Daily Power Consumption Plan Derivation via the Monte Carlo-Based Regression Tree Algorithm

Genrawan, Gh, Hoendarto*
Institute Technology of Sepuluh Nopember Surabaya, Indonesia

Ahmad, As, Saikhu
Institute Technology of Sepuluh Nopember Surabaya, Indonesia

Rv Hari, Rvhg, Ginardi
Institute Technology of Sepuluh Nopember Surabaya, Indonesia

ABSTRACT

Various methods and algorithms are used in predicting electricity consumption. Data-based methods can produce mathematical models with efficiency and good accuracy and do not require professional knowledge. In this research, electricity consumption prediction will be performed based on training a Monte Carlo (MC) simulation on each leaf generated by the regression tree (RT) algorithm. The prediction no longer relies on the average of the samples contained in the leaf, but now relies on the sample probabilities. Often the regression tree algorithm gives overfitting results, so training each leaf will eliminate this. The dataset from Trapeznikov Institute of Control Sciences (TICS), Russia will be used to train and test the proposed method were obtained from because they were adequately recorded. The proposed Monte Carlo Regression Tree (MCRT) algoritm is used to train monthly data and tested on different months' data. The results are used to make predictions of daily trend usage to determine if there is any irregularity in electricity consumption.

CCS CONCEPTS

• **Theory of computation** → Theory and algorithms for application domains; Machine learning theory; Models of learning.

KEYWORDS

Monte Carlo, Regression trees, Prediction, Electricity Consumption, campus building

ACM Reference Format:

Genrawan, Gh, Hoendarto, Ahmad, As, Saikhu, and Rv Hari, Rvhg, Ginardi. 2024. Daily Power Consumption Plan Derivation via the Monte Carlo-Based Regression Tree Algorithm. In *2024 10th International Conference on Computing and Artificial Intelligence (ICCAI 2024), April 26–29, 2024, Bali Island, Indonesia.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3669754.3669816

1 INTRODUCTION

Population growth and the effects of global climate change have caused the use of electrical energy to increase day by day. Coupled with so many inventions of electronic equipment which certainly requires electrical energy to drive it. Also, the establishment of office buildings, including campus buildings that require electrical energy for the activities of its occupants [1]. According to a report from the International Energy Agency [2], buildings energy demand increased by around 4 per cent from 2020 to 135 EJ – the largest increase in the last 10 years and CO2 emissions from buildings operations have reached an all-time high of around 10 GtCO2, around a 5 per cent increase from 2020. (IEA, 2022).

Some recent studies show that office buildings including universities account for about 40% of the electrical energy demand and greenhouse gas emissions worldwide, due to increasing population and affluence. To improve building energy efficiency solutions such as smart control, wastage detection, and demand management strategies have been developed to address this increasing challenge [3–5].

One of the best ways to save electricity consumption in buildings is through efficiency measures. The development of optimal control strategies is attracting increasing attention and is becoming one of the main solutions to achieve building energy efficiency. To execute an optimal control strategy, forecasting building energy consumption is generally an important component to enable the control system to determine the optimal control settings to achieve energy or cost savings. Precise prediction of building energy consumption is essential for the development of effective building energy management strategies to allow building energy consumption to be determined in advance so that the system can detect unreasonable load changes over time and thus avoid energy wastage [6, 7].

The accuracy of energy consumption prediction will directly affect the overall control performance and energy savings. The research results [8, 9] show that significant control performance improvement and more efficient operation can be achieved by providing more accurate predictions and using predictive controller model formulation. Data-driven approaches are machine learning techniques that focus on learning mathematical models to represent the dynamics of a data set, efficiently generating mathematical models [10, 11]. It is easier to use than model-based approaches that rely on highly specialized knowledge and significant time and effort. Furthermore, this approach is not generalized, so the derived model can be different for each different building structure [12–14].

2 RELATED WORK

Research conducted by [15] offers an improvement in the building consumption profile to show a better profile of the electricity load using statistical methods, Monte Carlo and probability distributions. It is to identify weaknesses and possible improvements in building energy management with the best scenario to optimize demand using Monte Carlo.

*PhD Candidate

ICCAI 2024, April 26–29, 2024, Bali Island, Indonesia

ACM ISBN 979-8-4007-1705-5/24/04
https://doi.org/10.1145/3669754.3669816

This research implemented Decision Tree (DT), Random Forest (RF), and Gradient Boosted Regression Tree (GBRT) methods on residential buildings as the second largest electricity consumers in Turkey. The goal is to detect the factors that determine household electricity consumption in Turkey. Since the GBRT method provides the lowest Root Mean Squared Error (RMSE), the impact of each variable on electricity consumption is analyzed by this method. With the help of Partial Dependence Plots (PDP) provided by the GBRT method, the impact of each categorical and continuous variable is presented [16].

The prediction specifying quantile regression models based on Gradient Boosted Regression Trees proved to improve accuracy over multiple linear regression models at the cost of more complexity. However, it still has simpler specification and tuning compared to other machine learning approaches [17]. The next research [18] states Boosted regression tree (BRT) is a combination of a boosting method and a decision tree algorithm that is repeated to evaluate and correct inaccuracies at each step. As a result, BRT is robust to missing values and outliers, and performs well with large datasets and many observations. The approach evolves the regression tree (RT) incrementally, using an encoded descent function to evaluate and improve the accuracy at each step. Machine learning techniques are used to solve this regression problem. Boosting Regression tree creates an incremental succession of trees in a regression situation, then selects the best tree by using a randomly differentiable loss function. The number of trees required for optimal prediction is determined by a combination of these two factors [19].

The proposed Monte-Carlo simulation [20] has the main advantage of better distributional results. The results are more robust compared to previous studies, as a wide range of system configurations are explored, and the results can be used for grid planning and enable an integrated view of the energy system coupled with other sectors. In this study, synthetic time series based on demographic data, and scenarios are used, and the proposed simulation methodology is based on individual behavior, so there is no need to use multidimensional factors as often used to determine grid dimensioning. Furthermore, the research utilizes Markov Chain Monte Carlo approach and the Seasonal Autoregressive Integrated Moving Average (SARIMA) model to estimate short- and long-term wind power. The framework draws Markov Chain time series data on wind energy to maintain stochasticity and realize the probability transition matrix [21]. To improve the power system resilience evaluation framework and analyze the impact of multiple micro-grid availability on power system resilience. The proposed quantitative resilience measure can be used in decision-making by power system operators and planners for future planning and capacity enhancement schemes. To accommodate the impact of multiple microgrids, a discrete multi-time transition model of the power system in response to extreme events. The system state probabilities (normal, microgrid, and emergency) are determined using time-independent transition matrices and correspond to homogeneous Markov chains [22].

3 PROPOSED METHOD

In this section, an electrical energy consumption prediction algorithm based on the Monte Carlo (MC) algorithm combined with Regression Tree (RT) will be discussed. Specifically, for each leaf of the binary tree formed using the RT algorithm, an MC model will be inserted into it, with the aim of producing more accurate predictions. Furthermore, a case study used to validate the accuracy of the predictive model of the proposed algorithm will be discussed. In this case, the accuracy of the predictive model will be tested on a common dataset from a Russian university.

3.1 Model Regression Tree Definition

Suppose that $X \in \mathbb{R}^{m\times n}$ is a data set containing information of measured electrical parameters of a building, with m and n denoting the number of samples and parameters contained in the data set, respectively. In particular, the proposed algorithm aims to generate a prediction of energy consumption per unit time, based on the information of $\boldsymbol{p} = [p(1)\ p(2) \cdots p(m)]^\intercal \in \mathbb{R}^m$ which are the active power parameters of a building using the MC and RT algorithms. This prediction algorithm will be divided into 3 main parts: (1) definition of RT model, (2) insertion of MC model in each RT leaf and (3) prediction of electricity consumption data in this study.

Suppose from the data set X state that $X = [X_e\ X_t]$ where $X_e = [\boldsymbol{e}_1\ \boldsymbol{e}_2 \ldots \boldsymbol{e}_{\boldsymbol{n}_1}]; X_e \in \mathbb{R}^{m\times n_1}$ is a set of electrical parameter data with $\boldsymbol{e_i} = [e_i(1)\ e_i(2) \ldots e_i(m)]^\intercal; \boldsymbol{e_i} \in \mathbb{R}^m; i = 1, 2, 3, \ldots, n$ are the electrical parameters $i - th$. Furthermore, $X_t = [\boldsymbol{y}\ \boldsymbol{m}\ \boldsymbol{d}\ \boldsymbol{h}\ \boldsymbol{m_i}] \in \mathbb{Z}^{m\times 5}$ is a data set that states the time information of the data set of X, where $\boldsymbol{y}, \boldsymbol{m}, \boldsymbol{d}, \boldsymbol{h}$, and $\boldsymbol{m_i}$ sequentially expresses year, month, day names (Monday, Tuesday, so on), hour and minute information. By definition, therefore $\boldsymbol{p} \in X_e$ as, $\boldsymbol{p}$ is the active power information that is part of the electrical parameters.

In this case, the information of the active power will be affected by the current and voltage. The voltage derived from an electrical substation should be a constant value in the range of $220\ V \pm 11\ V$ which is a standardized value of state-owned power plants (PLN) so it does not need to be modeled. Thus, modeling is sufficient for the electric current. For example $I = [\boldsymbol{I}_1\ \boldsymbol{I}_2\ \boldsymbol{I}_3] \in \mathbb{R}^{m\times 3}$ is the set of current data from phase 1 to 3. For each phase electric current available, a RT model can be derived dependent X_t to estimate the electric current consumption at each time, following the equation 1):

$$\hat{I}_j(k) = f_{RT_j}(\boldsymbol{x_t}(k)) \tag{1}$$

with $\hat{I}_j(k)$ prediction from the phase electric current $j - th$, $\boldsymbol{x_t}(k) = [y(k)\ m(k)\ d(k)\ h(k)\ m_i(k)]; \boldsymbol{x_t}(k) \in X_t$ is the time information vector measured at the time k and $f_{RT_j} : \mathbb{R}^5 \to \mathbb{R}$ represents an RT function to predict the phase electric current $j - th$. In this case, the RT model for predicting the electric power consumption of a building can be expressed as equation 2):

$$\hat{p}(k+1) = f_{RT_p}\left(\boldsymbol{x_t}(k), \hat{I}_1(k), \hat{I}_2(k), \hat{I}_3(k)\right) \tag{2}$$

where $\hat{p}(k+1)$ is a prediction of electric power consumption at the time of $k+1$ and f_{RT_p}: $\mathbb{R}^8 \to \mathbb{R}$ is an RT function to predict active power based on 3-phase current and time information.

3.2 Monte Carlo on the Regression Tree Model

In order to improve the accuracy of the model formed in step 1.1, then for each leaf of the f_{RT_p} will be inserted an MC model to estimate the active power. Suppose that $X_p = [X_t\ I]$ is a data set related to the active power measurement of a building. Furthermore,

Table 1: Model Predictive Accuracy for classical and MC-based RT algorithm

	8th October 2021	
Algorithm	RMSE	NRMSE
Classical RT	125.39	93.54
MCRT	124.56	93.87
	10th November 2021	
Algorithm	RMSE	NRMSE
Classical RT	243.47	91.85
MCRT	243.14	91.86

assume that $\boldsymbol{T}$ is a binary tree model generated by the function f_{RT_p} using the data X_p. Thus, for each leaf $l = 1, 2, \ldots, L$ there will be L small datasets, R_l as such that $U_L^{l=1} R_l = X_p$. The main purpose of this stage is to insert an MC model in each of the following $R_l \in \boldsymbol{T}$.

Suppose that the $l(\bar{\boldsymbol{x}}_{\boldsymbol{t}}(k)) = \{l : \bar{\boldsymbol{x}}_{\boldsymbol{t}}(k) \in R_l\}$ is a function to determine which leaf is occupied by the data $\bar{\boldsymbol{x}}_{\boldsymbol{t}}(k)$ at tree $\boldsymbol{T}$ with $\overline{\boldsymbol{x}_{\boldsymbol{t}}}(k) = [\, \boldsymbol{x}_{\boldsymbol{t}}(k)\ \hat{I}_1(k)\ \hat{I}_2(k)\ \hat{I}_3(k)]$, then $\forall R_l \in \boldsymbol{T}, \exists \boldsymbol{p}_{\boldsymbol{l}} = \{\boldsymbol{p} : \bar{\boldsymbol{x}}_{\boldsymbol{t}}(k) \in R_l\}; \boldsymbol{p}_{\boldsymbol{l}} \in \boldsymbol{p}$. In this case, it is possible to create MC simulations for $\boldsymbol{p}_{\boldsymbol{l}}$. Suppose that $\tilde{\boldsymbol{p}}_{\boldsymbol{l}}$ is $\boldsymbol{p}_{\boldsymbol{l}}$ which has been sorted from the smallest to the largest. Hence in this case, it is possible to group the values of the $\tilde{\boldsymbol{p}}_{\boldsymbol{l}}$ based on its quartile values, for example, a random variable can be defined as A_l in equation 3):

$$A_l = \begin{cases} 1\ jika\ p_l \le Q_{l1} & ; P(A_l = 1) = p_{l1} \\ 2\ jika\ Q_{l1} < p_l \le Q_{l2} & ; P(A_l = 2) = p_{l2} \\ 3\ jika\ Q_{l2} < p_l \le Q_{l3} & ; P(A_l = 3) = p_{l3} \\ 4\ jika\ p_l \ge Q_{l3} & ; P(A_l = 4) = p_{l4} \end{cases} \tag{3}$$

with Q_i states the quartile $i-th$ from $\tilde{\boldsymbol{p}}_{\boldsymbol{l}}$. Further, a new parameter can be defined, $\boldsymbol{b}_{\boldsymbol{l}} = [b_l(1)\ b_l(2)\ b_l(3)\ b_l(4)]^\intercal \in \mathbb{R}^4$ as a vector with information of the average for the data contained in each data segment. For example $b_l(1)$ is the average of all values $p_l \in \tilde{\boldsymbol{p}}_{\boldsymbol{l}}$ associated with $A_l = 1$, or, with slight abuse of notation, $b_l(1) = \bar{\boldsymbol{a}}_{\boldsymbol{l}}(1); \boldsymbol{a}_{\boldsymbol{l}}(1) = \{p_l \in \tilde{\boldsymbol{p}}_{\boldsymbol{l}} | p_l \le Q_{l1},\ \}$, dengan $\bar{\boldsymbol{a}}_{\boldsymbol{l}}(1)$ declare average from $\boldsymbol{a}_{\boldsymbol{l}}(1)$.

3.3 Electricity Power Consumption Prediction

As shown above, it is possible to predict the value of the active power at a certain time. For a random number set $\boldsymbol{x} \in [0, 1]^{\boldsymbol{n_t}}$ with n_t denotes the sum of random numbers, let's say $\boldsymbol{x}_{\boldsymbol{pl}}$ as an index vector equation 4), thus $\forall x \in \boldsymbol{x}$:

$$x_{p_l} = \begin{cases} 1\ if\ x < p_{l1} \\ 2\ if\ p_{l1} \le x < p_{l2} \\ 3\ if\ p_{l2} \le x < p_{l3} \\ 4\ if\ x \ge p_{l3} \end{cases} \tag{4}$$

which in this case, the prediction of active power at the time of $\boldsymbol{x}_{\boldsymbol{t}}(k)$ can be expressed as equation 5):

$$\hat{p}\,(\mathrm{k}) = \sum_{k=1}^{n_t} \frac{b_l\left(x_{p_l}(k)\right)}{n_t} \tag{5}$$

where $\hat{p}(k)$ is the prediction of the power at time k.

Figure 1: V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences

4 RESULTS AND DISCUSSION

Experiments were conducted using Matlab to process datasets from the V.A. Trapeznikov Institute of Control Sciences (TICS) of the Russian Academy of Sciences Figure 1 available for 2 years 2020 and 2021. The results shown use feeder data of 41 June and July 2021 as training data and use data of 8th October 2021 and 10 November 2021 as test data. The dataset were preprocessed before applying the proposed algorithm in order to clean the dataset from several missing data points that occurred randomly. Table 1 shows the model predictive accuracy for both algorithms, while Figure 2 shows the plot of the electrical power consumption on selected dates.

From Table 1. We can see that the proposed algorithm can predict the dynamics of the power consuption precisely. As in Figure 2, it can be seen that the MCRT predicted power consumption closely resemble the true power consumption data provided by TICS. In particular, the average accuracy on both dates is 92.86 for the MCRT algorithm and 92.69 for the classical RT algorithm. From this result, we can see that assigning MC algorithm into each leaf of the RT algorithm can potentially increase the model predictive accuracy of the RT algorithm.

Figure 3 shows the daily power consumption prediction with the proposed algorithm on the selected dates. From Figure 3 we can see

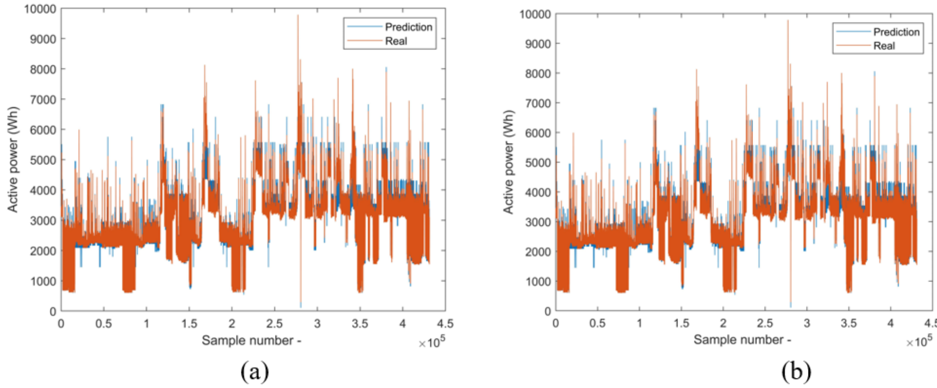

Figure 2: Monthly Power Consumption for October (a) and November (b) 2021

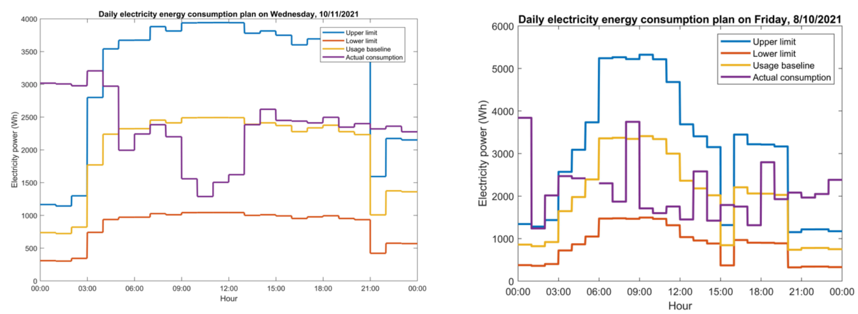

Figure 3: Electrical power consumption plan for TICS on 8th October and 10th November 2021

that most of the true power consumption dynamics are contained inside the bound defined by the proposed algorithm. There are some violations to the power consumption at during 9 pm up to 3 am. This is mostly due to the limitation of the dataset, where the power consumption during the last 2 months were slightly lower during the summer, where on the October and November there were more usage on the electrical power due to usage of the heating appliances on the colder temperature during the autumn. These results show the possibility of applying the proposed algorithm for the purpose of building power consumption monitoring. I.e. it is possible to combine this algorithm with remote control devices, in order to check whether there are power usage outside of active hour, and then turn it of remotely via the application of Internet of things.

5 COMPETING INTEREST

The authors state that none of the financial or interpersonal issues they have dealt with have had a discernible effect on the research reported in this paper.

ACKNOWLEDGMENTS

All people who have made a major contribution to the work detailed in the book but who do not fit the criteria for copyright are acknowledged in the credits and have provided us with written permission to do so. Technical aid, writing and editing support, and general support are a few examples.

REFERENCES

[1] X. Cao, X. Dai, and J. Liu, "Building energy-consumption status worldwide and the state-of-the-art technologies for zero-energy buildings during the past decade," *Energy Build*, vol. 128, pp. 198–213, 2016, doi: 10.1016/j.enbuild.2016.06.089.

[2] United Nations Environment Programme, 2022 Global Status Report for Buildings and Construction: Towards a Zero-emission, Efficient and Resilient Buildings and Construction Sector, vol. 224. 2022. [Online]. Available: https://wedocs.unep.org/handle/20.500.11822/41133

[3] A. Li, F. Xiao, C. Fan, and M. Hu, "Development of an ANN-based building energy model for information-poor buildings using transfer learning," *Build Simul*, vol. 14, no. 1, pp. 89–101, 2021, doi: 10.1007/s12273-020-0711-5.

[4] H. H. Hosamo, M. S. Tingstveit, H. K. Nielsen, P. R. Svennevig, and K. Svidt, "Multiobjective optimization of building energy consumption and thermal comfort based on integrated BIM framework with machine learning-NSGA II," *Energy Build*, vol. 277, p. 112479, 2022, doi: 10.1016/j.enbuild.2022.112479.

[5] T. Liu, Z. Tan, C. Xu, H. Chen, and Z. Li, "Study on deep reinforcement learning techniques for building energy consumption forecasting," *Energy Build*, vol. 208, 2020, doi: 10.1016/j.enbuild.2019.109675.

[6] K. Amasyali and N. M. El-Gohary, "Real data-driven occupant-behavior optimization for reduced energy consumption and improved comfort," *Appl Energy*, vol. 302, Nov. 2021, doi: 10.1016/j.apenergy.2021.117276.

[7] X. Zhou, W. Lin, R. Kumar, P. Cui, and Z. Ma, "A data-driven strategy using long short term memory models and reinforcement learning to predict building electricity consumption," *Appl Energy*, vol. 306, Jan. 2022, doi: 10.1016/j.apenergy.2021.118078.

[8] E. A. Hailu, "Energy Efficiency Analysis And Auditing Of Ethiopian University Buildings: Case Of Buildings In Arba Minch Institute Of Technology," *INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH*, vol. 6, no. 09, 2017, [Online]. Available: www.ijstr.org

[9] W. Guo, L. Che, M. Shahidehpour, and X. Wan, "Machine-Learning based methods in short-term load forecasting," *Electricity Journal*, vol. 34, no. 1, p. 106884, 2021, doi: 10.1016/j.tej.2020.106884.

[10] Y. Dong, X. Ma, and T. Fu, "Electrical load forecasting: A deep learning approach based on K-nearest neighbors," *Appl Soft Comput*, vol. 99, p. 106900, 2021, doi: 10.1016/j.asoc.2020.106900.

[11] G. Hafeez, K. S. Alimgeer, and I. Khan, "Electric load forecasting based on deep learning and optimized by heuristic algorithm in smart grid," *Appl Energy*, vol. 269, no. March, p. 114915, 2020, doi: 10.1016/j.apenergy.2020.114915.

[12] K. Li, Z. Ma, D. Robinson, W. Lin, and Z. Li, "A data-driven strategy to forecast next-day electricity usage and peak electricity demand of a building portfolio using cluster analysis, Cubist regression models and Particle Swarm Optimization," *J Clean Prod*, vol. 273, p. 123115, 2020, doi: 10.1016/j.jclepro.2020.123115.

[13] K. Amasyali and N. M. El-Gohary, "A review of data-driven building energy consumption prediction studies," *Renewable and Sustainable Energy Reviews*, vol. 81, no. January 2016, pp. 1192–1205, 2018, doi: 10.1016/j.rser.2017.04.095.

[14] Z. Liu *et al.*, "A review of data-driven smart building-integrated photovoltaic systems: Challenges and objectives," *Energy*, vol. 263, no. PE, p. 126082, 2023, doi: 10.1016/j.energy.2022.126082.

[15] J. N. Castillo, V. F. Resabala, L. O. Freire, and B. P. Corrales, "Modeling and sensitivity analysis of the building energy consumption using the Monte Carlo method," *Energy Reports*, vol. 8, pp. 518–524, Dec. 2022, doi: 10.1016/j.egyr.2022.10.198.

[16] D. Guven and M. O. Kayalica, "Analysing the determinants of the Turkish household electricity consumption using gradient boosting regression tree," *Energy for Sustainable Development*, vol. 77, no. September, p. 101312, 2023, doi: 10.1016/j.esd.2023.101312.

[17] G. Díaz, J. Coto, and J. Gómez-Aleixandre, "Prediction and explanation of the formation of the Spanish day-ahead electricity price through machine learning regression," *Appl Energy*, vol. 239, no. February, pp. 610–625, 2019, doi: 10.1016/j.apenergy.2019.01.213.

[18] G. Wang *et al.*, "Predication of entropy generation rate in a concentrating photovoltaic thermal system with twisted tube turbulator using Boosted regression tree algorithm," *Case Studies in Thermal Engineering*, vol. 53, p. 103729, 2024, doi: 10.1016/j.csite.2023.103729.

[19] P. Sharma and B. B. Sahoo, "Precise prediction of performance and emission of a waste derived Biogas–Biodiesel powered Dual–Fuel engine using modern ensemble Boosted regression Tree: A critique to Artificial neural network," *Fuel*, vol. 321, no. February, p. 124131, 2022, doi: 10.1016/j.fuel.2022.124131.

[20] S. Arens, S. Schlüters, B. Hanke, K. Von Maydell, and C. Agert, "Monte-Carlo Evaluation of Residential Energy System Morphologies Applying Device Agnostic Energy Management," *IEEE Access*, vol. 10, pp. 7460–7475, 2022, doi: 10.1109/ACCESS.2021.3138549.

[21] F. S. Al-Duais and R. S. Al-Sharpi, "A unique Markov chain Monte Carlo method for forecasting wind power utilizing time series model," *Alexandria Engineering Journal*, vol. 74, pp. 51–63, Jul. 2023, doi: 10.1016/j.aej.2023.05.019.

[22] A. Younesi, H. Shayeghi, A. Safari, and P. Siano, "Assessing the resilience of multi microgrid based widespread power systems against natural disasters using Monte Carlo Simulation," *Energy*, vol. 207, Sep. 2020, doi: 10.1016/j.energy.2020.118220.