JOURNAL OF DATA SCIENCE | Vol.2024:41
eISSN:2805-5160

Precision Prediction of Household Electricity Consumption Through Data-Driven Model

Sadaf Hashmi[1], Solomon Jebaraj[2], Utpalkumar B Patel[3], Bhumika[4]

[1]Department of ISME, ATLAS SkillTech University, Mumbai, Maharashtra, India
[2]Department of Computer Science and Information Technology, Jain (Deemed to be University), Bangalore, Karnataka, India
[3]Computer Science and Engineering, Parul Institute of Technology, Parul University, Vadodara, Gujarat, India
[4]Centre for Research Impact and Outcome, Institute of Engineering and Technology, Chitkara University, Rajpura, 140401, Punjab, India

Email: sadaf.hashmi@atlasuniversity.edu.in[1], solomon.j@jainuniversity.ac.in[2], utpalkumar.patel8853@paruluniversity.ac.in[3], bhoomika@chitkara.edu.in[4]

Abstract

An effective strategy for managing energy and sustainability is the accurate forecasting of household electricity consumption. A new challenge arises in consumption patterns for traditional models, which face difficulties in variability and data variety. This study aims to bridge the gap by proposing a novel technique called the Mountain Gazelle optimizer-driven Malleable Random Forest technique (MG-MRF), for improving electricity consumption prediction. This has enabled MG-MRF to model different consumption patterns as well as manage variability in the data. The study collected extensive datasets from different households, and those datasets had to undergo preprocessing to ensure integrity. Evaluation results of the approach further underscore the potential of MG-MRF to give accurate and dependable predictions, consequently allowing informed decision-making for the consumption of energy. The proposed method outperformed the traditional models with a prediction accuracy of 98.2%, precision of 94%, recall of 90%, and an f1-score of 92%. This study emphasizes the importance of adaptive modeling techniques in understanding and predicting household electricity usage, enabling the development of more effective energy management strategies. The experimental results advocate and contribute to sustainable energy practices by raising consumer awareness regarding their electrical consumption.

Keywords

Mountain Gazelle optimizer-driven Malleable Random Forest (MG-MRF); Electricity Consumption Prediction; Adaptive Modeling; Energy Management

Submission: 24 July 2024; **Acceptance**: 28 October 2024

Introduction

As people demand increasingly higher amounts of electricity worldwide, coupled with a rapidly growing awareness of renewable energy, forecasting household electricity usage has become essential. In such regard, Alzoubi (2022) stated that better electricity usage forecasting is associated with better resource allocation. The increase in grid stability can be utilized for better planning of the energies production, resulting in reduced environmental impacts, as argued by Liaqat et al. (2021).

Traditional approaches like time-series models are no longer sufficient to deal with the uncertainty and variability that characterize household energy use, which is mainly dependent on weather, user behavior, and appliance efficiency, as noted by Al Misba et al. (2023). New advances in data-driven methods have enabled the use of more sophisticated models that, through a simple structure can capture subtle patterns of electricity use, according to Wang et al. (2021).

Hamdoun et al. (2021) emphasized that the capabilities of machine learning (ML) algorithms when combined with large datasets, provide greater precision through the analysis of temporal dependency. Some methods, such as random forest (RF) and long short-term memory (LSTM), demonstrated improved self-reliance and accuracy in predictions, which can further enhance the energy management at individual and systemic levels, according to Zhou et al. (2022).

According to Gebremeskel et al. (2021), utility firms will benefit from better balancing of supply and demand, reduced peak energy production costs, and enhanced public education on appliance use for energy conservation. Zaidan et al. (2022) argue that the intensification of global efforts to optimize energy consumption accurate predictions of household electricity usage crucial for creation of an efficient and sustainable environment.

This paper aims to predict household electricity consumption using the Mountain Gazelle optimizer-driven Malleable Random Forest (MG-MRF) model, which would ultimately provide insight into efficient energy management and further promote the use of sustainable electricity.

The study is broadly classified as follows: Section 2 related works based on the prediction of household's electrical consumption. In Section 3, data collection and preprocessing together with the proposed MG-MRF approach are discussed. The results are presented in Section 4, where the performances of different models are shown. The study is concluded with some insightful observations and possible future developments in Section 5.

Methodology

The method first gathers the house electricity consumption dataset. Then, data preprocessing using min-max normalization is performed to improve uniform scaling. After that, the Mountain Gazelle optimizer-driven Malleable Random Forest (MG-MRF) model is developed, and the results are obtained. Figure 1 depicts the overall flow of the research.

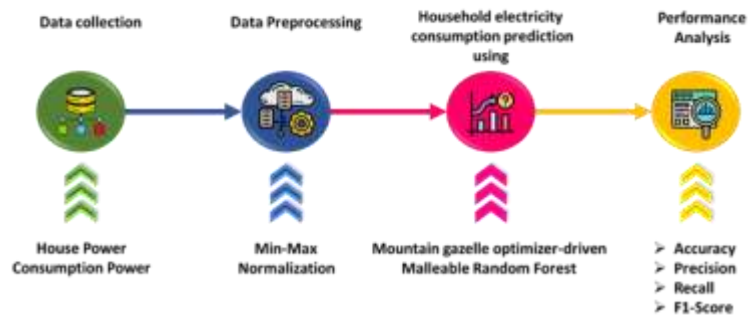

Figure 1. Overall flow of the research

1. Dataset

The household electricity consumption dataset consists of measurements at minute intervals of reactive and active power, energy sub-metering and, voltage values in a single household. It includes variables like global reactive power, active power, sub-metering, and voltage data from different appliances.

2. Min-max normalization for data preprocessing

Min-max normalization is the most commonly utilized method for data preprocessing. In this technique, each component's basic prediction is placed at a value of 0, the highest possible value is placed at a value of 1, and any additional value is located at a decimal fall between 0 and 1. Equation (1) represents the general formula for calculating min-max normalization.

$$W_{scaled} = \frac{(w - w_{min})}{(w_{max} - w_{min})} \tag{1}$$

where, w_{min} is the minimum possible value of W, and w_{max} is the maximum possible value of W.

3. Mountain Gazelle optimizer-driven Malleable Random Forest (MG-MRF)

The mountain gazelle optimizer-driven malleable random forest (MG-MRF) model for electricity consumption in a household is one of the advanced predictive tools, in which dynamic consumption patterns are applied to optimize model parameters for conventional RF using the mountain gazelle optimizer. This fusion encourages efficiencies and accuracies that make prediction of electric consumption a simple task. Further, the MG-MRF model can easily handle large amounts of data, and it often uses a real-time smart grid system that would enhance the efforts toward energy management.

3.1. Malleable Random Forest (MRF)

MRF adapts to any characteristic of the input data while at the same time improving its prediction of electricity consumption by a household. Improvement of adaptability towards changing usage, patterns along with robustness of calculation enhances prediction accuracy, which aids in effective energy management and helps households make informed decisions about power consumption.

To increase the system's classification accuracy, the suggested MRF model optimizes the decision tree node splitting algorithm through an adaptive parameter selection procedure. The different qualities will result in different decision trees when alternative node-splitting algorithms are chosen for the same data set. The accuracy of the MRF algorithm is found to be different.

Consequently, in addition to creating a new splitting rules for selecting and dividing node characteristics, the decision tree will be selected first, followed by choosing the best attribute to split the nodes. The node-splitting algorithm is divided into a linear combination. By dividing the sample set C using characteristics b, the node splitting formula of Equations (2) and (3) displays the $Gain$ and the $Gini$ index.

$$Gain(C,b) = Ent(C) - \sum_{u=1}^{U} \frac{|C^u|}{|C|} Ent(C^u) \tag{2}$$

$$Gini(C,b) = \sum_{u=1}^{U} \frac{|C^u|}{C} Gini(C^u) \tag{3}$$

Where C^u signifies that every sample in the C with a value of b^u on the attribute b is contained in the u branch node. Equation (4) defines the entropy $Ent(C)$ and Equation (5) defines the value of $Gini(C)$ that calculates the degree of inequality.

$$Ent(C) = -\sum_{l=1}^{|z|} o_l log_2 o_l \tag{4}$$

$$Gini(C) = \sum_{l=1}^{|z|} \sum_{l' \neq l} olol' = 1 - \sum_{l=1}^{|z|} ol^2 \tag{5}$$

Equation (6) provides the combination node splitting formula and adaptive parameter selection procedure. The goal of node splitting is to enhance the quality of the dataset after division.

$$G = \min_{\propto,\beta \epsilon Q} E\{C,b\} = \alpha Gini(C,b) - \beta Gain(C,b)$$

$$s.t. \begin{cases} \alpha + \beta = 1 \\ 0 \leq \alpha, \beta \leq 1 \end{cases} \tag{6}$$

Where the weight coefficient of variable splitting is denoted by the symbols $\alpha, and\ \beta$. G has a low value in the interim period. To find the best combination of parameters, the adaptive parameter selection method is used. The accuracy rate and the categorization error rate are utilized in the test to assess the efficiency of the MRF model. Equation (7) defines the sample $C's$ categorization error rate as follows:

$$F(e;C) = \frac{1}{n} \sum_{j=1}^{n} JJ(e(w_j) \neq z_j) \tag{7}$$

The rate of accuracy is described as follows in Equation (8).

$$acc(e;C) = \frac{1}{n} \sum_{j=1}^{n} JJ\left(e\left(w_j\right) = z_j\right) = 1 - F(e;C) \tag{8}$$

3.2. Mountain Gazelle optimizer (MG)

The Mountain Gazelle optimizer (MG) is an optimizer that applies mountain gazelle foraging behavior to determine household electricity consumption. This will seek out an efficient solution in the search space so that the parameters of predictive models are refined to enhance the accuracy and reliability in forecasting electricity usage patterns based on historical information and influence factors. The social hierarchy of the mountain gazelle herd served as the model for one of the created population-based on optimization algorithms known as the Mountain Gazelle optimizer (MG).

This algorithm was inspired by nature. Four mountain gazelle behaviors are included in the mathematical structure of the MG algorithm: bachelor male herds (BMH), maternity herds (MH), migration in search of food (MSF), and territorial solitary males (TSM). A resolution to the optimization issue (W) with C response variables is represented by each gazelle in the population.

The MG algorithm defines a significant amount of random numbers, with the following notations for each. $M(C)$, where C is the number of elements, defines vectors of random numbers derived from a normal distribution. Random integers in the range $[1, 2]$ are defined as q_j, and random numbers that follow an even distribution inside the range $[0, 1]$ are defined as q. Initially, the four coefficients must be defined to characterize four behaviors numerically, as indicated by Equation (9).

$$Cof = \begin{cases} b + 1 + q_1 \\ b.M_1(C) \\ q_2(C) \\ M_2(C).M_3(C)^2.\cos\left(2q_3.M_4(C)\right) \end{cases} \tag{9}$$

Where, $b = -1 + iter.\left(\frac{-1}{\max\_iter}\right)$. After that, vector E is defined by using Equation (10).

$$E = M_5(C).\exp\left(2 - iter.\frac{2}{\max\_iter}\right) \tag{10}$$

A simple normal random vector is left in the final iteration of the second half of the duplication of E, which initiates with values larger than 1 (depending on the extreme amount of interactions) and converges exponentially to 1. Equation (11) defines the young male herd coefficient vector, and all appropriate values are defined for that purpose.

$$BH = W_{ra}.q_1 + N_{pr}.q_2 \tag{11}$$

Where W_{ra} is a randomly chosen response from the final third of the population. For the chosen population, N_{pr} stands for mean percentage, averaged across all dimensions in the input vector. The algorithm's TSM component simulates the actions of mature male gazelles that construct and safeguard their territories. Equation (12) describes how it is employed in the algorithm to improve the exploitation ability.

$$TSM = W_1 - \left|\left(q_{j1}.BH - q_{j2}.W_s\right).E\right|Cof_q \tag{12}$$

Where Cof_q is the randomly chosen coefficient from Equation (9), W_s is the agent that is currently being updated, and W_1 is the best solution that has been discovered so far. The second behavior, MH, is being characterized by females and their offspring, reflecting an equilibrium between exploration and exploitation in the algorithm. Premature convergence is avoided and diversity in the solution space is ensured by this system expressed in equation (13).

$$MH = BH + Cof_q + \left(q_{j3}.W_1 - q_{j4}.W_{rand}\right).Cof_q \tag{13}$$

Where W_{rand} refers to a population-wide solution, that was randomly selected. In order to simulate BMH's performance, the parameter $Dist$ should be calculated as presented in Equation (14).

$$Dist = |W_s - W_1|(2q_6 - 1) \tag{14}$$

The young male gazelles are represented by the third behavior, BMH, which is defined by Equation (15). BMH is used to stretch out the search space for proper exploration of an algorithm.

$$BMH = W_s - Dist + \left(q_{j5}.W_1 - q_{j6}.BH\right).Cof_q \quad (15)$$

Finally, Equation (16) is used to develop MSF. This algorithm has a mechanism of random search.

$$MSF = (ka - va).q_7 + ka \quad (16)$$

Where, ka and va denote the parameter space's lower and upper bounds, respectively.

Results and Discussion

The proposed technique has been implemented using Python (v3.10) on Windows 11 OS. A great capacity for executing ML algorithms is delivered by the system's high-performance IRIS graphics card and Intel Core i7 processor. Here, the proposed Mountain Gazelle optimizer-driven Malleable Random Forest (MG-MRF) is compared with existing methods, such as Transformer model + K-means clustering (TM+KMC), LSTM, and K-means (Zhang et al., 2021) using metrics like precision (%), accuracy (%), f1-score (%), and recall (%).

Figure 2 demonstrates the power consumption of household appliances, highlighting that the AC has consumed almost 80%, while the fan exhibits the lowest at 30%. Other appliances, including TV, washing machine, lights, and fridge, has consumed 70%, 50%, 40%, and 60%, respectively.

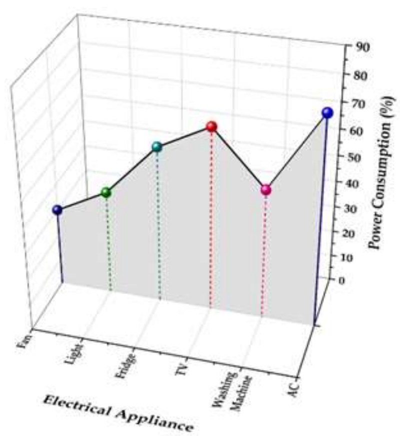

Figure 2. Power consumption output of the appliances

1. Accuracy

The quantity of accurate predictions made over the total number of instances defines accuracy. This provides information on the overall effectiveness of the model constructed for predicting household electricity consumption. The output of accuracy is depicted in Figure 3 and Table 1. The outcome demonstrates that the proposed MG-MRF (98.2%) approach outperforms the existing methods, such as TM+KMC (97%), LSTM (97%), and K-means (96%).

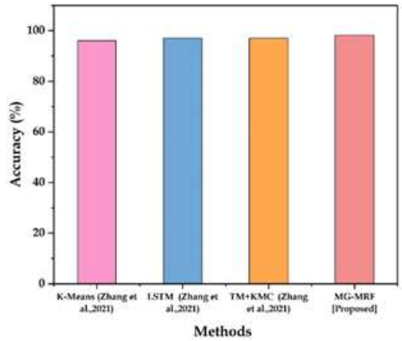

Figure 3. Accuracy

2. Precision

The metric measures the ratio of the true positives among all positive predictions made. In other words, this metric indicates how accurately the model aligns with actual power consumption instances and the rate of false positives. Figure 4 and Table 1 illustrate the output of precision. The result shows that the suggested MG-MRF (94%) method performs superior than existing methods, such as TM+KMC (80%), LSTM (74%), and K-means (82%).

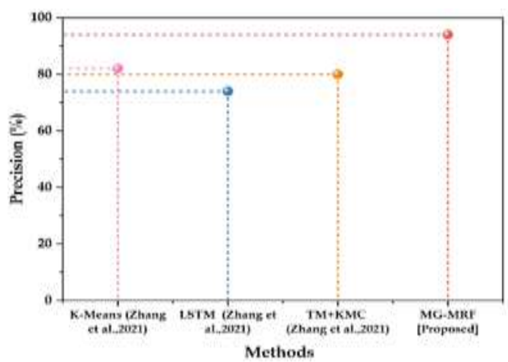

Figure 4. Precision

3. Recall

The metric measures true positive predictions in proportion with the overall positive situation. This is calculated so that every relevant instance of household electricity consumption can be detected correctly. The recall output is given by Figure 5 and Table 1. The outcome demonstrates that the suggested MG-MRF (90%) technique performs well than the existing methods, such as TM+KMC (66%), LSTM (60%), and K-means (28%).

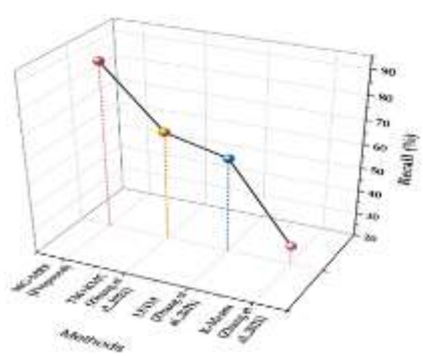

Figure 5. Precision

4. F1-Score

This combines both precision and recall into one metric by balancing them together. This is a very vital measure in determining the goodness of fit for the model in estimating the electricity consumption levels for the household. Figure 6 and Table 1 present the result of f1-score. The outcome depicts that the proposed MG-MRF (92%) framework outperforms the existing approaches, such as TM+KMC (72%), LSTM (66%), and K-means (42%).

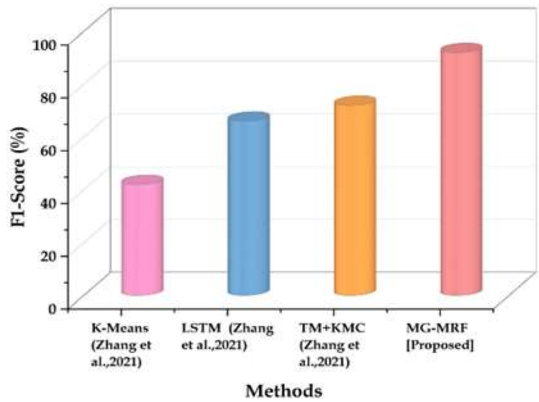

Figure 6. F1-Score

Table 1. Comparison of metrics

Methods	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
K-means (Zhang et al., 2021)	96	82	28	42
LSTM (Zhang et al., 2021)	97	74	60	66
TM+KMC (Zhang et al., 2021)	97	80	66	72
MG-MRF [Proposed]	98.2	94	90	92

References

Al Misba, W., Mavikumbure, H. S., Rajib, M. M., Marino, D. L., Cobilean, V., Manic, M., & Atulasimha, J. (2023). Spintronic physical reservoir for autonomous prediction and long-term household energy load forecasting. *IEEE Access*, *11*, 124725–124737. https://doi.org/10.1109/ACCESS.2023.3326414

Alzoubi, H. A. (2022). Machine learning for intelligent energy consumption in smart homes. *International Journal of Computations, Information and Manufacturing (IJCIM)*, 2(1). https://doi.org/10.54489/ijcim.v2i1.75

Banik, R., Das, P., Ray, S., & Biswas, A. (2021). Prediction of electrical energy consumption based on machine learning technique. *Electrical Engineering*, *103*(2), 909–920. https://doi.org/10.1007/s00202-020-01126-z

Chi, D. (2022). Research on electricity consumption forecasting model based on wavelet transform and multi-layer LSTM model. *Energy Reports*, *8*, 220–228. https://doi.org/10.1016/j.egyr.2022.01.169

Gao, F., Chi, H., & Shao, X. (2021). Forecasting residential electricity consumption using a hybrid machine learning model with online search data. *Applied Energy*, *300*, 117393. https://doi.org/10.1016/j.apenergy.2021.117393

Gebremeskel, D. H., Ahlgren, E. O., & Beyene, G. B. (2021). Long-term evolution of energy and electricity demand forecasting: The case of Ethiopia. *Energy Strategy Reviews*, *36*, 100671. https://doi.org/10.1016/j.esr.2021.100671

Hamdoun, H., Sagheer, A., & Youness, H. (2021). Energy time series forecasting—analytical and empirical assessment of conventional and machine learning models. *Journal of Intelligent & Fuzzy Systems*, *40*(6), 12477–12502. https://doi.org/10.3233/JIFS-201717

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

Liaqat, R., Sajjad, I. A., Waseem, M., & Alhelou, H. H. (2021). Appliance level energy characterization of residential electricity demand: Prospects, challenges and recommendations. *IEEE Access*, *9*, 148676–148697. https://doi.org/10.1109/ACCESS.2021.3123196

Liu, C., Sun, B., Zhang, C., & Li, F. (2020). A hybrid prediction model for residential electricity consumption using holt-winters and extreme learning machine. *Applied Energy*, *275*, 115383. https://doi.org/10.1016/j.apenergy.2020.115383

Saoud, L. S., Al-Marzouqi, H., & Hussein, R. (2022). Household energy consumption prediction using the stationary wavelet transform and transformers. *IEEE Access*, *10*, 5171–5183. https://doi.org/10.1109/ACCESS.2022.3140818

Shao, X., Kim, C. S., & Sontakke, P. (2020). Accurate deep model for electricity consumption forecasting using multi-channel and multi-scale feature fusion CNN–LSTM. *Energies, 13*(8), 1881. https://doi.org/10.3390/en13081881

Shen, M., Lu, Y., Wei, K. H., & Cui, Q. (2020). Prediction of household electricity consumption and effectiveness of concerted intervention strategies based on occupant behavior and personality traits. *Renewable and Sustainable Energy Reviews, 127*, 109839. https://doi.org/10.1016/j.rser.2020.109839

Wang, Z., Hong, T., Li, H., & Piette, M. A. (2021). Predicting city-scale daily electricity consumption using data-driven models. *Advances in Applied Energy*, *2*, 100025. https://doi.org/10.1016/j.adapen.2021.100025

Zaidan, E., Abulibdeh, A., Alban, A., & Jabbar, R. (2022). Motivation, preference, socioeconomic, and building features: New paradigm of analyzing electricity consumption in residential buildings. *Building and Environment*, *219*, 109177. https://doi.org/10.1016/j.buildenv.2022.109177

Zhou, X., Lin, W., Kumar, R., Cui, P., & Ma, Z. (2022). A data-driven strategy using long short term memory models and reinforcement learning to predict building electricity consumption. *Applied Energy, 306*, 118078. https://doi.org/10.1016/j.apenergy.2021.118078

Zhang, J., Zhang, H., Ding, S., & Zhang, X. (2021). Power consumption predicting and anomaly detection based on transformer and K-means. *Frontiers in Energy Research, 9*, 779587. https://doi.org/10.3389/fenrg.2021.779587