E-ISSN 2224-2333
P-ISSN 2222-9930

Pakistan Journal of Engineering Technology and Science (PJETS)
Volume 12, Issue.01, PP. 117-124, June 2024
https://doi.org/10.22555/pjets.v12i1.1065

A Systematic Empirical Evaluation of Machine Learning Algorithm on Energy Prediction

Ibtehaj Ul Haq [1], Waqar Khan [2], Yusra Saeed [2*], Syed Sajjad Hussain Rizvi [2*], Shamim Akhtar [3], Pardeep Kumar [4]

[1] Commercial Bank International, UAE,
[2] SZABIST University, Karachi,
[3] Majmaah University Saudi Arabia
[4] QUEST Nawabshah
**Corresponding Author:* yusra.saeed@szabist.edu.pk

Abstract:

The energy crisis has alerted all scientists and researchers in every field. The evident increase in the usage of electronic items causes high electricity consumption. Especially in residential households and high-rise apartments, buildings show high electrical loads. This study is designed to forecast the energy need using AI prediction models and find the most efficient model to predict the future loads of any household considering its surroundings, like weather, air pressure, room temperature and others. This is a comprehensive comparative study that provides a comprehensive comparison between regression models. The finding of the study is that by using the optimum model, we can achieve the best results in predicting electricity load.

Keywords: *Energy Prediction, ML Algorithms, Energy Optimization, Performance Compression of ML Algorithm*

I. INTRODUCTION

Energy consumption has emerged as the biggest challenge at present. High-rise buildings and increasing dependency on electronic gadgets have plunged electricity utilization [1]. In a study on residential buildings held in the UK, [2] home appliances on standby were causing a 10.2% rise in overall consumption. Prediction models are used to find and predict abnormal electricity consumption and predict energy consumption for load management. The increase in world energy consumption has already caused energy exhaustion, and the impacts on the global environment are unprecedented.

International Energy Agency pointed to horrifying environmental effects demonstrated through data [3]. Emission of greenhouse gases has plunged in the last decade, as proven by scientific research; energy consumption has played an evident role in such an increase [4]. Building designs were created via 3D modeling over the past few decades, which led to building information models (BIM) [5]. After the involvement of technology in the construction arena, researchers were keen to minimize the energy use in buildings, for which they have used multiple strategies experimenting on building structures and materials [6]. The number of electrical gadgets and their use by the consumer are the main factors in electricity consumption. Consumers' use of appliances leaves recordable signals in the domestic environment [7]. Other factors like humidity, noise, and natural light also affect readings [8] [9].

II. LITERATURE REVIEW

This section refers to past research in the field, highlights the foundations of the study in this field, and gets through the advanced research through the data and different methods.

A. Modeling and Appliance Consumption

Data were collected in the Pacific Northwest from 454 domestic units and 140 commercial units [10]. The data description is referred to from [11]. Collected information demonstrates, and it is also stated by the researcher that appliances like refrigerators show flat readings a whole day, but other appliances like juicers and dishwasher shows a rise in the evening [8]. Markov Chain Monte Carlo technique was employed to devise the model to generate synthetic occupancy data to generate synthetic data and is suggested in [12] to evaluate lighting, heating, and energy consumption of appliances. A benchmark model, Energy Plus, to evaluate a building's energy consumption was represented in [13]. DOE's Building Technologies Program (BTP), Pacific Northwest National Laboratory (PNNL), Lawrence Berkeley National Laboratory (NBNL), and National Renewable Energy Laboratory (NBEL) [14] jointly proposed the Energy-Plus model. In [15] suggested the building's estimated hourly energy consumption determined by a coefficient by monthly bills, which will relieve the building's dynamic simulation. A study [16] in 2014 profiled major domestic-use appliances like refrigerators and washing machines to estimate daily energy consumption by each of them [17]. We have discussed models that suggest various methods of building a structure to help reduce energy consumption and predict future electricity bills. Problems with energy use during the operational phase will be discussed in detail.

B. Household Energy Consumption Prediction

This part of the paper demonstrates research on energy forecasting to determine the parameters, models, and methods that are vital for load prediction. Forecasting, ANN (Artificial Neural Networks), multiple regression, SVM [18], time series approaches [19], and prediction models [20] are used to forecast the energy needs in most studies. The most work has considered the parameters like time of day, the outdoor temperature of the day, the month of the year, weekend & holiday occurrences, previous day consumption, rainfall index, global solar radiation measure, wind speed index, and occupancy as attributes [21] [22]. The impact of weather in Wales and England was studied as a research variable to monitor monthly electricity demand [23].

The variation in demand from 91 to 95% was explained by studying heating and cooling degrees during days [24]. In recent years, electrical energy has been an active study in the use of patterns in buildings [25]. The temporal daily variation in the distribution of each house was found. Another study examined electricity data from 1628 houses and apartments [26]. A list of the variables was extracted from this data and studied by researchers: weather, location (ZIP code), the age of the building, Natural light input, floor area, number of residents, and salary levels. Scholars found that the most important were location, floor area, and weather [27]. Refrigerators and Entertainment appliances are also important variables in determining daily minimum energy usage [28] [29].

The minimum efficiency of appliances is recorded when the house is occupied often. Individual item consumption prediction system is discussed in [30]. Data such as past consumption, season and month, day, and hour are used in the model. Data for the last 24 hours is suggested to be most crucial for efficient prediction [31]. In Ottawa, Canada, 23 houses were monitored at a one-minute resolution for their Air Conditions, furnace fans, and other major electric appliances [32]. Another study in the UK, in which the use and ownership of appliances were studied, was an odds ratio analysis of 183 dwellings [33]. The author observed that high energy demand was observed in households with more than 30 electronic appliances. Data was collected through the survey, not from any measurement, which is the main limitation of the study. This paper explores the impact of energy consumption in the household. High-rise buildings can cause higher energy wastage.

The study highlights several types of research to avoid energy wastage and usage patterns in the daily household. Linear Regression is a supervised prediction algorithm. It models the relationship between one or more dependent variables and independent variables [34]. By observing the set of values of explanatory variables, linear regression is called to be a fit prediction model. Even after developing a model, if any additional and extra variable values are added without merging it, then still it is a fitted model for prediction [35] [36]. LSA (least squares approach) is used and fit for prediction, which is used by the Linear Regression model [36].

III. DATASET

The dataset used in this research paper is known as the Appliances energy prediction dataset. This dataset contains the measurements of energy consumption in a household along with the various environmental and weather-related attributes. This dataset is divided into two different files: training and testing. Total of the 34 columns in the training and testing dataset.

Data was recorded in [9] ten ten-minute time stamps to achieve the desired accuracy. The following table shows detailed column descriptions to help understand the dataset. Appliance energy consumption and light energy consumption are two output variables. In the original dataset [9], these two columns were used separately. However, we merged both columns by summating their values to minimize the noise in the dataset, as light energy contained multiple zero values. Temperature is calculated in the Celsius unit and was measured in different parts of the house. Weather readings were also measured. Readings were taken from the nearest airport. The same place was used to get wind pressure in the area. The timestamp was also broken into multiple columns before applying an algorithm. The date-time column was divided into the day, month, hour, and minutes columns. Table 1 illustrates the attribute and the corresponding unit of a dataset.

Table 1: Attribute of the dataset

Col#	Attribute of dataset	Unit
1	Appliances energy consumption	Wh
2	Light energy consumption	Wh
3	T1, Temperature in the kitchen area	◦C
4	RH1, Humidity in the kitchen area	%
5	T2, Temperature in the living room area	◦C
6	RH2, Humidity in the living room area	%
7	T3, Temperature in the laundry room area	◦C
8	RH3, Humidity in the laundry room area	%
9	T4, Temperature in the office room	◦C
10	RH4, Humidity in the office room	%
11	T5, Temperature in bathroom	◦C
12	RH5, Humidity in bathroom	%
13	T6, Temperature outside the building	◦C
14	RH6, Humidity outside the building	%
15	T7, Temperature in the ironing room	◦C
16	RH7, Humidity in the ironing room	%
17	T8, Temperature in teenager room	◦C
18	RH8, Humidity in teenager room	%
19	T9, Temperature in parents' room	◦C
20	RH9, Humidity in parents' room	%
21	To, Temperature outside	◦C
22	Pressure (from Chièvres weather station)	mm Hg
23	Rho, Humidity outside	%
24	Windspeed	m/s
25	Visibility (from Chièvres weather station)	km
26	Tdewpoint	◦C
27	Random Variable 1 (RV 1)	Nil
28	Random Variable 2 (RV 2)	Nil
29	Number of seconds from midnight (NSM)	s
30	Week status (weekend (1) or a weekday	nil
31	Day of Week	nil

IV. RESULTS AND ANALYSIS

The study aims to examine data acquired from domestic units to predict energy consumption and find the most accurate predicting algorithm and model in the given circumstances. All the models were trained using the MATLAB tool. Models were executed in the Matlab tool one by one, and the following readings were observed. Calculated RMSE, MSE, ASE, and prediction time in Matlab. Fine Tree, Medium Tree, Coarse Tree, and Bagged Tree ranked in the top, respectively. By observing root mean square error, mean square error, and prediction time, the fine tree can be declared a winner in overall comparison.

The fine tree provided the most efficient prediction results regarding the energy prediction in the household. Table 2 presents the comparative results of different prediction models, including metrics such as Root Mean Square Error

(RMSE), R-squared, Mean Squared Error (MSE), Mean Absolute Error (MAE), Prediction Speed, Training Time, and Average Testing Accuracy. The detailed table to analyze the result of all models is as follows. The following table will help decide the winner of all models.

Table 2: Comparative results of different prediction models

	Root Mean Square Error (RMSE)	R-Squared	Mean Squared Error (MSE)	Mean Abso-lute Error (MAE)	Prediction Speed (obs/sec) x10K	Training Time (sec)	Average Testing Ac-curacy (%)
Linear Regression	6.90E-16	1	4.76E-27	5.66E-14	35	16.1	52.8386
Interactions Linear Reg	8.36E-14	1	6.99E-27	7.02E-14	7.4	6.4551	52.2506
Robust Linear	4.10E-14	1	1.68E-27	3.48E-14	97	2.5037	52.575
Fine Tree	0.11217	0.74	0.00216	0.034539	4.4	7.9593	97.2019
Medium Tree	0.11217	1	0.012581	0.09698	130	2.7893	89.6796
Coarse Tree	0.11217	1	0.012581	0.09698	170	2.3276	76.9465
Linear SVM	1.0796	0.99	1.1655	0.93509	14	35.682	50.7097
Quadratic SVM	1.0889	0.99	1.1857	0.9432	21	17.416	52.575
Fine Gaussian SVM	1.4612	0.99	2.1352	1.2873	0.35	40.989	57.9278
Medium Gaussian	0.88112	1	0.77637	0.70256	3.2	32.62	55.1095
Coarse Gaussian	0.51022	1	0.23033	0.35369	17	15.056	51.4193
Boosted Tree	1.2404	0.99	1.5386	1.0687	7.5	73.239	58.0495
Bagged Tree	0.53114	1	0.28211	0.37494	3.9	62.572	74.2092

As per the above table, Linear Regression shows better results as it's RMSE is the least. But tree algorithms like a Medium tree and Coarse tree prediction show better results than all others and their RMSE is also very low.

A. Testing Graphs

In this section, the testing results of the machine-learning algorithm are presented on the benchmark dataset, Fig 1 to Fig 13. In the test graph, the blue line indicates the actual/true value, and the red line depicts the predicted value. It can be inferred from the graph that the predicted value tracks the true value of the data. Specifically, the Fine Tree and Medium Tree are scientifically outperforming in terms of testing accuracy, i.e., 97% and 90%, respectively. Given that, most of the other variants are submitting a test accuracy of marginally higher than 55%.

This significant difference strongly advocates for the Fine Tree and Medium Tree. Moreover, the same rationale is verified from the empirical result in Table 2. Figure 1 shows the testing graph fine tree, Figure 2 shows testing graph medium tree, Figure 3 shows Testing graph Coarse tree, Figure 4 shows Testing graph Bagged tree, Figure 5 shows Testing graph Boosted tree, Figure 6 shows Testing graph Coarse Gauss SVM, Figure 7 shows Testing graph Medium Gauss SVM, Figure 8 shows Testing graph Fine Gauss SVM, Figure 9 shows Testing graph Quadratic SVM, Figure 10 shows Testing graph Linear SVM, Figure 11 shows Testing graph interaction linear regression, Figure 12 shows Testing graph Robust Linear regression and Figure 13 shows Testing graph Linear regression.

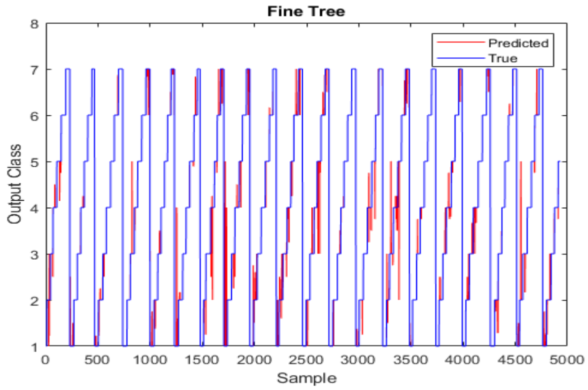

Figure 1: Testing Graph Fine Tree

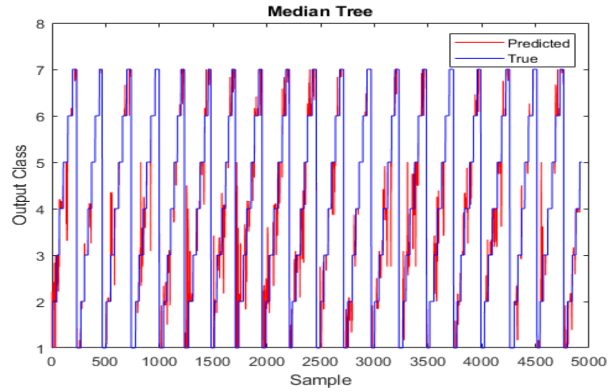

Figure 2: Testing Graph Medium Tree

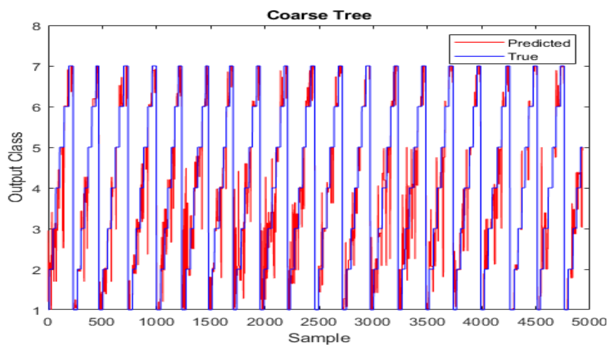

Figure 3: Testing Graph Coarse Tree

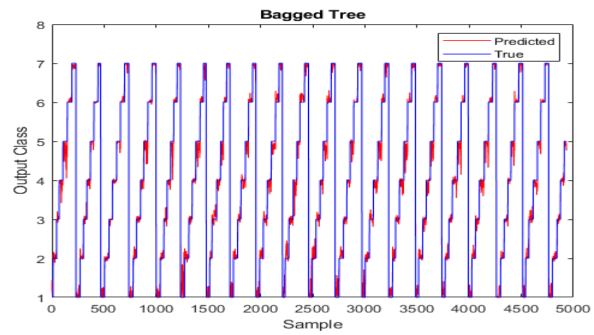

Figure 4: Testing Graph Bagged Tree

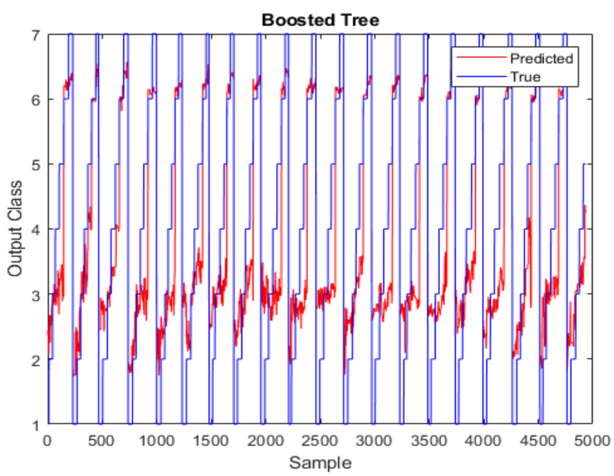

Figure 5: Testing Graph Boosted Tree

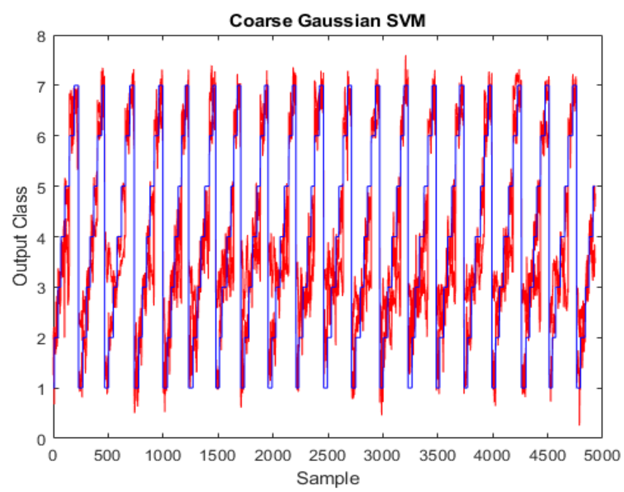

Figure 6: Testing Graph Coarse Gauss SVM

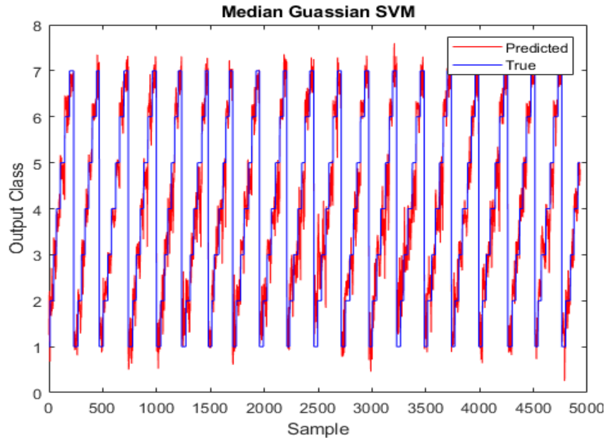

Figure 7: Testing Graph Medium Gauss SVM

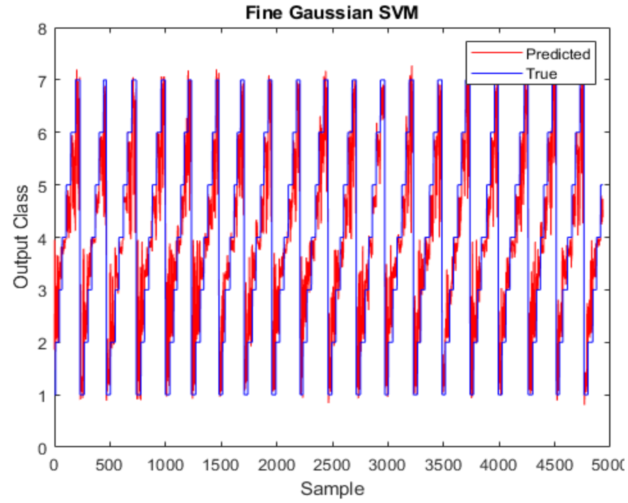

Figure 8: Testing Graph Fine Gauss SVM

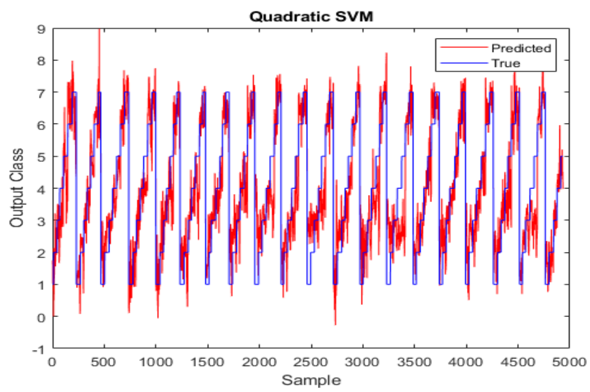

Figure 9: Testing Graph Quadratic SVM

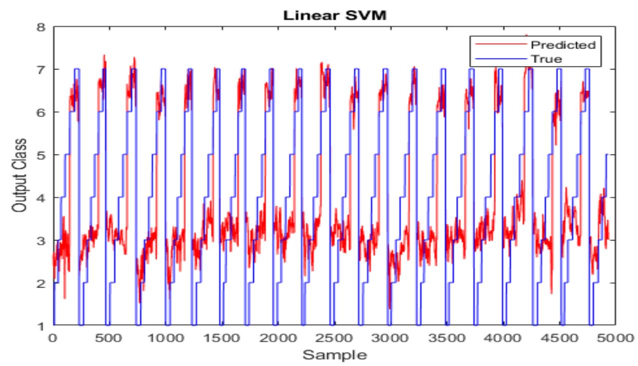

Figure 10: Testing Graph Linear SVM

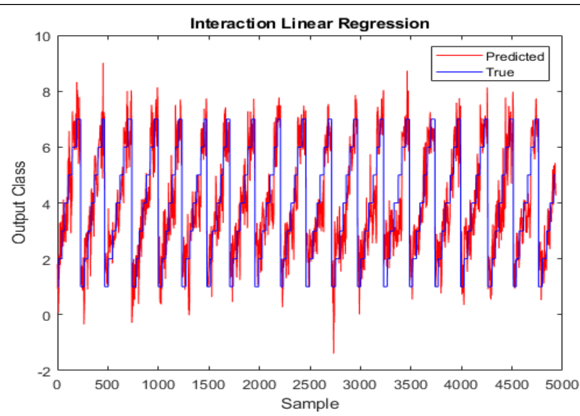

Figure 11: Testing Graph Interaction Linear Regression

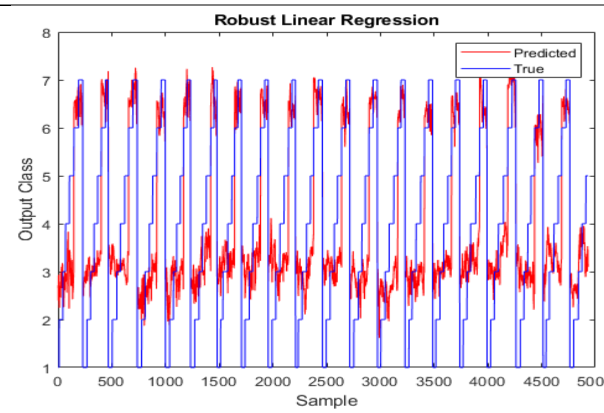

Figure 12: Testing graph Robust Linear regression

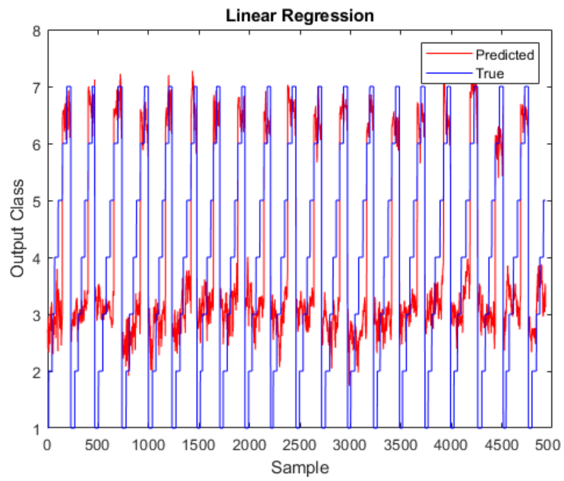

Figure 13: Testing graph Linear regression

V. CONCLUSION

The study demonstrates the effectiveness of tree-based machine learning models in predicting household energy consumption. In this study, we have tested several regression models and used Matlab 2018 for this purpose. We used data collected from multiple households in the UK in 2016 collected through the months on a daily basis. The study found that the fine tree algorithm is best for the purpose. It can predict a 98% accurate electricity load under the given circumstances and environment. The correct electricity load prediction can save a valuable amount of energy and reduce the energy wastage caused by electronics appliances in houses. These findings highlight the potential of machine learning algorithms in enhancing energy management practices, ultimately contributing to more sustainable and efficient energy usage in residential buildings. Future research should build on these findings to develop more robust and generalizable energy prediction models. Future studies should consider expanding the dataset to include data from different regions and time periods to improve the generalizability of the findings. Also, investigating the use of advanced machine learning models, such as deep learning and ensemble methods, to further improve prediction performance. In addition to this, additional features, such as occupancy patterns and appliance-specific energy consumption, should be explored to enhance the predictive power of the models. Moreover, developing real-time energy prediction systems that can provide immediate feedback and recommendations for energy optimization can also be considered for future work areas.

REFERENCES

[1] Rahman, Abidur, Omar Farrok, and Md Mejbaul Haque. "Environmental impact of renewable energy source based electrical power plants: Solar, wind, hydroelectric, biomass, geothermal, tidal, ocean, and osmotic." Renewable and Sustainable Energy Reviews 161 (2022): 112279.

[2] McKenna, Eoghan, et al. "Explaining daily energy demand in British housing using linked smart meter and socio-technical data in a bottom-up statistical model." Energy and Buildings 258 (2022): 111845.

[3] Zheng, Jiajia, Yongjie Dang, and Ullah Assad. "Household energy consumption, energy efficiency, and household income–Evidence from China." Applied Energy 353 (2024): 122074..

[4] Omrany, Hossein, et al. "The uptake of City Information Modelling (CIM): A comprehensive review of current implementations, challenges and future outlook." Smart and Sustainable Built Environment 12.5 (2023): 1090-1116.

[5] Omrany, Hossein, et al. "The uptake of City Information Modelling (CIM): A comprehensive review of current implementations, challenges and future outlook." Smart and Sustainable Built Environment 12.5 (2023): 1090-1116.

[6] Xu, Hong, et al. "Prediction of thermal energy inside smart homes using IoT and classifier ensemble techniques." Computer Communications 151 (2020): 581-589.

[7] Das, Sujit, and Elizabeth Mao. "The global energy footprint of information and communication technology electronics in connected Internet-of-Things devices." Sustainable Energy, Grids and Networks 24 (2020): 100408.

[8] Wen, Lulu, et al. "Demand side management in smart grid: A dynamic-price-based demand response model." IEEE Transactions on Engineering Management (2022).

[9] Kamel, Ehsan, Shaya Sheikh, and Xueqing Huang. "Data-driven predictive models for residential building energy use based on the segregation of heating and cooling days." Energy 206 (2020): 118045.

[10] McPherson-Hathaway, Braydi. "Comparing BEopt (EnergyPlus) Energy Predictions to Measured Circuit Level Energy Consumption of 12 Similar Small Energy-Efficient Single-Family Residences." (2021).

[11] Bharati, Subrato, Prajoy Podder, and M. Mondal. "Visualization and prediction of energy consumption in smart homes." International Journal of Hybrid Intelligent Systems 16.2 (2020): 81-97.

[12] Buttitta, Giuseppina, and Donal P. Finn. "A high-temporal resolution residential building occupancy model to generate high-temporal resolution heating load profiles of occupancy-integrated archetypes." Energy and Buildings 206 (2020): 109577.

[13] Arjunan, Pandarasamy, Kameshwar Poolla, and Clayton Miller. "EnergyStar++: Towards more accurate and explanatory building energy benchmarking." Applied Energy 276 (2020): 115413.

[14] Im, Piljae, et al. "Empirical validation of building energy modeling for multi-zones commercial buildings in the cooling season." Applied Energy 261 (2020): 114374.

[15] Satre-Meloy, Aven, Marina Diakonova, and Philipp Grünewald. "Daily life and demand: an analysis of intra-day variations in residential electricity consumption with time-use data." Energy Efficiency 13 (2020): 433-458.

[16] Zhang, Minhui, and Qin Zhang. "Grid parity analysis of distributed photovoltaic power generation in China." Energy 206 (2020): 118165.

[17] Lu, Chujie. "Enhancing real-time nonintrusive occupancy estimation in buildings via knowledge fusion network." Energy and Buildings 303 (2024): 113812.

[18] Chen, Siliang, et al. "Lifelong learning with deep conditional generative replay for dynamic and adaptive modeling towards net zero emissions target in building energy system." Applied Energy 353 (2024): 122189.

[19] Bibri, Simon Elias, et al. "Smarter eco-cities and their leading-edge artificial intelligence of things solutions for environmental sustainability: A comprehensive systematic review." Environmental Science and Ecotechnology 19 (2024): 100330.

[20] Gopinath, Rajendiran, et al. "Energy management using non-intrusive load monitoring techniques–State-of-the-art and future research directions." Sustainable Cities and Society 62 (2020): 102411.

[21] Wang, Zhe, et al. "Predicting city-scale daily electricity consumption using data-driven models." Advances in Applied Energy 2 (2021): 100025.

[22] Kristensen, Martin Heine, Rasmus Elbæk Hedegaard, and Steffen Petersen. "Long-term forecasting of hourly district heating loads in urban areas using hierarchical archetype modeling." Energy 201 (2020): 117687.

[23] Son, Hyojoo, and Changwan Kim. "A deep learning approach to forecasting monthly demand for residential–-sector electricity." Sustainability 12.8 (2020): 3103.

[24] Islam, Abu Reza Md Towfiqul, Itmam Ahmed, and Md Siddiqur Rahman. "Trends in cooling and heating degree-days overtimes in Bangladesh? An investigation of the possible causes of changes." Natural Hazards 101.3 (2020): 879-909.

[25] Kewo, Angreine, Pinrolinvic DK Manembu, and Per Sieverts Nielsen. "A Rigorous Standalone Literature Review of Residential Electricity Load Profiles." Energies 16.10 (2023): 4072.

[26] Solà, María del Mar, et al. "Promoting energy efficiency at household level: a literature review." Energy Efficiency 14 (2021): 1-22. [27] M. A. a. D. J. S. Hart, "Quantifying the influence of land-use and surface characteristics on spatial variability in the urban heat island.".

[27] Trotta, Gianluca. "An empirical analysis of domestic electricity load profiles: Who consumes how much and when?." Applied Energy 275 (2020): 115399.

[28] Lee, Sangyoon, and Dae-Hyun Choi. "Energy management of smart home with home appliances, energy storage system and electric vehicle: A hierarchical deep reinforcement learning approach." Sensors 20.7 (2020): 2157.

[29] Jin, Yuan, et al. "Appliance use behavior modelling and evaluation in residential buildings: A case study of television energy use." Building Simulation. Vol. 13. Tsinghua University Press, 2020.

[30] Hsieh, Sun-Yuan, et al. "Utilization-prediction-aware virtual machine consolidation approach for energy-efficient cloud data centers." Journal of Parallel and Distributed Computing 139 (2020): 99-109.

[31] Proedrou, Elisavet. "A comprehensive review of residential electricity load profile models." IEEE Access 9 (2021): 12114-12133.

[32] Ma, Ping, et al. "Review of family-level short-term load forecasting and its application in household energy management system." Energies 16.15 (2023): 5809.

[33] Kumar, Vijendra, et al. "The state of the art in deep learning applications, challenges, and future prospects: A comprehensive review of flood forecasting and management." Sustainability 15.13 (2023): 10543.

[34] Harrell, J. F. E. "Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis. Cham: Springer International Publishing; 2015." Imprint: Springer (2023).

[35] Li, Sai, et al. "Estimation and inference for high-dimensional generalized linear models with knowledge transfer." Journal of the American Statistical Association (2023): 1-12.

[36] Tang, Wan, Hua He, and Xin M. Tu. Applied categorical and count data analysis. CRC Press, 2023.