An Effective Non-Intrusive Load Monitoring (NILM) for Residential Appliances using Wavelet Transform and Clustering

Nobert Kyereboah Mensah
University of Mines and Technology, Tarkwa

Hamidu Abdel-Fatao
University of Mines and Technology, Tarkwa

Yao Yevenyo Ziggah
University of Mines and Technology, Tarkwa

Solomon Nunoo
University of Mines and Technology, Tarkwa

ABSTRACT
In recent years, energy consumption has gained attention due to the impact of human activities on the environment as well as the economic burden imposed by energy consumption on end users. Understanding how energy is consumed at the appliance level is crucial yet challenging with conventional aggregate energy data from utility companies. Non-Intrusive Load Monitoring (NILM) emerges as a solution, enabling the estimation of individual appliance energy usage from aggregate data. This study proposes an innovative approach to cluster residential appliances within NILM systems, aiming to identify appliance usage patterns and facilitate predictive modeling using only aggregate consumption data. Adopting wavelet transformation for data denoising, coupled with K-Means and Self-Organising Map (SOM) clustering algorithms, the methodology identifies appliances with similar consumption patterns. A comparative evaluation is performed to assess the performance of these clustering methods, considering aspects such as accuracy, computational time, and interpretability of results. To validate the proposed methodology, real-world data from the REFIT dataset, offering detailed information on the energy consumption of individual household appliances, is employed. The study not only aids in the selection of clustering algorithms for NILM but also provides insights into the strengths and limitations of each algorithm in the context of energy management within smart grid systems.

General Terms
Comparative Analysis, Algorithms, Energy Conservation

Keywords
Non-Intrusive Load Monitoring (NILM); Wavelet Transform, K-Means, Self-Organising Map (SOM), REFIT dataset.

1. INTRODUCTION
The increasing global demand for energy poses serious challenges for the environment, natural resources, and climate change. Energy conservation is crucial to reduce environmental pressures and cost [1][2]. In this context, domestic and residential energy usage assumes a critical role. With households accounting for a substantial amount of overall energy consumption, efforts to promote energy conservation at the residential level are important. The global usage of energy is increasing at an alarming rate, leading to a strain on natural resources and significant contributions to climate change. To mitigate these effects, it is essential that energy conservation is prioritised. By reducing energy consumption, its negative impact on the environment and cost implications can be reduced. The importance of energy disaggregation has grown in recent years for residential consumers, power companies, and appliance manufacturers, however, many residential consumers have a limited understanding of their energy usage or the consumption of specific appliances [3].

One of the key challenges in energy conservation is understanding how energy is being used. This is particularly difficult in residential facilities, where aggregate energy consumption is often presented. Additionally, power companies need accurate estimates of future energy usage to implement more efficient generation strategies, such as load-dependent energy generation, smart grids, dynamic pricing models, or energy conservation efforts. Appliance manufacturers can also benefit from a detailed understanding of patterns of their appliances' usage to create more energy-efficient products or new power applications, such as home automation, activity sensing, and healthcare. Thus, monitoring appliance-level energy information allows for a detailed understanding of energy consumption patterns and helps identify areas for potential energy savings. The electrical loads can be monitored through a distributed approach, where each appliance has its own sensor, or by disaggregating the building-level energy consumption profile, which is known as Non-Intrusive Load Monitoring (NILM).

Hart's work on non-intrusive load monitoring (NILM) has helped to address this challenge. NILM is a technique that uses aggregate power consumption data to identify the power consumption of individual appliances. However, one of the key challenges in NILM is identifying patterns in the data. Clustering is a powerful tool that can be used to group similar data points together and identify patterns. This is particularly important when dealing with large and complex datasets.

Research on NILM has seen significant growth in recent years in three main areas: feature selection and extraction for different loads, development of detection and classification algorithms, and acquisition of power datasets that aid in developing and evaluating NILM systems. A comprehensive review of existing NILM approaches can be found in [4-7] *etc*.

According to a review of NILM techniques in [8] these approaches can be divided into two main categories: supervised and unsupervised methods. Supervised methods require labelled data to train a classifier, which can then recognise

appliance usage patterns from the aggregated load measurement. While these methods have shown good performance in load disaggregation, they require additional instrumentation and human effort to set up. In contrast, unsupervised NILM systems do not require pre-training and are thus expected to have a wider range of applicability and less intrusion. This paper focused on unsupervised NILM techniques specifically for low-frequency sampling rates, where the sampling rate is typically 1 to 3 samples per second. Despite the lower amount of information provided by low-frequency sampling data, it requires low-cost hardware and is a more feasible approach in terms of cost. The main task in this case was to detect appliance state transitions (e.g. ON or OFF) from the power measurements.

In the context of low-frequency sampling rates, one of the main goals of NILM is to accurately detect transitions in appliance states (such as turning on or off) using only the aggregate power measurements to use event detection [9]. This process can be approached in two main ways: state-based methods and event-based methods [10]. Each approach offers unique advantages and employs different techniques for accurate appliance monitoring and analysis.

Event-based Non-Intrusive Load Monitoring (NILM) is a technique used to determine the power consumption of individual appliances within a household by analysing the aggregated electrical signal [11]. In NILM, one of the key steps is identifying the operation state of appliances, such as whether they are turned on or off, from the captured electrical load data. There are two main approaches to this task: the state-based and operational or event-based approach.

State-based NILM approaches do not rely on event detectors, instead, they represent each appliance operation using a state machine with distinct state transitions based on appliance usage patterns [12]. They are since when the appliance turns ON/OFF or changes running states, it creates different edge measurements with a probability distribution that matches that appliance. State-based NILMs typically rely on Hidden Markov Models (HMM) and their various adaptations [13-16].

One of the main benefits of the state-based approach is that it provides a continuous estimate of the state of the appliances, allowing for detailed analysis of the energy consumption patterns over time. In addition, this approach can capture subtle changes in the power consumption signal that may not be detected by the event-based approach. However, state-based approaches are limited by the need for expert knowledge to set apriori values for each appliance state via long periods of training. They also have high computational complexity [17]; and do not have a good way to handle the fact that states may stay unchanged for long time intervals [18]. Despite these limitations, the state-based approach has been widely used in NILM and has been shown to be effective in many applications. Despite the demonstrated capabilities of the state-based approach, this paper adopted the event-based or operational state approach due to its advantages over the state-based approach.

The event-based approach focuses on the state transition edges generated by appliances and uses change detection algorithms to identify the start and end of an event [19][20]. This approach relies on the fact that power consumption in a home is constantly changing, with significant "steps" in the data indicating that an event, such as an appliance turning on or off, has occurred. The event-based approach is simpler and more computationally efficient than the state-based approach, making it more suitable for real-time implementation. In addition, the event-based approach is less sensitive to measurement error and electrical interference, as it only considers the changes in the power consumption signal. This makes it a more reliable and robust method for identifying the state of the appliances in NILM. Approximations such as active power, increasing/falling edge, *etc* are extracted, after which the extracted appliances signature is analysed to classify the event based on the appliance and its power consumption estimated.

Figure 1 shows the total (real) power consumption against time for a single-family home over a busy forty-minute period. During this interval, the total load shows a great deal of activity, due mainly to cooking. Four different-sized step changes are clearly present, providing distinctive signs of the refrigerator, two oven elements, and a stove burner element. Therefore, the identification of load signatures is highly dependent on the following operational states. The operational state of appliances is a key factor in Non-Intrusive Load Monitoring (NILM) and is used to classify different types of appliances. Fig 1 shows the appliance power consumption over time.

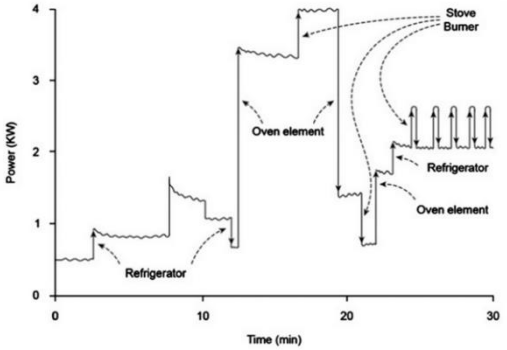

Figure 1. Appliance Power Consumption over Time

The operational state categorisation is a method of grouping appliances based on their behaviour and characteristics to better understand their energy consumption patterns. The categorisation proposed by [21] includes four main types of operational states:

1. Appliances with ON and OFF operation states. Examples of such devices include table lamps, toasters, etc.
2. Multi-state appliances with a finite number of operating states are also referred to as Finite State Machines (FSM). Consumer appliances belonging to this category include washing machines, stove burners, etc. The switching pattern of these appliances is also repeatable which makes it easier for the disaggregation algorithm to identify their operation.
3. Appliances belonging to this category are also known as Continuously Variable Devices (CVD) because of their variable power draw characteristics with no fixed number of states. The power drill and dimmer lights are examples of CVDs with no repeatability in their power draw characteristics. Hence it is very challenging for the NILM methods to disaggregate this type of appliance from the aggregated load measurements.
4. Appliances that remain active throughout weeks or days consuming energy at a constant rate and are

therefore referred to as permanent consumer devices [22]. An example of such appliances is a hardwired smoke detector, telephone sets, cable TV receivers, etc.

This paper proposes a data analysis procedure for Non-Intrusive Load Monitoring (NILM) which aims to estimate the energy consumption of individual appliances in a household. The proposed procedure includes the use of wavelet transform for denoising the aggregate energy consumption data and the application of both K-Means and Self-Organising Map (SOM) clustering algorithms to group similar patterns in the denoised data. The performance of the two clustering algorithms is then evaluated and compared in terms of accuracy, computational time, and interpretability of results. The proposed method will be tested on real-world data from the REFIT dataset, which provides detailed information on the energy consumption of individual appliances in a household. The results of this study will inform the selection of clustering algorithms for NILM and provide insight into the advantages and limitations of each algorithm for energy management in smart grid systems.

1.1 Contributions of the Study

The contributions of this study are as follows:

1. A data analysis procedure for Non-Intrusive Load Monitoring (NILM) that includes the use of wavelet transform for denoising the aggregate energy consumption data and the application of both K-Means and Self-Organising Map (SOM) clustering algorithms to group similar patterns in the appliances.
2. A comparative analysis of the performance of the two clustering methods in terms of accuracy, computational time, and interpretability of results.
3. An evaluation of the proposed method on real-world data from the REFIT dataset, which provides detailed information on the energy consumption of individual appliances in a household.

The stated contribution is expected to add significant knowledge to the field of NILM and help in improving the understanding of energy consumption at the appliance level, which is essential for effective energy management.

In the subsequent sections of the paper will include the following: a description of the dataset used in this study, methodology, feature extraction, and selection, comparative study of the clusters, findings and conclusions.

1.2 REFIT NILM Dataset

The REFIT Electrical Load Measurements dataset consists of data collected from 20 households located in the Loughborough area from 2013 to 2015. The collection of this dataset was supported by the Engineering and Physical Sciences Research Council (EPSRC) through the Personalised Retrofit Decision Support Tools for UK Homes using Smart Home Technology project, which is a collaboration among the Universities of Strathclyde, Loughborough, and East Anglia.

Each household was equipped with 10 power sensors, including a current clamp for the household aggregate and 9 Individual Appliance Monitors (IAMs). Only active power in Watts was recorded, with data being collected at approximately 6 to 8-second intervals. The subset of appliances being monitored was selected based on the largest consumers in UK households as documented by the Department of Energy and Climate Change.

The data is stored in a CSV (Comma Separated Values) format and includes the following columns: Date and Time (in YYYY-MM-DD and HH:mm:ss format respectively), UNIX timestamp (in UTC), Aggregate, Appliance1, Appliance2, Appliance3, ..., Appliance9. Data was only recorded when there was a change in load, and missing values were forward filled with the last recorded value unless the gap was greater than 2 minutes, in which case it was filled with 0. The sensors were not synchronised, and the collection script polled every 6-8 seconds, so the sensor may have updated anywhere within the last 6-8 seconds.

The dataset includes a few periods of missing data due to household internet failures, hardware failures, and network routing issues. All the data has been cleaned to forward-fill NaN values and remove values exceeding 4000 Watts on the IAM datastreams, as this is above the maximum possible power draw. Table 1 and 2 show sample of the dataset.

Table 1. Description of Appliances

Appliance	Description
App1	Fridge, Hotpoint, RLA50P
App2	Freezer(1), Beko, CF393APW
App3	Freezer(2), Unknown, Unknown
App4	Washer Dryer, Creda, T522VW
App5	Washing Machine, Beko, WMC6140
App6	Dishwasher, Bosch, Unknown
App7	Computer, Lenovo, H520s
App8	Television Site, Toshiba, 32BL502b
App9	Electric Heater, GLEN, 2172

2. METHODOLOGY

This paper aims to analyse the energy consumption patterns of different household appliances and identify potential patterns that can be used to improve energy efficiency. The study employs unsupervised clustering techniques to examine the energy consumption data obtained from the appliances. The methodology consists of two main steps: feature extraction and selection, followed by unsupervised clustering.

In the first step, statistical feature extraction was performed to analyse the dataset. This involved computing the statistics for each appliance, including the minimum, maximum, standard deviation, and mean values. Additionally, day sections were defined to characterise the time intervals, as energy consumption patterns are expected to vary throughout the day.

In the second step, unsupervised clustering techniques were applied to the dataset. Specifically, Wavelet Transformation was employed to denoise the dataset after which the K-Means, and Self-Organising Maps (SOM) were used to cluster the energy consumption data. The results of these clustering techniques were analysed to identify any potential energy consumption patterns among the appliances.

Table 2. Sample of the REFIT Dataset

Time	Unix	Aggregate	App1	App2	App3	App4	App5	App6	App7	App8	App9
10/09/2013 13:06	1381323977	523	74	30	69	20	25	18	40	35	15
10/09/2013 13:06	1381323991	526	75	30	69	20	25	18	40	35	15
10/09/2013 13:06	1381324006	540	74	30	68	20	25	18	40	35	15
10/09/2013 13:07	1381324021	532	74	30	68	20	25	18	40	35	15
10/09/2013 13:07	1381324035	540	74	30	69	20	25	18	40	35	15
10/09/2013 13:07	1381324038	539	74	30	69	20	25	18	40	35	15
10/09/2013 13:07	1381324050	537	74	30	69	20	25	18	40	35	15
10/09/2013 13:07	1381324052	537	74	30	69	20	25	18	40	35	15

The methodology described in this study has the potential to provide valuable insights into energy consumption patterns, which can be used to improve energy efficiency and reduce energy costs. By identifying patterns in energy consumption, households can better manage their energy usage and reduce their overall environmental impact.

2.1 Feature Extraction and Selection

In this study, statistical feature extraction was performed on the power consumption data of various appliances in the household. The objective was to extract features that would aid in the classification of appliances based on their power consumption profiles.

The minimum power consumption P_{min}, maximum power consumption standard deviation of power consumption, and mean power consumption were computed for each appliance. These features were calculated using Equations 1, 2, 3, and 4, respectively.

$$P_{min} = \min_{t=1}^{T} P(t) \quad (1)$$

where T is the total time and

t is the specific time

The maximum power consumption P_{max} for an appliance is given by:

$$P_{max} = \max_{t=1}^{T} P(t) \quad (1)$$

The standard deviation of power consumption σ_P for an appliance is given by:

$$\sigma_P = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(P(t)-\bar{P})^2} \quad (2)$$

where $\bar{P}$ is the mean power consumption, which is given by:

$$\bar{P} = \frac{1}{T}\sum_{t=1}^{T} P(t) \quad (3)$$

To calculate the statistics for each appliance in the dataset, an algorithm was developed that takes a dataframe and an appliance name as inputs and returns a dictionary of statistics for that appliance. The algorithm is outlined in Algorithm 2.

First, the dictionary A_c is created by mapping each appliance name to its corresponding label in the dataframe. Then, the label for the specified appliance is looked up and saved as the variable *label*. Using this label, the minimum value (Min_p), maximum value (Max_p), standard deviation (Std_p) dataframe. The algorithm is outlined in Algorithm 2.

The algorithm begins by creating an empty list to store the dictionaries of statistics for each appliance. Then, it loops through the nine appliances in the dataset, calls the first algorithm to compute the statistics for each appliance, and appends the resulting dictionary to A_s. Once all appliances have been processed, the algorithm creates a dataframe from A_s and reorders the columns to have 'App', 'Min', 'Max', 'Std', and 'Mean' in that order. The final output of the algorithm is the dataframe Df_s, which contains the statistics for all appliances in the dataset. Figure 2 depicts a flowchart of the methodology.

2.2 Unsupervised Clustering Techniques

In the field of non-intrusive load monitoring (NILM), clustering high-dimensional data is an effective technique for identifying appliance usage patterns and for preprocessing data for further analysis. Among the categories of clustering techniques, partition-based (e.g., K-Means) and hierarchy-based (e.g., SOM) algorithms are commonly used for clustering NILM data [23-25].

However, managing noisy, incomplete, or sampled data is often necessary when using these clustering techniques, and this can be particularly challenging with high-dimensional data. As such, it is important to carefully choose the appropriate clustering technique for a given task to achieve optimal separation of the data into distinct clusters, as the performance of different techniques can vary significantly for different applications and data types.

In this study, the discrete wavelet transform (DWT) as a denoising tool was used to preprocess high-dimensional NILM data, and then compare the performance of two popular unsupervised clustering techniques, K-Means, and SOM, in terms of their ability to separate appliance usage patterns into distinct clusters. The comparative analysis of these two clustering techniques was conducted to determine which technique is better suited for clustering high-dimensional NILM data.

Table 3. Day Sections

Section	Time Range
Early morning	4am - 8am
Morning	8am - 12pm
Noon	12pm - 4pm
Evening	4pm - 8pm
Night	8pm - 12am
Late night	12am - 4am

2.2.1 Wavelet Transform

Wavelet transformation is a mathematical tool that is used to analyze and process signals with complex and dynamic characteristics. It has been used in various fields, including signal processing, image processing, and data compression. Wavelets are functions that are used to decompose signals into different scales and frequencies, making them ideal for analyzing signals with varying frequencies and amplitudes.

In the context of Non-Intrusive Load Monitoring (NILM), wavelet transformation is a useful technique for processing and denoising the recorded signals before applying clustering algorithms like KMeans and SOM. In this paper, we utilize the Discrete Wavelet Transform (DWT) for preprocessing step, denoising the Non-Intrusive Load Monitoring (NILM) data before applying K-Means and SOM clustering algorithms. DWT is a mathematical tool that allows for the decomposition of a signal into different frequency bands, allowing the identification and removal of noise in the data is expressed in Equation 5. The DWT is a time-frequency representation technique that decomposes a signal into a series of wavelets that capture information at different scales. The decomposition is performed by convolving the signal with a set of basic functions, known as the wavelet basis, and scaling functions, known as the scaling basis. The DWT is computed by iteratively filtering the signal, first with a high-pass filter, and then with a low-pass filter, each followed by subsampling.

$$W(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \psi^* \left(\frac{t-b}{a} \right) dt \quad (4)$$

where $x(t)$ is the input signal, $\psi(t)$ is the wavelet function, c_i is the scale parameter, and b is the translation parameter.

In this paper, the DWT is used to decompose the NILM data into different frequency bands. The detail coefficients, which represent high-frequency components, are then thresholded to remove noise. The thresholding is performed using a soft thresholding function as expressed in Equation 6.

$$f(x, \lambda) = \text{sign}(x) \max(|x| - \lambda, 0) \quad (5)$$

where x is the input data, λ is the threshold, and sign (x) is the sign of *i.* The threshold is chosen using the universal threshold method.

After denoising, the NILM data using the DWT, K-Means, and SOM clustering algorithms are applied to the denoised data to identify the appliances in the data. Overall, the DWT provides an effective method for denoising NILM data, allowing for more accurate identification of the underlying appliances in the data.

2.2.2 Partition-based K-Means

K-Means is a well-known iterative clustering technique that aims to partition a given dataset into a specific number of clusters. The input dataset X is represented as a collection of time-domain features extracted from the historical sensor data. The dataset X contains n instances, and each instance x_i represents a specific data point. The primary goal of the K-Means algorithm is to divide the dataset X into clusters, such that $k<=n$. The algorithm proceeds by first selecting k initial centroids from the dataset. These centroids are typically selected randomly. Once the initial centroids are selected, each data point is assigned to the nearest centroid based on its similarity. The similarity measure used is often the Euclidean distance metric. The resulting data points are partitioned into k clusters, where each cluster is represented by its centroid. The centroid of a cluster C_j is calculated using Equation 7.

$$c_j = \frac{1}{n_j} \sum_{x_i \in C_j} x_i \quad (6)$$

Here, nj denotes the number of instances in cluster C_j, and x_i is an instance in X. The objective of the K-Means algorithm is to minimize the sum of squared distances between each data point and its assigned centroid, known as the Sum of Squares Error (SSE) expressed as Equation 8. The algorithm iteratively updates the centroids and reassigns the data points to the nearest centroid until convergence, where the centroids no longer change.

$$SSE = \sum_{i=1}^{k} \sum_{x_j \in C_i} \left| x_j - c_i \right|^2 \quad (7)$$

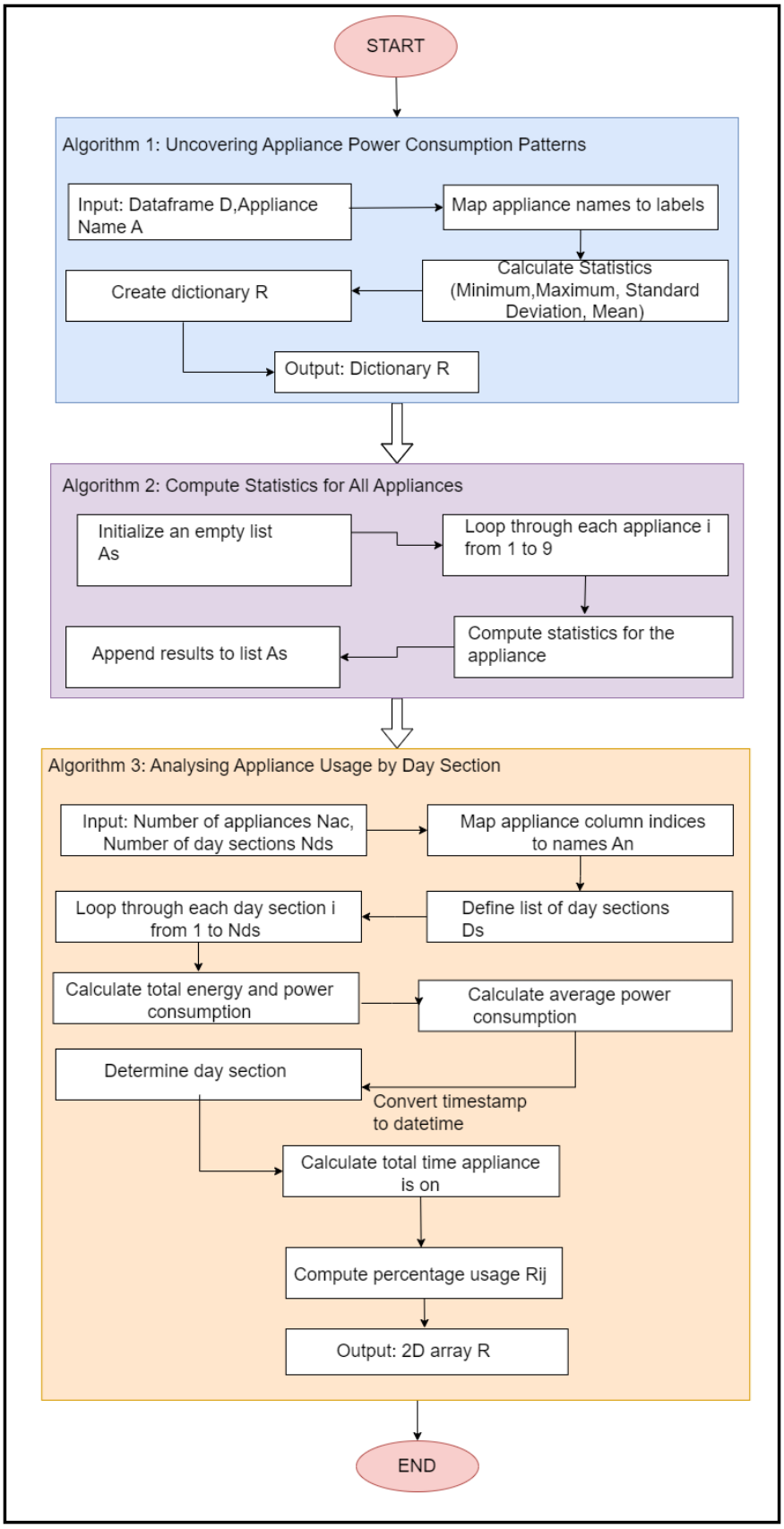

Fig 2 Flowchart of the Methods

K-Means is sensitive to the initial selection of cluster centroids. Therefore, it is essential to specify the optimal number of clusters k in advance. The time complexity of K-Means is expressed as Big O notation and is $O(nkt)$, where n is the number of data points in X, k is the number of clusters, and t is

the number of iterations. However, despite its popularity and efficiency, K-Means can suffer from several limitations, such as sensitivity to the initial selection of centroids and difficulties in handling high-dimensional data.

2.2.3 Self-Organising Maps (SOM)

Self-Organising Maps (SOMs) are a popular unsupervised machine learning technique used for clustering and visualization of high-dimensional data. In the context of Non-Intrusive Load Monitoring (NILM), SOMs can be used to identify electrical appliances and their usage patterns by analyzing the voltage and current signals obtained from the mains of a building.

The SOM algorithm works by mapping the input data onto a 2D grid of neurons, where each neuron represents a cluster or a codebook vector. The weights of the neurons are adjusted during the training process to minimize the quantization error, which is the difference between the input signal and the nearest neuron in the grid. The weight update equation for SOM can be expressed in Equation 9.

$$w_j(t+1) = w_j(t) + \alpha(t)h_{ij}(t)[x(t) - w_j(t)] \quad (9)$$

Where w_j (t) is the weight vector of neuron *j* at time *t*, *x(t)* is the input vector at time *t*, $\alpha(t)$ is the learning rate at time *t*, and $h_{ij}(t)$ is the neighbourhood function, which determines the influence of the neighbouring neurons on the weight update. The neighbourhood function in Equation 10 is typically defined as a Gaussian function of the distance between the neurons.

$$h_{ij}(t) = exp\left(-\frac{|r_l - r_j|^2}{2\sigma^2(t)}\right) \quad (8)$$

where r_l and r_j are the positions of neurons l and j in the 2D grid, and $\sigma(t)$ is the neighborhood radius at time t. The neighborhood radius is typically initialized to a large value and gradually decreased during the training process to allow the SOM to capture both global and local features of the input data. The distance between two vectors x and y is typically calculated using the Euclidean distance expressed in Equation 11.

$$Euclidean\ distance = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (9)$$

where n is the dimensionality of the vectors.

In NILM, the SOM algorithm can be used to cluster similar electrical appliances based on their usage patterns and to identify the transitions between different appliance states. The trained SOM can then be used to disaggregate the aggregate power signal into individual appliance-level signals, which can be used for energy management and appliance diagnostics.

2.2.4 Techniques for Inferring the Optimum Number of Clusters

Determining the optimal number of clusters is a critical step in clustering analysis, as it has a significant impact on the resulting clustering partitions. Choosing an appropriate number of clusters is essential to obtaining accurate and useful information from the dataset. However, selecting the optimal number of clusters is a challenging task, especially when prior knowledge of the dataset is not available. In addition, when dealing with high-dimensional datasets, the determination of an optimal number of clusters becomes more difficult. To address this challenge, various validation measures were developed to evaluate the quality of the clustering results and determine the optimal number of clusters The results obtained were assessed using five effective validation measures, namely the Calinski-Harabasz Index (CHI), Silhouette Index (SI), Davies-Bouldin Index (DBI), Gap Index (GI), and Krzanowski and Lai Index (KLI);

1. **Calinski-Harabasz Index (CHI)**: The CHI measures the ratio of the between-cluster dispersion to the within-cluster dispersion and is defined as:

$$CHI(k) = \frac{tr(B_k)}{tr(W_k)}\frac{N-k}{k-1} \quad (10)$$

Where tr (B_k) is the trace of the between-cluster scatter matrix, tr(W_k) is the trace of the within-cluster scatter matrix, *N* is the total number of data points, and *k* is the number of clusters.

The between-cluster scatter matrix B_k is the sum of the products of the cluster centroid differences and their transpose, scaled by the number of observations in each cluster. Mathematically, it is defined as:

$$B_k = \sum_{i=1}^{k} n_i(\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T \quad (11)$$

where n_i is the number of observations in cluster *i*, m_i is the centroid of cluster *i*, $\mathbf{m}$ is the overall centroid of the data, and *T* denotes the transpose operator.

The within-cluster scatter matrix W_k is the sum of the covariance matrices of all clusters, and is defined as:

$$W_k = \sum_{i=1}^{k}\sum_{\mathbf{x}_j \in C_i}(\mathbf{x}_j - \mathbf{m}_i)(\mathbf{x}_j - \mathbf{m}_i)^T \quad (12)$$

where C_i is the set of observations in cluster *i*, x_j is the *j*-th observation, and m_i is the centroid of cluster *i*.

2. **Silhouette Index (SI)**: The SI measures the quality of a clustering result based on the average distance between data points in a cluster and the distance between the data points in the nearest cluster. The SI for a single data point *i* is defined as:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (13)$$

Where *a(i)* is the average distance between point *i* and all other points in the same cluster, and *b(i)* is the minimum average distance between point *i* and all points in the nearest cluster.

To compute *a(i)*, the average distance between data point *i* and all other data points within the same C_k is calculated using the equation below:

$$a(i) = \frac{1}{|C_k| - 1}\sum_{j \in C_k, j \neq i} d(i,j) \quad (14)$$

where $|C_k|$ is the number of data points in cluster C_k, and *d (i, j)* is the distance between data points *i* and *j*.

To compute b*(i)*, we first calculate the average distance between point *i* and all other data points in each of the other clusters *Ci*:

$$b(i) = \min_{l \neq k} \frac{1}{|C_l|} \sum_{j \in C_l} d(i, j) \quad (15)$$

Where $|Cl|$ is the number of data points in cluster *Cl*.

Finally, the SI for the clustering result is computed as the average SI value over all data points in the dataset:

$$SI = \frac{1}{n} \sum_{i=1}^{n} s(i) \quad (16)$$

where n is the total number of data points. The SI ranges from -1 to 1, where a value of 1 indicates that the data point is well-matched to its own cluster and poorly matched to other clusters, while a value of -1 indicates that the data point is poorly matched to its own cluster and well-matched to other clusters. A value of 0 indicates that the data point is on the boundary between two clusters.

3. **Davies-Bouldin Index (DBI)**: The DBI measures the distance between clusters and the compactness of the clusters. The lower the DBI value, the better the clustering result. The DBI is defined as:

$$DBI(k) = \frac{1}{k} \sum_{i=1}^{k} \max_{j \neq i} \left(\frac{S_i + S_j}{d(c_i, c_j)} \right) \quad (17)$$

where S_i is the scatter of cluster *i*, C_i is the centroid of cluster *i*, and *d(ci,cj)* is the distance between the centroids of clusters *i and j*.

The scatter of a cluster $_i$ is defined as:

$$S_i = \frac{1}{n_i} \sum_{x \in C_i} d(x, c_i)^2 \quad (18)$$

where n_i is the number of data points in cluster *i*, C_i is the set of data points in cluster *i*, *x* is a data point, c_i is the centroid of cluster *i*, and *d(x,ci)* is the Euclidean distance between data point x and the centroid *ci* of cluster *i*.

The distance between the centroids of two clusters *i and j* is defined as:

$$d(c_i, c_j) = \| c_i - c_j \|_2 \quad (19)$$

where $\|.\|_2$ represents the Euclidean distance between two points. The DBI can be computed for a range of values of k and the optimal number of clusters can be determined by selecting the value of k that minimizes the DBI.

4. **Gap Index (GI)**: The GI compares the dispersion of data within a cluster with the dispersion of data in a reference data set to estimate the optimal number of clusters. The GI is defined as:

$$GI(k) = E_n^*(log(W_k)) - log(W_k) \quad (20)$$

where *k* is the number of clusters, $E_n^*(log(W_k))$ is the expected value of the log of the within-cluster dispersion for a random reference data set of size n, and W_k is the within-cluster dispersion of the original data set.

The expected value of the log of the within-cluster dispersion for a random reference data set of size n is computed as:

$$E^n(log(W_k)) = \frac{1}{B} \sum b = 1^B log(W_{kb}^{)}) \quad (21)$$

where B is the number of reference data sets, and W^*_{kb} is the within-cluster dispersion of the b_{th} reference data set.

The within-cluster dispersion is defined as:

$$W_k = \sum_{i=1}^{k} \frac{1}{2n_i} D_i \quad (22)$$

where n_i is the number of data points in cluster *i*, and D_i is the dispersion of data points in cluster *i*:

$$D_i = \sum_{j=1}^{n_i} dist(x_{ij}, c_i)^2 \quad (23)$$

where $dist(x_{ij}, c_i)$ is the Euclidean distance between data point X_{ij} and the centroid *Ci* of cluster *i*.

5. **Krzanowski and Lai Index (KLI)**: The KLI measures the quality of a clustering result based on the average distance between data points in a cluster and is defined as:

$$k(i) = \frac{1}{n_i - 1} \sum_{j=1, j \neq i}^{n_i} d(x_i, x_j) - \frac{2}{n_i} \sum_{j=1}^{n_i} d(x_i, c_i) \quad (24)$$

where *ni* is the number of points in the same cluster as point *i*, *d(xi,xj)* is the distance between data points *i* and *j*, and d(x_i, c_i) is the distance between data point *i* and the centroid c_i of its cluster.

The KLI for the entire data set is then defined as:

$$KLI(k) = \frac{1}{N} \sum_{i=1}^{N} k(i) \quad (25)$$

where *N* is the total number of data points and *k* is the number of clusters. The lower the KLI value, the better the clustering result.

2.3 Handling Class Imbalance in Multiclass Dataset

In the field of Non-Intrusive Load Monitoring (NILM), the impact of class imbalance on the performance of machine learning models has also been extensively studied [26]. Class imbalance can cause biased predictions towards the dominant appliances and hinder the ability of models to accurately detect and identify the appliances with lower frequency of occurrence [27]. This issue is further compounded in the case of NILM where the datasets are often highly multiclass in nature, consisting of multiple appliances with varying frequency of use.

To measure the extent of class imbalance in NILM datasets, previous studies have used various metrics, such as the Empirical Distribution (ED), Class Frequencies (CF), Imbalance Ratio (IR), Imbalance Degree (ID), and the Likelihood-Ratio based Imbalance Degree (LRID) [28]. The ED and CF metrics can provide informative insights into class imbalance, but they are not suitable for highly multiclass datasets [28]. For instance, when the number of classes is large, ED and CF may not provide a reliable measure of the extent of class imbalance.

The IR metric considers only the smallest minority and the largest majority classes, which may not be sufficient to capture the full distribution of classes in a dataset. As a result, IR may not provide an accurate measure of class imbalance in multiclass datasets. Therefore, other metrics like ID and LRID have been proposed to address the limitations of ED, CF, and IR. ID is an effective measure of class imbalance in multiclass datasets as it provides a reliable estimate of the imbalance degree for each class. LRID, on the other hand, uses a likelihood-ratio based approach to measure the imbalance degree, which has been found to be effective in identifying imbalanced classes in multiclass datasets.

2.3.1 *Imbalance Degree (ID)*

Imbalance degree (ID) is a metric used to measure the extent of class imbalance in a dataset [29]. It considers information from all classes, both minority and majority, and provides a summary of the difference between the minority and majority class distributions. The ID metric is calculated using the following equation:

$$ID = \frac{1}{1 + d_{pm,b}} \tag{26}$$

Where p^m is the distribution showing exactly *m* minority classes with the highest distance and *b* is the majority class distribution. The maximum distance between *b* and all possible p^m is given as d_{pmb}.

In a balanced dataset, where there are equal numbers of instances for each class, the ID value is 1. Values of ID greater than 1 signify an imbalanced dataset, with higher ID values indicating more severe class imbalance.

The Hellinger distance (HD) is often used in instantiating the ID metric as it produces significant and robust results Hellinger (1909). The HD is a measure of the similarity between two probability distributions and is defined as:

$$HD(p,q) = \sqrt{\frac{1}{2}\sum_{i=1}^{n}(\sqrt{p_i} - \sqrt{q_i})^2} \tag{27}$$

Where *p and q* are two probability distributions, and n is the number of classes. The use of HD in the ID metric provides a more reliable estimate of class imbalance in multiclass datasets.

2.3.2 *Likelihood-Ratio Imbalance Degree (LRID)*

The Likelihood-Ratio imbalance degree (LRID) was introduced as a high-resolution metric by Zhu *et al* to quantify the degree of imbalance in multiclass datasets. The LRID metric employs the LR test to generate a numerical value that encapsulates the dissimilarity between the actual class distribution p and the observed distribution b.

The LR test for the multinomial distribution (N, p) tests the hypothesis $H_0 : p = b$ against $H_1 : p \neq b$, and can be expressed as:

$$LR = 2\sum_{i=1}^{K} n_i \ln \frac{p_i}{b_i} \tag{28}$$

where K represents the number of classes, n_i denotes the number of samples in the *ith* class, p_i represents the estimated probability of the *ith* class, and bi is the observed probability of the *ith* class.

If the dataset is balanced, then p=b, and LR=0. In contrast, if the dataset is imbalanced, and p significantly differs from *b*, then LR> 0. A larger value of LR corresponds to a larger discrepancy between *p and b*.

LRID can be expressed using the LR statistic as:

$$LRID = \frac{1}{K}\sum_{i=1}^{K} \ln\left[1 + \frac{n_i}{(N - n_i)b_i}\right] \tag{29}$$

where *N* is the total number of samples in the dataset.

When $pi = n_i/N$, the LRID equation can be simplified as:

$$LRID = \frac{1}{K}\sum_{i=1}^{K} \ln\left[1 + \frac{1 - p_i}{p_i \times IR}\right] \tag{30}$$

where *IR* is the Imbalance Ratio and is defined as:

$$IR = \frac{\max_i b_i}{\min_i b_i} \tag{31}$$

By utilizing LRID to estimate the level of class imbalance, resampling techniques like SMOTE can be employed to generate synthetic instances of the minority class(es) and address the imbalanced data distribution.

3. RESULTS AND DISCUSSION

3.1 Appliance Power Consumption Patterns Analysis Result

Table 4 presents the statistical features extracted for each appliance, including Minimum (Min) and Maximum (Max) power consumption, Standard deviation (Std), and Mean power consumption. Figure 3 shows the minimum values for each appliance, Figure 4 shows the maximum values for each appliance, Figure 5 shows the standard deviation of values for each appliance, and Figure 6 shows the mean values for each appliance. Each plot is color-coded according to the corresponding appliance.

Table 4 Statistical Features for each Appliance

Appliance	Min	Max	Std	Mean
Fridge	0	972	30.33	14.4
Freezer (1)	0	926	27.46	15.4
Freezer (2)	0	1024	36.29	27.7
Washer Dryer	0	2048	20.4	0.21
Washing Machine	0	0	0	0
Dishwasher	0	2336	166.7	12.5
Computer	0	53	3.78	0.41
Television Site	0	57	9.93	3.31
Electric Heater	0	1	0.02	1

Looking at the minimum and maximum power consumption, we can see that the dishwasher and washing machine have the highest maximum power consumption values, while the Fridge, Freezer (1), and Freezer (2) have relatively lower maximum power consumption values. Interestingly, the washing machine

has a minimum power consumption of zero, which could indicate that it is switched off during certain periods.

The standard deviation of power consumption is highest for the dishwasher and washing machine, indicating that these appliances have a wider range of power consumption values compared to other appliances. On the other hand, the computer and television have the lowest standard deviation, indicating that their power consumption is more consistent.
Finally, we can see that the mean power consumption values are generally low for most appliances, except for the electric heater. This indicates that these appliances do not consume a lot of power on average, which is not surprising as they are mostly used for domestic purposes.

In summary, the statistical features provide insights into the power consumption profiles of different appliances. The observed differences in minimum and maximum power consumption, standard deviation, and mean power consumption values between appliances could be useful in distinguishing between them based on their power consumption profiles. The color-coding of each plot corresponds to the respective appliance.

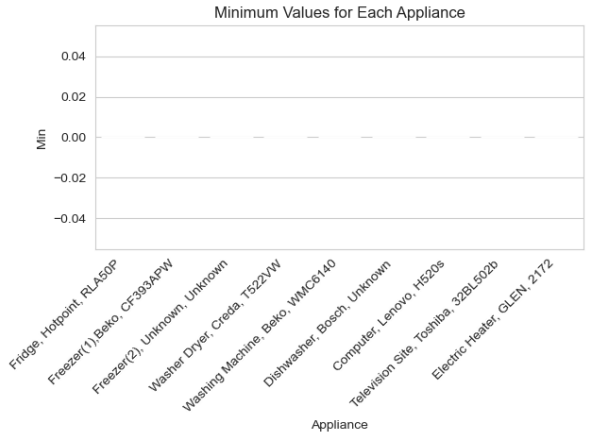

Figure 3 Minimum Values for Each Appliance

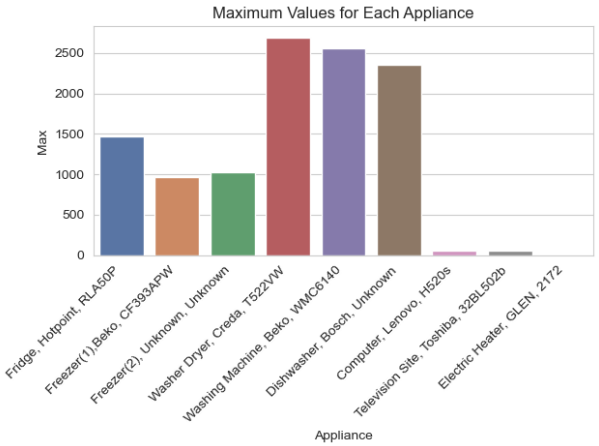

Figure 4 Maximum Values for Each Appliance

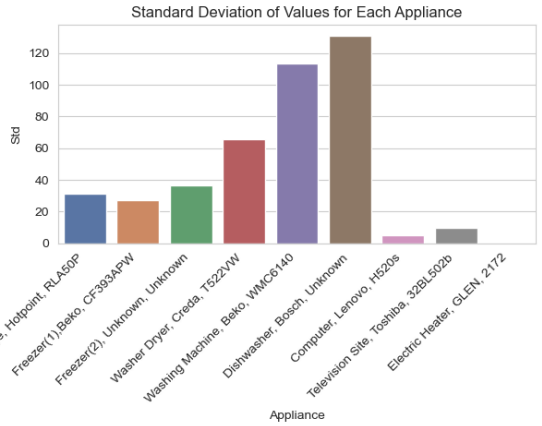

Figure 5 Standard Deviation of Values for Each Appliance

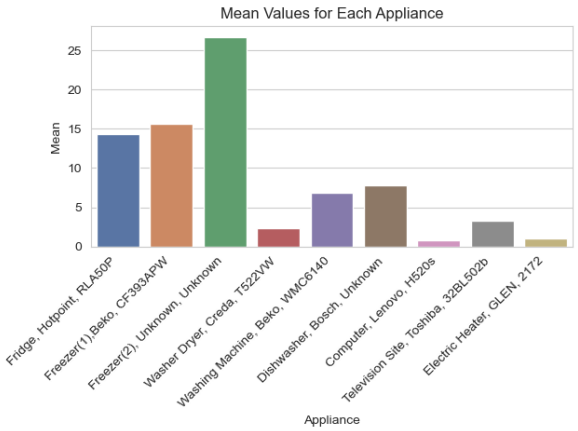

Figure 6 Mean Values for Each Appliance

3.2 Result from Analysing Appliance Usage by Day Section

Table 5 shows the energy consumption (in watts) of various household appliances at different times of the day. The appliances are categorised into different groups based on their functions, such as cooling (fridge, freezers), cleaning (washing machine, dishwasher), entertainment (television, computer), and heating (electric heater).

Looking at the data, it is observed that the energy consumption of the cooling appliances, such as the fridge and freezers, is relatively consistent throughout the day, with some minor fluctuations. This is expected since these appliances are designed to maintain a constant temperature, and the compressor cycles on and off to keep the temperature within a specified range.

The washing machine and dishwasher's energy consumption is highest during the afternoon and early evening, which corresponds to the time when people are typically at home and doing household chores. The computer and television's energy consumption are highest during the evening and late night, likely because this is when people are more likely to use these appliances for leisure or work.
The electric heater's energy consumption is almost constant throughout the day, as it is used for heating the room, and people typically keep it on for extended periods.

Table 5 Power Consumption of Appliances at Different Times of the Day

Appliance	Early Morning	Morning	Noon	Evening	Night	Late Night
Fridge	17.72	18.67	19.7	18.5	20.7	20.23
Freezer(1)	32.16	30.81	33	36.25	33.7	31.49
Freezer(2)	38.47	38.28	35.3	37.71	37.8	38.11
Washer Dryer	0	0.29	0.4	0.76	0.02	0
Washing Machine	0	1.8	3.03	3.65	0	0.01
Dishwasher	0.72	1.39	1.45	0	0	0
Computer	0	0.32	2.38	2.83	6.85	7.78
Television Site	8.68	0	3.17	4.41	16.5	30.86
Electric Heater	99.97	99.94	100	99.98	99.9	99.96

Overall, it is important to record these values in line with the day sections to understand the energy usage pattern of the household and identify opportunities to reduce energy consumption. By knowing the energy consumption of each appliance, households can make informed decisions about how and when to use them to save energy and reduce their utility bills.

In summary, the appliances can be grouped based on their functions and energy usage pattern. The cooling appliances have relatively consistent energy consumption throughout the day, while the washing machine and dishwasher's energy consumption is highest during the afternoon and early evening. The computer and television's energy consumption are highest during the evening and late night, and the electric heater's energy consumption is almost constant throughout the day.

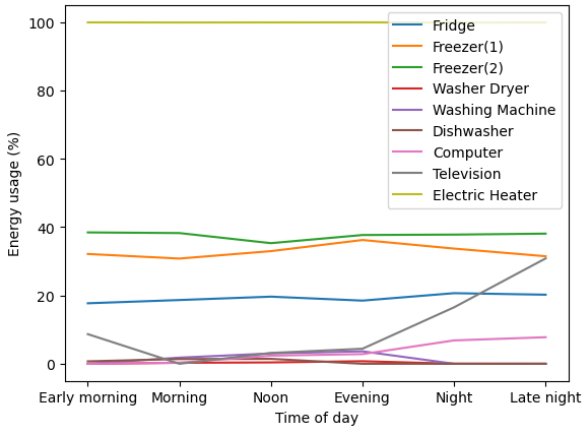

Figure 7 Electricity Consumption by Appliance Throughout the Day

Figure 7 shows the energy consumption of several household appliances during different times of the day. The appliances include a fridge, two freezers, a washer dryer, a washing machine, a dishwasher, a computer, a television, and an electric heater. The data is presented in kilowatt-hours (kWh) and is recorded in six time slots: early morning, morning, noon, evening, night, and late night. The electric heater has the highest energy consumption throughout the day, averaging around 99.95 kWh, followed by the two freezers, which have an average consumption around 35-38 kWh. The fridge and the washing machine have relatively low energy consumption, with an average consumption around 19-20 kWh and 3 kWh, respectively. The dishwasher has the lowest energy consumption, with an average consumption of less than 2 kWh. It also observes that the consumption of most appliances is higher during the evening and night, which is likely when they are being used more frequently. The electric heater, for instance, has a higher consumption during the night and late-night slots, while the television has a higher consumption during the evening and night slots. The fridge and the freezers have a relatively constant energy consumption throughout the day, with a slight increase during the night slot.

3.3 Class Imbalance in Multiclass Datasets Result

Based on the calculated values of ID (Equation (30)) and LRID (Equation (31)), it can be concluded that, the class distribution of the NILM dataset is highly imbalanced. The value of ID = 0.099 indicates that the minority class (the appliance with the lowest number of instances) is only 0.1% of the size of the majority class (the appliance with the highest number of instances), which indicates a significant class imbalance. The LRID value of 6.843 is a logarithmic measure of the same imbalance and indicates that the distribution is highly imbalanced. In summary, these results suggest that the class imbalance in the NILM dataset should be taken into consideration when developing classification models for this dataset.

3.4 Denoising of Appliance Power Consumption Data

The process of denoising the appliance power consumption data, a fundamental step in the data analysis procedure, is critically evaluated in this section. To assess the impact of this process, a comparison was made between the statistical features (mean, standard deviation, and range) of both the original and the denoised data.

As demonstrated in Table 7, it was observed that the denoising process effectively preserved the key statistical characteristics of the original data. This implies that the noise elements were significantly reduced without causing a meaningful alteration in the information content of the data.
For example, the 'Fridge, Hotpoint, RLA50P', which had an original mean power consumption of 17.54 (units), remained the same after the denoising process. The same can be observed

for the standard deviation and the range, indicating that the dispersion and spread of the data remained consistent post-denoising.
Such results were consistently observed across all the appliances in the dataset, highlighting the efficacy of the denoising process in preserving the integrity of the original data.

This preservation of the original data's statistical features is crucial in the context of clustering. Clustering algorithms, such as K-Means and Self-Organising Maps (SOM), rely heavily on the intrinsic patterns and structures present in the data. If the denoising process had significantly altered these, the ability of the clustering algorithms to identify accurate and meaningful clusters would have been compromised.

The denoised data, free from noise and irrelevant variations, would thus be a more reliable input for the clustering algorithms. It would increase the likelihood of forming meaningful clusters, which in turn would enhance the accuracy of the Non-Intrusive Load Monitoring (NILM).

3.5 Understanding Appliance Energy Usage Through K-Means Clustering

This section discusses the results obtained from the clustering analysis using the K-Means algorithm to gain insights into appliance energy usage patterns. The primary objective was to uncover hidden structures within the dataset and identify distinctive groups of energy consumption behaviors exhibited by household appliances. To accomplish this, systematic approach that involved determining the optimal number of clusters **(K)** through the elbow analysis and further evaluating the clustering results within the range of K=2 to K=5 was applied

The elbow analysis is a crucial step in finding an appropriate K value that captures the inherent data distribution while avoiding overfitting. By plotting the distortion (inertia) values across a range of K values as shown in Figure 8, a distinct "elbow point" at K=4 was identified. This point indicates a transition from rapid distortion reduction to a more gradual decline, implying that the data's structure is well-captured within 2 to 5 clusters.

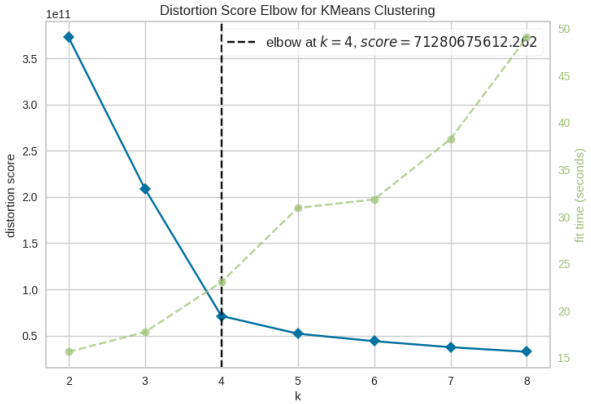

Figure 8 Davies-Bouldin Index (DBI) K-Means Evaluation Metrics

The analysis was extended within the range of K=2 to K=5 and evaluated the clustering quality using multiple metrics. The results are summarised in Table 6 and depicted in the accompanying graph in Figure 9 to 12.

Table 6 K-Means Clustering Evaluation Metrics

K	DBI	CHI	KLI	GI
2	0.20	8421682.67	80.22	-4.58
3	0.20	10252164.23	69.62	-4.47
4	0.19	24484187.55	60.55	-4.36
5	0.31	25767861.86	59.42	-4.36

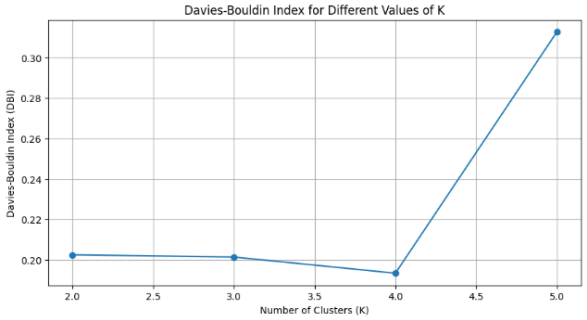

Figure 9 Davies-Bouldin Index (DBI) K-Means Evaluation Metrics

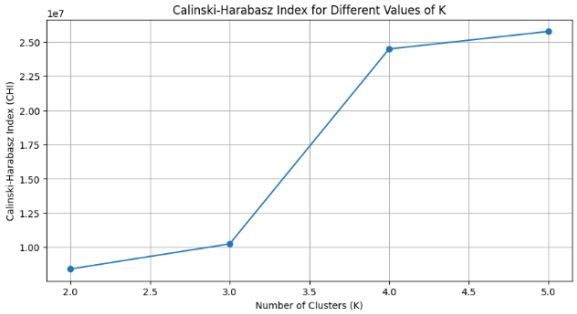

Figure 10 Calinski-Harabasz Index (CHI) K-Means Evaluation Metrics

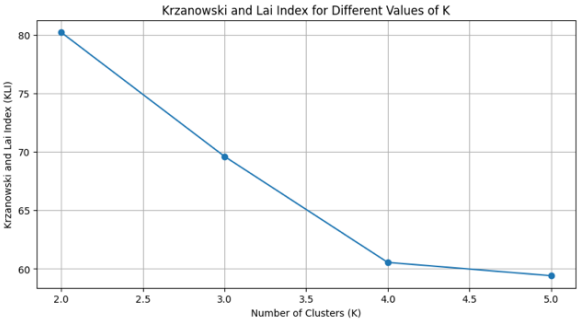

Figure 11 Davies-Bouldin Index (DBI) K-Means Evaluation Metrics

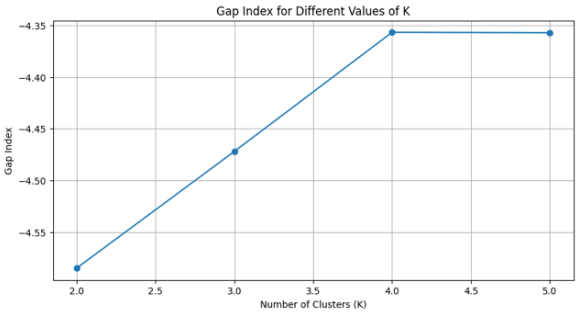

Figure 12 Gap Index (DBI) K-Means Evaluation Metrics

The DBI demonstrates a consistent decline as K increases, indicating improved cluster separation. Similarly, CHI and KLI both exhibit decreasing trends, suggesting better clustering quality. The GI also reinforces these observations. While K=2 yields a relatively low DBI score, K=4 emerges as a balanced choice across multiple metrics, aligning with the elbow point identified in the distortion plot.

Table 7 Comparison of Statistical Features Between Original and Denoised Appliance Power Consumption Data

Appliance	Original Mean	Original Std	Original Range	Denoised Mean	Denoised Std	Denoised Range
Fridge, Hotpoint, RLA50P	17.54	43.09	3584	17.54	43.09	3584
Freezer(1), Beko, CF393APW	16.55	28.84	3452	16.55	28.84	3452
Freezer(2), Unknown, Unknown	29.01	38	3657	29.01	38	3657
Washer Dryer, Creda, T522VW	1.84	56.11	3584	1.84	56.11	3584
Washing Machine, Beko, WMC6140	11.03	143.92	3072	11.03	143.92	3072
Dishwasher, Bosch, Unknown	11.16	156.79	2525	11.16	156.79	2525
Computer, Lenovo, H520s	2.47	11.93	2094	2.47	11.93	2094
Television Site, Toshiba, 32BL502b	5.8	13.17	3584	5.8	13.17	3584
Electric Heater, GLEN, 2172	69.48	255.8	2119	69.48	255.8	2119

3.6 Understanding Appliance Energy Usage through SOM Clustering

This section discusses the results obtained from the clustering analysis using the Self-Organising Maps (SOM) algorithm. The analysis process mirrors the K-Means approach and involves determining the optimal number of clusters (K) through evaluation metrics. We evaluate the SOM results for K=2 to K=5 to gain a comprehensive understanding of the clustering quality.

As depicted in the Table 8, our evaluation metrics provide insights into the clustering quality for each K value. Figure 13 to 16 show details on the individual metrics.

Table 8 SOM Clustering Evaluation Metrics

K	DBI	CHI	KLI	Gap
2	7.06	606,935.08	0.06	-4.76
3	4.42	566,347.49	0.29	-3.83
4	4.47	357,948.94	0.25	-3.57
5	3.32	1,253,110.49	0.32	-3.43

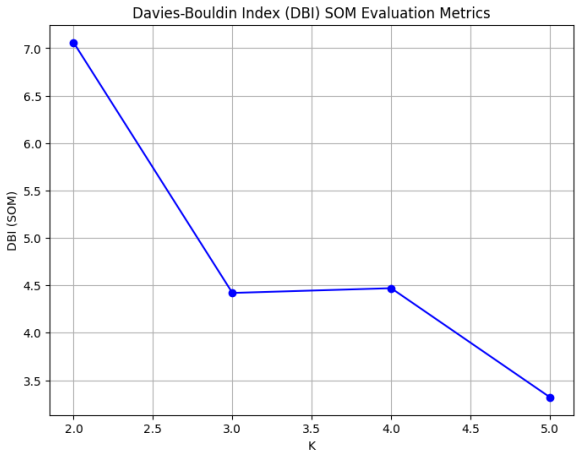

Figure 13 Davies-Bouldin Index (DBI) SOM Evaluation Metrics

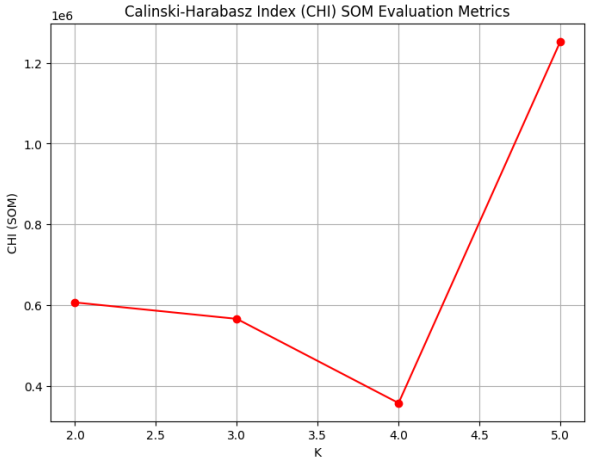

Figure 14 Calinski-Harabasz Index (CHI) SOM Evaluation Metrics

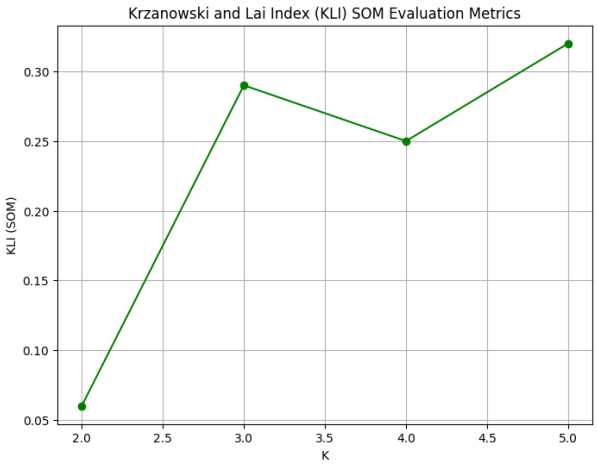

Figure 15 Davies-Bouldin Index (DBI) SOM Evaluation Metrics

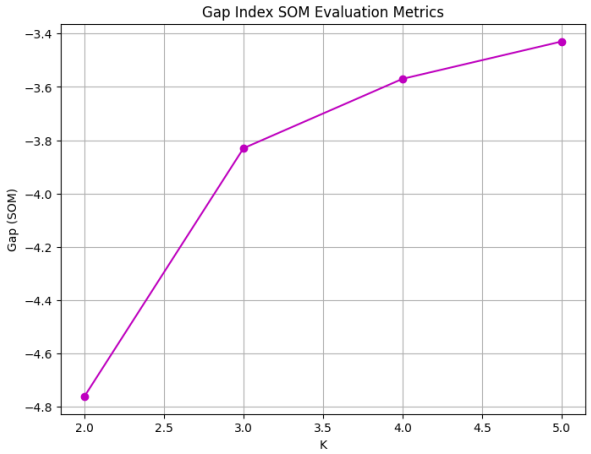

Figure 16 Gap Index (DBI) SOM Evaluation Metrics

3.7 Comparing K-Means and SOM Clustering Results

In this subsection, we present a comparative analysis of two clustering methods, K-Means and Self-Organising Map (SOM), as applied to Non-Intrusive Load Monitoring (NILM). The evaluation is based on several clustering quality metrics, including the Davies-Bouldin Index (DBI), Calinski-Harabasz Index (CHI), Krzanowski-Lai Index (KLI), and Gap Index (Gap).

The analysis of the Davies-Bouldin Index (DBI) on K-Means revealed that the DBI values varied between 0.19 and 0.31 across the range of tested values of K. The lowest DBI value was observed at K=4, indicating that K=4 resulted in the most compact and well-separated clusters. In contrast, the application of SOM clustering led to notably higher DBI values, ranging from 3.32 to 7.06, with K=5 producing the lowest DBI value.

Similarly, the examination of the Calinski-Harabasz Index (CHI) demonstrated that K-Means outperformed SOM, particularly at K=3 and K=4. This observation suggests that K-Means generated clusters with superior between-cluster variance and lower within-cluster variance. Conversely, SOM clustering resulted in significantly lower CHI values, indicating less favorable clustering outcomes in terms of compactness and separation.

Regarding the Krzanowski-Lai Index (KLI), the analysis indicated that K-Means achieved the lowest KLI value at K=4, implying the presence of the most compact clusters among the various tested values of K. In contrast, SOM consistently produced low KLI values across different K values, indicating well-separated clusters with minimal variation across different K values.

Lastly, in the case of the Gap Index (Gap), both K-Means and SOM exhibited relatively consistent Gap Index values across different K values. This suggests that the number of clusters selected did not significantly impact the gap statistic in either method.

The comparative analysis reveals that K-Means clustering outperforms SOM clustering in terms of cluster quality and compactness for the specific problem of NILM. K-Means consistently demonstrates lower DBI, higher CHI, and a trend toward lower KLI values, indicating more distinct, well-separated, and compact clusters. In contrast, SOM shows higher DBI values and lower CHI values, suggesting less favorable clustering results in terms of compactness and separation.

In conclusion, the comparative analysis within this subsection provides insights into the performance of clustering methods in NILM, highlighting the strengths and weaknesses of K-Means and SOM in this application.

4. CONCLUSIONS AND RECOMMENDATIONS

4.1 Conclusions

This research effort delved into a comprehensive analysis of appliance power consumption patterns within the realm of Non-Intrusive Load Monitoring (NILM). The findings and conclusions can be summarized as follows:

The exploration of power consumption patterns among various household appliances revealed considerable disparities in their behaviour. These distinctions encompassed features such as minimum and maximum power consumption, standard deviation, and mean power consumption. Notably, certain appliances exhibited high maximum power consumption values, while others maintained relatively lower figures. This analysis highlights the uniqueness of each appliance's power consumption profile.

The evaluation of class imbalance within the NILM dataset unveiled a significant disparity in the distribution of appliance classes. Metrics like Imbalance Degree (ID) and Likelihood-Ratio Imbalance Degree (LRID) indicated that the dataset's class distribution was highly skewed, with the minority class constituting a mere 0.1% of the majority class. This observation underscores the critical challenge of class imbalance in NILM applications.

The denoising process applied to appliance power consumption data proved to be highly effective. It successfully retained key statistical characteristics from the original data, including mean, standard deviation, and range. This preservation of statistical features is essential for downstream clustering analysis, enhancing the accuracy of Non-Intrusive Load Monitoring (NILM).

The utilisation of K-Means and Self-Organising Maps (SOM) clustering methods offered insights into appliance energy usage patterns. The optimal number of clusters (K) was identified through evaluation metrics. K-Means consistently demonstrated superior cluster quality and compactness, while SOM displayed less favorable clustering results in terms of data organisation.

4.2 Recommendations

Considering the research findings, several recommendations emerge:

1. Addressing Class Imbalance: The prevalence of class imbalance within NILM datasets underscores the importance of employing class imbalance mitigation techniques. Techniques like oversampling or undersampling can help balance the representation of appliances with low-frequency occurrences.

2. Data Quality Enhancement: The denoising techniques applied to data preprocessing should be

integrated as a standard practice. Striking the right balance between noise reduction and feature preservation is essential to enhance data quality for NILM applications.

3. Clustering Method Selection: The choice of clustering methods should be made judiciously, considering the specific characteristics of the dataset. K-Means clustering is recommended when compact and well-separated clusters are desired, while SOM clustering may find suitability in scenarios where cluster interpretability is of lesser concern.

4. Exploration of Advanced Techniques: Future research endeavors should explore advanced clustering techniques and hybrid approaches tailored to the unique challenges of NILM. This exploration could lead to more accurate appliance identification and improved energy monitoring.

5. Real-World Implementation: The insights gained from this study should be translated into practical applications, such as smart homes and energy conservation programs. Real-world household energy monitoring systems can significantly benefit from the techniques and methods discussed.

6. Data Diversity: Continued data collection efforts are pivotal for enhancing the generalizability of appliance identification models. The incorporation of diverse datasets from various sources and geographical locations will contribute to the development of more robust NILM solutions.

5. ACKNOWLEDGMENTS

This research was supported by the University of Mines and Technology. The authors would like to express their sincere gratitude to the institution for providing the necessary resources and support throughout this study.

6. REFERENCES

[1] Çimen, H., Bazmohammadi, N., Lashab, A., Terriche, Y., Vasquez, J.C. and Guerrero, J.M., 2022. An online energy management system for AC/DC residential microgrids supported by Non-Intrusive Load Monitoring. Applied Energy, 307, p.118136.

[2] Tanoni, G., Principi, E. and Squartini, S., 2024. Non-Intrusive Load Monitoring in industrial settings: A systematic review. Renewable and Sustainable Energy Reviews, 202, p.114703.

[3] Liaqat, R., Sajjad, I.A., Waseem, M. and Alhelou, H.H., 2021. Appliance level energy characterization of residential electricity demand: prospects, challenges and recommendations. IEEE Access, 9, pp.148676-148697.

[4] Angelis, G.F., Timplalexis, C., Krinidis, S., Ioannidis, D. and Tzovaras, D., 2022. NILM applications: Literature review of learning approaches, recent developments and challenges. Energy and Buildings, 261, p.111951.

[5] Huber, P., Calatroni, A., Rumsch, A. and Paice, A., 2021. Review on deep neural networks applied to Low-Frequency NILM. Energies, 14(9), p.2390.

[6] Verma, A., Anwar, A., Mahmud, M.A., Ahmed, M. and Kouzani, A., 2021. A comprehensive review on the NILM algorithms for energy disaggregation. arXiv preprint arXiv:2102.12578.

[7] Chavan, D.R. and More, D.S., 2022. A systematic review on low-resolution NILM: Datasets, algorithms, and challenges. Electronic Systems and Intelligent Computing: Proceedings of ESIC 2021, pp.101-120.

[8] Liu, B., Liu, W., Luan, W., Yu, Y., Zhao, B. and Wang, Y., 2024. Automatic appliance labeling for unsupervised NILM based on hierarchical decision-making. IEEE Transactions on Instrumentation and Measurement.

[9] Luan, W., Wei, Z., Liu, B. and Yu, Y., 2022. Non-intrusive power waveform modeling and identification of air conditioning load. Applied Energy, 324, p.119755.

[10] Gomes, L., Natário, D., Costa, A., Barros, J.P. and Campos-Rebelo, R., 2024. Event-Based Modeling of Input Signal Behaviors for Discrete-Event Controllers. Applied Sciences, 14(12), p.5289.

[11] Al-Khadher, O., Mukhtaruddin, A., Ridzuan Hashim, F., Azizan, M.M. and Mamat, H., 2023. An implementation framework overview of non-intrusive load monitoring. Journal of Sustainable Development of Energy, Water and Environment Systems, 11(4), pp.1-40.

[12] Kumar, P. and Abhyankar, A.R., 2023. A Time Efficient Factorial Hidden Markov Model-Based Approach for Non-Intrusive Load Monitoring. IEEE Transactions on smart Grid, 14(5), pp.3627-3639.

[13] Kim, H., Marwah, M., Arlitt, M., Lyon, G. and Han, J., 2011, April. Unsupervised disaggregation of low frequency power measurements. In Proceedings of the 2011 SIAM international conference on data mining (pp. 747-758). Society for Industrial and Applied Mathematics.

[14] Kolter, J.Z. and Jaakkola, T., 2012, March. Approximate inference in additive factorial hmms with application to energy disaggregation. In Artificial intelligence and statistics (pp. 1472-1482). PMLR.

[15] Makonin, S. and Popowich, F., 2015. Nonintrusive load monitoring (NILM) performance evaluation: A unified approach for accuracy reporting. Energy Efficiency, 8, pp.809-814.

[16] Mendoza-Zubieta, V., Gonzalez-Villaseñor, G.A., Vargas-Ortega, G., Gonzalez, B., Ramirez-Renteria, C., Mercado, M., Molina-Ayala, M.A. and Ferreira-Hermosillo, A., 2015. High prevalence of metabolic syndrome in a mestizo group of adult patients with primary hyperparathyroidism (PHPT). BMC endocrine disorders, 15, pp.1-7.

[17] Gopinath, R., Kumar, M., Joshua, C.P.C. and Srinivas, K., 2020. Energy management using non-intrusive load monitoring techniques–State-of-the-art and future research directions. Sustainable Cities and Society, 62, p.102411.

[18] Rehman, A.U., Lie, T.T., Vallès, B. and Tito, S.R., 2021. Non-invasive load-shed authentication model for demand response applications assisted by event-based non-intrusive load monitoring. Energy and AI, 3, p.100055.

[19] Kumar, M., Singh, P.K., Maurya, M.K. and Shivhare, A., 2023. A survey on event detection approaches for sensor based IoT. Internet of Things, 22, p.100720.

[20] Gallego, G., Delbrück, T., Orchard, G., Bartolozzi, C., Taba, B., Censi, A., Leutenegger, S., Davison, A.J., Conradt, J., Daniilidis, K. and Scaramuzza, D., 2020. Event-based vision: A survey. IEEE transactions on pattern analysis and machine intelligence, 44(1), pp.154-180.

[21] Wang, F., Li, K., Duić, N., Mi, Z., Hodge, B.M., Shafie-khah, M. and Catalão, J.P., 2018. Association rule mining based quantitative analysis approach of household characteristics impacts on residential electricity consumption patterns. Energy conversion and management, 171, pp.839-854.

[22] Hamid, O., Barbarosou, M., Papageorgas, P., Prekas, K. and Salame, C.T., 2017. Automatic recognition of electric loads analyzing the characteristic parameters of the consumed electric power through a Non-Intrusive Monitoring methodology. Energy Procedia, 119, pp.742-751.

[23] Tari, L., Nardone, A., Ferrigno, L., Monti, A. and Ponci, F., 2024, May. The impact of Electrical Signature Quality in NILM process: a preliminary analysis based on unsupervised approch. In 2024 IEEE International Workshop on Metrology for Industry 4.0 & IoT (MetroInd4. 0 & IoT) (pp. 88-93). IEEE.

[24] Cheng, X., Li, D., Hu, W. and Xu, J., 2024, May. A Stacking Ensemble Learning Method for Non-Intrusive Appliance Load Monitoring. In 2024 36th Chinese Control and Decision Conference (CCDC) (pp. 6173-6178). IEEE.

[25] Lu, L., Kang, J.S. and Yu, M., 2024. Event Detection Based on Robust Random Cut Forest Algorithm for Non-Intrusive Load Monitoring. Journal of Modern Power Systems and Clean Energy.

[26] Zhou, M., Shao, S., Wang, X., Zhu, Z. and Hu, F., 2022. Deep learning-based non-intrusive commercial load monitoring. Sensors, 22(14), p.5250.

[27] Zhang, J., Chen, X., Ng, W.W., Lai, C.S. and Lai, L.L., 2019. New appliance detection for nonintrusive load monitoring. IEEE Transactions on Industrial Informatics, 15(8), pp.4819-4829.

[28] Buabeng, A., Simons, A., Frempong, N.K. and Ziggah, Y.Y., 2021. A novel hybrid predictive maintenance model based on clustering, smote and multi-layer perceptron neural network optimised with grey wolf algorithm. SN Applied Sciences, 3(5), p.593.

[29] Zhu, R., Wang, Z., Ma, Z., Wang, G. and Xue, J.H., 2018. LRID: A new metric of multi-class imbalance degree based on likelihood-ratio test. Pattern Recognition Letters, 116, pp.36-42.