Analyzing Domestic Energy Behavior with a Multi-Dimensional Appliance-Level Dataset

Abdullah Alsalemi[a], Abbes Amira[b,a], Hossein Malekmohamadi[a], Kegong Diao[c]*

[a] Institute of Artificial Intelligence, De Montfort University, Leicester, UK
[b] Department of Computer Science, University of Sharjah, Sharjah, UAE
[c] Institute of Energy and Sustainable Development, De Montfort University, Leicester, UK
**Correspondence email: P2621877@my365.dmu.ac.uk*

Abstract: Data, in its purest nature, has an authority on the systems it accompanies by feeding an accurate representation of the observed reality. In energy efficiency, the underlying motivation for big data efforts revolves around the intrinsic need to understand end-user electric energy consumption and means to improve it. Hence, developing a rich, detailed, and realistic power consumption dataset entails a deliberate process of preparing the data collection environment, configuring proper Internet of Energy (IoE) sensors and managing the collected data. In this work, a novel power consumption dataset is presented in efforts to improve the state-of-the-art of energy efficiency research in buildings. The dataset is also accompanied by a two-dimensional (2D) counterpart produced using Gramian Angular Fields (GAF) that creates pictorial summaries from one-dimensional (1D) data. Data acquisition is carried out using the ODROID-XU4 edge computing hub, Home Assistant software, and a collection of smart plugs and sensors. A notable use case is presented to signify the merits of the data and its analysis tools to achieve computationally efficient classification.
Keywords: Appliance-level dataset, energy efficiency, artificial intelligence, data lake, internet of energy.

1. Introduction

The underlying motivation for big data efforts in research efficiency research revolves around the intrinsic motivation of understanding the rationale behind end-user electric energy consumption and means to improve it [1], [2]. Appearing as a simple question, the quest for finding the right algorithm, system, framework, and solution has been the overwhelming pursuit of many scholars in the field of engineering [3], computing [4], and sciences [5]. In recent years, a number of datasets have been published, with varying structure, size, and purpose [3], [6]–[8]. Such contributions have led the development of classification algorithms [9], anomaly detectors [10], appliance identification system [11], as well as recommenders to improve energy efficiency behavior in buildings [12].

In the same path, developing a rich, detailed, and realistic power consumption dataset entails a deliberate process of preparing the data collection environment, configuring proper Internet of Things (IoT) sensors and communication modules, as well as managing the collected data in a secure, privacy-preserving manner using edge computing technology. Not only that, but creating a bridge where data can be pre-processed, post-processed, and classified using a variety of programs and algorithms in real-time or near real-time manger is an important challenge to address in order to optimize the impact of the produced dataset.

Enriching the data with supporting parameters can aid in improving the accuracy of the perception of reality and therefore lead to better outcomes. Examples includes supplementing the database with ambient environmental conditions such as temperature, humidity, and occupancy. Such multi-parameter representation of energy consumption brings the attention to the concept of micro-moments [13], time-based contextual encapsulations of energy behaviors throughout the duration of the dataset. Energy micro-moments (EMMs) captures the normal power consumption, switch on/off actions, and contextual events such as operating an appliance while outside the environment and excessive consumption.

On the other hand, the research and computing community have debated the most suitable manner of storing, accessing, and processing the collected data to optimize for computational efficiency. In the grand scheme, the fusion of the aforementioned concepts and techniques has brought upon a new Internet of Energy (IoE), where IoT interests with energy research to produce more favorable outcomes. Notwithstanding the great benefits of accumulating large datasets, an increased burden on the processing and Machine Learning (ML) side results in slow and sluggish data management that consumes excessive time and computational resources. When one cannot compromise on data fidelity by reducing the number of data points, parameters, or frequency, innovating in data post-processing efficiency becomes only prudent.

This brings the discussion to the notion of dimensional transformation. The authors delve into the Gramian Angular Fields (GAF) transforms that converts time-series data into polar-coordinates based pictorial alternatives [14], [15]. GAFs summarize a large number of one-dimensional (1D) timestamped datapoints into brief, color coded two-dimensional (2D) images that can speed up data modeling training and also reveal novel patterns not shown in the original format.

Therefore, in this work, a novel power consumption dataset that is presented in efforts to support the research community's efforts to improve the state-of-the-art of energy efficiency research in buildings. The dataset is collected at a domestic household from number of common appliances and takes advantage ambient condition sensing data to bring more color and fidelity to each data reading. Moreover, the paper discusses the data collection method, post-processing, and an impact use case on how the data can be exploited to classify energy consumption patterns. Hence, the main contributions of this paper are:

1. Presenting a novel, appliance level dataset for energy efficiency research that combines power consumption readings with ambient environmental conditions such as temperature, humidity, and occupancy;
2. Developing a post-processing workflow that enhances computational efficiency by utilizing GAF transformations; and
3. Showcasing an impact use case of the collected dataset for energy data classification.

The remainder of this paper is organized as follows. Section 2 describes the methods used to collect the data as well as GAF post-processing techniques. Section 3 reports on the collected dataset results and discussed the benefits and limitations of using GAFs in ML-based classification. Section 4 concludes the work.

2. Methods

In this section, the workflow of data collection is succinctly presented and divided into (a) the data hub setup, (b) smart plugs and sensors configuration, (c) and post-processing techniques. Fig. 1 shows the main data collection setup components.

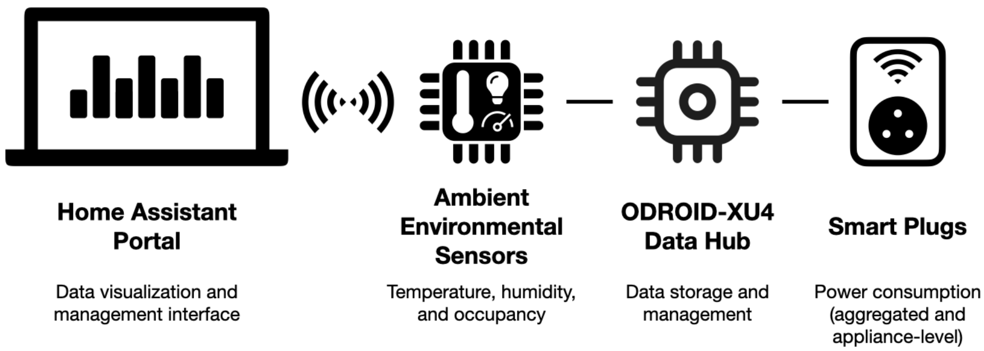

Fig. 1. Overview of the Data Collection Setup.

2.1. ODROID-XU4 Data Hub

To start, the data is collected at a small-size home environment. Before installing sensors and meters, a hub device is required as a data storage and management location. In this work, we have employed the ODROID-XU4 high-performance edge platform due to its cost-efficiency, high-performance in data management and processing operations, as well as its computability with the needed software. The ODROID-XU4 is connected the household local network using an Ethernet cable to ensure connection stability. On the board, Home Assistant, an open-source smart home management software, is installed and configured to store all the data is a lightweight Structured Query Language (SQLite) format.

2.2. Smart Plug and Sensors Configuration

Following the data hub setup, sensors are installed and configured in different locations of the household. Firstly, smart plugs are attached to common appliances including a kettle, TV, toaster, computer, fridge, etc. Table 1 lists the connected appliances. Each smart plus is calibrated separately with a power meter to mitigate any biases in readings. Also, ambient condition sensors, namely temperature & humidity and occupancy sensors are placed in strategic locations to capture contextual information that helps understand energy consumption behavior further. Table 2 lists the sensors used and their corresponding locations. Similarly, the temperature and humidity sensor are calibrated against reference meters to ensure accuracy. The smart plugs and sensors are installed in the living room, kitchen, and study room.

Table 1. Appliances Used, Their Locations, and Connected Smart Plugs.

Connected Appliances	Location	Sensor
Kettle	Kitchen	LocalBytes Power Monitoring Smart Plug
Toaster		
Fridge		
Washing Machine		
TV	Living Room	
Computer Setup 1 (Laptop + LCD Monitor)	Study Room	
Computer Setup 2 (Laptop + Docking Station + LCD Monitor)		

Table 2. Ambient Environmental Conditions Sensors and Their Corresponding Locations.

Location	Sensors
Kitchen	• SONOFF SNZB-03 Motion Sensor
Living Room	• SONOFF SNZB-02 Temperature & Humidity Sensor • SONOFF SNZB-03 Motion Sensor
Study Room	• SONOFF SNZB-02 Temperature & Humidity Sensor

Once all smart plugs and sensors are installed and configured, they are tested and validated using various tests for a few days. The tests include wireless connectivity tests, occupancy verification tests, and data integrity tests. Once confirmed, the data collection officially starts, where data acquisition frequency averages between 1-3 sec. This allows the data to grow very large very quickly. In the span of three months, the data has reached more than 4 million datapoints.

2.3. Post-Processing

As continuous readings are accumulated into Home Assistant's SQLite database file, the data grows quite large, exceeding more than 2.5 GB. In this case, conventional data import methods may fail on computing systems with average virtual memory. Hitherto in this work, we outline an alternative technique to data import and processing. Firstly, the dataset is split into smaller manageable fragments, and then each fragment is imported separately. Secondly, the fragments are combined in one Python Pandas DataFrame object, which can handle large-size data. Thirdly, the data is cleaned and restructured to only have the necessary columns for further post-processing (i.e., timestamp, appliance name, power consumption, temperature, humidity, and occupancy). Fourthly, the data is exported as a CSV file, after which the file size of the database is reduced to several megabytes.

As described earlier, 1D cartesian-based data need to be transformed into the polar coordinates system to run the GAF transform [16], [17] and thus, the Gramian Angular Summation Field (GASF) using the cosine function, as in Eq. (2) [18]:

$$GASF = \begin{pmatrix} \cos(\emptyset_1 + \emptyset_1) & \cdots & \cos(\emptyset_1 + \emptyset_N) \\ \vdots & \ddots & \vdots \\ \cos(\emptyset_N + \emptyset_1) & \cdots & \cos(\emptyset_N + \emptyset_N) \end{pmatrix} \quad (2)$$

where $\emptyset$ the is derived in Eq. (3):

$$\{\emptyset = \arccos(\tilde{x}_l), \qquad 0 < \tilde{x}_l < 1 \in x \quad (3)$$

Hence, the CSV file is fed into the GAF generator created in [19] to convert the 1D time-series data into heat map-like 2D images. GAF images are generated for each appliance in 1-hour segments as well as each ambient environmental parameter (e.g., temperature). The end result compromises a compact collection of GAF images that summarize energy power consumption for a multi-million, multi-month, and multi-appliance dataset. In the following section, we will describe the advantages of such concise representations in terms of ML execution speed as well as the potential wealth of insight that can be impacted from the use 2D Deep Learning (DL) algorithms.

3. Results & Discussion

In this section, the data collection results are reported in terms of specifications in addition to the discussion of a potential use case of data in terms of DL-based data classification.

3.1. 1D Home Assistant Data

As shown in Fig. 2, the Home Assistant portal is shown. It depicts real-time readings of all the smart plugs and sensors in the household, with the ability to view usage details for each entry for the last 24 hours when clicked on. It also worth depicting a sample of the readings collected from the household's environmental sensors and appliances in Fig. 3 and Fig. 4 respectively.

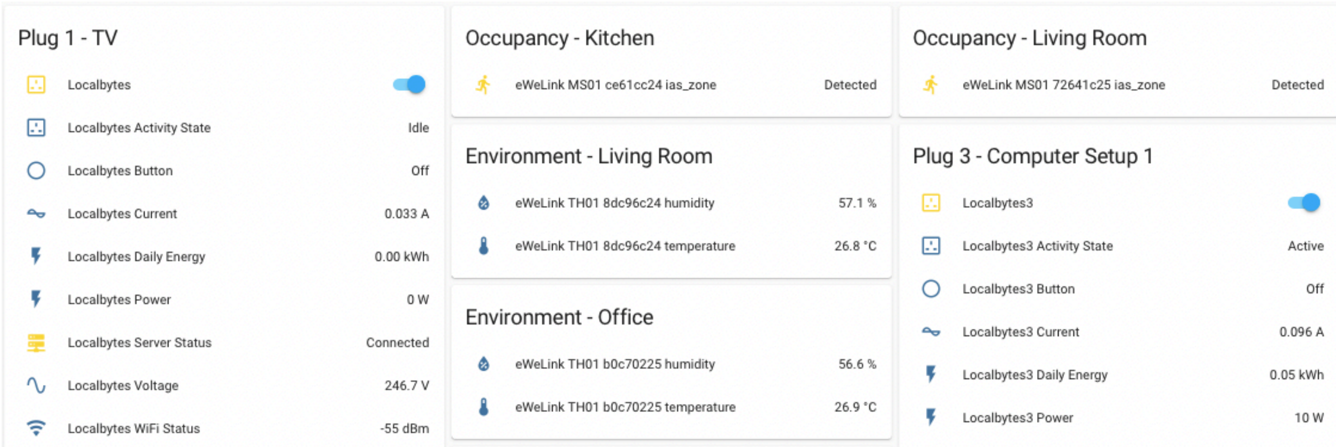

Fig. 2. The Home Assistant Dashboard.

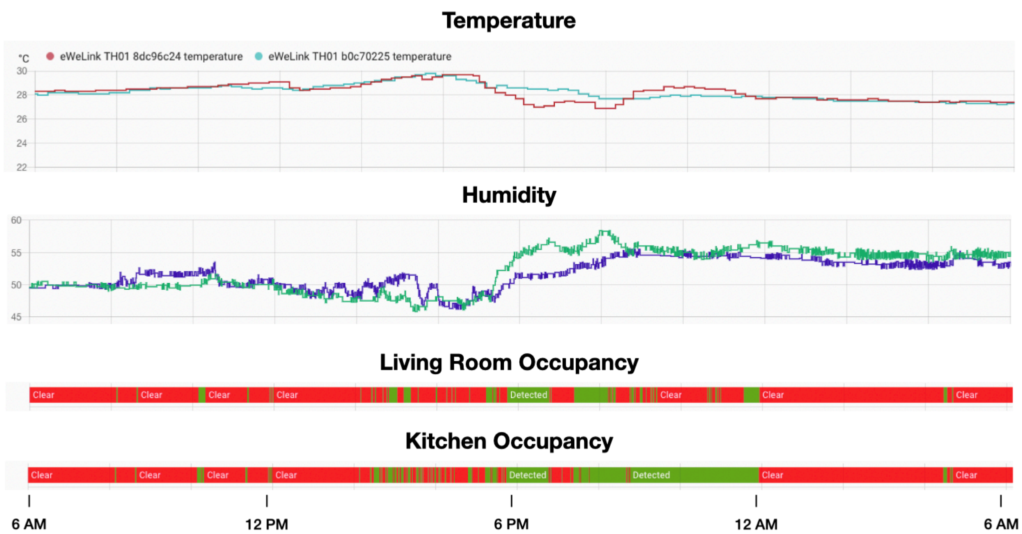

Fig. 3. Ambient Environmental Data Samples.

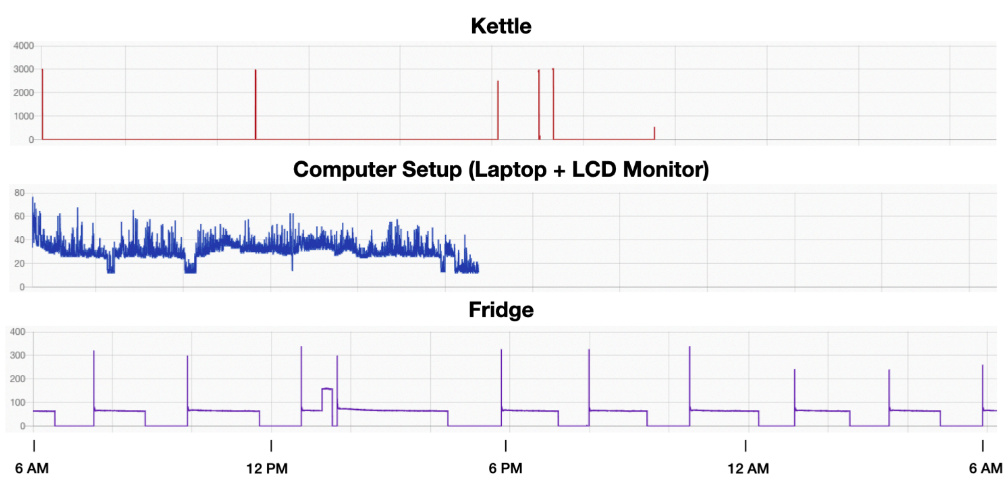

Fig. 4. Appliance-Level Power Consumption (W) Data Samples.

As mentioned, the smart plugs are compact enough to be conveniently used on a daily basis without compromising on the lifestyle of the end-user. However, the smart plugs themselves consume an average of 1.2 W, so when multiplied with the total number of connected appliances, not to mention the power consumption of the ODROID-XU4, they incur a subtle, yet noticeable addition to the overall yearly consumption. Therefore, it is imperative to use the smart plugs deftly to identify areas of improvements in daily-to-day consumption habits, apply them to reduce improve overall efficiency, and then consider removing redundant smart plug setups and use smart meter readings. On the other hand, smaller sensors use 3V CR2450 batteries that needs replacement every few months, and hence their reoccurring costs should be considered in the calculation of energy savings.

3.2. 2D GAF Data

As described earlier, the post-processing workflow in this work involves converting 1D data into 2D images using GAF transformation in order to open new opportunities for data classification and obtaining novel insights that help improve energy efficiency behavior in buildings.

To illustrate an example of the merits of 2D GAF data, we will discuss the case of power consumption classification. Because of the 2D nature of the data, a plethora of DL methods can be employed, e.g., Convolutional Neural Networks (CNN), Long-Short Term Memory (LSTM), etc. to classify the data and gain useful insights with significantly lower computational time [20].

As shown in Fig. 5, sample GAFs are classified based on a modified EfficientNet-B0 classifier (CNN with transfer learning). The algorithm, explained in detail in [20], classifies the data based on average normalized values of the GAF data, i.e., a GAF is normal is its average normal value within its duration window is less than a specified abnormality threshold, and vice versa. It is worth mentioning that the model is trained on a Linux instance while the testing has been carried out on an ODROID-XU4 board to signify the merits of 2D GAF classification in terms of classification computational performance, e.g. the GAF conversion can convert 70 million data points are converted into ~1,600 images with 648×648 resolution. Such advantage permits (a) enhanced privacy-preserving data processing (especially detailed power consumption) as the data collection and processed are carried on the edge (i.e., the ODROID-XU4) rather offloading to a 3rd party cloud service; and (b) real-time or near real-time useful end-user feedback due to the high computational efficiency of the whole workflow encompassing data acquisition, post-processing, and classification.

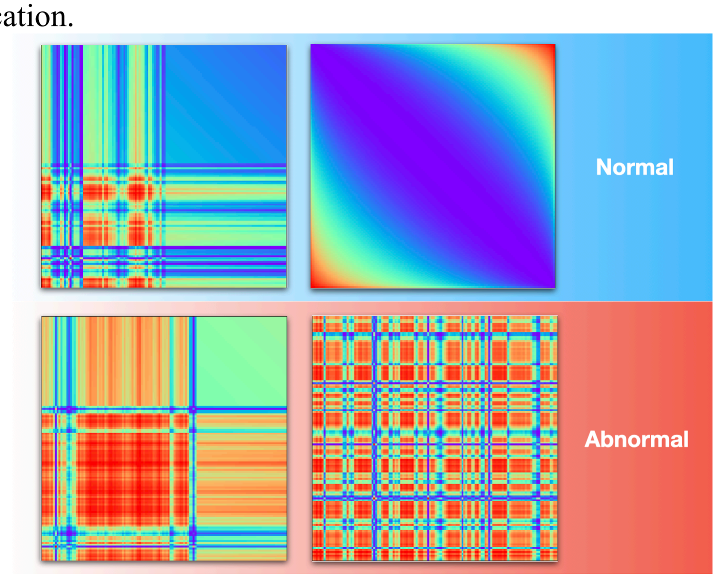

Fig. 5. Sample GAF Images Classified as Either Normal or Abnormal.

It is worth noting that the importance of the using GAFs to visualize and process large data shines when attempting to identify patterns specific to each appliance. Such findings can support appliance identification efforts by utilizing a visual 2D fingerprint for each appliance, e.g., as shown in the repeating pattern of the fridge consumption in Fig. 4. With that in mind, it is in the authors future work directions to explore and evaluate the notion of 1D-2D correlation further.

Despite the upshots of the proposed dataset and its 2D GAF classification use case, the work involves a number of limitations. Chiefly, the very large file size of the collected data (2.5+ GB generated from 5 appliances and 4 sensors in 3 months) can strongly impede long-term data acquisition for multiple years, given the limited local storage capacities of the ODROID-XU4. This should be addressed by deploying an SQLite data compression program on regular basis. Moreover, the accumulated extraneous power consumption of the connected smart plugs as well as the sensor battery replacements costs should be wisely considered in the overall cost-savings intended by the developed energy efficiency system. Lastly, despite GAFs powerful abilities to absorb substantial amounts of data in compact images, it does not fully replace the need to deliberately analyze the time-series representation in its full-scale.

4. Conclusions

In this paper, a novel multi-dimensional power consumption dataset is presented that in comprises multiple domestic appliances as well as micro-moment based ambient conditions. The data collection and workflow are carried out using the ODROID-XU4 edge computing hub, Home Assistant software, and a collection of smart plugs and sensors. Acquired data is converted into 2D GAF images for faster and richer classification. A sample use case signifies near real-time computational efficiency of binary GAF classification. In conclusion, this work aims to pave a path towards the use of multi-parameter power consumption datasets coupled with modern data analysis tools to improve the landscape of energy efficiency.

References

[1] A. Alsalemi, C. Sardianos, F. Bensaali, I. Varlamis, A. Amira, and G. Dimitrakopoulos, "The Role of Micro-Moments: A Survey of Habitual Behavior Change and Recommender Systems for Energy Saving," *IEEE Systems Journal*, vol. 13, no. 3, Art. no. 3, Sep. 2019.

[2] M. Bourdeau, X. qiang Zhai, E. Nefzaoui, X. Guo, and P. Chatellier, "Modeling and forecasting building energy consumption: A review of data-driven techniques," *Sustainable Cities and Society*, vol. 48, p. 101533, Jul. 2019, doi: 10.1016/j.scs.2019.101533.

[3] A. Alsalemi *et al.*, "Boosting Domestic Energy Efficiency Through Accurate Consumption Data Collection," in *2019 IEEE SmartWorld, Ubiquitous Intelligence Computing, Advanced Trusted Computing, Scalable Computing Communications, Cloud Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, Aug. 2019, pp. 1468–1472. doi: 10.1109/SmartWorld-UIC-ATC-SCALCOM-IOP-SCI.2019.00265.

[4] I. Varlamis *et al.*, "Using big data and federated learning for generating energy efficiency recommendations," *Int J Data Sci Anal*, May 2022, doi: 10.1007/s41060-022-00331-2.

[5] C. Boomsma, R. V. Jones, S. Pahl, and A. Fuertes, "Do psychological factors relate to energy saving behaviours in inefficient and damp homes? A study among English social housing residents," *Energy Research & Social Science*, vol. 47, pp. 146–155, Jan. 2019, doi: 10.1016/j.erss.2018.09.007.

[6] J. Kelly and W. Knottenbelt, "The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes," *Scientific Data*, vol. 2, p. 150007, Mar. 2015, doi: 10.1038/sdata.2015.7.

[7] K. Anderson, A. Ocneanu, D. R. Carlson, A. Rowe, and M. Bergés, "BLUED : A Fully Labeled Public Dataset for Event-Based Non-Intrusive Load Monitoring Research," Beijing, China, 2012.

[8] Y. Himeur, K. Ghanem, A. Alsalemi, F. Bensaali, and A. Amira, "Artificial intelligence based anomaly detection of energy consumption in buildings: A review, current trends and new perspectives," *Applied Energy*, vol. 287, p. 116601, Apr. 2021, doi: 10.1016/j.apenergy.2021.116601.

[9] C. Benavente-Peces and N. Ibadah, "Buildings Energy Efficiency Analysis and Classification Using Various Machine Learning Technique Classifiers," *Energies*, vol. 13, no. 13, Art. no. 13, Jan. 2020, doi: 10.3390/en13133497.

[10] Y. Himeur, A. Alsalemi, F. Bensaali, and A. Amira, "Smart power consumption abnormality detection in buildings using micromoments and improved K-nearest neighbors," *International Journal of Intelligent Systems*, vol. 36, no. 6, Art. no. 6, 2021, doi: 10.1002/int.22404.

[11] M. Chen and C. Lin, "Standby Power Management of a Smart Home Appliance by Using Energy Saving System With Active Loading Feature Identification," *IEEE Transactions on Consumer Electronics*, vol. 65, no. 1, Art. no. 1, Feb. 2019.

[12] C. Sardianos *et al.*, "The emergence of explainability of intelligent systems: Delivering explainable and personalized recommendations for energy efficiency," *International Journal of Intelligent Systems*, vol. 36, no. 2, Art. no. 2, 2021, doi: 10.1002/int.22314.

[13] A. Alsalemi *et al.*, "Achieving Domestic Energy Efficiency Using Micro-Moments and Intelligent Recommendations," *IEEE Access*, vol. 8, pp. 15047–15055, 2020, doi: 10.1109/ACCESS.2020.2966640.

[14] J. CHEN and X. WANG, "Non-intrusive Load Monitoring Using Gramian Angular Field Color Encoding in Edge Computing," *Chinese Journal of Electronics*, vol. 31, no. 4, Art. no. 4, 2022.

[15] A. Alsalemi, A. Amira, H. Malekmohamadi, K. Diao, and F. Bensaali, *Elevating Energy Data Analysis with M2GAF: Micro-Moment Driven Gramian Angular Field Visualizations*. International Conference on Applied Energy, 2021. Accessed: Dec. 27, 2021. [Online]. Available: https://dora.dmu.ac.uk/handle/2086/21303

[16] Y.-Y. Hong, J. J. F. Martinez, and A. C. Fajardo, "Day-Ahead Solar Irradiation Forecasting Utilizing Gramian Angular Field and Convolutional Long Short-Term Memory," *IEEE Access*, vol. 8, pp. 18741–18753, 2020, doi: 10.1109/ACCESS.2020.2967900.

[17] K. P. Thanaraj, B. Parvathavarthini, U. J. Tanik, V. Rajinikanth, S. Kadry, and K. Kamalanand, "Implementation of Deep Neural Networks to Classify EEG Signals using Gramian Angular Summation Field for Epilepsy Diagnosis," *arXiv:2003.04534 [cs, eess]*, Mar. 2020, Accessed: Sep. 22, 2021. [Online]. Available: http://arxiv.org/abs/2003.04534

[18] Z. Wang and T. Oates, "Imaging Time-Series to Improve Classification and Imputation," presented at the Twenty-Fourth International Joint Conference on Artificial Intelligence, Jun. 2015. Accessed: Sep. 22, 2021. [Online]. Available: https://www.aaai.org/ocs/index.php/IJCAI/IJCAI15/paper/view/11082

[19] A. Alsalemi, A. Amira, H. Malekmohamadi, K. Diao, and F. Bensaali, *Elevating Energy Data Analysis with M2GAF: Micro-Moment Driven Gramian Angular Field Visualizations*. International Conference on Applied Energy, 2021. Accessed: Dec. 27, 2021. [Online]. Available: https://dora.dmu.ac.uk/handle/2086/21303

[20] A. Alsalemi, A. Amira, H. Malekmohamadi, and K. Diao, "Lightweight Gramian Angular Field classification for edge internet of energy applications," *Cluster Comput*, Aug. 2022, doi: 10.1007/s10586-022-03704-1.