Enhancing Predictive Performance of Non-Intrusive Load Monitoring Through Systematic Feature Extraction*

[1]N. K. Mensah, [1]S. Nunoo, [1]H. Abdel-Fatao, and [1]Y. Y. Ziggah
[1]University of Mines and Technology, Tarkwa, Ghana

Mensah, N. K., Nunoo, S., Abdel-Fatao, H. and Ziggah, Y. Y. (2023), "Enhancing Predictive Performance of Non-Intrusive Load Monitoring through Systematic Feature Selection", *Ghana Journal of Technology*, Vol. 7, No. 2, pp. 35-39

Abstract

Non-Intrusive Load Monitoring (NILM) is a highly effective method for maximizing energy efficiency by analysing recorded voltage and current measurements to determine appliance-level electricity consumption. Real-time power consumption information provided by NILM enables consumers to make informed decisions to save energy and resources. However, the challenge lies in extracting meaningful features for accurate appliance classification, and existing research in this area is limited. To address this gap, this study focuses on enhancing the predictive performance of NILM through the combination of various electrical features. A dataset derived from Intrusive Load Monitoring (ILM) is utilised, with emphasis placed on selecting the most significant electrical characteristics. Two experiments are conducted, with the first employing only the root mean square current (IRMS) as a feature and the second incorporating six electrical characteristics. Six machine learning classification algorithms are applied to each experiment, and their results are compared in terms of accuracy, precision, recall, and F-measure. The findings demonstrate that utilizing the six extracted features, including current, voltage, and day section, outperforms the standalone IRMS feature. This comparative analysis highlights the effectiveness of these six feature sets for NILM in achieving improved classification accuracy. In conclusion, this study emphasizes the importance of feature extraction in NILM and provides evidence of the superior performance obtained by incorporating multiple electrical features. The results contribute to understanding efficient appliance classification for NILM, enabling enhanced energy management and conservation. Future research may explore additional datasets and advanced techniques to further optimize appliance classification in NILM systems.

Keywords: Non-Intrusive Load Monitoring, Energy Consumption, Feature Extraction, Machine Learning, Time Series Analysis

1 Introduction

The industrial sector is known for its high energy consumption, accounting for approximately 40% of the total. To provide a more specific perspective, the construction industry alone uses about 4% of the world's energy output and contributes significantly to energy-related greenhouse gas emissions, amounting to one-third (Robert and Kummert, 2012). Moreover, the residential sector, particularly, is estimated to be responsible for 27% of global energy consumption and 17% of CO2 emissions (Nejat et al., 2015). Considering these statistics, it's projected that energy demand will double by the year 2050 (Larcher and Tarascon, 2015; Xing et al., 2011; Vickers, 2017). To address both current and future energy requirements, it becomes imperative to embrace solutions that promote the construction of more energy-efficient buildings and enhance energy utilization in existing structures.

According to the International Electrotechnical Commission (IEC), the most significant component in overcoming energy challenges will be the intelligent and cost-effective use of electricity as the principal energy source (Gungor *et al*., 2013). Over the years, studies have been conducted to ascertain how energy can be utilized more efficiently in workplaces, buildings, and households. Currently, there are two main strategies commonly employed to enhance energy efficiency within the construction sector: Intrusive Load Monitoring (ILM) and Non-Intrusive Load Monitoring (NILM). While the ILM approach has demonstrated its effectiveness, it is associated with high costs and time requirements. Conversely, NILM offers a more promising and practical solution. This is primarily because NILM leverages modern monitoring systems that influence user behavior, encouraging more responsible energy consumption practices (Gopinath et al., 2020).

Despite advancements that have been made in designing more energy-efficient buildings, efficient consumption of energy in residential settings remains a challenge (Himeur *et al*., 2022). This can largely be attributed to the fact that consumers lack a readily available frame of reference to ascertain the amount of power consumed by individual appliances in their homes which invariably has a direct impact on their overall power consumption. The most common form of power consumption feedback for most residential customers is a monthly power bill. Studies have shown that supplying customers with real-time power use data, on an aggregate basis, can help them adjust their behavior and save 10-15% on energy expenses (Rode, 2021). Moreover, access to disaggregated power consumption data can further enable users to save better.

The evolution of electricity systems towards energy conservation and sustainability is driven by three

**Manuscript received June 7, 2023*
Revised version accepted September 25, 2023

fundamental concepts: a) Decarbonization: This concept signifies the shift from a centralized power generation model to a distributed one, where numerous users simultaneously act as consumers and producers. This shift is facilitated by the increasing integration of Renewable Energy Sources (RES) into the grid. The outcome is a more environmentally friendly approach to power generation., b) Decentralisation: This term characterizes the transition from a centralised power generation system to a distributed one, where a diverse range of users play dual roles as consumers and producers of electricity. and c) Digitalisation of the grid: In this phase, the grid is enhanced with smart devices and IoT (Internet of Things) technology. These technologies enable comprehensive monitoring and control of the grid, resulting in the generation of vast amounts of data. This data can then be harnessed by AI algorithms for various purposes, further optimising the electricity system.

NILM stands to gain significant advantages from the ongoing digitalisation of the grid, offering a solution that enhances energy efficiency and serves the interests of both consumers and utility providers. NILM, encompasses a range of techniques designed to evaluate the electrical power consumption of individual appliances. This assessment is achieved by collecting current and/or voltage measurements from a limited number of points within a building's power distribution system. Through this approach, NILM contributes to a more efficient and informed management of energy consumption, benefiting both end-users and utility companies. As a result, NILM can provide consumers with real-time power consumption feedback, allowing them to make better decisions that save resources and money. Power companies, appliance manufacturers, and other parties could utilise the data to improve electricity usage efficiency and better understand how electricity is used.

The stated advantages of the NILM present it as a viable alternative to ILM. ILM is exemplified in Fig. 1 (Kahl, 2019) measuring units on each relevant appliance (Hart, 1992), whereas NILM shown in Fig. 2 (Kahl, 2019) aims for a single intelligent sensor at the aggregated signal, typically at the electric cabinet based on appliance-specific properties in the current and voltage data. The intelligent sensor is typically equipped with state-of-the-art artificial intelligence to recognize appliance statuses, classes, and consumption in real or near-real-time. Disaggregating energy from a single-point meter via Artificial Intelligence (AI) algorithms is a low-cost option that may be put into the user's residence with minimal external interference.

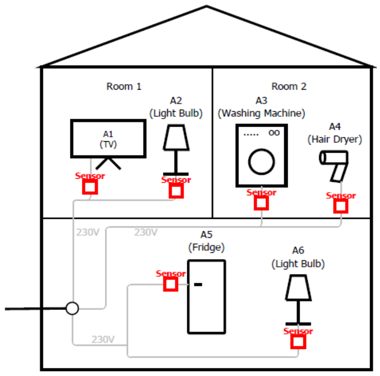

Fig. 1 Intrusive Load Monitoring

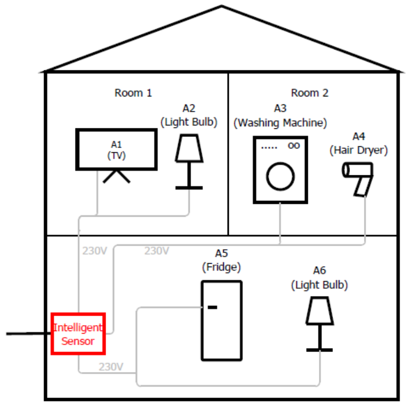

Fig. 2 Non-Intrusive Load Monitoring

NILM, which was initially introduced by Hart (1991), involves a technique for disaggregating electrical loads. This is achieved by analyzing the distinctive power consumption patterns associated with individual appliances within the overall aggregated data (Hamid, 2017). The mathematical definition of this problem is as follows:

$$P(t) = \sum_{j=1}^{L} p_j(t) + e(t), \tag{1}$$

where P_j denotes the power consumption of the j_{th} individual appliance at the time, *t*, *L* denotes the total number of appliances, $e(t)$ denotes noise, and *P(t)* denotes the aggregated consumption measured in the main electrical panel at time, *t*. The term load signature is employed to refer to the unique energy consumption of each electrical appliance.

Hart (1992) conducted research to show how different electrical appliances produce different power signatures, such as active power, current, and voltage. The research demonstrated how on-off events may be used to describe the use of particular appliances. The NILM model has since been improved by several researchers who have explored this concept.

Feature extraction, which uses signal processing techniques to extract features from voltage (V) and current (I) measurements, is a critical stage in

NILM. The ultimate purpose of the feature extraction step is to create a signature (using a feature or a collection of characteristics) that can uniquely identify each device. The distinctiveness of the appliance signature compared to that of other devices determines the performance of any NILM system. As a result, identifying such a signature is critical for increasing the load discrimination capability of the NILM system. However, developing an effective algorithm to recognize and measure the power consumed by each device in the circuits is one of the key obstacles.

Previous studies have employed either steady-state analytic approaches (Hart, 1992; Drenker and Kader, 1999; Marchiori, 2010; Dong *et al*., 2012; He *et al*., 2012) or transient methods (Leeb *et al*., 1995; Cole and Albicki, 1998; Shaw *et al*., 2008; Wang and Zheng, 2011; Chang *et al.*, 2011) to study this topic. The active/reactive power is measured and then utilised to classify the loads using Steady-state analysis methodologies. Transient analysis, on the other hand, extracts information from the transient following the switching of the loads and uses them to identify the operating appliances.

The major disadvantage of transient analysis is that it necessitates a high sampling rate and a big memory size in order to collect signal sequences. These criteria inevitably add to the complexity of the metering system, as well as its expenses. Steady-state analysis, on the other hand, does not necessitate a high sampling rate or a big memory size, making it a particularly practical approach for low-cost monitoring systems.

Despite the fact that NILM has been the focus of research for over two decades, there has yet to be a systematic selection of the different electrical properties proposed for successful load discrimination. One of the major challenges in NILM is determining the most meaningful set of electrical parameters to classify appliances (Patel *et al*., 2007).

A review of other works reveals that most works only extract voltage and current to classify the appliances. The challenge here is that the voltage and current are very few features and thus may not give a real scenario of the nature of the appliances classified. Additionally, two or more appliances may have the same current which can lead to a misrepresentation of the actual energy consumption by the individual appliances. A selection of more features is the more efficient approach to achieving very competitive classification results. This study, therefore, seeks to overcome the limitation of other works by considering six electrical features including the day section (the number of hours of the day the appliance is utilised or the duration in which the appliance is used). Root mean square of current used, root mean square of voltage used, maximum power, minimum power, power factor and day section. The main contributions of this paper are to:

1) recommend a set of electrical features which can improve the predictive performance of machine learning algorithms in NILM; and

2) model five (5) Machine Learning algorithms using the extracted electrical features to classify the appliances.

The rest of the paper is laid out as follows: Section 2 examines the most recent research in the field of NILM. The proposed method is examined in Section 3, which includes details on the data, features, and predictive algorithm used. Section 4 has the conclusion and recommendations, which is followed by a detailed discussion in Section 5. In Section 6, conclusions are drawn and future work directions are suggested.

1.1 Literature Review

This section summarizes the most significant research initiatives and related solutions that have been presented to assist with appliance feature extraction and classification.

In the past few decades, NILM has gained significant attention in the field of energy efficiency due to its ability to provide real-time information on energy consumption. The NILM process involves extracting a set of features from recorded voltage and current measurements, followed by applying a classification algorithm to determine the electrical consumption of appliances. While the classification of appliances is a critical step in the NILM process, there has not been sufficient research on the systematic extraction of the most meaningful feature set for appliance classification.

The literature on feature extraction and classification of appliances for NILM is vast. One of the earliest works in this area was conducted by Hart (1992), who laid the foundation for present-day active studies into feature extraction techniques and disaggregation models related to NILM applications.

Yan *et al*. (2019) proposed a Bayesian-based classification approach to detect loads from power signals, while Dong *et al*. (2012) developed an event-based method for decoupling loads from power signals. Machlev et al. (2019) and Ghosh et al. (2020) put forward novel objective functions aimed at modeling voltage, current, and power signals. They also introduced innovative optimization techniques, including non-dominated sorting genetic algorithm II, and an artificial bee colony optimization-based approach for the purpose of load classification.

Numerous studies have employed various Machine Learning (ML) techniques to classify loads based on extracted power, voltage, and current features. For instance, Ghosh *et al.* (2019) introduced a load identification process grounded in fuzzy rule-based methods. This approach involved determining the harmonic impedances of residential loads using voltage and current signals. In a similar vein, Sadeghianpourhamami *et al.* (2017) identified home loads using a recursive feature elimination method. They followed this by employing a Random Forest (RF) classifier based on the most effective steady-state and transient features derived from current and power signals. Liu *et al.* (2019) adopted yet another strategy, utilising an RF classifier. Their approach involved the extraction of 45 time-series features from voltage and current signals for load classification.

In Hassan *et al*. (2013), the research team utilized six distinct waveshape features obtained from Voltage-Current (V-I) patterns of residential loads. They applied a range of classification algorithms, including feed-forward Artificial Neural Networks (ANN), hybrid ANN-DE (Differential Evolution), Support Vector Machine (SVM), and adaptive boost algorithms to categorize these loads. Meanwhile, Chang *et al*. (2015) adopted a feature extraction method based on the Hellinger distance algorithm. They then employed a particle swarm optimization-tuned Artificial Neural Network (ANN) classifier to classify loads based on these features. Du *et al*. (2015) pursued a different approach by mapping V-I trajectories to binary grid cells. Afterward, they identified loads using extracted graphical signatures and implemented a Support Vector Machine (SVM) classifier for the classification task. Gulati et al. (2016) introduced a load classification method relying on the K-Nearest Neighbor (KNN) algorithm. They extracted features from radio frequency inference signals using an eight-fit Gaussian mixture model and a k-peak finder to aid in the classification process. In the study by Gillis *et al.* (2017), a KNN classifier was employed for load detection. They extracted features from current signals using a novel set of higher-order orthogonal wavelets, contributing to their unique approach to load classification.

Recent advancements in Deep Learning (DL) techniques for large-scale data analysis have spurred the development of load identification strategies that make use of various DL methods. For instance, Quek *et al*. (2019) employed a 1-D convolutional stacked long short-term memory (LSTM) recurrent neural network to recognize low-voltage DC loads. Liu *et al*. (2018) took a unique approach by classifying loads based on images depicting their Voltage-Current (V-I) trajectories. They utilized transfer learning from a pre-trained AlexNet Convolutional Neural Network (CNN) for this task. In another study by Liu *et al.* (2021), a sequence-to-point learning approach in CNN, based on transfer learning, was applied. In this method, the CNN module was initially trained on a single appliance and then used to detect additional appliances within the same and different domains. De Baets *et al.* (2018) employed a six-layer CNN model to identify loads based on their V-I trajectory images. Kong *et al.* (2019) utilized a multi-layered CNN module to detect loads by analyzing energy consumption patterns. These advancements reflect the increasing use of DL techniques in load identification research.

While there has been a significant amount of research conducted in the field of NILM, there still remains a lack of research on systematic feature extraction for appliance classification. The selection of appropriate features is crucial to the success of NILM, as they directly affect the performance of the classification algorithms (Angelis *et al*., 2022).
Furthermore, many existing studies focus on using a single feature, such as Root Mean Square Current (I_{rms}), to classify appliances. However, it is widely acknowledged that a combination of multiple features can improve the accuracy and reliability of the classification results. Therefore, this study aims to investigate the effectiveness of a set of six electrical characteristics, including current, voltage, frequency, active power, reactive power, and harmonic distortion, for appliance classification in NILM.

To evaluate the effectiveness of the proposed feature set, two experiments were conducted using a dataset derived from ILM data. In the first experiment, only the I_{rms} feature was used for classification, while in the second experiment, all six electrical characteristics were considered. Six machine-learning classification algorithms were applied to each experiment, and the results were compared.
The findings of this study demonstrate that the use of the six extracted features significantly improves the performance of NILM in terms of accuracy, precision, recall, and F-measure. The comparative analysis performed reveals that these six feature sets are the most efficient features to consider for NILM.

In conclusion, this study contributes to the ongoing research efforts to enhance the performance of NILM. The proposed set of six electrical features provides a more effective approach to feature extraction for appliance classification, which can improve the accuracy and reliability of NILM applications. Furthermore, this study highlights the importance of considering multiple features for appliance classification, rather than relying on a single feature such as I_{rms}

2 Methodology

The proposed methodology consists of four consecutive steps, as illustrated in Fig. 3. This section provides a detailed description of the data processing, outlining the specific steps involved. Additionally, the feature extraction procedure is presented in depth, discussing the specific techniques utilized. Finally, the selection of the most suitable predictive model is justified based on the characteristics of the dataset and research objectives.

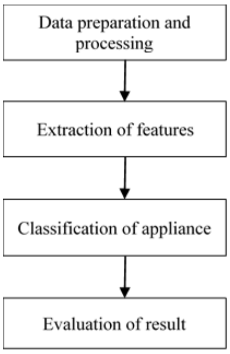

Fig. 3 Block Diagram Showing the Steps in Classifying Appliances Using NILM

2.1 Data Preparation and Processing

The Indraprastha Institute of Information Technology (IAWE) in India provided access to the IAWE data collection. This dataset comprises aggregated and submetered electricity and gas values from 33 residential sensors with a one-second resolution. The data collection contains 73 days worth of observations from a single residence in Delhi, India.

Energy consumption signals for six unique devices from house 1 were picked and preprocessed from the IAWE dataset for the purpose of appliance classification using KNN, Logistic Regression (LR), SVM, Decision Tree (DT), Random Forest (RF), and Naive Bayes (NB). The selected devices were an Air conditioner, Computer, Clothes iron, Washing machine, Television, and Fridge. For each device, its energy consumption signal was cropped into a 30-day window, and any activity that occurred within 60 seconds was used to extract the features employed in this research. The features extracted included RMS Current, RMS Voltage, Maximum Power, Minimum Power, Power Factor, and Day Section.
The sample raw and processed energy consumption data for the Fridge are shown in Fig. 4 and Fig. 5.

2.2 Simulation Parameters

Feature extraction is a critical step in the NILM process, as it involves selecting the most meaningful electrical characteristics from the collected data to be used for appliance classification. There are several methods that can be used for feature extraction, including:

Statistical methods: These methods use statistical measures such as mean, standard deviation, skewness, and kurtosis to extract features from the data. These methods are simple and easy to implement, but they may not capture all the relevant information from the data.

Time-domain methods: These methods use time-domain characteristics such as root mean square, power factor, and harmonic distortion to extract features from the data. These methods are useful for capturing the temporal variations in the data but may not be as effective for capturing the frequency-domain characteristics of the data.

Frequency-domain methods: These methods use frequency-domain characteristics such as the power spectrum and the frequency content of the data to extract features. These methods are useful for capturing the frequency-domain characteristics of the data but may not be as effective for capturing the temporal variations in the data.

Machine learning methods: These methods use machine learning algorithms such as decision trees, neural networks, and support vector machines to extract features from the data. These methods are more complex than the previous methods but can be more effective in extracting relevant features from the data.

In this study, we used a combination of time-domain and frequency-domain methods for feature extraction. We first calculated the current root mean square (I_{rms}) and Voltage Root Mean Square (V_{rms}) from the data, which are time-domain features commonly used in NILM research. The max power (P_{max}), min power (P_{min}), and power factor (P_f), which are frequency-domain features that capture the variations in the power consumption of the appliances, were also calculated. In the final step, the data was divided into six distinct time slots, representing various segments of the day. These time slots include the day section, early morning, morning, noon, evening, night, and late night, allowing for a more detailed and time-specific analysis of the data. This feature captures the consumption pattern of the appliances based on the time of the day, which can be useful for identifying the appliances that are in use during specific times of the day. By using these features, temporal and frequency-domain characteristics of the data were captured, which improved the predictive performance of the NILM technique. Table 1 depicts the consumption-related feature usage identified in this section. In the following section, the features proposed in this paper are elaborated.

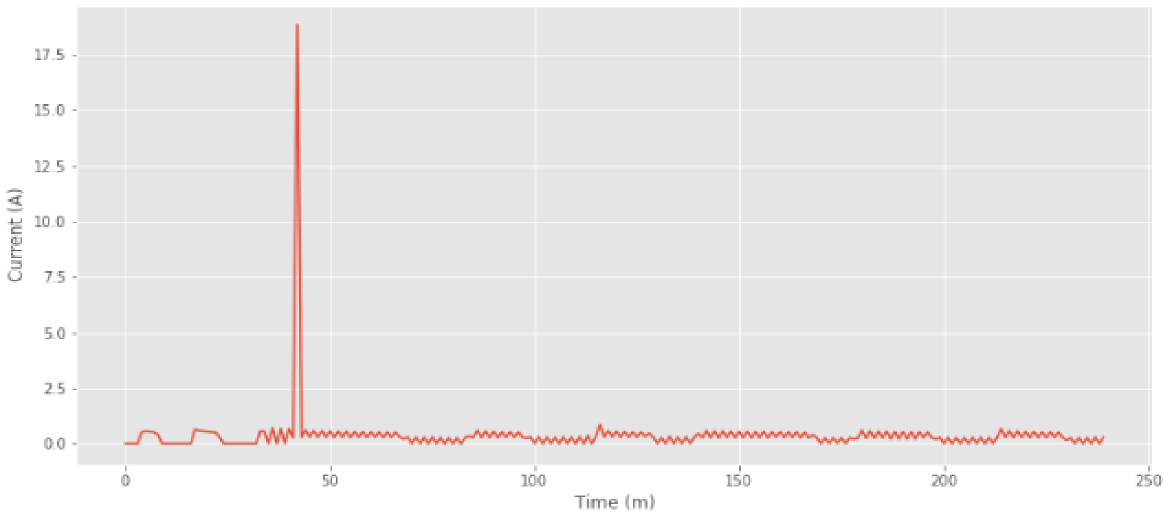

Fig. 4 Raw Data for Fridge

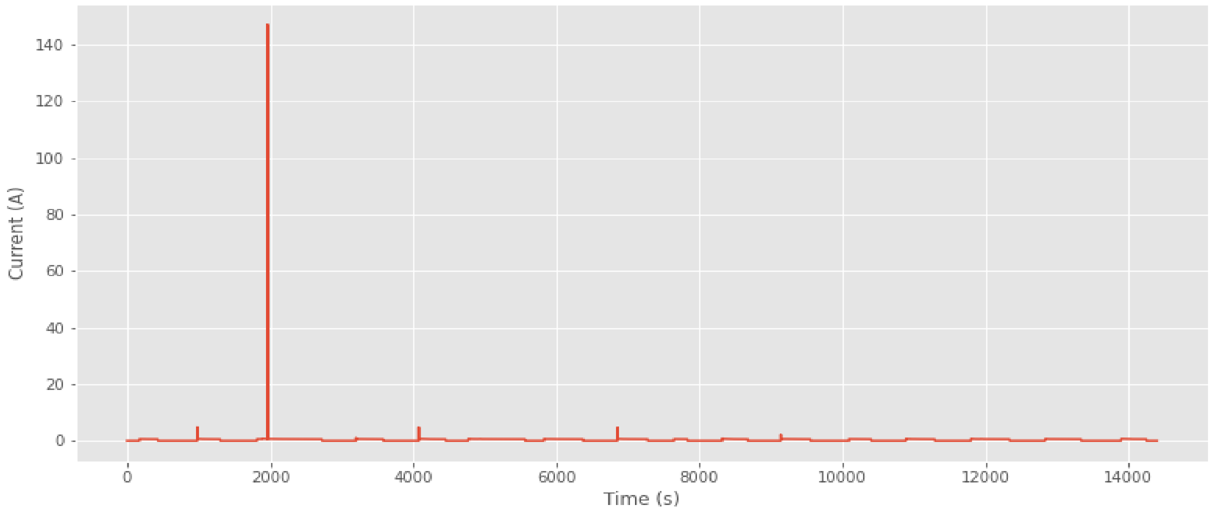

Fig. 5 Processed Data for Fridge

Table 1 Extracted Features

Notation	Description
I_{rms}*	Current root mean square
V_{rms}	Voltage root mean square
P_{max}	Maximum value of the power
P_{min}	Minimum value of the power
D_s	Day section
P_f	Power Factor

The six features used in the study, along with the formula used to calculate them and the reason for their selection are:

1) Current Root Mean Square (I_{rms}): This feature is calculated as the root mean square of the current data over a certain period of time. I_{rms} is a commonly used feature in NILM research as it is a good indicator of the energy consumed by an appliance. By using I_{rms}, we can get a better understanding of how much energy an appliance is consuming at a given time. The formula is:

$$Irms = \sqrt{\frac{1}{T}\sum_{i=1}^{T}(I(i)^2)} \qquad (2)$$

where T is the number of samples and $I(i)$ is the current value at time i.

2) Voltage Root Mean Square (V_{rms}): This feature is calculated as the root mean square of the voltage data over a certain period of time. V_{rms} is a commonly used feature in NILM research as it is a good indicator of the power consumed by an appliance. By using V_{rms}, we can get a better understanding of how much power an appliance is consuming at a given time. The formula is:

$$Vrms = \sqrt{\frac{1}{T}\sum_{i=1}^{T}(V(i)^2)} \qquad (3)$$

where T is the number of samples and $V(i)$ is the voltage value at time i.

3) *Max Power*: This feature is calculated as the maximum power value over a certain period of

time. The max power feature is important as it represents the maximum energy consumption of an appliance. By using max power, we can get a better understanding of how much energy an appliance is consuming at its peak consumption. The formula is:

$$P_{max} = \max(P(i)) \quad (4)$$

where P(i) is the power value at time *i*.

4) Min Power: This feature is calculated as the minimum power value over a certain period of time. The min power feature is important as it represents the minimum energy consumption of an appliance. By using min power, we can get a better understanding of the base consumption of an appliance. The formula is:

$$P_{min} = \min(P(i)) \quad (5)$$

where is the power value at time.

5) Power factor: This feature is calculated as the ratio of real power to apparent power. Power factor is an important feature as it captures the relationship between the real and apparent power, which can be helpful in identifying the nature of the load. By using power factor, we can get a better understanding of how much of the energy consumed by an appliance is being used effectively. The formula is:

$$P_f = \frac{P}{S} \quad (6)$$

where is the real power and is the apparent power.

6) Day Section: This feature is used to partition the data into 6 time slots, which correspond to different times of the day: early morning, morning, noon, evening, night, and late night. This feature is used to capture the consumption pattern of the appliances based on the time of the day. By using day section, we can get a better understanding of when an appliance is in use and how it affects the overall energy consumption. Table 2 describes the Day section and its time range.

Table 2 Extracted Features

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

In this study, different appliances were classified using NILM techniques. The most essential step in the NILM process is feature extraction. Six features (I_{rms}, V_{rms}, P_{max}, P_{min}, P_f, and D_s) were extracted from the IAWE dataset to enhance the predictive performance of the NILM.

After extracting the features, the next step included classifying the different appliances using the extracted features. Six machine learning algorithms were used to classify the appliances. The algorithms used included DT, SVM, KNN, NB and the RF.

2.2.1 Decision Tree (DT)

Each branch of a decision tree can be viewed as an if-then statement. A decision tree is constructed using a hierarchical approach. The branches are generated by dividing the dataset into subsets based on the most significant characteristics. The final classification occurs at the decision tree's leaves. Unlike other algorithms, a decision tree is easy to understand and visualize, requires little data preparation, and handles numerical and categorical data quite well. The hyper-parameters max_depth, min_samples, samples_leaf, max_features assist tuning decision trees for better results.

2.2.2 Random Forest (RF)

Random Forest is a machine learning algorithm that utilizes an ensemble of decision trees. It combines insights from multiple predictors by employing bagging, a technique that involves training each tree on a random sample of the original dataset. This method provides better generalization compared to a single decision tree, but it can be less interpretable due to the model's multiple layers. Common hyperparameters used for tuning Random Forest include n_estimators, max_features, max_depth, min_samples_split, min_samples_leaf, and bootstrap.

2.2.3. Support Vector Machine (SVM)

The Support Vector Machine (SVM) is a machine learning algorithm that identifies the optimal way to classify data based on its position relative to a positive/negative class border. It achieves this by finding a hyperplane that maximizes the margin or distance between data points of different classes. Similar to decision trees and random forests, SVM, also known as the Support Vector Classifier (SVC), can be employed for both classification and regression tasks. One notable advantage of SVM is its memory efficiency. It utilizes a subset of training points, called support vectors, in the decision function, making it suitable for datasets with a large number of samples. Common hyperparameters for tuning SVM include C, kernel, and gamma.

2.2.4. Naive Bayes (NB)

Naive Bayes is founded on Bayes' Theorem, which is a method for calculating conditional probability based on prior knowledge and the naive assumption that each attribute is independent of one another. The greatest advantage of NB is that, unlike the majority of machine learning algorithms, it works relatively well even with minimal amounts of training data. Gaussian Naive Bayes is a classification algorithm based on the normal distribution. The common hyper-parameters are priors, var_smoothing.

2.2.5. K-Nearest Neighbour (KNN)

This method depicts each data point in an n-dimensional space characterized by n attributes. Additionally, it computes distances between points and assigns unknown data labels based on the closest observed data labels. KNN-based classification is a form of lazy learning because it does not seek to build a generic internal model; rather, it just retains examples of the training data. Classification is determined by the simple majority vote of each point's k nearest neighbors. Due to the fact that KNN retains all previous instances and requires examining the complete dataset to classify a new test point, the nominal training but extensive testing duration of KNN incurs equivalent memory and computing costs. This approach is simple, robust to noisy training data, and successful with massive data. The common hyper-parameters are n_neighbours, weights, leaf_size, *P*.

KNN is a supervised learning technique consisting of a given labeled dataset that contains training sets and aims to reflect the correlation between x and y. The purpose of KNN is to discover a function: $h \times y$, such that a new test point x, $h(x)$ may confidently deduce the corresponding output y. In KNN categorization, a new test point is assigned to the category with the highest availability among its k nearest neighbors based on the number of votes cast by its neighbors. If $k = 1$, then a new point is given to the category of its only nearest neighbor.

These algorithms were selected for their popularity and effectiveness in classification tasks (Rafati *et al*., 2022). The hyperparameters for each algorithm will be tuned to achieve the best performance. After applying these algorithms, the best algorithm will be selected based on the performance metrics.

2.3 Fine-Tuning Classification of Appliances Algorithms Through Parameter Optimization

Fine-tuning classification refers to the process of adjusting the parameters of a machine learning model in order to improve its performance on a specific task. Hyperparameters are parameters of a machine learning algorithm that are not learned from data and need to be set before training. Optimizing hyperparameters is important because it can significantly improve the performance of the model. There are different types of optimization methods, such as manual search, random search, and grid search. In this research, we used grid search to optimize the hyperparameters of the classification algorithms. Grid search is a systematic way of exploring a range of hyperparameters by defining a grid of values to search over.

Performing optimization is important because it helps to identify the best combination of hyperparameters for the model, which in turn can improve the accuracy of the classification. In this study, the grid search was adopted to optimize the hyperparameters of the six classification algorithms. The methodology approach involved defining a parameter grid for each algorithm and using GridSearchCV, a function from the scikit-learn library, to perform a grid search. The best combination of hyperparameters was selected based on the highest accuracy score. The results were summarized in terms of the best accuracy score and the best combination of hyperparameters for each algorithm.

Table 3 Machine Learning Algorithm and Related Parameters

Method	Parameter	Value
Random Forest	Splits	5
	No. of Estimators	10
Support Vector	C	100
	Gamma	0.9
	Kernel	RBF
Decision Tree	Criterion	Entropy
	Maximum Depth	8
Gradient Boosting	Learning rate	0.01
	Maximum Depth	3
	No. of Estimators	500
Logistic Regression	C	1000
	Penalty	12
	Solver	lbfgs
KNN	Metric	Manhattan
	No. of neighbours	3

Table 3 presents the machine learning algorithms used in the study along with their corresponding parameters and values that were fine-tuned through optimization. The algorithms and their parameters were selected based on their potential to perform accurate classification of appliances.

For the RF algorithm, the number of splits was set to 5, and the number of estimators was set to 10. For the SVM algorithm, the value of C was set to 100, gamma was set to 0.9, and the kernel was set to RBF. For the DT algorithm, the criterion was set to Entropy, and the maximum depth was set to 8. For the GB algorithm, the learning rate was set to 0.01, the maximum depth was set to 3, and the number of estimators was set to 500. For the LR algorithm, the value of C was set to 1000, the penalty was set to 12, and the solver was set to lbfgs. For the KNN algorithm, the metric was set to Manhattan, and the number of neighbors was set to 3.

In summary, fine-tuning classification through hyperparameter optimization is an important step in improving the performance of machine learning models. Grid search is a commonly used method for hyperparameter optimization because it allows for a systematic exploration of a range of hyperparameters. In this study, we used grid search to optimize the hyperparameters of several classification algorithms, which led to improved accuracy in the classification of appliances.

2.3.1. Evaluation of Results

Although the accuracy metric is often used to evaluate classification performance, it can lead to a misleading conclusion, especially for outlier detection where the minority is typically underrepresented and thus becomes a skewed performance measure for the majority class (Truong *et al*., 2013). Therefore, we are using F-measure and G-mean as classifier performance evaluation metrics.

2.3.2. F-measure and G-mean

Consider a two-class problem labeled TP and TN. As a confusion matrix, Table 4 illustrates the overall performance of the classifier.

Table 4 Confusion Matrix

	Predicted Positive	**Predicted Negative**
Actual Positive	True Positives (TP)	False Negatives (FN)
Actual Negative	False Positives (FP)	True Negatives (TN)

In the confusion matrix in Table 4, TP (true positive) represents a group of correctly identified positive samples, while TN (true negative) represents the group of correctly identified negative samples. On the other hand, FP (false positive) is a collection of positive points that are incorrectly identified as negative, and FN (false negative) is a collection of negative points that are incorrectly detected as positive.

The overall accuracy or error rate can be calculated using the confusion matrix. However, for unbalanced learning tasks, precision and recall are more informative metrics. They can be calculated based on the confusion matrix as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (7)$$

$$Error\ Rate = 1 - Accuracy \quad (8)$$

$$F - measure = \frac{2 \times Precision \times Recall}{TP + FP + TN + FN} \quad (9)$$

$$G - mean = \sqrt{Precision + Recall} \quad (10)$$

where Precision and Recall can be evaluated based on the confusion matrix, which is defined as follows:

$$Precision = \frac{TP}{TP + FP} \quad (11)$$

$$Recall = \frac{TP}{TP + FN} \quad (12)$$

Since the precision value represents the exactness of the classifier and the recall value represents its F-measure and G-mean, which are the mixture of precision and recall, are more comprehensive and extensively employed in unbalanced learning.

2.3.4 Use of F-measure and G-mean

In this research, the focus was on utilizing more comprehensive evaluation metrics, namely F-measure and G-mean, as performance indicators for classifiers, rather than relying solely on accuracy. These metrics were chosen to address the challenges posed by unbalanced learning tasks, where the minority class is typically underrepresented.

By incorporating precision and recall, the F-measure captures the delicate balance between correctly identifying positive samples and minimizing both false positives and false negatives. This metric provides a more nuanced and balanced assessment of the classifier's performance in classification tasks. Additionally, the G-mean metric considers both the true positive rate and the true negative rate, providing a holistic measure of performance that considers both classes. This metric offers a broader perspective on the classifier's effectiveness in handling different class distributions.

Adopting F-measure and G-mean as evaluation metrics ensures a more accurate evaluation of the classifier's performance in addressing class imbalance issues. It allows for a comprehensive understanding of the classifier's effectiveness in the specific classification task at hand.

3 Results and Discussion

This section discusses the results that were obtained.

3.1 Classification of Appliance

Machine learning techniques were employed to classify the appliances used in this study. Since the dataset was a labelled dataset, supervised machine learning algorithms were employed. Extensive review of literature also affirmed the use of supervised machine learning techniques to classify the electrical appliances in this study. This is largely because, supervised machine learning techniques outperformed unsupervised machine learning techniques in the various study conducted by other researchers (Alghawazi *et al.*, 2022). Also, it is possible to analyze and understand the logic behind their classification process, which is, not always possible with other techniques. For example, in neural networks, it is difficult to understand what happens during the classification as well as which features play the most important role for the device identification, since they act as a black box. Six supervised ML algorithms; Support vector Machine, Random Forest, Decision Trees, Gradient Boosting, K-Nearest Neighbour and Logistic Regression were used and their accuracies were compared.

Table 5 Experiment 1 - Classification of Appliances Using Current Root Mean Square

Model	Accuracy	Error Rate	Precision	Recall	F-Measure	G-Mean	Time/sec
KNN	0.9359	0.0641	0.9292	0.929	0.9292	0.9571	252s
Logistic Regression	0.6041	0.3959	0.6205	0.6205	0.6205	0.7572	220.7
Decision Tree	0.9351	0.0649	0.9264	0.926	0.9264	0.9554	1.5s
Random Forest	0.9997	0.0003	0.9994	0.999	0.9994	0.9996	239.2s
SVM	0.9101	0.0311	0.9122	0.902	0.9113	0.9201	14421
Gradient Boosting	0.9247	0.0010	0.9201	0.911	0.9212	0.9510	52211

Table 6 Experiment 2 - Classification of Appliances Using Extracted Features

Classifier	Accuracy	Error Rate	Precision	Recall	F-Measure	G-Mean	Time/sec
KNN	0.9988	0.0012	0.9979	0.9979	0.9979	0.9987	380s
Logistic Regression	0.9942	0.0058	0.9915	0.992	0.9915	0.9949	13.7s
Decision Tree	0.9994	0.0006	0.9994	0.9994	0.9994	0.9996	2.7
Random Forest	0.9204	0.0796	0.9203	0.9203	0.9203	0.9516	207.9
SVM	0.9988	0.0012	0.9973	0.9973	0.9973	0.9984	13464
Gradient Boosting	0.96547	0.0008	0.9773	0.9754	0.9942	0.9522	49902

3.2 Evaluation of Classification Algorithms

In Experiment 1, the classification was based solely on the current root mean square (Irms) feature. The algorithms performances varied, with Random Forest achieving the highest accuracy of 0.9997, while Logistic Regression had the lowest accuracy of 0.6041. It is important to note that Logistic Regression demonstrated a relatively lower performance compared to other algorithms. The precision and recall values ranged from 0.6205 to 0.9994, indicating varying degrees of success in correctly classifying the appliances. Random Forest achieved high precision, recall, F-measure, and G-mean values, suggesting its effectiveness in accurately classifying appliances based on a single feature.

In Experiment 2, multiple extracted features were used for appliance classification. The results showed that utilizing multiple features generally improved the classification performance compared to using only the Irms feature. Decision Tree achieved the highest accuracy of 0.9994, while Logistic Regression had an improved accuracy of 0.9942 in this experiment. The precision and recall values were high across the algorithms, ranging from 0.9915 to 0.9994, indicating their effectiveness in accurately classifying appliances using multiple features. The F-measure and G-mean values further supported the algorithms' overall performance.

Comparing the two experiments, it is evident that using multiple features led to improved classification accuracy in most cases. This highlights the importance of considering multiple aspects of the electrical signal when attempting to classify appliances accurately. While Experiment 1 relied solely on the Irms feature, Experiment 2 incorporated additional features, resulting in enhanced performance.

However, it is important to note that the choice of algorithm also played a significant role in the classification performance. Different algorithms showed varying degrees of success, emphasizing the need to carefully select the appropriate algorithm based on the specific characteristics of the dataset and the appliances being classified.

Furthermore, it is worth considering the computational time required by the algorithms. Experiment 2 generally exhibited longer execution times compared to Experiment 1 due to the increased complexity of utilizing multiple features. This aspect should be considered when deploying these classification models in real-time scenarios or with large datasets.

In conclusion, the combination of Experiment 1 and Experiment 2 demonstrated the effectiveness of utilizing multiple features for appliance classification. The results indicated that incorporating multiple features generally improved the accuracy, precision, recall, F-measure, and G-mean values compared to relying on a single feature.

The choice of algorithm remains crucial in achieving optimal classification performance. These findings contribute to the understanding of machine learning techniques for appliance classification and can guide the selection of appropriate approaches for similar tasks in the future.

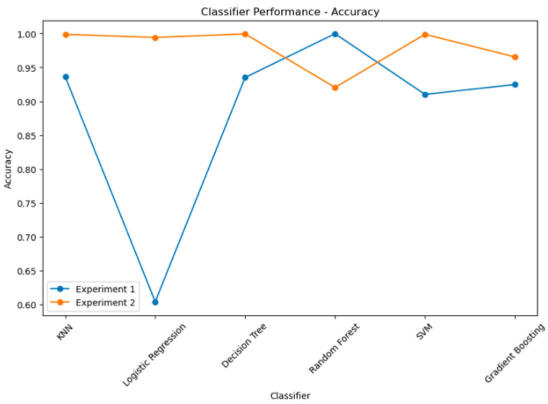

Fig. 6 Classification Accuracy of Appliances in Experiment 1 and 2

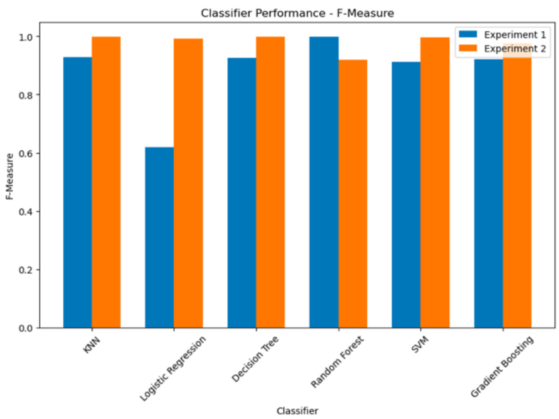

Fig. 7 F-Measure of Appliances Classification in Experiment 1 and 2

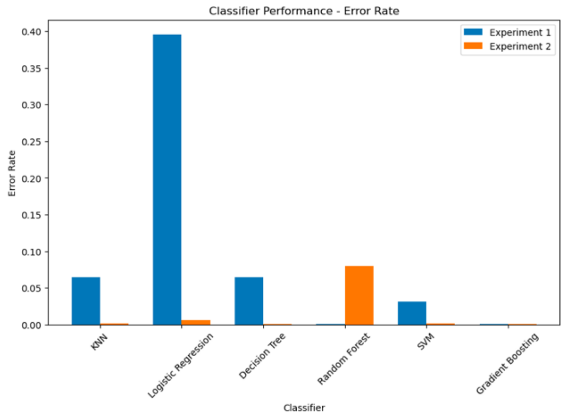

Fig. 8 Error Rate of Appliances Classification in Experiment 1 and 2

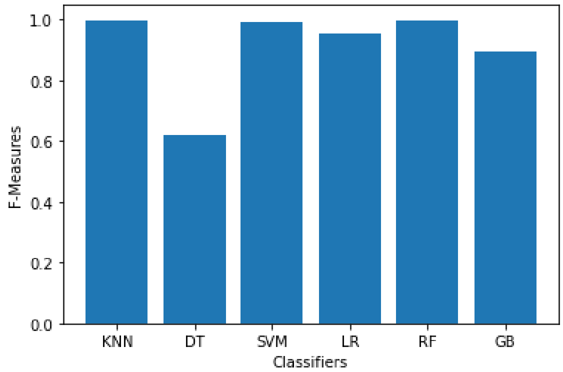

Fig. 9 Recall of Appliances Classification in Experiment 1 and 2

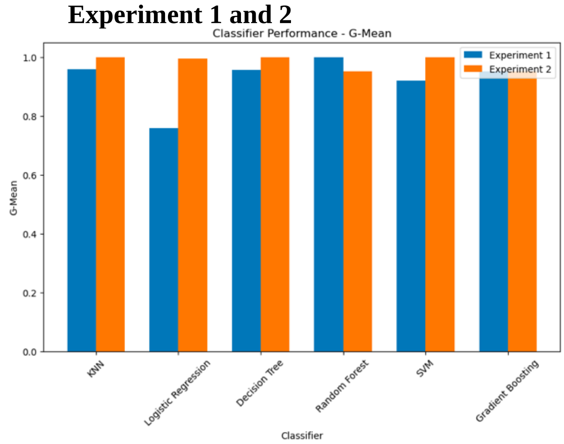

Fig. 10 G-Mean of Appliances Classification in Experiment 1 and 2

4 Conclusions and Recommendations

This research aimed to enhance the predictive performance of Non-Intrusive Load Monitoring (NILM) through the combination of various electrical features using the IAWE dataset. Two experiments were conducted, with Experiment 1 utilizing only the current root mean square (Irms) feature, and Experiment 2 incorporating six electrical characteristics: Irms, voltage root mean square (Vrms), max power, min power, power factor, and day section. The performance of six machine learning classification algorithms was evaluated and compared in both experiments.

4.1. Conclusions

The results indicated that the application of the six extracted features significantly outperformed the standalone use of Irms in terms of accuracy, precision, recall, and F-measure. Specifically, K-Nearest Neighbors (KNN), Decision Tree (DT), and Random Forest (RF) classifiers demonstrated superior performance compared to Logistic Regression, Support Vector Machine (SVM), and Naive Bayes.

Moreover, the KNN classifier consistently achieved the highest overall accuracy and F-measure in both experiments, suggesting its effectiveness for NILM applications. When all six features were utilized, the overall accuracy and F-measure were further improved, underscoring the importance of feature extraction in NILM.

4.2. Recommendations

Based on the findings of this research, several recommendations can be made:

1) Feature Extraction: Future NILM studies should consider incorporating the six electrical features (Irms, Vrms, max power, min power, power factor, and day section) for feature extraction. This comprehensive set of features has demonstrated superior performance in accurately classifying appliances.

2) Classifier Selection: The KNN classifier exhibited the best overall performance in terms of accuracy and F-measure. Therefore, it is recommended to utilize the KNN classifier for NILM applications. However, further experimentation and comparison with other classifiers can be explored to determine the most suitable classifier for specific scenarios.

3) Performance Evaluation: Future research should focus on evaluating additional machine learning algorithms and advanced techniques to further improve the performance of appliance classification in NILM. Techniques such as deep learning approaches may provide valuable insights and enhance accuracy.

4) Dataset Expansion: To validate the findings and generalize the results, it is recommended to incorporate additional datasets from diverse sources and geographical locations. This would enable a more comprehensive evaluation of the proposed methodology and its applicability in different settings.

Overall, this study highlights the significance of feature extraction and classifier selection in improving the performance of NILM. By adopting the recommended approaches, practitioners and researchers can achieve more accurate appliance classification, leading to efficient energy management and conservation.

References

Alghawazi, M., Alghazzawi, D. and Alarifi, S. (2022), “Detection of sql injection attack using machine learning techniques: a systematic literature review”, *Journal of Cybersecurity and Privacy*, Vol. 2, No. 4, pp. 764 - 777.

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*, p. 111951.

Chang, H. H, Chen, K. L., Tsai, Y. P. and Lee, W. J. (2011), “A New Measurement Method for Power Signatures of Non-Intrusive Demand Monitoring and Load Identification”, *IEEE Transactions on Industry Applications*, Vol. 48, pp. 764 – 771.

Chang, H. H., Lee, M. C., Chen, N., Chien, C. L. and Lee, W. J. (2015), “Feature Extraction Based Hellinger Distance Algorithm for Non-Intrusive Aging Load Identification in Residential Buildings”, *IEEE Transactions on Industry Applications*, pp. 1–8.

Cole, A. I. and Albicki, A. (1998), “Data Extraction for Effective Non-Intrusive Identification of Residential Power Loads” IMTC/98 Conference Proceedings. *IEEE Instrumentation and Measurement Technology Conference*. Where Instrumentation is Going (Cat. No. 98CH3622-2), Vol. 2, pp. 812 – 815.

De Baets, L., Ruyssinck, J., Develder, C., Dhaene, T. and Deschrijver, D. (2018), “Appliance Classification using VI Trajectories and Convolutional Neural Networks”, *Energy and Buildings*, Vol. 158, pp. 32–36.

Dong, M., Meira, P. C., Xu, W. and Freitas, W. (2012), “An Event Window Based Load Monitoring Technique for Smart Meters”, *IEEE Transactions on Smart Grid*, Vol. 3, pp. 787 – 796.

Drenker, S. and Kader, A. (1999), “Non-Intrusive Monitoring of Electric Loads”, *IEEE Computer Applications in Power*, Vol. 12, pp. 47 – 51.

Du, L., He, D., Harley, R. G., and Habetler, T. G. (2015), “Electric Load Classification by Binary Voltage–Current Trajectory Mapping”, *IEEE Transactions on Smart Grid*, Vol. 7, pp. 358 – 365.

Ghosh, S., Chatterjee, A and Chatterjee D. (2019), Improved Non-Intrusive Identification Technique of Electrical Appliances for a Smart Residential System”. *IET Generation, Transmission & Distribution*, Vol. 13, pp. 695 – 702.

Ghosh, S., Manna, D., Chatterjee, A. and Chatterjee, D. (2020), “Remote Appliance Load Monitoring and Identification in a Modern Residential

System with Smart Meter Data", IEEE *Sensors Journal,* Vol. 21, pp. 5082 – 5090.

Gillis, J. M., Chung, J. A. and Morsi, W. G. (2017) "Designing New Orthogonal High-Order Wavelets for Nonintrusive Load Monitoring", *IEEE Transactions on Industrial Electronics,* Vol. 65, pp. 2578 – 2589.

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,* Vol. 62, pp. 102411.

Gulati, M., Singh, V. K., Agarwal, S. K. and Bohara, V. A. (2016), "Appliance Activity Recognition using Radio Frequency Interference Emissions", *IEEE Sensors Journal,* Vol. 16, pp. 6197 – 6204.

Gungor V. Cagri, Sahin Dilan, Kocak Taskin, Ergut Salih, Buccella Concettina, Cecati Carlo and Hancke Gerhard P. (2013), "A Survey on Smart Grid Potential Applications and Communication Requirements," *IEEE Transactions on Industrial Informatics*, vol. 9, no. 1, pp. 28–42.

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 International Conference on Technologies and Materials for Renewable Energy, Environment and Sustainability,* Vol. 119, pp. 742 - 751.

Hart G. W. (1992), "Nonintrusive Appliance Load Monitoring", *Proceedings of the IEEE*. Vol. 80, pp. 1870 – 1891.

Hassan, T., Javed, F. and Arshad, N. (2013), "An Empirical Investigation of VI Trajectory Based Load Signatures for Non-Intrusive Load Monitoring", *IEEE Transactions on Smart Grid.*, Vol. 5, pp. 870–878.

He, D., Du, L., Yang, Y., Harley, R. and Habetler, T. (2012), "Front End Electronic Circuit Topology Analysis for Model-Driven Classification And Monitoring of Appliance Loads in Smart Buildings", *IEEE Transactions on Smart Grid*, Vol. 3, pp. 2286–2293.

Himeur, Y., Alsalemi, A., Bensaali, F., Amira, A. and Al‐Kababji, A., 2022. "Recent trends of smart nonintrusive load monitoring in buildings: A review, open challenges, and future directions", *International Journal of Intelligent Systems*, 37 (10), pp. 7124–7179.

Kahl, M. (2019), "Machine Learning for Non-Intrusive Load Monitoring", *Ph.D. thesis,* Techn-ische Universität München, Vol. 196 pp. 112–123.

Kong, W., Dong, Z. Y., Wang, B., Zhao, J. and Huang, J. (2019), "A Practical Solution for Non-Intrusive Type II Load Monitoring Based on Deep Learning and Post-processing", *IEEE Transactions on Smart* Grid, Vol. 148–160, pp 11.

Larcher, D. and Tarascon, J. M. (2015), "Towards Greener and More Sustainable Batteries for Electrical Energy Storage", *Nature Chemistry, Vol 7, pp 19 – 29.*

Leeb, S. B., Shaw, S. R. and Kirtley, J. L. (1995), Transient Event Detection in Spectral Envelope Estimates for Non-Intrusive Load Monitoring", *IEEE Transactions On Power Delivery,* Vol. 10, pp. 1200–1210.

Liu, H., Wu, H. and Yu, C. (2019), "A Hybrid Model for Appliance Classification Based on Time Series Features, *Energy and Buildings*, Vol. 196 pp. 112–123.

Liu, Y., Wang, X., You, W. (2018), "Non-Intrusive Load Monitoring by Voltage–Current Trajectory Enabled Transfer Learning", *IEEE Transactions on Smart Grid,* Vol. 10, pp. 5609–5619.

Liu, Y, Zhong, L, Qiu, J., Lu, J. and Wang W. 2021. Unsupervised Domain Adaptation For Non-Intrusive Load Monitoring Via Adversarial And Joint Adaptation Network, *IEEE Transactions on Industrial Informatics,* Vol. 18 pp. 266–277.

Machlev, R., Belikov, J., Beck, Y. and Levron, Y. (2019), "Mo-NILM: A Multi-Objective Evolutionary Algorithm for NILM Classification", *Energy and Buildings*, Vol. 199, pp. 134–144.

Marchiori, A., Hakkarinen D., Han, Q. and Earle, L. (2010), "Circuit-level Load Monitoring for Household Energy Management, *IEEE Pervasive Computing,* Vol. 10, pp. 40–48.

Nejat, P., Jomehzadeh, F., Taheri, M. M., Gohari, M. and Majid, M. Z. A. (2015), "A Global Review of Energy Consumption, CO_2 Emissions and Policy in The Residential Sector (With An Overview Of The Top Ten Co2 Emitting Countries*)", Renewable and Sustainable Energy Reviews*, Vol 43, pp 843–862.

Patel, S. N., Robertson, T., Kientz, J. A., Reynolds, M. S. and Abowd, G. D. (2007), "At the flick of a switch: Detecting and classifying unique electrical events on the residential power line (nominated for the best paper award)", *In UbiComp 2007: Ubiquitous Computing: 9th International Conference, UbiComp 2007, Innsbruck, Austria, September 16-19, 2007 Proceedings 9*, pp. 271-288.

Quek, Y. T., Woo, W. L. and Logenthiran, T. (2019), Load Disaggregation Using One-Directional Convolutional Stacked Long Short-Term Memory Recurrent Neural Network", *IEEE Systems Journal,* Vol 14, pp. 1395–1404.

Rafati, A., Shaker, H. R. and Ghahghahzadeh, S. (2022), "Fault Detection and Efficiency Assessment for HVAC Systems Using Non-Intrusive Load Monitoring: A Review", Energies 2022, 15 (1), p. 341.

Robert, A. and Kummert, M. (2012), "Designing Net-Zero Energy Buildings for the Future Climate, Not for the Past", *Building and Environment.* Vol. 55, pp. 150–158.

Rode, I. (2021), "Enhancing Residential Energy Conservation: A Review of Financial and Non-Financial Policy Instruments"

Sadeghianpourhamami, N., Ruyssinck, J., Deschrijver, D., Dhaene, T. and Develder, C. (2017), "Comprehensive Feature Selection For Appliance Classification In NILM", *Energy and Buildings,* Vol. 151, pp. 98–106.

Shaw, S. R., Leeb, S. B., Norford, L. K. and Cox, R. W. (2008), "Nonintrusive Load Monitoring and Diagnostics in Power Systems", *IEEE Transactions on Instrumentation and Measurement*. Vol. 57, pp. 1445–1454.

Truong, N. C., McInerney, J., Tran-Thanh, L., Costanza, E. and Ramchurn, S. D. (2013), "Forecasting Multi-Appliance Usage for Smart Home Energy Management".

Vickers, N. J. (2017), "Animal Communication: When I'm Calling You, Will You Answer Too?", *Current Biology*, Vol 27, pp. R713–R715.

Wang Z. and Zheng G. (2011), "Residential Appliances Identification and Monitoring by a Non-Intrusive Method", *IEEE transactions on Smart Grid.* Vol. 3, pp. 80–92.

Xing Y, Hewitt N, Griffiths P. (2011), "Zero Carbon Buildings Refurbishment -A Hierarchical Pathway", *Renewable and Sustainable Energy Reviews*. Vol. 15, pp. 3229–3236.

Yan, D., Jin, Y., Sun, H., Dong, B., Ye, Z. and Li, Z, Yuan, Y. (2019), "Household Appliance Recognition Through a Bayes Classification Model", *Sustainable Cities and Society*, Vol. 46, pp. 101393.

Authors

N. K. Mensah is presently enrolled in a PhD program in Computer Science and Engineering, while concurrently serving as a Postgraduate Assistant in the Department of Computer Science and Engineering at UMaT. He obtained his BSc in Computer Science and Engineering from the University of Mines and Technology. In addition to his academic pursuits, he holds the position of Senior Software Engineer and leverages his expertise to streamline business processes and reduce reliance on manual labor. His research focuses on Non-Intrusive Load Monitoring.

S. Nunoo is an Associate Professor in Electrical Engineering at the University of Mines and Technology, Tarkwa, Ghana. He received the PhD in Electrical Engineering at Universiti Teknologi Malaysia, Skudai in 2015, the MPhil degree in Electrical and Electronic Engineering from the University of Mines and Technology (UMaT), Tarkwa in 2007, and the BSc degree in Electrical Engineering from Western University College of KNUST, now UMaT in 2004. His research interest is energy management and signal processing for wireless communications with emphasis on adaptive filtering and compressive sampling.

H. Abdel-Fatao is a Lecturer in the Computer Science and Engineering Department at the University of Mines and Technology, Tarkwa, Ghana. He holds a PhD in Computer and Information Sciences from the University of South Australia, Adelaide. His research interests include GPS trajectory data mining, artificial intelligence, robotics and embedded computing, software and web development, bio-informatics.

Y. Y. Ziggah is a Lecturer at the Geomatic Engineering Department of the University of Mines and Technology (UMaT). He holds a BSc in Geomatic Engineering from Kwame Nkrumah University of Science and Technology, Kumasi, Ghana. He obtained his Master of Engineering and Doctor of Philosophy degrees in Geodesy and Survey Engineering from China University of Geosciences (Wuhan), P. R. China. His research interests include artificial intelligent application in engineering, geodetic coordinate transformation, gravity field modelling, height systems and geodetic deformation modelling.