Deep learning based predictive analysis of energy consumption for smart homes

Sangeeta Malik[1] · Sitender Malik[2] · Ishmeet Singh[1] · Harsh Vardhan Gupta[1] · Sidhant Prakash[1] · Rachna Jain[3] · Biswaranjanjan Acharya[4] · Yu-Chen Hu[5]

Abstract

Predicting energy consumption has become crucial to creating a sustainable and intelligent environment. With the aid of forecasts of future demand, the distribution and production of energy can be optimized to meet the requirements of a vastly growing population. However, because of the varied types of energy consumption patterns, predicting the demand for any household can be difficult. It has recently gained popularity with social Internet of Things-based smart homes, smart grid planning, and artificial intelligence-based smart energy-saving solutions. Although there are methods for estimating energy consumption, most of these systems are based on one-step forecasting and have a limited forecasting period. Several prediction models were implemented in this paper to address the problem mentioned above and achieve high accuracy, including the baseline model, the Auto-Regressive Integrated Moving Average (ARIMA) model, the Seasonal Auto-Regressive Integrated Moving Average (SARIMAX with eXogenous factors) model, the Long Short-Term Memory (LSTM) Univariate model, and the LSTM Multivariate model.

1 Introduction

The Internet of Things (IoT) is a novel paradigm of interconnectivity that has recently garnered momentum in the field of modern telecommunications. The IoT ecosystem is expected to usher in an era of the ubiquitous presence of uniquely identifiable physical objects or "things" (referred to as IoT devices or smart devices) connected via the Internet to measure, report and, in some cases, perform actions autonomously [1].

There are tens of IoT use-case applications and services across many industries, including but not limited to E-Health, Smart Building, Smart Manufacturing, Smart Agriculture, Connected Vehicles, Environmental Monitoring, and Home Automation & Security (HAS)/ Smart Home. These IoT services are becoming increasingly popular with users, partly aided by products and services from leading technology pioneers like Apple, Google, and

---

Extended author information available on the last page of the article

Amazon. To mention a few, they include HAS (such as Apple HomePod, Google Home, and Amazon Echo), video surveillance (such as Google Nest Cam, and Netgear Arlo), and wearable applications (such as FitBit). Many of these services are still in the early stages of development and have not yet received thorough literary characterization. In smart homes, there are numerous IoT applications. In contrast to a normal home, which has functional furniture and housing for habitation, a smart home aims to offer services that improve the user's comfort and enjoyment. A smart home is a sizable collection of IoT applications that go beyond simply being a place to live and give consumers access to new services and control over daily activities and leisure pursuits [2].

Despite the growth of the sector and its technologies, many experts claim that there are problems with network congestion and electricity usage in the IoT space. The methods available to solve these challenges are inadequate, and modern smart home services lead to indiscriminate network and energy use. Due to battery issues with IoT devices and a sharp rise in electrical energy usage, numerous research projects on alternate energy have recently been undertaken. However, this innovative energy production necessitates lengthy, extensive research; as a result, it is a very speculative, forward-looking enterprise. These issues demand autonomous and effective network processing and a technique that can reduce excessive energy use by precisely operating IoT depending on user usage patterns [3].

The usage data of IoT users must be studied to solve these issues, and an intelligent manager is needed as a platform that fully controls and manages this analyzed usage data. Here, the phrase "intelligent manager" refers to a manager who offers services that build a tailored space for the user in line with the objective of a smart home and network technology to reduce energy usage, as opposed to an IoT platform that merely connects IoT devices. IoT platforms currently offer intelligent services, but most of these process massive amounts of data and perform mathematical computations more rapidly and correctly than a person. A smart home intelligent manager controls groups of devices and provides customized services for each user using algorithms that learn from and analyze user data. This enables smart homes to prepare for a variety of features in a wide range of circumstances and environments. These forecasts and preparations allow the IoT applications to be monitored and controlled in line with the researched data, hence reducing network utilization and energy waste. Additionally, automated services anticipate the user's feelings, thoughts, and needs to create a comfortable environment.

Several prediction models, such as the baseline model, ARIMA basic model, ARIMA Dynamic model, Seasonal Auto-Regressive Integrated Moving Average (SARIMA) model, SARIMAX model, LSTM Univariant and LSTM Multivariant model have been implemented in this paper. The applied prediction models predict the near future and real-time electric energy consumption of Heating, Ventilation, and Air Conditioning (HVAC) for equipment, furnaces, fridges, dishwashers, wine cellars, etc.) in smart homes.

The paper introduces several prediction models, namely ARIMA, SARIMAX, LSTM Univariate, and LSTM Multivariate, without delving into the specific rationale behind their selection. To address this, it is essential to underscore that energy consumption inherently exhibits time series patterns, making time series models particularly apt for prediction tasks. The choice of these models is grounded in their capacity to capture and analyze temporal dependencies within the data.

ARIMA and SARIMAX These models are renowned for their effectiveness in capturing the temporal dynamics and seasonality of time series data. ARIMA focuses on auto-regression,

integrated differencing, and moving averages, while SARIMAX extends this approach by incorporating exogenous variables, providing a robust framework for time series forecasting.

LSTM Univariate and LSTM Multivariate Long Short-Term Memory networks excel in modeling sequential dependencies over extended time intervals. The 'Univariate' variant operates on a single variable, suitable for unidimensional time series data. In contrast, the 'Multivariate' LSTM accommodates multiple correlated variables, offering a more comprehensive understanding of complex energy consumption patterns.

By employing this diverse set of models, we aim to harness their unique strengths in deciphering different facets of energy consumption behavior, ensuring a comprehensive comparative analysis that enriches the insights into the predictive capabilities of each model.

The objectives of this paper are as follows:

- Development of an IoT-driven smart home case study environment.
- To do a predictive analysis of the energy consumption patterns between an IoT-enabled home using deep learning techniques.
- To study the effect and response of external factors, such as weather conditions, power supply, hardware architecture, etc., on the IoT system's energy consumption.

This paper makes an addition to the field by addressing the drawbacks of conventional methods, including the need for improved predictive accuracy, the difficulty of accounting for a variety of environmental factors, and the inability to accurately capture dynamic patterns in energy usage. Our research intends to considerably improve the accuracy and dependability of energy consumption predictions by recognizing and resolving these issues, providing insightful information for energy optimization tactics and smart home applications. There are many state-of-the-art models, such as transformers. However, we did not find a particular transformer for our use case, and it was not feasible to train a transformer; hence, we used more conventional models such as LSTM Multivariate, ARIMA, etc.

The remaining sections are organized as follows: The development of the IoT-driven building case study environment and the IoT device measurement data are described in Section 2. The problem of predicting electric energy consumption is discussed in Section 3. Section 4 presents the electric energy prediction results obtained from the prediction models and their comparisons, discussions, and applications. In Section 5, the conclusion is presented.

2 Literature survey

This section focuses on a selection of publications by researchers on appliances and socio-economic factors that aid in recognizing the diverse data and methodologies that have been used to identify appliances' energy usage. Table 1 summarizes various research works [4, 6, 11–26] on electric power consumption.

Aswathi Balachandran et al. [4] used machine learning techniques for energy consumption analysis. To analyse the variable energy consumption, the datasets are trained with the

Table 1 Summary of the studies on electrical power consumption

Year	Ref	Name of method	Dataset collection	Results	Limitations
2018	[11]	Both linear and non-linear regression techniques	Self-generated dataset from the data collected from various sensors	Multi-step time-series model using the ARIMA algorithm was successfully developed	Requires a long historical horizon, especially for seasonal products.
2018	[12]	Regression model (SVR, LR, RF, MLP)	2017's UCI Machine Learning Repository, from the University of Mons.	Prediction of the energy consumption of different appliances	Model re-training is needed for new and bigger data
2019	[4]	kNN algorithm	Dataset with load at four different times of day	kNN algorithm-based classification of energy consumption	Lack of proper dataset
2019	[13]	Convolutional neural network and long short-term memory	Individual household electric power consumption dataset from the UCI machine learning repository	Predict electrical energy consumption by extracting features from variables that affect power consumption.	Determining the optimal hyperparameters requires relatively large efforts by trial and error.
2019	[14]	Tuning machine learning models	Ecotect and EnergyPlus dataset	Addresses the deficiency in using ML methods for estimating building energy loads through an exhaustive analysis of prevalent ML models fitting over energy simulation data.	External factors for data analysis are not considered.
2020	[15]	Ensemble learner	Electric energy consumption info of a city presented by SGCC, Jan. 2018–Feb. 2019	For power consumption prediction, combine the ranking method and iterative optimization.	Due to the iterative learning process, computational complexity is very high.
2020	[16]	Exponential model and Elman Recurrent Neural Network (RNN)	Real-time dataset Weather station from the roof of Clemson University.	Predicted electrical power consumption using power loads in the future.	It cannot handle real-time energy consumption prediction.
2020	[17]	Deep residual neural network	Dataset from Pecan Street 2015–2018.	Deep learning-based load forecasting for the near future.	Re-training of the model is required for new data.
2020	[18]	SOM clustering algorithm and deep autoencoder,	Single house electric consumption dataset from Dec-2006 to Nov- 2010 in the monthly energy consumption data from Gainesville, USA 2006–2010	Cluster-based electric energy consumption prediction using smart sensors.	High computational complexity due to the clustering process.

Table 1 (continued)

Year	Ref	Name of method	Dataset collection	Results	Limitations
2020	[19]	MEC-TLL, LSTM, and transfer learning,	Collected data from the Smart buildings located in Korea Jan-2016 to Dec-2018	Multiple electric energy forecasting using transfer learning	Due to the iterative learning process, computational complexity is very high.
2020	[6]	Hierarchical deep reinforcement learning	2018's Detailed State Data from U.S. Energy Information Administration.	A two-level distributed deep reinforcement learning algorithm is proposed to reduce the cost of electricity by scheduling the energy consumption of two controllable home appliances.	Unable to work with multiple agents
2020	[20]	Agent-based modeling	Self-generated dataset	Predicted energy consumption incorporating building parameters, occupant behavior, etc.	External factors for data analysis are not considered.
2022	[21]	Hybrid model and Grey nonlinear models.	Maximum Electricity prediction for Slovakia.	Better performance was observed with the data obtained from the official system of Slovakia by the Hybrid model.	Unable to Analyse the Data for External factors.
2022	[24]	Multiagent-based secure REM using Q-learning	2020's PJM Data Miner and Residential gas prices	Proposed Q-MSEM scheme to reduce enery consumption with a notch of discomfort in DRP.	Unable to work with real-time energy prices and fault detection
2020	[25]	Ensemble of HMMs generalized into a novel mixed-dependency model.	Washington State University Smart Department and generated dataset from [31]	Proposed a mixed dependency model to deal with the complexity of multiple residents' activities in a smart home.	Less effective than deep learning approaches
2022	[26]	Probabilistic neural network	The MHEALTH (Mobile Health) dataset	Presented a new Effective Resource Allocation Methodology for Fog environment	The model needs to be tested for new and bigger data.

k-Nearest Neighbour (kNN) algorithm so that they are clustered. This algorithm can make suggestions for future homes, and their energy consumption can be estimated. Josephat Kalezhi et al. [5] proposed an Energy Management System (MES) capable of regulating appliances' energy consumption in a home environment using IoT infrastructure. The system uses various sensors to continuously track multiple situations, using the information to operate appliances. The power usage of specific appliances can also be observed. The outcomes demonstrate that the system reacts quickly to alterations in the home environment. Sangyoon Lee et al. [6] proposed a hierarchical deep reinforcement learning algorithm for the scheduling of energy consumption of smart home appliances and distributed energy resources, including an energy storage system and an electric car is presented. Two agents interact in the proposed deep reinforcement learning method to schedule the ideal residential energy use in an effective manner.

Regarding the forecast of energy consumption, a substantial quantity of literature has been produced representing various attempts to anticipate the energy required by various equipment. Elkonomou described [7] an artificial neural network-based approach for prediction. A series of tests were conducted using the multilayer perceptron model to determine which architecture provided the best generalization. Input and output data were utilized in the training, validation, and testing processes. Gray and Chrispin Alfred's [1] thesis sought to examine and better understand the energy consumption of IoT network applications and services by creating energy consumption models and energy-efficient network architectures for the delivery of IoT services. To accomplish this purpose, they used a few case studies, including video surveillance services and HAS, two of the most well-known and commonly used IoT services. Gonçalves Ivo et al. [8] proposed an Energy Management System (EMS) to minimize cost and customer dissatisfaction.

A sustainable, energy-efficient intelligent street lighting system is proposed by Zhon Chen et al. [9] that uses IoT sensors, renewable energy sources, an effective decision-making module, and a dimming system to aid in energy conservation. We use a variety of sensors and actuators that detect motion and light, transmitting that data to the master control unit over the ZigBee network and to determine traffic flow and ambient status. For hourly building energy forecast, Wang Zeyu et al. [10] suggested a homogenous ensemble approach using Random Forest (RF). The method was used to forecast how much power would be used per hour in two educational facilities in North Central Florida. To find out how parameter setting affected the model's ability to predict outcomes, the RF models trained with various parameter settings were compared. The results demonstrated that RF was not overly sensitive to the number of variables and that the empirical number of variables is preferable because it is faster and more accurate. These models were compared head-to-head to demonstrate the superiority of RF over Regression Tree (RT) and Support Vector Regression (SVR) in predicting building energy consumption.

Geetha et al. [22] proposed using an RF-supervised learning model to forecast power consumption and identify the level of peak demand. This model outperforms the existing models in terms of accuracy, stability and generalization.The forecast says that the boom of electric vehicles will increase the electricity demand globally by 3% for the upcoming year. This forecast analysis helps the service providers and the government understand the customers' lifestyles. The prediction of electricity consumption is a vital foundation for smart energy management. Yağanoğlu et al. [23] proposed a study that aims to develop a smart classroom concept that provides energy savings and air conditioning based on the analysis of the environmental parameters in the classroom environments in real-time. Measured values are then analyzed for anomaly detection and energy saving. Tests were performed for seven different scenarios, and the best accuracy and sensitivity were calculated as 98%.

It shows that the proposed IoT-based smart classroom architecture is suitable for creating smart classrooms. The IoT integrates objects used in daily life with digital platforms by connecting to the Internet. The amount of data captured increases in direct proportion to the number of sensors on the objects. In this study, embedded systems, the IoT and artificial intelligence applications have been implemented to monitor and regulate the physical conditions in university classrooms.

Table 1 provides a summary of the studies on electrical power consumption discussed in the article and their key findings. It includes the study's focus, methods used, and main conclusions. It provides a useful overview of the current state of research on electrical power consumption and can be used as a reference for further study and analysis. The literature survey summarises and highlights the following points:

1. The energy consumption of the appliances represents a significant portion of the residential sector's accumulated electricity demand. It is essential for the smart home's power management [27].
2. Due to the growing number of electrically powered machines, it is crucial to identify the primary energy consumers.
3. The consumed energy can be broken down into components, including heating, ventilation, and air conditioning systems.
4. The energy consumption patterns of appliances can vary considerably.
5. External factors have a greater impact on appliances in highly insulated buildings [28].

3 Methodology

3.1 System architecture

Machine learning has allowed computer systems to automatically learn without being explicitly programmed. The authors have used three machine learning algorithms ('baseline', 'ARIMA basic', 'long short-term memory networks''). The architecture diagram depicts a high-level overview of the system's major components and significant interdependencies.

The flowchart, as shown in Fig. 1, comprehensively represents our research project's system architecture. It outlines the sequential execution flow and highlights the major steps involved in refining raw data and predicting future data. The six crucial steps in the process are as follows:

1. The flow of the process used to refine raw data and predict data is depicted in the architecture diagram.
2. The next step is pre-processing the collected raw data into an understandable format. This involves handling missing values, data normalization, and feature extraction to prepare the dataset for model training.
3. The data must be trained by separating the dataset into train data and test data. This ensures that the model's performance is evaluated on unseen data, providing a reliable assessment of its capabilities.
4. The data is evaluated with the application of a machine learning algorithm, which is the baseline Model, RNNs, and ARIMA algorithm. Each model's classification accuracy and forecasting performance are measured to gauge its effectiveness.
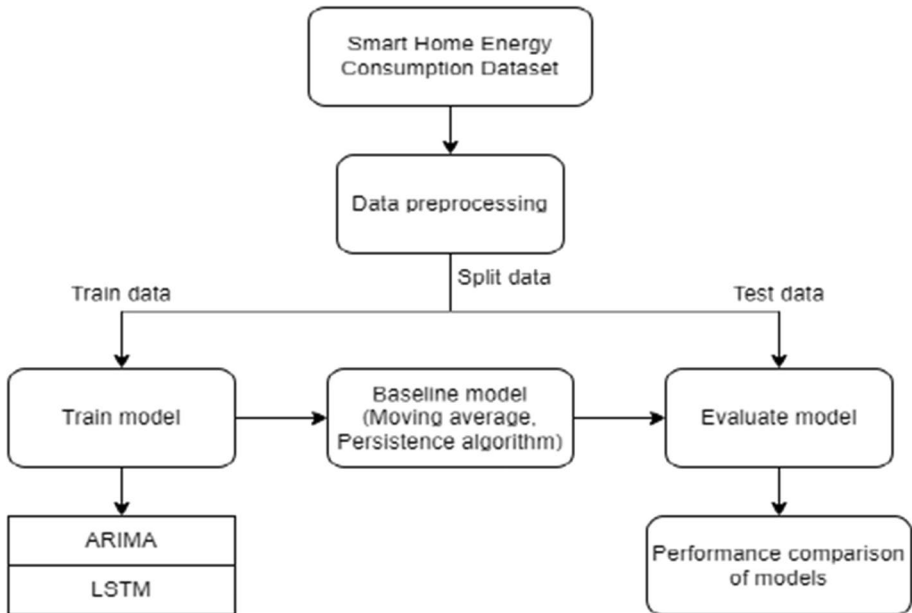
Fig. 1 Block diagram of the proposed technology

5. After training the data with these algorithms, we test the models on the same dataset to validate their predictive capabilities. This step is crucial for determining the models' robustness and suitability for real-world applications.
6. Finally, the result of these algorithms is compared based on classification accuracy.

4 Experimental results and discussion

Predicting energy consumption remains difficult despite the abundance of research studies employing various techniques and models, owing to the extent of consumption-influencing factors (e.g., physical properties of the building, installed equipment, weather conditions, and energy-use behavior of building occupants). To address the difficulties, this paper proposes a novel approach and provides a summary of forecasting models that incorporate occupant behavior and weather conditions.

The prediction models can generate information about the patterns that drive energy consumption, enabling designers to incorporate more accurate input parameters into energy models. In other words, the suggested model's results can be integrated with those of existing building simulation models to generate relevant operating variables. For instance, a simulation model can use the anticipated energy consumption to infer occupancy and building operational data. Tenants and building owners can use this information to manage the building more efficiently. Using Root Mean Squared Error (RMSE), models were determined to be directly comparable to data-based energy values. The findings suggest that electricity demand can be predicted using machine learning algorithms, deploying

renewable energy, planning for high/low load days, minimizing energy waste, recognizing anomalies in consumption trends, and quantifying energy and cost-saving actions.

One potential application of these findings is in the field of smart grid management. Using machine learning algorithms to predict electricity demand, utilities can better plan for and manage integrating renewable energy sources into the grid. This can help to minimize energy waste and ensure that the grid remains stable and reliable. Additionally, by identifying anomalies in consumption trends, utilities can quickly identify and address any issues that may arise, such as equipment failures or network outages. Another potential application is in the commercial and industrial sectors. By using machine learning to predict electricity demand, businesses can take proactive measures to reduce their energy consumption and costs. For example, they can schedule energy-intensive tasks during periods of low demand and take action to maximize energy efficiency.

While the use of ARIMA and LSTM models in energy prediction is not unprecedented, this study introduces several methodological distinctions and novel aspects. The employment of a Multivariate LSTM model sets this study apart. The model captures intricate relationships that univariate models may overlook by incorporating multiple predictor variables, including environmental factors and household characteristics. This study offers a comprehensive comparative analysis of various models, including ARIMA, SARIMAX, LSTM Univariate, and LSTM Multivariate. The rationale for the selection of each model and a detailed evaluation using multiple metrics distinguish this work from others.

4.1 Descriptions of dataset

The dataset used in this study for smart home energy prediction was gathered from an open source. To compute and analyse appliances' energy usage, this data includes information for 503,910 instances with 32 attributes, as shown in Table 2. The prediction dataset for this energy of appliances has 32 input attributes and a date (time) value. This CSV file includes the readings of household appliances in kW from a smart meter for a period of 350 days, together with local weather data for that location.

In our dataset, the given attributes are described as follows: *use* represents the total energy consumption of the household, *gen* represents the total energy generated using solar or other power generation resources. The dishwasher, furnace, wine cellar, fridge, etc., represents the energy consumption of various appliances. The attributes like pressure, wind-speed, cloud-cover, dewpoint, etc., cover the weather-related parameters. Figure 2 represents the correlation between weather parameters.

4.2 Performance measures

Performance evaluation measures serve as essential parameters for conducting a comparative analysis of different deep learning techniques, aiming to identify the best algorithm or method for predicting energy consumption in smart homes. These evaluation measures provide insights into the relative performance of various models and assist in determining the most effective approach.

To establish an RMSE threshold, we employ the moving average technique. The moving average is a statistical method used to analyze and smoothen data fluctuations over a specific period. It calculates the average value of a series of data points within a defined window or interval, commonly applied in time series analysis. By utilizing the moving average as a baseline, we can set an RMSE threshold that other models should strive to exceed.

Table 2 Explanation of attributes

Data variables	SI units
Time	Date-time
Use	kilowatt(kw)
Gen	kilowatt(kw)
House Overall	kilowatt(kw)
Dishwasher	kilowatt(kw)
Furnace 1	kilowatt(kw)
Furnace 2	kilowatt(kw)
Home office	kilowatt(kw)
Fridge	kilowatt(kw)
Wine cellar	kilowatt(kw)
Garage door	kilowatt(kw)
Kitchen 12	kilowatt(kw)
Kitchen 14	kilowatt(kw)
Kitchen 38	kilowatt(kw)
Barn	kilowatt(kw)
Well	kilowatt(kw)
Microwave	kilowatt(kw)
Living room	kilowatt(kw)
Solar	kilowatt(kw)
Temperature	℃
Humidity	%
Visibility	meter(m)
Apparent temperature	℃
Pressure	Pa
Wind speed	m/s
Cloud cover	%
Wind bearing	degree
Precip intensity	%
Dew point	℃
Precip probability	%

Consequently, the RMSE values of alternative models can be compared against the RMSE of the moving average model, enabling a relative performance assessment.

In addition to RMSE, various metrics are employed to evaluate prediction findings. These metrics include Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Absolute Scaled Error (MASE), and R-squared (R2) score. These measurements offer distinct perspectives on predictive models' accuracy, error, and goodness of fit. By considering these metrics, we gain a comprehensive understanding of the predictive power of each model. These metrics are defined as:

- **mean_squared_error (MSE)**: The mean_squared_error function computes mean square error, a risk metric corresponding to the expected value of the squared (quadratic) error or loss.

  If $\hat{y}_i$s the predicted value of the *ith* sample, and y_i is the corresponding true value, then the mean squared error estimated over $n_{samples}$ is defined as:
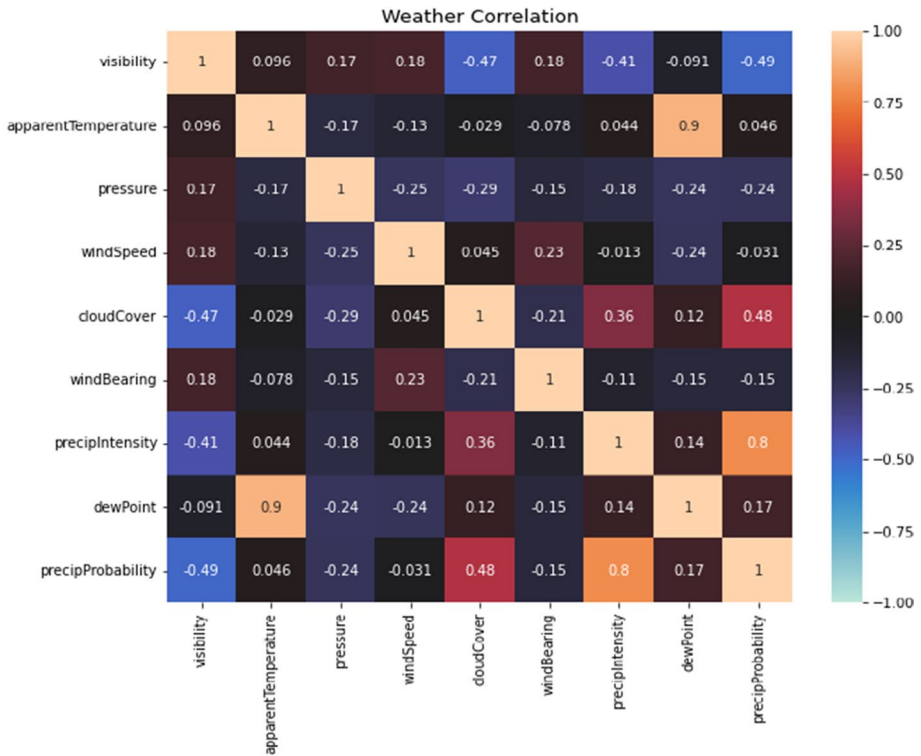
Fig. 2 Correlation between weather parameters

$$MSE(y, \widehat{y}) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} \left(y_i - \widehat{y}_i\right)^2 \tag{1}$$

- **root_mean_squared_error (RMSE)**: RMSE, or Root Mean Square Error, is a metric used to measure the average difference between predicted values and actual values. It quantifies the accuracy of a predictive model, with lower values indicating better performance.

$$MSE(y, \widehat{y}) = \sqrt{\frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} \left(y_i - \widehat{y}_i\right)^2} \tag{2}$$

- **r2_score**: The r2_score function computes the coefficient of determination, usually denoted as R2. It represents the proportion of variance (of y) explained by the model's independent variables. It provides an indication of goodness of fit and, therefore, a measure of how well unseen samples are likely to be predicted by the model through the proportion of explained variance.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(y_i - \widehat{y}_i\right)^2}{\sum_{i=1}^{n} \left(y_i - \overline{y}\right)^2} \tag{3}$$

- **mean_absolute_error (MAE)**: The mean_absolute_error function computes mean absolute error, a risk metric corresponding to the expected value of the absolute error loss or l1-norm loss. If $\hat{y}_i$s the predicted value of the *ith* sample, and y_i is the corresponding true value, then the mean absolute error estimated over $n_{samples}$ is defined as:

$$MAE(y,\hat{y}) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} \left(|y_i - \hat{y}_i| \right) \tag{4}$$

- **mean_absolute_percentage_error (MAPE)**: The mean absolute percentage error, also known as Mean Absolute Percentage Deviation (MAPD), is an evaluation metric for regression problems. The idea of this metric is to be sensitive to relative errors. It is, for example, not changed by a global scaling of the target variable. It is given as follows:

$$MAPE(y,\hat{y}) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} \left(\frac{|y_i - \hat{y}_i|}{\max(\in, |y_i|)} \right) \tag{5}$$

Utilizing these evaluation metrics, in conjunction with the moving average as a baseline, allows for a comprehensive comparative analysis of alternative models in predicting energy consumption in smart homes. The results obtained from this analysis contribute to informed decision-making and the selection of optimal predictive models.

The evaluation metrics used in this study include MAE, MASE, RMSE, and Coefficient of Determination (R2 Score). MAE measures the average absolute errors between predicted and actual values, MASE provides a scaled error considering the mean absolute error and a scaled mean absolute error of the naïve model. RMSE quantifies the square root of the average squared differences between predicted and actual values. R2 Score, or the Coefficient of Determination, indicates the proportion of the variance in the dependent variable (energy consumption) that is predictable from the independent variables (predictor variables).

In simpler terms, achieving a higher R2 Score signifies a better predictive power of the model. A higher R2 Score indicates that a larger proportion of the variation in energy consumption can be explained by the predictor variables, making the model more reliable and effective in capturing the underlying patterns in the data.

4.3 The baseline model

A baseline model is the simplest effort at modeling; it provides us with a reference metric that we can use throughout the development process. This baseline model is frequently a heuristic (rule-based) model but could also be a straightforward machine learning model. When plotting time series data, these fluctuations may prevent us from clearly understanding the plot's peaks and valleys. As a result, we use the rolling average concept to create the time series plot so that the data's value is evident.

The rolling average, also known as the moving average, is the simple mean of the previous 'n' values. It can help us detect trends that would otherwise be difficult to identify. Additionally, they can be utilized to identify long-term trends. The rolling average can be calculated by adding the previous 'n' values and dividing by 'n' itself. In contrast, the rolling average's first (n-1) values would be NaN in this case. Figure 3 represents the rolling
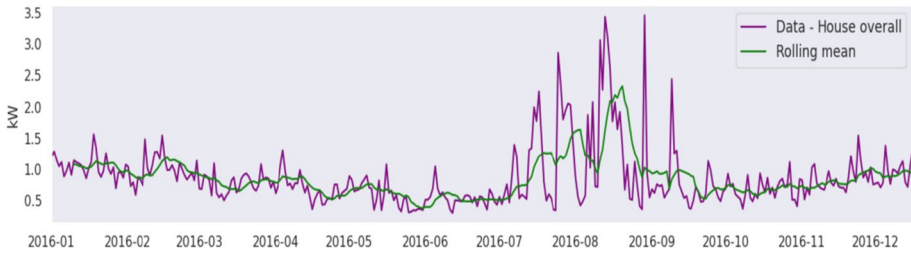
Fig. 3 Results of the rolling mean and the data house overall for the baseline model

mean and the data house overall graph of 2016 with respect to the energy consumption and time for the baseline model, with Table 3.

4.4 ARIMA model

Class of models that "describes" a given time series based on its own past values, that is, its own lags and the lagged forecast errors, so that the resulting equation can be used to predict future values. ARIMA models can be used to model any "non-seasonal" time series that exhibits patterns and is not random white noise [29].

Making the time series stationary is the first step in developing an ARIMA model. In ARIMA, the term 'Auto Regressive' refers to a linear regression model that employs its own lags as predictors. As you are aware, linear regression models perform optimally when the predictors are uncorrelated and independent. The most prevalent method is differentiation. That is, deduct the prior value from the current value. Depending on the complexity of the series, multiple differencing may be required at times. Therefore, d represents the minimal number of differentiations required to make the series stationary. Figure 4 depicts the step-by-step process of the ARIMA model.

Starting with the dataset, we first perform a stationary check and apply series transformation techniques. Next, we analyze the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) to determine the appropriate Auto Regressive (AR) and Moving Average (MA) components for the ARIMA model. Afterward, the time series is split into training and testing sets. In one branch, we fit an ARMA model to the training data; in another, we employ a SARIMA model. Once the models are fitted, we evaluate their performance using metrics such as R2, MASE, MAPE, MAE, and RMSE. By following the systematic approach depicted in the flowchart, we construct an accurate ARIMA model for

Table 3 Metrics representing the predictive power of the baseline model

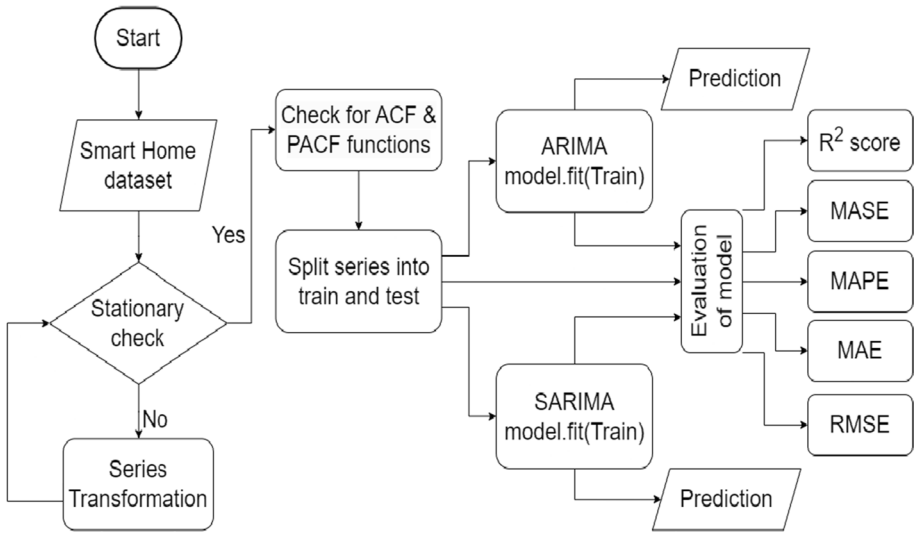| Metric	Value
RMSE	0.304
MAE	0.194
MAPE	0.233
MASE	0.790
R2 Score	-0.203

Fig. 4 Flowchart of the ARIMA model

time series forecasting, enabling us to make reliable predictions about future values based on past data.

Augmented Dickey Fuller (ADF) The ADF test's null hypothesis is that the time series is non-stationary. If the p-value of the test is less than the significance level (0.05), then the null hypothesis is rejected, and it is concluded that the time series is indeed stationary. To be stationary, the critical values at 1%, 5%, and 10% confidence intervals should be as close as possible to the Test Statistics. Figure 5 represents the output of the forecasting model on test Data of 2016 for the ARIMA model. From the stats model, we have used AD-fuller, whose test results are shown in Table 4: The augmented dickey fuller test operates on the statistic that yields a negative number and rejects the null hypothesis. In addition, Table 5 represents the results of the ARIMA model.
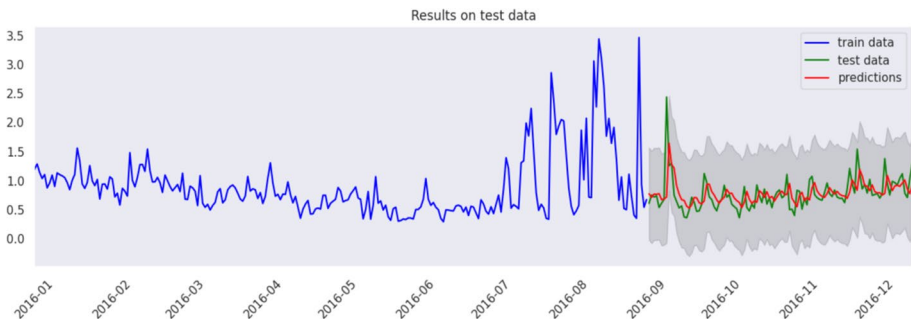

Fig. 5 Results of the rolling mean and the data house overall for the ARIMA model

Table 4 Dick Fuller test result for the ARIMA model

Metric	Value
Test statistic	-2.988003
p-value	0.036030
#Lags Used	10.000000
Number of observations used	340.000000
Critical value (1%)	-3.449730
Critical value (5%)	-2.870079
Critical value (10%)	-2.571319
dtype	Float64

Table 5 Metrics representing the predictive power of the ARIMA model

Metric	Value
RMSE	0.263
MAE	0.174
MAPE	0.226
MASE	0.710
R^2 Score	0.095

Dependent on this negative number, the greater the magnitude of the negative number, the greater the confidence of unit root presence at some level of the time series. There are several parameters in the ARIMA function that we can set for the optimum result, so we have set the values as follows for our prediction:

arima_model = auto_arima (train, start_p = 0, d = 0, start_q = 0, max_p = 5, max_d = 5, max_q = 5, start_P = 0, D = 1, start_Q = 0, max_P = 5, max_D = 5, max_Q = 5, m = 12, season-al = True, error_action = 'warn', trace = True, surpress_warnings = True, stepwise = True, random_state = 20, n_fits = 50).

4.5 SARIMAX model

The ARIMA class of models is extended by the seasonal auto-regressive integrated moving average with eXogenous factors or SARIMAX. ARIMA models logically consist of the MA term and the AR term. In the former, the value at any given time is simply the weighted sum of earlier values. The SARIMAX model's step-by-step process is depicted in Fig. 4 as a flowchart.

The latter model incorporates past residuals instead of weighted sums to represent the same value (confer. time series decomposition). SARIMA models can be expressed in several ways. One of them is Time Series Forecasting, which is defined as:

$$SARIMA(p, d, q) \times (P, D, Q, S)$$

where *p* denotes non-seasonal AR order, d denotes non-seasonal differencing, q denotes non-seasonal MA order, P denotes seasonal AR order, D denotes seasonal differencing, Q denotes seasonal MA order, and S denotes the length of the repeating seasonal pattern.

SARIMAX results are shown in Table 6, while Table 7 represents the metrics used for predicting the power of the SARIMAX, and Fig. 6 represents the predictions made by the SARIMAX on test data.

4.6 LSTM model

LSTM stands for Long short-term memory networks employed in deep learning. Many RNNs can learn long-term dependencies, particularly in problems involving sequence prediction. Traditional forecasting techniques like ARIMA and HWES are still widely used and effective, but they do not have the same level of broad generalizability as memory-based models like LSTM [30]. The LSTM fixes the serious short-memory problem that recurrent neural networks experience. The LSTM controls whether to keep, forget, or ignore data points using a series of "gates," each with its own RNN based on a probabilistic model. Using the ReLu function as an activation function and training the LSTM model with the total number of trainable parameters: 10,451.

Table 6 SARIMAX results

Parameter name	Value
Ljung-Box (L1) (Q)	0.01
Prob(Q)	0.93
Heteroskedasticity	12.34
Prob (H) (two-sided)	0.00
Jarque-Bera (JB)	2993.30
Prob (JB)	0.00
Skew	2.85
Kurtosis	19.18

Table 7 Metrics representing the predictive power of the SARIMAX model

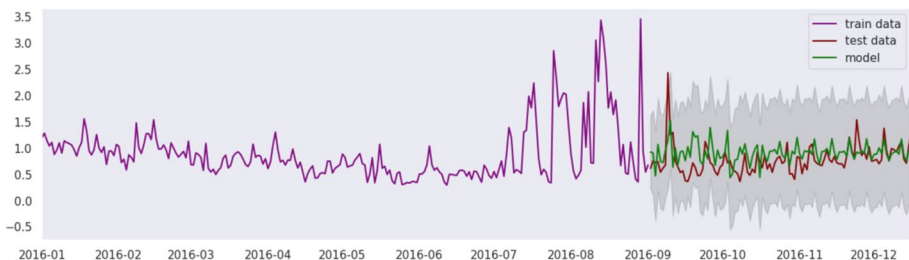| Metric	Value
RMSE	0.317
MAE	0.243
MAPE	0.361
MASE	0.989
R^2 Score	-0.312

Fig. 6 Results of the rolling mean and the data house overall for the SARIMAX mode

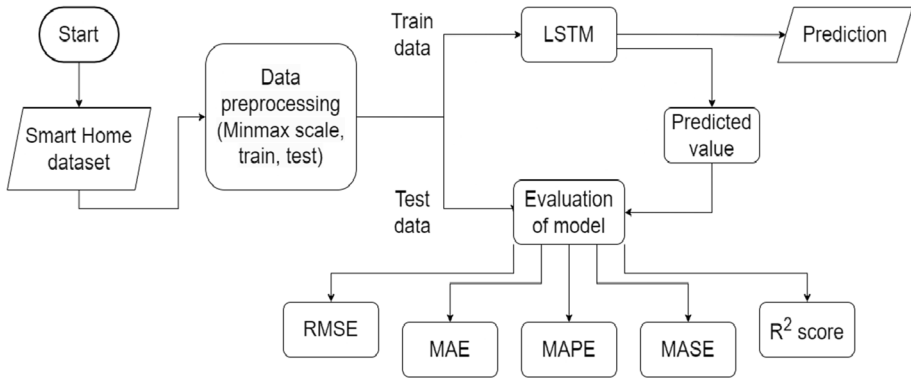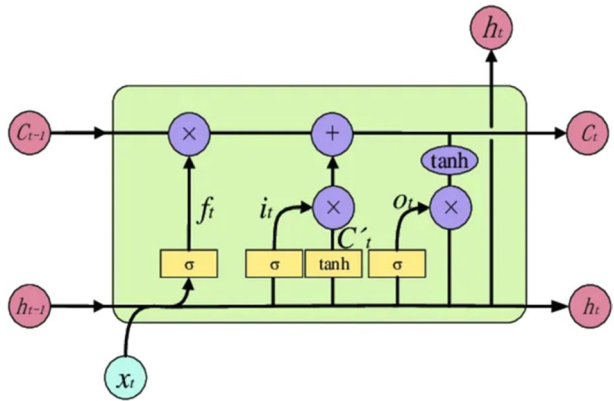**Fig. 7** Architecture for LSTM [32]

Fig. 8 Flowchart for LSTM working

To provide a visual representation of the LSTM Univariate model's working, we have added a architectural diagram (Fig. 7) and flow chart (Fig. 8) depicting the step-by-step process for the LSTM Univariate model's predictions. This flow chart outlines the key stages involved, from data preprocessing to the LSTM model training and prediction. Additionally, it showcases how the LSTM model's performance is evaluated by comparing its predictions to the actual values in the test dataset, using various evaluation metrics such as RMSE, MAE, MAPE, MASE, and R2 score. This flow chart is a valuable tool to illustrate the LSTM model's methodology and contribution to our time series forecasting analysis.

Figure 9 represents the LSTM Univariate model's prediction. Forecasting time series is the process of estimating a time series' future value using historical data. Numerous time series problems can be resolved by looking forward to one step. Multi-step time series forecasting simulates the distribution of a signal's future values over a forecast horizon. This approach forecasts multiple output values simultaneously, a forecasting technique used to forecast the future path of a gradually rising sine wave. Additionally, many of these variables depend on one another, complicating statistical modelling. Table 8 represents the results for LSTM.

A univariate forecast model simplifies this complexity to a single factor. Other dimensions are taken into consideration. While a multivariate model can account for
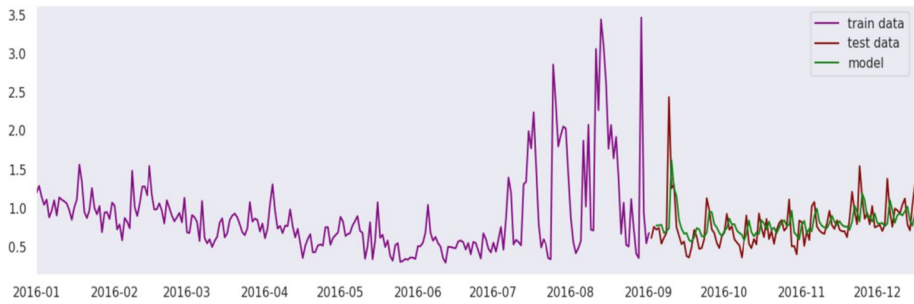
Fig. 9 Results of the LSTM Univariate model's prediction

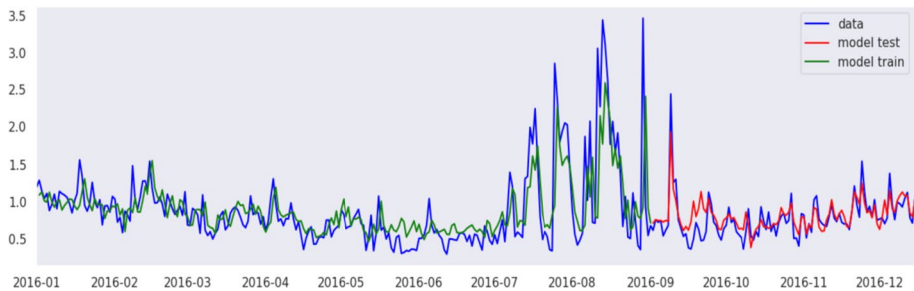| Metric	LSTM univariate	LSTM multivariate
RMSE	0.262	0.164
MAE	0.174	0.120
MAPE	0.306	0.179
MASE	0.989	0.546
R^2 Score	0.112	0.639

Table 8 Metrics representing the predictive power of the LSTM model

Fig. 10 Results of the LSTM Multivariate model's prediction

only a portion of the influencing factors, it can still account for multiple factors concurrently. Figure 10 represents the LSTM Multivariate model's prediction. To obtain more comprehensive information on LSTM, it is recommended to refer to the paper [33].

4.7 Practical insights and applications

The obtained results showcase the LSTM Multivariate model's predictive power and offer valuable practical insights for energy optimization techniques and load management in smart homes. The following points highlight the key practical implications:

1. Optimal Energy Consumption Patterns: The LSTM Multivariate model, by demonstrating superior predictive capabilities, provides a foundation for identifying optimal energy consumption patterns. Insights derived from the model can guide the development of energy-efficient strategies within smart homes.
2. Resource Allocation Strategies: Leveraging the predictive accuracy of the model, resource allocation strategies can be refined. This includes better management of energy storage systems, scheduling of home appliances, and efficient use of electric vehicles, contributing to overall load management.
3. Real-time Decision Support: The continuous and reliable predictions offered by the LSTM Multivariate model empower homeowners with real-time decision support. This aids in making informed choices regarding energy consumption, contributing to both cost savings and environmental sustainability.
4. Adaptability to Various Smart Home Infrastructures: The model's adaptability allows extrapolation of results to similar infrastructures such as offices, schools, and colleges. This extends the practical applicability beyond smart homes, enabling a broader implementation scope.

4.8 Results summary

Table 9 summarizes the results for the models as discussed in this paper. The prediction models can be applied to data sets of other environmental conditions to check the responses.

5 Conclusion and future scope

The lower values of MAE, MASE, and RMSE imply higher accuracy of a regression model. However, a higher R2 score value is considered desirable. Both RMSE and R-squared measure how well a linear regression model fits a given dataset. RMSE indicates how accurately a regression model can predict the absolute value of a response variable, whereas R-squared indicates how well the predictor variables can explain the variation in the response variable.

Table 9 shows that the LSTM Multivariate model has demonstrated its potential as an effective predictive model, achieving the highest predictive power of all tested models. This research has shown that the LSTM Multivariate model could be a valuable tool for

Table 9 Summary results for the predictive models

Metric	Model				
	Baseline	ARIMA	SARIMAX	LSTM Univariate	LSTM Multivariate
RMSE	0.304	0.263	0.317	0.262	**0.164**
MAE	0.194	0.174	0.243	0.174	**0.120**
MAPE	0.233	0.226	0.361	0.306	**0.179**
MASE	0.790	0.710	0.989	0.989	**0.546**
R^2 Score	-0.203	0.095	-0.312	0.112	**0.639**

researchers and practitioners in carrying out predictions related to the energy consumption of smart homes. The results can further be used for various energy optimization techniques and load management. Based on the results obtained, we aim to implement technologies like Integration with Technologies (e.g., Blockchain, Cloud Computing), Anomaly detection, and Application to Different Environmental Conditions. Integrating technologies like blockchain and cloud computing aims to improve data security and continuous prediction reliability. Blockchain ensures secure and tamper-resistant data storage, while cloud computing provides a scalable and accessible platform for data collection.

Anomaly detection entails the identification and management of atypical or erroneous data. To tackle this issue, we suggest incorporating a mechanism that can effectively identify and eliminate anomalous data points during the training phase of the model. Moreover, employing a technique of initially training models on varied datasets and subsequently refining them to suit particular environmental circumstances could augment their ability to adapt. The process of adapting prediction models to different environmental conditions requires the identification of common characteristics, the implementation of reliable data preprocessing techniques, and the integration of domain-specific expertise. Transfer learning strategies and a design for resilience to environmental variability are crucial. Validation protocols should encompass diverse datasets while also considering ethical considerations.

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Declarations

Ethical approval No animals were involved in this study. All applicable international, national, and/or institutional guidelines for the care and use of animals were followed.

Conflict of interest The authors declare that they do not have any conflict of interests that influence the work reported in this paper.

References

1. Gray CA, Acces M (2018) Energy consumption of Internet of Things applications and services. Minerva Acces. http://hdl.handle.net/11343/224197. Accessed 15 Aug 2023
2. Domb M (2019) 'Smart home systems based on Internet of Things', IoT and Smart Home Automation [Working Title]. IntechOpen. https://doi.org/10.5772/intechopen.84894
3. Jo H, Yoon YI (n.d.) Intelligent smart home energy efficiency model using artificial tensorflow engine - human-centric computing and Information sciences, SpringerLink. https://doi.org/10.1186/s13673-018-0132-y
4. Balachandran A, Ramalakshmi, Venkatesan M, Lakshmi K, Jahnavi, Jothi V (2019) Energy consumption analysis and load management for Smart Home. 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), pp 46–49. https://doi.org/10.1109/ICOEI.2019.8862734
5. Kalezhi J, Ntalasha D, Chisanga T (2018) Using internet of things to regulate energy consumption in a home environment. IEEE PES/IAS PowerAfrica 2018:551–555. https://doi.org/10.1109/PowerAfrica.2018.8521060
6. Lee S, Choi D-H (2020) Energy management of smart home with home appliances, energy storage system and electric vehicle: a hierarchical deep reinforcement learning approach. Sensors 20(7):2157. https://doi.org/10.3390/s20072157
7. Ekonomou L (2010) Greek long-term energy consumption prediction using artificial neural networks. Energy 35(2):512–517. https://doi.org/10.1016/j.energy.2009.10.018

8. Gonçalves I, Gomes Á, Henggeler Antunes C (2019) Optimizing the management of Smart Home Energy Resources under different power cost scenarios. Appl Energy 242:351–363. https://doi.org/10.1016/j.apenergy.2019.03.108

9. Chen Z, Sivaparthipan CB, Muthu B (2022) IOT based smart and intelligent smart city energy optimization. Sustain Energy Technol Assess 49:101724. https://doi.org/10.1016/j.seta.2021.101724

10. Wang Z, Wang Y, Zeng R, Srinivasan RS, Ahrentzen S (2018) Random forest based hourly building energy prediction. Energy Build 171:11–25. https://doi.org/10.1016/j.enbuild.2018.04.008

11. Shorfuzzaman M, Hossain MS (2021) Predictive analytics of energy usage by IOT-based smart home appliances for Green Urban Development. ACM Trans Internet Technol 22(2):1–26. https://doi.org/10.1145/3426970

12. Somu N, Ramamritham K (2020) A hybrid model for building energy consumption forecasting using long short term memory networks. Appl Energy 261. https://doi.org/10.1016/j.apenergy.2019.114131

13. Kim T-Y, Cho S-B (2019) Predicting residential energy consumption using CNN-LSTM neural networks. Energy 182:72–81. https://doi.org/10.1016/j.energy.2019.05.230

14. Seyedzadeh S, Pour Rahimian F, Rastogi P, Glesk I (2019) Tuning machine learning models for prediction of building energy loads. Sustain Cities Soc 47:101484. https://doi.org/10.1016/j.scs.2019.101484

15. Liang K, Liu F, Zhang Y (2020) Household power consumption prediction method based on selective ensemble learning. IEEE Access 8:95657–95666. https://doi.org/10.1109/access.2020.2996260

16. Bedi G, Venayagamoorthy GK, Singh R (2020) Development of an IOT-driven building environment for prediction of electric energy consumption. IEEE Internet Things J 7(6):4912–4921. https://doi.org/10.1109/jiot.2020.2975847

17. Kiprijanovska I, Stankoski S, Ilievski I, Jovanovski S, Gams M, Gjoreski H (2020) HousEEC: day-ahead household electrical energy consumption forecasting using deep learning. Energies 13(10):2672. https://doi.org/10.3390/en13102672

18. Ullah A et al (2020) Deep learning assisted buildings energy consumption profiling using smart meter data. Sensors 20(3):873. https://doi.org/10.3390/s20030873

19. Le T, Vo MT, Kieu T, Hwang E, Rho S, Baik SW (2020) Multiple electric energy consumption forecasting using a cluster-based strategy for transfer learning in Smart Building. Sensors 20(9):2668. https://doi.org/10.3390/s20092668

20. Hajj-Hassan M, Awada M, Khoury H, Srour I (2020) A behavioral-based machine learning approach for predicting building energy consumption. Construction Research Congress 2020. https://doi.org/10.1061/9780784482865.109

21. Pavlicko M, Vojteková M, Blažeková O (2022) Forecasting of electrical energy consumption in Slovakia. Mathematics 10(4):577. https://doi.org/10.3390/math10040577

22. Geetha R, Ramyadevi K, Balasubramanian M (2021) Prediction of domestic power peak demand and consumption using supervised machine learning with smart meter dataset. Multimed Tools Appl 80:19675–19693. https://doi.org/10.1007/s11042-021-10696-4

23. Yağanoğlu M et al (2023) Design and validation of IOT based Smart Classroom. Multimed Tools Appl. https://doi.org/10.1007/s11042-023-15872-2

24. Kumari A, Tanwar S (2021) Multiagent-based secure energy management for multimedia grid communication using Q-learning. Multimed Tools Appl 81(25):36645–36665. https://doi.org/10.1007/s11042-021-11491-x

25. Tran SN, Ngo T-S, Zhang Q, Karunanithi M (2020) Mixed-dependency models for multi-resident activity recognition in smart homes. Multimed Tools Appl 79:31–32. https://doi.org/10.1007/s11042-020-09093-0

26. Talaat FM (2022) Effective prediction and resource allocation method (EPRAM) in fog computing environment for smart healthcare system. Multimed Tools Appl 81(6):8235–8258. https://doi.org/10.1007/s11042-022-12223-5

27. Arghira N, Hawarah L, Ploix S, Jacomino M (2012) Prediction of appliances energy use in Smart homes. Energy 48(1):128–134. https://doi.org/10.1016/j.energy.2012.04.010

28. Smith J, Johnson A (2022) End-to-end time series analysis and forecasting: a trio of SARIMAX, LSTM, and prophet - Part 1. Towards data science. [Online]. Available: https://towardsdatascience.com/end-to-end-time-series-analysis-and-forecasting-a-trio-of-sarimax-lstm-and-prophet-part-1-306367e57db8. Accessed 15 Aug 2023

29. Elsaraiti M, Ali G, Musbah H, Merabet A, Little T (2021) Time series analysis of electricity consumption forecasting using Arima model. 2021 IEEE Green Technologies Conference (GreenTech). https://doi.org/10.1109/greentech48523.2021.00049

30. Doe J (2021) Introduction to Long Short-Term Memory (LSTM). Analytics Vidhya. [Online]. Available: https://www.analyticsvidhya.com/blog/2021/03/introduction-to-long-short-term-memory-lstm/. Accessed 15 Aug 2023
31. Alemdar H, Durmaz Incel O, Ertan H, Ersoy C (2013) ARAS human activity datasets in multiple homes with multiple residents. Proceedings of the ICTs for improving Patients Rehabilitation Research Techniques. https://doi.org/10.4108/icst.pervasivehealth.2013.252120
32. Smagulova K, James AP (2019) A survey on LSTM memristive neural network architectures and applications. Eur Phys J Spec Top 228:2313–2324. https://doi.org/10.1140/epjst/e2019-900046-x
33. Eswarsai (2021) Exploring different types of LSTM, Available: https://medium.com/analytics-vidhya/exploring-different-types-of-lstms-6109bcb037c4. Accessed 15 Aug 2023

Authors and Affiliations

Sangeeta Malik[1] · Sitender Malik[2] · Ishmeet Singh[1] · Harsh Vardhan Gupta[1] · Sidhant Prakash[1] · Rachna Jain[3] · Biswaranjanjan Acharya[4] · Yu-Chen Hu[5]

✉ Yu-Chen Hu
  ychu@thu.edu.tw

  Sangeeta Malik
  sangeeta@msit.in

  Sitender Malik
  sitendra@msit.in

  Ishmeet Singh
  singhishmeet16@gmail.com

  Harsh Vardhan Gupta
  harsh.madhav.gupta@gmail.com

  Sidhant Prakash
  sidhant7prakash@gmail.com

  Rachna Jain
  rachnajain@bpitindia.com

  Biswaranjanjan Acharya
  biswaacharya@ieee.org

1 Computer Science and Engineering, Maharaja Surajmal Institute of Technology, New Delhi, India

2 IT Department, Maharaja Surajmal Institute of Technology, New Delhi, India

3 IT Department, Bhagwan Parshuram Institute of Technology, New Delhi, India

4 Department of Computer Engineering-AI & BD, Marwadi University, Rajkot, Gujarat, India

5 Department of Computer Science, TungHai University, No. 1727, Sec. 4, Taiwan Boulevard, Xitun District, 407224 Taichung City, Taiwan, Republic of China