has poor R^2 performance making it less reliable than RF and KNN. KNN shows little less testing accuracy on the other hand DT demonstrates the largest gap in testing and training. ARIMA and Linear Regression less effective than RF. Loss of training and testing are lowest for RF (0.15, 0.56) thus is more effective in addition to its suitability for residential, commercial as well as mixed-use buildings.

Table 3 Comparison of MAE Values for Compared model with Existing ML Methods

Model / Method	**MAE**	**R^2 Score**	**Training Accuracy**	**Testing Accuracy**	**Training Loss**	**Testing Loss**	**Building Type**	**Existing Methods**
Random Forest (RF)	0.3	0.62	95%	80%	0.15	0.56	Residential, Commercial	ARIMA, Naive Model
K-Nearest Neighbor (KNN)	0.4	0.61	92%	75%	0.61	0.67	Residential, Commercial	Linear Regression, Moving Average
Decision Tree (DT)	0.42	0.6	90%	70%	0.2	0.51	Residential, Mixed-Use	Naive Model, ARIMA
ARIMA	9.39	0.3-0.7	93%	92%	0.23	0.45	Residential	Naive Model, Linear Regression
Linear Regression	3.43	0.2-0.5	94%	77%	0.17	0.38	Residential, Small Commercial	ARIMA, Moving Average

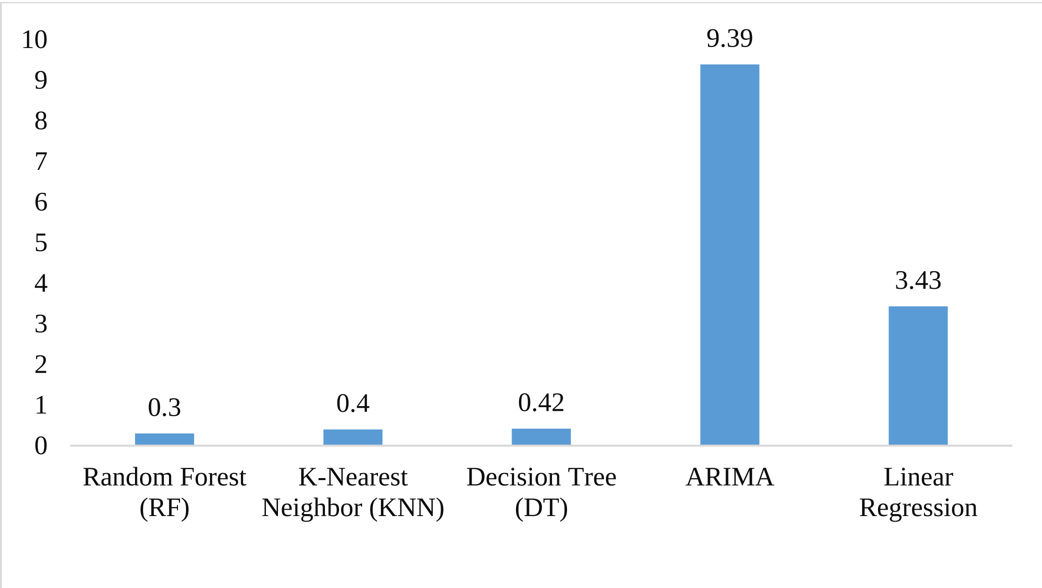

Fig.7 Comparison of MAE values

The lowest MAE of RF in Fig.7 indicates that its predictions are the most accurate because of its ability to cope with noise and variability resulting in the best fit for dataset thus outperforming ARIMA and Linear Regression models. With moderate MAE the KNN's emphasis on proximity to neighbors can contribute to inaccuracies if the dataset contains outliers or non-linear shifts. DT perform better that KNN but less that RF as it can model non-linear acquaintances its overfitting causes more prediction errors. Existing ML models ARIMA (9.39) and Linear Regression (3.43) have substantially larger MAE values showing poor performance utilizing this dataset.

3.3 Train test of RF, KNN and DT

In Table 4 comparison of actual to predicted values of electrical load from the three ML models are shown only for 10 data points out of 15,000 data points . Random forests (RF) actual values are ensemble that comprise a voting committee consisting of n binary decision trees. Each tree is built from a casually picked subset of the original training data. This is because single decision trees are susceptible to overfitting on training data. The next step is to average the forecasts of each tree to get the final prediction as the RF output values. Random Forest (RF) provides excellent training accuracy and reasonable testing accuracy though it could benefit from better model generalization (as indicated by its low R^2 score). Reading 173 shows a real value of 0.0166 but RF predicts 1.875 which is the most precise of the models. The KNN classifier detects the nearest neighbor and then provide the predicted values corresponding to all actual values after training and testing with KNN algorithm.

Table 4 Random Forest, KNN and Decision Tree Output Values

Reading	Actual	RF Predicted	KNN Predicted	DT Predicted
173	0.0166	1.875	0.190	0.0211
192	0.0114	0.900	0.117	0.0072
133	0.0163	1.253	1.383	0.0107
129	0.0131	0.762	1.186	0.0046
126	0.0175	1.634	0.128	0.0177
140	0.0133	1.042	0.102	0.0086
184	0.0201	1.468	0.135	0.0143
186	0.0181	1.370	1.496	0.0133
176	0.0192	0.958	0.096	0.0096
177	0.0137	1.804	0.143	0.0222

Fig.8 shows the profile where the actual values as categorised from previous load data are followed by the predicted trend line for peak phase and off-peak values.

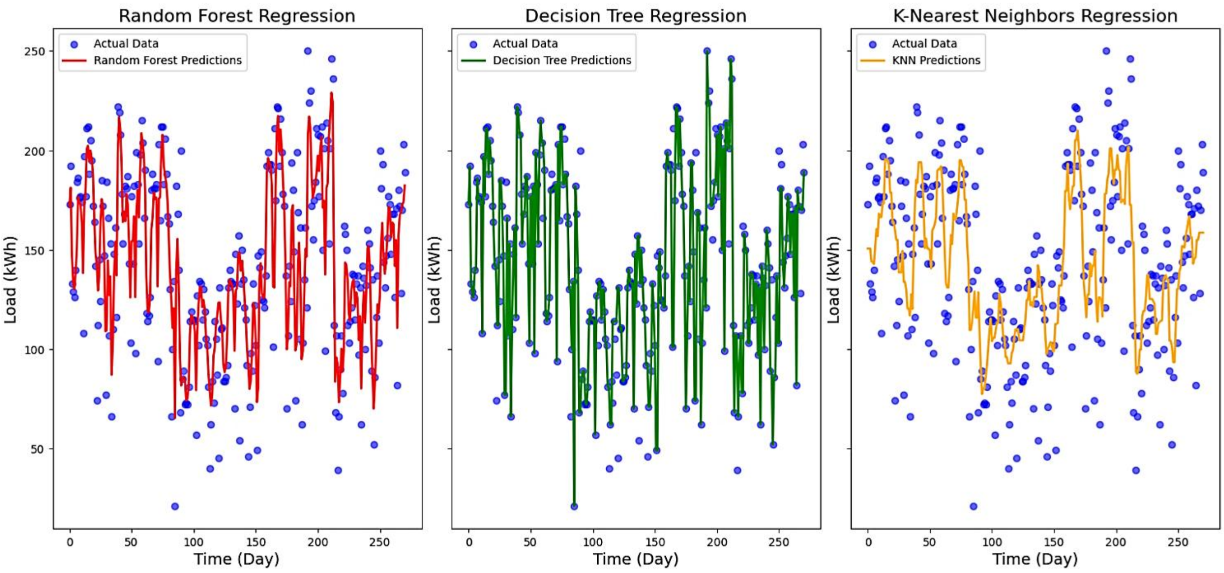

Fig. 8 Plot of actual data and predicted Load values by RF, KNN and DT algorithm

The result of actual vs predicted load values demonstrates that the forecasting curve nearly approaches the actual one thus the Random Forest algorithm predicts the electric load series fairly in Fig.8. The nearest neighbor of the actual value of k can decide the class of prediction and minimise the false predictions. The model evaluates different values to find optimal value. The KNN predicted values almost follow the actual load line obtained based on historical data fed to the model except at a few points where factors like noise or insufficient training data. The overall trend follows the peak-demand periods and low-demand phases predicted for the K-closest vectors. The pattern of the profile in Fig.8 helps to show that the decision tree model predicted load values across different magnitudes. Deviation of the predicted vector from the actual and might compromise in models accuracy for forecasting.

3.4 Outlier detection, Error prediction and Statistical parameters Analysis of RF, KNN and DT

While there is a general tendency of increasing real values with readings certain high numbers do not match expected values. The scatter figure indicates that forecasts typically cluster around actual values with some significant outliers. The outliers are identified in the dataset demonstrated in fig. 9. The points shown in red in the graphs are possible outliers because their residuals are more than two standard deviations apart from the mean.

In RF the residuals are densely concentrated around zero demonstrating accurate predictions with few errors. Because RF is an ensemble it averages predictions from several decision trees rendering it less susceptible to

outliers. KNN depends on the distance from neighbor data points and thus gets skewed by far values. In DT the outliers spread largely. Thus RF is best fit with robust predictions and minimum sensitivity for extreme values.
The error distribution in Fig.10 shows error distribution comparison (x-axis represent error values and y-axis represent the frequency of occurrence). It is observed that the prediction error narrows around zero in RF indicating RF has more consistent and stable predictions and offers better accuracy as compared to the two other models distributed widely with 0-50 error values for KNN shows less reliable prediction with moderate errors but does not always yield optimal performance. Moderate spread in DT leads to large errors which means it frequently overestimates energy load. So its investigated that the RF performs better with a more balanced and stable error distribution.

$$\text{Prediction error }\% = \frac{|\text{Expt.Value} - \text{Pred.Value}|}{\text{Expt.Value}} \times 100 \quad (4)$$

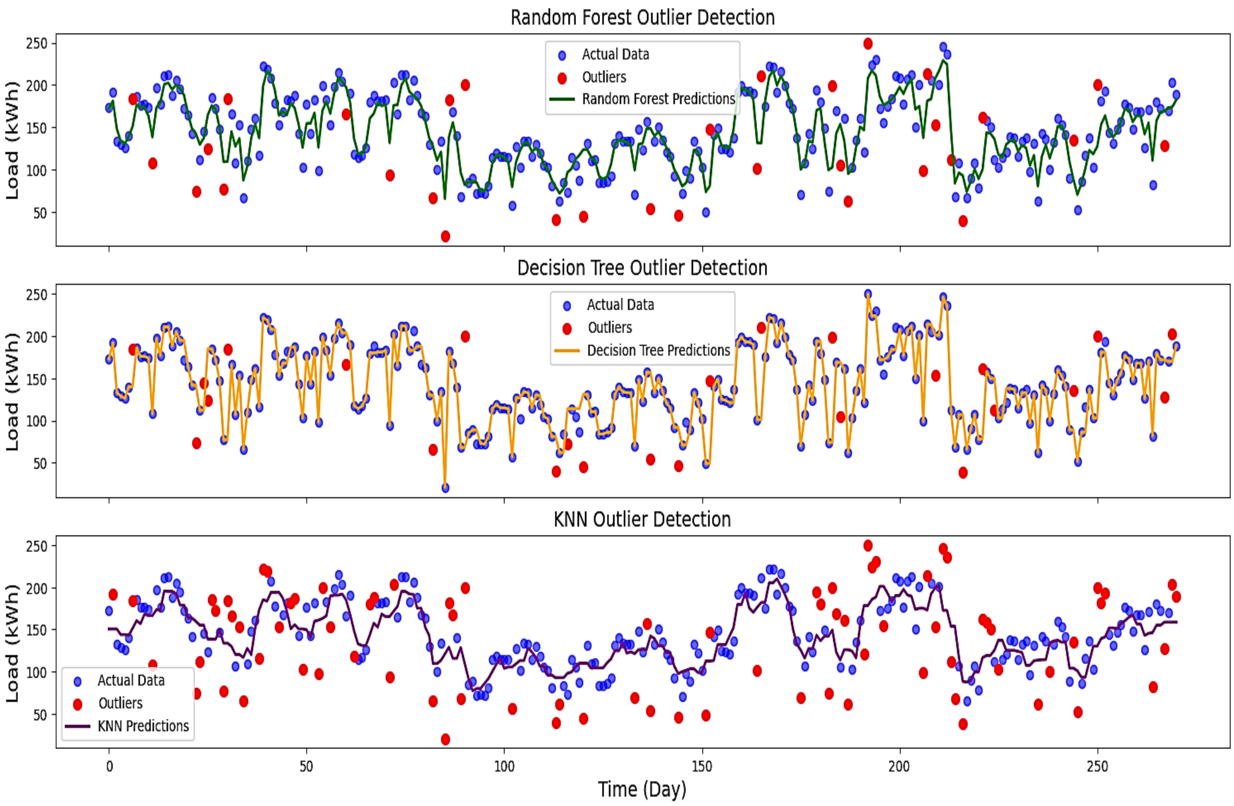

Fig.9 Outlier detection

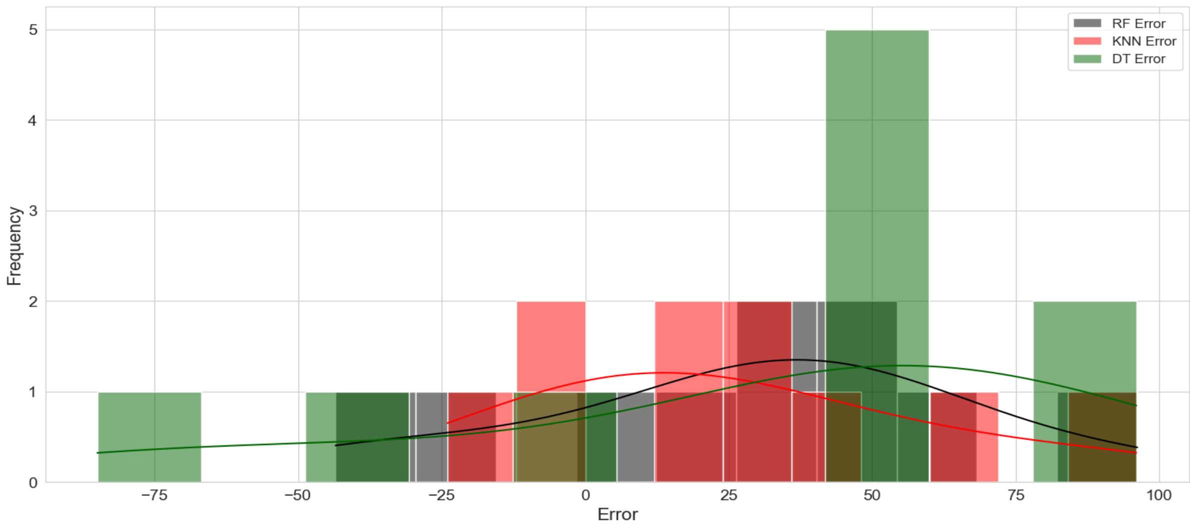

Fig.10 Prediction of error over time

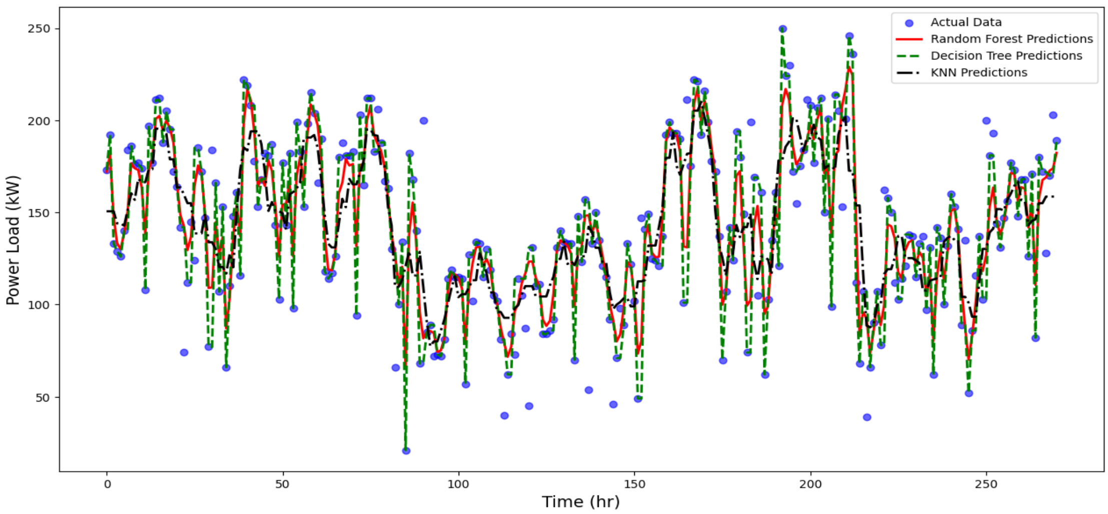

Fig.11 Actual vs predicted Load Scatter plot for ML techniques

The graphic in Fig.11 compares the three ML models in this study. In RF there are authentic predictions for normal and extreme electrical load values showcasing adaptability. DT produces moderate accuracy but is less accurate than RF as for noisy data points it deviates from diagonal giving mispredictions for extreme data points. KNN is more applicable to uniform distribution datasets as it is more sensitive to outliers. Prediction accuracy is observed in the order of RF > DT > KNN. RF is more robust because of its collaborative approach across various scenarios.

Table 5 Statistical Parameters Analysis

Statistic	Reading	Actual	Predicted	Residual
Count	10	10	10	10
Mean	888.80	1.91	1.57	0.34
Std Dev	388.02	1.92	1.40	0.84
Min	127.00	0.01	0.00	-0.65
25%	688.25	0.16	0.19	0.00
Median	989.00	1.76	1.70	0.6
75%	1137.00	2.39	2.30	0.44
Max	1360.00	5.15	4.27	2.49

According to Table 5 the RF, KNN and DT process the same number of data points for comparison. As per mean value the average load value is presented by 888.80, actual value 1.91 that correspond to the predicted value 1.57 the low because they demonstrate normalised value in ML models. The residual value 0.34 is small which demonstrates the better performance of models with very low error weightage. In standard deviation the range of load values is variable and is high as 388.02. The fluctuation has been smoothed out by reduced variability of actual (1.92) and predicted (1.40). Due to the reduction in variance and overfitting Random Forest technique achieved smallest residuals as it handles extreme values better the deviation in predictions. K-Nearest Neighbor model struggles with outliers and higher residual variance indicates its sensitivity to noise in the dataset thus effecting its performance. The large residuals for few data points make Decision Tree model prone to overfitting thus reducing robustness for complex distributions results in larger errors. Random Forest is the most reliable alternative for electrical load forecasting because of its consistently accurate and precise forecasts.

3.5 Assessment and Evaluation of Model Efficacy

In evaluating metrics like Mean Absolute Error (MAE) and the R^2 score provides insight into a model's accuracy, stability and reliability for electrical load forecasting in Demand Side Management (DSM) and how it helps assess model efficacy.

Table 6 Values of Evaluation Parameters

Parameter	Random Forest (RF)	K-Nearest Neighbor (KNN)	Decision Tree (DT)
MAE	0.3	0.4	0.42
R^2 score	0.62	0.61	0.6

Our analytical models exhibited differing degrees of precision in predicting load with the ensuing findings for Mean Absolute Error (MAE) and R2 coefficient as presented in Table 6. In Fig. 12 the Random Forest model exhibits superior performance regarding Mean Absolute Error (MAE) at 0.3 implying it yields the most precise predictions in this scenario. Nonetheless the markedly low R^2 value of 0.62 suggests a failure to capture a significant portion of the inherent variance within the dataset which may indicate potential overfitting or a deficiency in generalization capabilities.

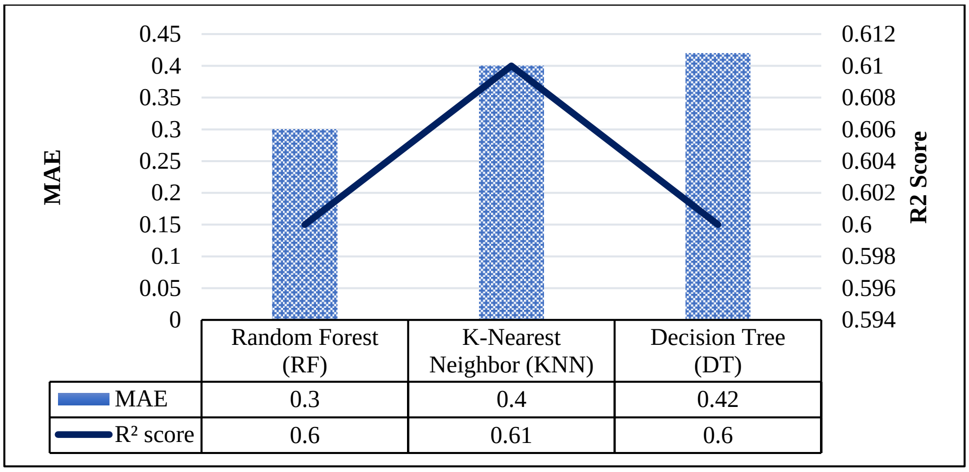

Fig. 12 Evaluation of Prediction models

In contrast, the K-Nearest Neighbor algorithm demonstrates a higher MAE of 0.4 compared to Random Forest implying a lower overall accuracy however its R^2 value of 0.61 is marginally improved suggesting it accounts for a somewhat greater extent of data variance. The Decision Tree method ranks the lowest in performance. It records the highest MAE at 0.42 among the three evaluated models and the R^2 value signifies inadequate performance as this model performs worse than a rudimentary baseline alternative. Although Random Forest is the most accurate model (MAE) there is room for improvement in its capacity to capture underlying variation as evidenced by the low R^2 value. Tuning the KNN and Decision Tree models can enhance MAE and R^2.

4. Conclusion

The results demonstrate a novel data-driven comparison approach that is highly effective in forecasting the energy consumption of the residential community in institutional buildings. The advanced machine learning architecture was able to capture the intricate non-linear relationships and temporal dependencies inherent in the energy usage data resulting in remarkably accurate predictions. The suggested load forecasting models (RF, KNN and DT) show higher accuracy. The Mean Absolute Error (MAE) at 0.42 for DT followed by the Random Forest model at 0.3. Random Forest and KNN models demonstrate improved prediction accuracy with higher R^2 values indicating better variance explanation. When the MAE values of the suggested models are compared to those of conventional models (e.g. Linear Regression, SVM, Fine Tree and Gaussian Process Regression) they considerably outperform existing techniques particularly in multi-building configurations. Random Forest, KNN and Decision Tree models produced much lower MAE values (0.3, 0.4 and 0.42 respectively) compared to higher MAEs from traditional methods in single-building cases. The concept incorporates the integration of weather data, dynamic prediction topologies that can execute machine learning algorithms in real time as well as an operational layer for dynamic load forecasting. This method predicts hourly power consumption based on real-time data which aids demand-side management and grid dependability.

The suggested forecasting system works well not just for individual buildings but also for foreseeing loads over multiple buildings which is essential for scaling to larger grid regions or complicated infrastructure. By delivering high-accuracy forecasting this model assists utility providers with developing proactive data-driven load balancing alternatives diminishing the probability of transmission line overload and enabling environmentally friendly energy management practices in the field of smart grid optimisation and sustainable energy management.
Further work can be extended to multiple regions or countries to validate to be applicable across diverse settings. Comparison with other models like Long short-term memory, Gradient Boosting Machines or Neural Networks can reveal computational efficiency.

5. Reference

[1] Li, L. L., Zhao, X., Tseng, M. L., & Tan, R. R. "Short-term wind power forecasting based on support vector machine with improved dragonfly algorithm," *Journal of Cleaner Production*, 242, pp. 118447, 2020. https://doi.org/10.1016/j.jclepro.2019.118447

[2] Fan, G. F., Guo, Y. H., Zheng, J. M., & Hong, W. C. "Application of the weighted k-nearest neighbor algorithm for short-term load forecasting," *Energies*, 12(5), pp. 916, 2019. https://doi.org/10.3390/en12050916

[3] Mosavi, A., Salimi, M., Faizollahzadeh Ardabili, S., Rabczuk, T., Shamshirband, S., & Varkonyi-Koczy, A. R. "State of the art of machine learning models in energy systems, a systematic review," *Energies*, 12(7), pp. 1301, 2019. https://doi.org/10.3390/en12071301

[4] Ahmad, N., Ghadi, Y., Adnan, M., & Ali, M. "Load forecasting techniques for power system: Research challenges and survey," *IEEE Access*, 10, pp. 71054–71090, 2022. https://doi.org/10.1109/ACCESS.2022.3187839

[5] Fathi, S., Srinivasan, R., Fenner, A., & Fathi, S. "Machine learning applications in urban building energy performance forecasting: A systematic review," *Renewable and Sustainable Energy Reviews*, 133, pp. 110287, 2020. https://doi.org/10.1016/j.rser.2020.110287

[6] Alfares, H. K., & Nazeeruddin, M. "Electric load forecasting: Literature survey and classification of methods," *International Journal of Systems Science*, 33(1), pp. 23–34, 2002. https://doi.org/10.1080/00207720110067421

[7] Sharma, B., Rizwan, M., & Anand, P. "Optimal design of renewable energy based hybrid system considering weather forecasting using machine learning techniques," *Electrical Engineering*, 105(6), pp. 4229–4249, 2023. https://doi.org/10.1007/s00202-023-01945-w

[8] Sharma, S. K., Palwalia, D. K., & Shrivastava, V. "Performance analysis of grid-connected 10.6 kW (commercial) solar PV power generation system," *Applied Solar Energy*, 55, pp. 269–281, 2019. https://doi.org/10.3103/S0003701X19050128

[9] Khan, S. U. R., Hayder, I. A., Habib, M. A., Ahmad, M., Mohsin, S. M., Khan, F. A., & Mustafa, K. "Enhanced machine-learning techniques for medium-term and short-term electric-load forecasting in smart grids," *Energies*, 16(1), pp. 276, 2022. https://doi.org/10.3390/en16010276

[10] Hussain, S., Sharma, S. K., & Lal, S. "Feasible synergy between hybrid solar PV and wind system for energy supply of a green building in Kota (India): A case study using iHOGA," *Energy Conversion and Management*, 315, pp. 118783, 2024. https://doi.org/10.1016/j.enconman.2024.118783

[11] Khan, P. W., Byun, Y. C., Lee, S. J., & Park, N. "Machine learning based hybrid system for imputation and efficient energy demand forecasting," *Energies*, 13(11), pp. 2681, 2020. https://doi.org/10.3390/en13112681

[12] Heydari, A., Nezhad, M. M., Pirshayan, E., Garcia, D. A., Keynia, F., & De Santoli, L. "Short-term electricity price and load forecasting in isolated power grids based on composite neural network and gravitational search optimization algorithm," *Applied Energy*, 277, pp. 115503, 2020. https://doi.org/10.1016/j.apenergy.2020.115503

[13] Sushama, D., Wankhade, B., & Patil, R. "Machine learning-based electrical load prediction using regression learners: A performance evaluation," *IEEE Conference Proceedings*, pp. 1595–1601, 2024. https://doi.org/10.1109/icoici62503.2024.10696398

[14] Díaz-Bedoya, D., González-Rodríguez, M., Clairand, J. M., Serrano-Guerrero, X., & Escrivá-Escrivá, G. "Forecasting univariate solar irradiance using machine learning models: A case study of two Andean cities," *Energy Conversion and Management*, 296, pp. 117618, 2023. https://doi.org/10.1016/j.enconman.2023.117618

[15] Huang, R., Wei, C., Wang, B., Yang, J., Xu, X., Wu, S., & Huang, S. "Well performance prediction based on Long Short-Term Memory (LSTM) neural network," *Journal of Petroleum Science and Engineering*, 208, pp. 109686, 2022. https://doi.org/10.1016/j.petrol.2021.109686

[16] Cai, W., Yang, J., Yu, Y., Song, Y., Zhou, T., & Qin, J. "PSO-ELM: A hybrid learning model for short-term traffic flow forecasting," *IEEE Access*, 8, pp. 6505–6514, 2020. https://doi.org/10.1109/ACCESS.2019.2963784

[17] Sahoo, A. K., Pradhan, C., & Das, H. "Performance evaluation of different machine learning methods and deep-learning-based convolutional neural network for health decision making," *Nature-Inspired Computing and Data Science*, pp. 201–212, 2020. https://doi.org/10.1007/978-3-030-33820-6_8

[18] Gandhi, U. D., Kumar, P. M., Babu, G. C., et al. "Sentiment analysis on Twitter data by using convolutional neural network (CNN) and Long Short-Term Memory (LSTM)," *Wireless Personal Communications*, 2021. https://doi.org/10.1007/s11277-021-08580-3

[19] Wang, S., Shao, C., Zhang, J., et al. "Traffic flow prediction using bi-directional gated recurrent unit method," *Urban Informatics*, 1, pp. 16, 2022. https://doi.org/10.1007/s44212-022-00015-z

[20] Li, L., Su, X., Bi, X., Lu, Y., & Sun, X. "A novel transformer-based network forecasting method for building cooling loads," *Energy and Buildings*, 296, pp. 113409, 2023. https://doi.org/10.1016/j.enbuild.2023.113409

[21] Zhao, L., Han, Y., Hawbani, A., Wan, S., Guo, Z., & Guizani, M. "MEDIA: An incremental DNN-based computation offloading for collaborative cloud-edge computing," *IEEE Transactions on Network Science and Engineering*, 2023. https://doi.org/10.1109/TNSE.2023.3335345

[22] Cui, Z., Wang, W., Jiang, W., Guo, J., & Liu, Y. "High-precision identification and prediction of low-voltage load characteristics in smart grids based on hybrid deep learning framework," *International Journal of Low-Carbon Technologies*, 19, pp. 2656–2666, 2024. https://doi.org/10.1093/ijlct/ctae221

[23] Michael, W., Emanuele, O., Alfredo, N., Travis, S., & Sonia, L. "Day-ahead electric load forecast: A comprehensive LSTM-EMD methodology and several diverse case studies," *Forecasting*, 5(1), pp. 297–314, 2023. https://doi.org/10.3390/forecast5010016

[24] Jain, A., & Gupta, S. C. "Evaluation of electrical load demand forecasting using various machine learning algorithms," *Frontiers in Energy Research*, 12, 2024. https://doi.org/10.3389/fenrg.2024.1408119

[25] Jasmine, J., Germin Nisha, M., & Prasad, R. "Enhancing smart grid reliability with advanced load forecasting using deep learning," *Electrical Engineering*, 2025. https://doi.org/10.1007/s00202-024-02946-z

[26] Singh, R. A., Kumar, R. S., Bajaj, M., et al. "Machine learning-based energy management and power forecasting in grid-connected microgrids with multiple distributed energy sources," *Scientific Reports*, 14, pp. 19207, 2024. https://doi.org/10.1038/s41598-024-70336-3

[27] Khan, M. A., Sabahat, Z., Farooq, M. S., et al. "Optimizing smart home energy management for sustainability using machine learning techniques," *Discover Sustainability*, 5, pp. 430, 2024. https://doi.org/10.1007/s43621-024-00681-w

[28] Azeem, A., Ismail, I., Jameel, S. M., & Harindran, V. R. "Electrical load forecasting models for different generation modalities: A review," *IEEE Access*, 9, pp. 142239–142263, 2021. https://doi.org/10.1109/ACCESS.2021.3120731

[29] Thirunavukkarasu, K., & Raju, L. "Enhanced energy efficiency smart buildings through LoRaWAN and adaptive machine learning techniques," *Wireless Networks*, 31, pp. 2875–2892, 2025. https://doi.org/10.1007/s11276-025-03915-5

[30] Moghimi, S. M., Gulliver, T. A., Chelvan, I. T., & Teimoorinia, H. "Load optimization for connected modern buildings using deep hybrid machine learning in island mode," *Energies*, 17(24), pp. 6475, 2024. https://doi.org/10.3390/en17246475

[31] Moghimi, S. M., Gulliver, T. A., Chelvan, I. T., & Teimoorinia, H. "Resource optimization for grid-connected smart green townhouses using deep hybrid machine learning," *Energies*, 17(23), pp. 6201, 2024. https://doi.org/10.3390/en17236201

[32] Sharma, S. K., Palwalia, D. K., & Shrivastava, V. "Distributed generation integration optimization using fuzzy logic controller," *AIMS Energy*, 7(3), pp. 337–348, 2019. https://doi.org/10.3934/energy.2019.3.337