Low Complexity Non-Intrusive Load Disaggregation of Air Conditioning Unit and Electric Vehicle Charging

Attique Ur Rehman[1], Tek Tjing Lie[1*], Brice Vallès[2], and Shafiqur R. Tito[3]
[1] Auckland University of Technology, [2]Genesis Energy Ltd., [3]Manukau Institute of Technology
Auckland, New Zealand
*tek.lie@aut.ac.nz

***Abstract*— Energy monitoring is inevitable towards achieving energy efficiency and conservation. Load disaggregation is one of the techniques towards effective energy monitoring. In the said domain, Non-Intrusive Appliance Load Monitoring (NIALM) is an attractive method where aggregated load data are acquired from a single metering point and segregated appliance level load is estimated using effective software techniques. This paper presents a low complexity event-based NIALM technique based on supervised machine learning. In this paper, the emphasis is on the disaggregation of Air Conditioning (AC) unit and Electric Vehicle (EV) charging loads due to their high significance for the overall power grid stability improvement. A comprehensive digital simulation has been carried out to validate the performance of the proposed approach and intended appliances are aptly classified having an outcome of 97% for same Data ID and 95% for different Data ID in terms of precision, recall, and f-score performance metrics.**

Keywords*— Smart Meters, Event Detection, Feature Engineering, Supervised Machine Learning, Non-Intrusive Appliance Load Monitoring

I. INTRODUCTION

Electricity which plays a key role in this era, has gone through a rapid transformation in terms of technologies in the last few decades. Today, with the growing emergence of prosumers[1] and microgrids, the amount of electricity not produced by large traditional power plants is ever increasing[2]. At the same time, high consumption loads like AC units and EV charging, become a significant load element for the existing power grid analysis [1, 2]. In the said scenario, smart grid offers many promising solutions like energy monitoring and demand response towards overall system efficiency and stability.

With the worldwide deployment of smart meters, today it is more viable to monitor and understand the energy consumption and the corresponding human behavior [3]. A comprehensive review of smart meter's applications, methodologies and challenges have been recently presented in [4]. Today, smart meters are a core component of energy monitoring and more specifically of Non-Intrusive Appliance Load Monitoring (NIALM) systems. The latter is used to monitor and acquire the aggregate household load data for further processing towards energy disaggregation. The NIALM system was first presented by Hart [5, 6] and is intended to identify the individual appliances' consumption data from a given information of aggregated load consumption data. Recently numerous NIALM methods have been proposed but in general a traditional NIALM comprises three main components namely, *Data acquisition*, *Feature Extraction,* and *Appliance Classification* [7]. A comprehensive overview of these components is presented in [8-10]. Fig. 1 presents the categorical as well as hierarchical details of a traditional NIALM system.

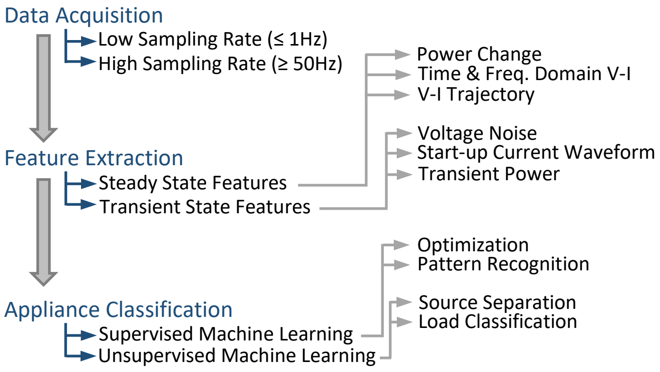

Fig. 1. Non-Intrusive Appliance Load Monitoring System

Further, based on the classification techniques, most of the proposed NIALM methods are categorized to either event-based or non-event based. A comparative overview of these methods is presented in [11]. Recently many NIALM algorithms are proposed based on different approaches including but not limited to Optimization [12], Hidden Markov Models [13], Neural Networks [14], and Machine Learning [9, 15, 16]. Further, in terms of machine learning and energy domain, [17] provides a comprehensive review of different classification algorithms for renewable energy applications including appliance load monitoring.

This paper proposes a low-complexity NIALM approach for significant load elements, i.e., AC and EV charging status inference. The proposed approach relies on event-based low-sampling NIALM system. The rest of the paper is organized as follows: Section II explains the details of the proposed event detection, feature extraction, and supervised machine learning. Section III gives a brief overview of the digital simulation studies and the corresponding results. Finally, the paper is concluded in Section IV.

---

[1] Customers who generate their own electricity, and can also utilize electric vehicles, battery storage etc.

[2] According to International Energy Agency, in 2013 the renewables accounted for around 22% of worldwide electricity generation and it is predicted to reach at least 26% increase in 2020.

II. PROPOSED APPROACH

A. Problem Statement

A time-series power load curve monitored at a metering point can be considered as an algebraic sum of n numbers of appliances' load, as shown in Eq. (1)

$$P_{agg}(t) = \sum_{i=1}^{n} P_i(t) \tag{1}$$

In this paper, $P_i(t)$ is defined in Eq. (2),

$$P_i(t) = P_{AC}(t) + P_{EV}(t) + n(t) \tag{2}$$

where *n(t)* is the measurement noise that is comprised of acquisition noise as well as other appliances' load consumption that are not in the scope of this paper. The task of the proposed NIALM approach is to identify the state of individual appliance load, i.e., $P_{AC}(t)$ and $P_{EV}(t)$ with the only information of aggregated load, i.e., $P_{agg}(t)$.

B. Event Detection and Feature Extraction

The aggregated power level depends on the individual appliances' events and changes whenever an appliance turns ON or OFF. Generally, an event is a portion of a signal that deviates from a previous steady state and lasts till the next steady state [18]. Event detection algorithms are used to detect all the turning ON and OFF events of appliances within the aggregated load consumption data. Recently many event detection algorithms have been proposed [19-21]. For this paper, the event detection algorithm named Mean Sliding Window (MSW) algorithm [22], with some further improvement in terms of post-processing step, has been adopted. The description of the improved MSW algorithm is as follows:

Improved MSW Algorithm

```
Input: Aggregated Load Data
Output: Starting and Ending Time Indices of Detected Events

1. Obtain aggregated load consumption data
2. Process the load data by means of filtering techniques, i.e., median filter
3. Select the sliding window width
4. Compute iteratively the mean of the load consumption curve and the corresponding difference
5. Select the threshold value for event detection
6. Compute the signal representing the steady-state and transient states by mean of pre-selected threshold value.
7. Use derivation function in order to compute edges for the extraction of starting and ending time indices of the detected events
8. Post-processing, i.e., event approval and delay correction due to window width
```

The resulting consecutive starting and ending time instances are linked together to obtain the transient portion of the signal referred as an event (either turning-ON or turning-OFF[3]). Later, features are extracted from each detected event, i.e., transient portion of the pre-processed aggregated load consumption data. The feature set '$\mathcal{F}$' used for classification in later stage of this research study is comprised of geometrical features (transient width 'τ_{width}'), power changes (peak to peak power '$P_{peak\ to\ peak}$'), and statistical features (standard deviation 'σ', variance 'σ^2', and mean value 'μ') as shown in Eq. (3).

$$\mathcal{F} = \{\tau_{width}, P_{peak\ to\ peak}, \sigma,\ \sigma^2,\ \mu,\ \} \tag{3}$$

C. Classification

The extracted features from each of the detected events are used as inputs to the classification stage. In this paper, supervised machine learning has been used to identify the appliances. Supervised machine learning requires a training phase where both the aggregated and labeled ground-truth data are required for training purpose. Numerous supervised machine learning algorithms such as Support Vector Machine (SVM), k-Nearest Neighbor (kNN), Decision Tree, Random Forest, and Regression are available in the literature. This paper adopted kNN for inference purposes because it is one of the simplest and widely used supervised machine learning algorithm. The kNN is a non-parametric method, which remembers the complete training data and performs classification only if the features of the testing data exactly match one of the training data [23].

Three key elements of this method are a set of labeled objects, a distance or similarity metric to compute the distance between objects, and the value of k, i.e., number of nearest neighbors. To identify an unlabeled object, the distance of this object to the labeled objects is computed by identifying its k-nearest neighbors, and the class labels of these nearest neighbors are then used to determine the class label of the object. The kNN is briefly discussed in [23], where a high-level summary of kNN algorithm is as follows [23]:

k-Nearest Neighbor Algorithm

```
INPUT
$D$, set of $k$ training objects, and test object $z = (\boldsymbol{x}', y')$

PROCESS
Compute $d(\boldsymbol{x}', \boldsymbol{x})$, the distance between $z$ and every object, $(\boldsymbol{x}, y) \in D$.
Select $D_z \subseteq D$, the set of $k$ closest training objects to $z$.

OUTPUT
$y' = argmax_v \sum_{(x_i, y_i) \in D_z} I(v = y_i)$
```

where v is a class label, y_i is the class label for the i^{th} nearest neighbors, and $I(.)$ is an indicator function that returns the value 1 if its argument is true and 0 otherwise [23].

III. SIMULATIONS AND RESULTS

To validate the proposed approach, comprehensive digital simulation studies have been carried out. In terms of data acquisition, the aggregated and ground-truth[4] data have been acquired from a real-world dataset known as Dataport[5] [24]. Dataport is the largest energy disaggregation data source owned

---

[3] Appliance turning-ON and turning-OFF are considered as distinct events having their own starting and ending time instances respectively.

[4] Ground-truth or labelled data is mandatory for supervised machine learning.

[5] https://dataport.cloud/

and operated by Pecan Street Inc. Dataport comprises of electricity data of 722 houses in the United States of America that includes 501 single-family homes, 35 town homes, 3 mobile homes, and 183 apartments in different cities [25]. Each household in the Dataport comprises of aggregated average power consumption data along with individual appliances' power consumption data and the corresponding ground-truth power signals respectively. Further details in terms of installed meters along with the installation dates, number of appliances, monitored building construction date, and different house sizes are given in [25].

During this research work, MATLAB® is used as a simulation tool for event detection and feature extraction. In terms of event detection, as previously mentioned, the MSW algorithm is used with some improvement in terms of post-processing steps. The parameters used for the event detection are kept the same as presented in [22] except the input data sample size which has been increased by 15 times to support the effectiveness of the proposed method. The said parameters are presented in Table I.

TABLE I. EVENT DETECTION PARAMETERS

Data Granularity	1/60 Hz
Data ID	26
Data Timeframe	18th June 2014 – 02nd July 2014
No. of Data Samples	21600
Pre-Processing Technique	Median Filtering
Window Width	5 Samples
Threshold Value	250 W

Based on the parameters presented in Table I, simulation studies are carried out resulting in the starting and ending time instances of the detected events within the preprocessed acquired aggregated load data. Fig. 2 presents a portion (for better visualization) of the corresponding results.

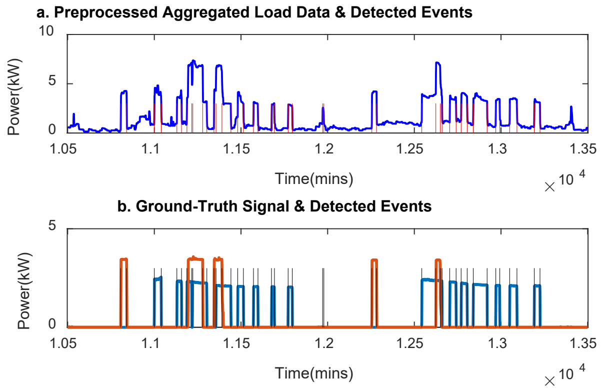

Fig. 2. (a) Preprocessed acquired aggregated load data and corresponding detected events in blue and red color respectively, (b) Ground-truth signal of AC and EV appears in dark cyan and orange color respectively along with the events detected by MSW algorithm appearing in black color

It is obvious from the results presented in Fig. 2 that mostly high consumption peaks in the preprocessed aggregated load data are effectively detected comparatively to the lower variation in the aggregated load data. It is anticipated and essential due to the pre-defined parameters presented in Table I, mainly threshold value and data granularity [26] of 250W and 1/60 Hz respectively. Further, from the presented results, it is clear that the detected events[6] are well aligned with the ground-truth power signals of the appliances, i.e., AC and EV.

For parameters presented in Table I, a total of 323 events were detected by MSW algorithm. For later stage of classification via kNN algorithm, features need to be extracted from the detected event. Here, five different features, as discussed in Eq. (3), are extracted for each detected event that lead to a feature set having 323×5 entries. A fragment of this feature set is presented in Table II.

TABLE II. EXTRACTED FEATURES OF THE DETECTED EVENTS

Indices	τ_{width}	$P_{peak\ to\ peak}$	σ	σ^2	μ	**Labels**
16	2	1.836	1.055	1.114	3.051	1
55	1	-0.832	0.588	0.346	1.56	0
6780	3	3.37	1.583	2.507	5.802	2
6797	5	-4.892	1.950	3.803	4.162	3
21021	3	3.429	1.626	2.646	2.838	2
21045	2	3.039	1.531	2.346	6.093	1
21058	2	-2.852	1.543	2.381	4.998	3
21083	2	3.041	1.753	3.076	6.645	2
21103	2	-2.525	1.377	1.896	5.309	0
21160	3	-1.164	0.495	0.245	1.864	3

The first column of Table II represents the starting time indices of the detected events. The last column presents the labels representing AC and EV turning ON and OFF obtained from the Dataport and will be used later to train the supervised machine learning classifier kNN.

For classification purposes, Scikit-learn which is a machine learning library for Python programming language is used as a simulation tool. Scikit-learn comprises numerous classification, regression, and clustering machine learning algorithms. Within the scope of this paper, kNN is used as a classification algorithm.

The extracted features and label data are the key inputs to the kNN algorithm. The classifier is trained via the aforementioned extracted features and label data. It is noteworthy that due to false positive (event) detection by MSW algorithm, the corresponding features were labeled as undefined NaN values. The said NaN valued features are not taken as an input for classifier training. Once the classifier is trained, it is ready to classify the testing data. In this paper, to validate the results, the trained classifier is tested rigorously on the same as well as different data ID at different timeframes respectively. The details of the training as well as testing data and the corresponding event detection result are presented in Table III.

TABLE III. CLASSIFIER TRAINING AND TESTING DATA PARAMETERS

	Training Data	**Testing Data**	
Data ID	**26**	**26**	**3036**
Timeframe	18.06.2014~ 02.07.2014	01.08.2014~ 04.08.2014	18.06.2014~ 21.06.2014
Duration	15 Days	04 Days	04 Days
No. of Samples	21600	5760	5760
Detected Events (MSW)	323	99	231
Classifier Input Features	315*5	96*5	220*5

---

[6] For the results starting time instances of the detected events are considered because it is the starting time that initiates the event regardless whether an appliance is turning ON or OFF.

The performance of the classification model is evaluated using the well-known performance metrics namely, *Precision*, *Recall,* and *F-Score* [21] as shown in Eq. (4).

$$F\text{-}Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

where,

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

The *Precision* and *Recall* along with the used terminologies of *True Positive*, *False Positive*, and *False Negative* are well defined in [22].

Before classification of the individual appliances' status in each testing data via kNN, the selection of *k*-value is of utmost importance for best classification accuracy. To select the optimal *k*-value for each testing data, elbow method is applied independently for each testing data. Based on the said method, the obtained optimal *k*-values for Data ID 26 and 3036 are 3 and 2 respectively. Fig. 3 presents the corresponding result of elbow method for Data IDs 26 and 3036 in terms of *k*-value vs. misclassification error rate.

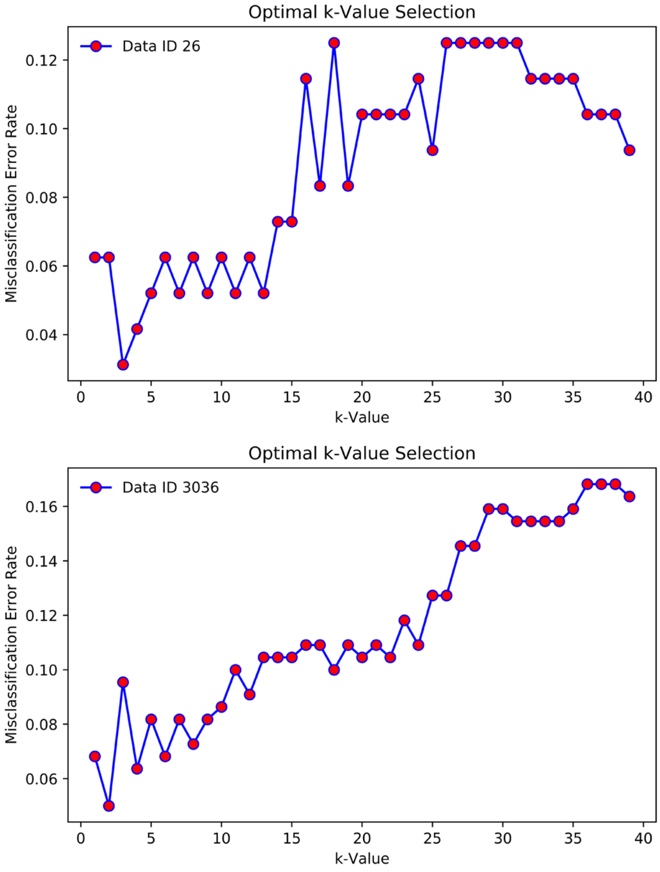

Fig. 3. Selection of Optimal *k*-value for Data ID 26 and 3036

Based on the optimal *k*-values, kNN classifier is used to classify the states of individual appliances, i.e., AC and EV charging for different Data IDs. Due to space contraint, results presented in this paper are limited to only one testing data, i.e., Data ID 26. Fig. 4 presents the confusion matrix in form of a heatmap for individual appliances' predicted status by kNN vs the ground-truth status of the corresponding appliance.

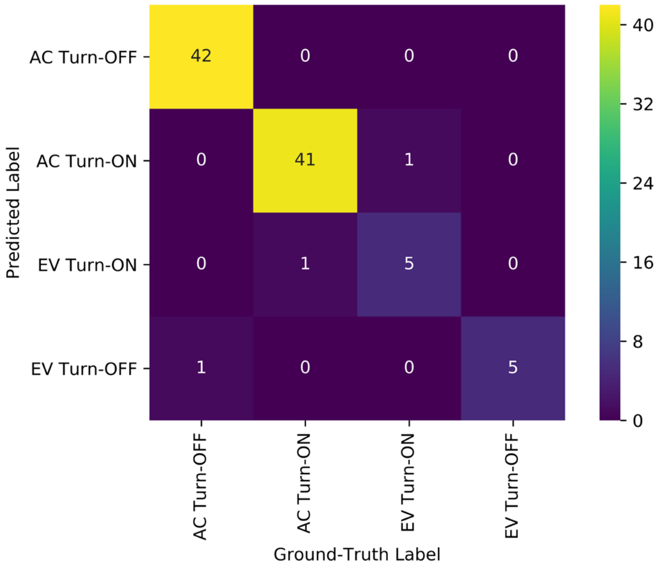

Fig. 4. Confusion Matrix for Data ID 26

The results presented in Fig 4 are further explored in Fig. 5 where the classification of individual appliances' states by kNN vs the ground-truth is presented date-wise.

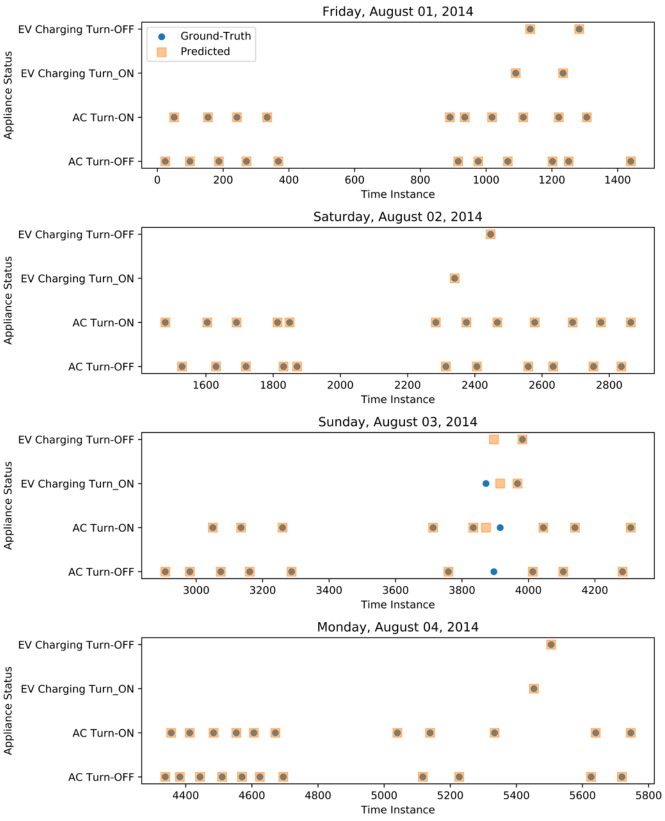

Fig. 5. Date-wise Classification Results for Data ID 26

From Fig. 4, it is observed that most of the appliances' states are classified precisely except a few misclassifications where AC turn-ON and turn-OFF are misclassified as EV turn-ON and OFF respectively. Similarly, an EV turn-ON is predicted incorrectly as an AC turn-ON. It is further concluded from Fig.

5 that all of the aforementioned misclassifications occurred on day three, i.e., Sunday, August 03, 2014.

Likewise, the appliances' states for another testing data, i.e., Data ID 3036 are classified using kNN. Table IV presents the corresponding performance results for individual appliances' states in terms of performance metrics namely, precision, recall, and f-score.

TABLE IV. INDIVIDUAL APPLIANCES' CLASSIFICATION RESULTS

Test Data		**Precision**	**Recall**	**F-Score**
Data ID 26	AC OFF	100 %	98 %	99 %
	AC ON	98 %	98 %	98 %
	EV ON	83 %	83 %	83 %
	EV OFF	83 %	100 %	91 %
Data ID 3036	AC OFF	94 %	96 %	95 %
	AC ON	96 %	95 %	96 %
	EV ON	100 %	75 %	86 %
	EV OFF	75 %	75 %	75 %

From Table IV, it is observed that the classification results regarding appliances' individual status are promising, where AC turn-OFF achieved 100%, 98%, and 99% in terms of precision, recall, and f-score respectively. Further, it is also observed that due to less number of EV turning-ON and turning-OFF events, the classifier is not trained well enough for the said events leading to low classification results relative to classification results of AC. But despite the fact, EV turn-OFF attained 83%, 100%, and 91% in terms of precision, recall, and F-score respectively.

Further, in terms of overall classification results for testing data, i.e,. Data ID 26 and 3036, the results are promising attaining 97% and 95% respectively for all performance metrics, i.e., precision, recall, and f-score. The corresponding results are presented in Table V.

TABLE V. OVERALL CLASSIFICATION RESULTS

	Testing Data ID	
	26	**3036**
Precision	97 %	95 %
Recall	97 %	95 %
F-Score	97 %	95 %

IV. CONCLUSION

This paper introduces a low complexity NIALM approach for AC and EV charging. The proposed NIALM approach is event-based and works well for low sampling data granularity. This paper also proposed improvements of MSW algorithm in terms of post-processing step in improving event detection accuracy. In terms of feature extraction, this paper utilizes a hybrid feature set based on power changes, geometrical, and statistical features of the detected events. Based on the extracted features of detected events, a supervised machine learning namely kNN algorithm (with optimal *k*-values using elbow method) is applied to aptly classify the states of the appliances under consideration. The overall approach is evaluated using different performance metrics and it is concluded that the proposed approach performs well. In future, other supervised machine learning algorithms will be implemented to investigate their performance in the NIALM domain, more specifically to identify the best classifier for low sampling data granularity.

ACKNOWLEDGMENT

The authors gratefully acknowledge the contribution of Callaghan Innovation for promoting the academia-industry linkages and provision of R&D Student-Fellowship Grant. The authors also acknowledge the research facilities provided by Genesis Energy Ltd and Auckland University of Technology, New Zealand.

REFERENCES

[1] L. W. Davis and P. J. Gertler, "Contribution of air conditioning adoption to future energy use under global warming," *Proceedings of the National Academy of Sciences,* p. 201423558, 2015.

[2] K. Clement-Nyns, E. Haesen, and J. Driesen, "The impact of charging plug-in hybrid electric vehicles on a residential distribution grid," *IEEE Transactions on power systems,* vol. 25, no. 1, pp. 371-380, 2010.

[3] K. Perez, W. Cole, M. Baldea, and T. Edgar, "Meters to models: Using smart meter data to predict and control home energy use," in *ACEEE 2014 Summer Study on Energy Efficiency in Buildings*, 2014.

[4] Y. Wang, Q. Chen, T. Hong, and C. Kang, "Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges," *arXiv preprint arXiv:1802.04117,* 2018.

[5] G. W. Hart, "Nonintrusive appliance load monitoring," *Proceedings of the IEEE,* vol. 80, no. 12, pp. 1870-1891, 1992.

[6] G. W. Hart, *Nonintrusive Appliance Load Data Acquisition Method: Progress Report*. MIT Energy Laboratory, 1984.

[7] L. Yu, H. Li, X. Feng, and J. Duan, "Nonintrusive appliance load monitoring for smart homes: Recent advances and future issues," *IEEE Instrumentation & Measurement Magazine,* vol. 19, no. 3, pp. 56-62, 2016.

[8] A. Zoha, A. Gluhak, M. A. Imran, and S. Rajasegarar, "Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey," *Sensors,* vol. 12, no. 12, pp. 16838-16866, 2012.

[9] R. Bonfigli, S. Squartini, M. Fagiani, and F. Piazza, "Unsupervised algorithms for non-intrusive load monitoring: An up-to-date overview," in *Environment and Electrical Engineering (EEEIC), 2015 IEEE 15th International Conference on*, 2015, pp. 1175-1180: IEEE.

[10] C. Klemenjak and P. Goldsborough, "Non-Intrusive Load Monitoring: A Review and Outlook," *arXiv preprint arXiv:1610.01191,* 2016.

[11] Y. F. Wong, Y. A. Şekercioğlu, T. Drummond, and V. S. Wong, "Recent approaches to non-intrusive load monitoring techniques in residential settings," in *Computational Intelligence Applications In Smart Grid (CIASG), 2013 IEEE Symposium on*, 2013, pp. 73-79: IEEE.

[12] D. Egarter and W. Elmenreich, "Load disaggregation with metaheuristic optimization," *Energieinformatik, Karlsruhe, Germany,* 2015.

[13] J. Z. Kolter and T. Jaakkola, "Approximate inference in additive factorial hmms with application to energy disaggregation," in *Artificial Intelligence and Statistics*, 2012, pp. 1472-1482.

[14] D. de Paiva Penha and A. R. G. Castro, "Convolutional neural network applied to the identification of residential equipment in non-intrusive load monitoring systems," in *3rd International Conference on Artificial Intelligence and Applications*, 2017, pp. 11-21.

[15] H. Altrabalsi, V. Stankovic, J. Liao, and L. Stankovic, "Low-complexity energy disaggregation using appliance load modelling," *AIMS Energy,* vol. 4, no. 1, pp. 884-905, 2016.

[16] J. Liao, G. Elafoudi, L. Stankovic, and V. Stankovic, "Power disaggregation for low-sampling rate data," in *2nd International Non-intrusive Appliance Load Monitoring Workshop, Austin, TX*, 2014.

[17] M. Pérez-Ortiz, S. Jiménez-Fernández, P. A. Gutiérrez, E. Alexandre, C. Hervás-Martínez, and S. Salcedo-Sanz, "A review of classification problems and algorithms in renewable energy applications," *Energies,* vol. 9, no. 8, p. 607, 2016.

[18] B. Wild, K. S. Barsim, and B. Yang, "A new unsupervised event detector for non-intrusive load monitoring," in *Signal and Information Processing (GlobalSIP), 2015 IEEE Global Conference on*, 2015, pp. 73-77: IEEE.

[19] M. Liu, J. Yong, X. Wang, and J. Lu, "A new event detection technique for residential load monitoring," in *Harmonics and Quality of Power (ICHQP), 2018 18th International Conference on*, 2018, pp. 1-6: IEEE.

[20] M. N. Meziane, P. Ravier, G. Lamarque, J.-C. Le Bunetel, and Y. Raingeaud, "High accuracy event detection for Non-Intrusive Load

Monitoring," in *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*, 2017, pp. 2452-2456: IEEE.

[21] A. A. Girmay and C. Camarda, "Simple event detection and disaggregation approach for residential energy estimation," in *Workshop on Non-Intrusive Load Monitoring (NILM), 2016 Proceedings of the 3rd International*, 2016.

[22] A. U. Rehman, T. T. Lie, B. Valles, and S. R. Tito, "Low Complexity Event Detection Algorithm for Non-Intrusive Load Monitoring Systems," in *2018 IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia)*, 2018, pp. 746-751: IEEE.

[23] X. Wu *et al.*, "Top 10 algorithms in data mining," *Knowledge and information systems,* vol. 14, no. 1, pp. 1-37, 2008.

[24] "Pecan Street Inc. Dataport 2019,".

[25] O. Parson *et al.*, "Dataport and NILMTK: A building data set designed for non-intrusive load monitoring," in *Signal and Information Processing (GlobalSIP), 2015 IEEE Global Conference on*, 2015, pp. 210-214: IEEE.

[26] G. Elafoudi, L. Stankovic, and V. Stankovic, "Power disaggregation of domestic smart meter readings using dynamic time warping," in *Communications, Control and Signal Processing (ISCCSP), 2014 6th International Symposium on*, 2014, pp. 36-39: IEEE.