

A systematic approach in appliance disaggregation using k-nearest neighbours and naive Bayes classifiers for energy efficiency

Chuan Choong Yang · Chit Siang Soh · Vooi Voon Yap

Received: 20 January 2016 / Accepted: 6 August 2017 / Published online: 21 August 2017
© Springer Science+Business Media B.V. 2017

Abstract One of the ways to achieve energy efficiency in various residential electrical appliances is with energy usage feedback. Research work done showed that with energy usage feedback, behavioural changes by consumers to reduce electricity consumption contribute significantly to energy efficiency in residential energy usage. In order to improve on the appliance-level energy usage feedback, appliance disaggregation or non-intrusive appliance load monitoring (NIALM) methodology is utilized. NIALM is a methodology used to disaggregate total power consumption into individual electrical appliance power usage. In this paper, the electrical signature features from the publicly available REDD data set are extracted by the combination of identifying the ON or OFF events of appliances and goodness-of-fit (GOF) event detection algorithm. The k-nearest neighbours (k-NN) and naive Bayes classifiers are deployed for appliances' classification. It is observed that the size of the training sets effects classification accuracy of the classifiers. The novelty of this paper is a systematic approach of NIALM using few training examples with two generic classifiers (k-NN and naive

Bayes) and one feature (power) with the combination of ON-OFF based approach and GOF technique for event detection. In this work, we demonstrated that the two trained classifiers are able to classify the individual electrical appliances with satisfactory accuracy level in order to improve on the feedback for energy efficiency.

Keywords Energy efficiency · Non-intrusive appliance load monitoring · Few training examples · k-nearest neighbour · Naive Bayes · Goodness-of-fit

Introduction

One of the ways to achieve energy efficiency is with energy usage feedback (Darby 2006; Fischer 2008; Froehlich 2009; Faruqi et al. 2010). Research work done showed that with energy usage feedback, behavioural changes by customers to improve energy efficiency through reducing electricity consumption are from 4 to 12% in the United States of America (Ehrhardt-Martinez et al. 2010). In Europe, researches showed that energy feedback can guide consumers in behavioural change in reducing energy consumption (Behavioural Insights Team 2011; European Environment Agency 2013). In order to improve on the appliance-level energy usage feedback, non-intrusive appliance load monitoring (NIALM) methodology is utilized (Froehlich et al. 2011; Carrie Armel et al. 2013). NIALM disaggregates the sum of power consumption of single point into individual electrical appliance power consumption

C. C. Yang (✉) · C. S. Soh · V. V. Yap
Faculty of Engineering and Green Technology, Universiti Tunku Abdul Rahman, Jalan Universiti, Bandar Barat, 31900 Kampar, Perak, Malaysia
e-mail: yangcc@utar.edu.my

C. S. Soh
e-mail: sohcs@utar.edu.my

V. V. Yap
e-mail: yapvv@utar.edu.my

(Hart 1992; Zoha et al. 2012). The energy or power consumption for individual electrical appliances can be determined from the disaggregated data (Chang et al. 2014).

The NIALM methodology can be implemented as follows:

- Step 1: acquisition of the electrical signature data.
- Step 2: extraction of feature from the signature data and,
- Step 3: classify the appliance either through supervised or unsupervised learning (Zoha et al. 2012; Wong et al. 2013).

After the acquisition of the electrical signature data, the features from these data can be extracted. The electrical appliance features can be categorized into three main types namely steady-state signatures, transient signatures and other types of signatures (Zoha et al. 2012; Wong et al. 2013; Zeifman and Roth 2011). In our proposed work, the steady-state signature data is utilized. The power change methodology is used to analyze the steady-state signature data in order to extract useful features (Hart 1992; Marchiori et al. 2011; Marceau and Zmeureanu 2000). With the power change methodology, the features are the real power and the reactive power. These features are used to track the switching of an electrical appliance operation state, that is, ON or OFF (Hart 1992). George Hart proposed the principle of “Switch Continuity Principle” (Hart 1992) which is applied in this work. This principle states that “in any small enough time interval, the number of appliances which change state to be usually zero, sometimes one and very rarely more than one”. The advantage of power change methodology is that high power load electrical appliances can be easily identified. Furthermore, this methodology requires low sampling rate. However, the low power appliances cannot be easily identified due to the overlapping of real and reactive power in the plot of reactive power versus the real power feature space (Zoha et al. 2012; Yang et al. 2015).

After the features are extracted from the signature data, the event detection is performed. In current NIALM methodology, there are two approaches for event detection, that is, event based or non-event based. In the event-based approach, the edge detection algorithm is utilized to detect events in a power utilization signature. The event detection algorithm tracks the changes of power usage between the switching ON or

OFF conditions to determine whether an event has occurred in steady-state signatures (Wong et al. 2013; Jazizadeh et al. 2014; Anderson et al. 2012; Jin et al. 2011). The non-event-based methodology takes into consideration every data in the total power consumption for deduction (Wong et al. 2013). Examples of non-event based are hidden Markov model (HMM) with various variation of extension of HMM (Wong et al. 2013; Kim et al. 2011; Kolter and Jaakkola 2012; Li et al. 2014; Basu et al. 2015a). The event detection methodology used in this work is the goodness-of-fit (GOF) technique (Jin et al. 2011; Cochran 1952). Researchers had reported good performance from this GOF technique for the event detection algorithm (Yang et al. 2015; Jin et al. 2011; Yang et al. 2014; Cochran 1952; Baets et al. 2017).

The third step in NIALM methodology is to classify the individual appliances from the extracted features and event detection results (Anderson et al. 2012). One of the classification groups is the supervised learning. This technique utilizes an offline training platform to build a data of information for future prediction. Some common supervised learning techniques are artificial neural networks (ANN) (Zoha et al. 2012; Chang et al. 2014; Chang 2012; Chang et al. 2010; Chang et al. 2008; Roos et al. 1994; Yoshimoto et al. 2000; Srinivasan et al. 2006), support vector machines (SVM) (Li et al. 2012; Figueiredo et al. 2011; Lin et al. 2010a; Kato et al. 2009; Patel et al. 2007; Kramer et al. 2012; Zeifman 2012), k-nearest neighbor (k-NN) (Figueiredo et al. 2011; Kramer et al. 2012; Alasalmi et al. 2012; Rahimi et al. 2012; Berges et al. 2011; Berges et al. 2010; Gupta et al. 2010; Saitoh et al. 2010; Berges et al. 2009; Duda et al. 2000; Cover and Hart 1967; Chahine et al. 2011; Giri et al. 2013a; Giri et al. 2013b) and naive Bayes classifier (Marchiori et al. 2011; Chahine et al. 2011; Giri et al. 2013a; Zeifman 2012; Lin et al. 2010b; Marchiori and Han 2009; Barker et al. 2014; Meehan et al. 2014). There is also research work done in supervised learning by utilizing a disaggregation algorithm which minimizes the training by performing the initial model building (Makonin et al. 2015). Another type of supervised learning is the deep neural network which utilizes a lot of training data as the algorithm has many parameters to be trained (Kelly et al. 2015).

On the other hand, the unsupervised learning does not require training before classification starts (Wong et al. 2013; Parson et al. 2014). One of the methods in unsupervised learning is clustering (Yang et al. 2015;

Wang and Zheng 2012; Shao et al. 2012; Goncalves et al. 2011). There is also a research work done by combining a one-time supervised training process over available labeled appliance data sets with an unsupervised training method over unlabeled household aggregate data (Parson et al. 2014). Another type of unsupervised learning is the application of the Hierarchical Dirichlet Process Hidden Semi-Markov Model (HDP-HSMM) factorial structure which provides Bayesian inference and modeling of the duration (Johnson and Willisky 2013).

There is also work done on the semi-supervised learning with very few labeled training examples which utilizes two classifiers separately on two attribute data sets. Each set is independent of the other given the class label (Zhou et al. 2007). There is also work done by Basu et al. (2015b) in comparing both the k-NN classifier with HMM methodology. In this work, the results show that the k-NN classifier with multiple distance-based metrics frequently perform better than the HMM. There was also research work by the same researchers (Basu 2014), done based on one feature which is the active power which performs with satisfactory accuracy.

In our proposed work, the supervised learning through k-NN and naive Bayes classifier is utilized. The k-NN algorithm is chosen as one of the classifier in our work because we only utilize one feature to classify the appliance. This methodology was also utilized by Gupta et al. (2010). There was also work done utilizing k-NN for training with few examples, for classifying individual electrical appliances through V-I trajectory based load signature with multi-stage classification algorithm (Yang et al. 2016). The data used in this work was from the REED data set House#3 (Kolter and Johnson 2011).

There was also work done utilizing a developed data set of four types of electrical appliances, and satisfactory classification accuracy can be achieved using few training examples with naive Bayes classifier and one feature (the average power) (Yang et al. 2017).

Furthermore, the work by Spiegel et al. (Spiegel and Albayrak 2014) utilized the REED data set with four classifiers. It was shown in their work that Naive Bayes has the highest classification accuracy and the lowest classification accuracy was the 1-NN (Spiegel and Albayrak 2014). Therefore, in our work we utilize the Naive Bayes and k-NN classifiers to benchmark our classification results with the work of Spiegel and Albayrak (2014). Apart from benchmarking purpose,

naive Bayes classifier is chosen because of the classifier's simplicity and robustness and requires small training data examples (Wu et al. 2008).

Once the classification algorithm is implemented, the results of the accuracy of the classification are obtained. In order to provide a more meaningful comparison between algorithms, various researchers reported the use of confusion matrix (CM) (Zoha et al. 2012; Alasalmi et al. 2012; Barker et al. 2014; Spiegel and Albayrak 2014). A CM is a matrix that illustrates the predicted and actual classifications (Kohavi and Provost 1998). In addition, to have consequential comparison with various NIALM methodologies, the use of a common data set is crucial (Zoha et al. 2012).

The novelty of this paper is the systematic approach of energy disaggregation using few training examples with two generic classifiers (k-NN and naive Bayes) and one feature (power) with the combination of ON-OFF based approach and goodness-of-fit (GOF) technique for event detection. Recent interesting work from researchers also shows that satisfactory classification accuracies can be achieved using few training examples from the REED data set (Yang et al. 2016) or from the developed data set (Yang et al. 2017).

In this paper, we investigated one of the problems associated with event-based load disaggregation, that is, the lack of training samples to train classification model. In actual practice, obtaining real appliance signature data can be a daunting task. In a real-world setting, this would require users to manually label the samples of power signatures for individual appliances at homes. This apparent lack of user friendliness can be intimidating to users and is a hindrance to commercialization of the technology.

We proposed that one way to circumvent this problem is to develop classification models that perform well with few training examples. Training classifier with few training examples is an area of interest in the machine learning community. Salperwyck et al. has reported a detailed study on the performance of different classifiers using few training examples (Salperwyck and Lemaire 2011). Their results show that it is possible to obtain good accuracy using small training sets if classification models with properly tuned parameters are trained. One of the classification methods recommended in their paper is the naive Bayes classifier (Salperwyck and Lemaire 2011). Furthermore, work performed by Forman and Cohen (2004) shows that one of the classifier used, naive Bayes, also performed well with few

training examples. Motivated by their results, this paper aims to investigate the feasibility of energy disaggregation using few training examples to be used with two classifiers, kNN and naive Bayes for satisfactory accuracy.

The paper is organized as follows: in section “**Methodology**”, discussion of the methodology used in our work. In section “**Results and discussion**”, the results of the classification algorithm implementation are discussed. Finally, conclusions are drawn in section “**Conclusions**”.

Methodology

As mentioned in section “**Introduction**”, the objective of this paper is to investigate the feasibility of using few training examples to develop classification model for NIALM. We utilize the k-NN and naive Bayes classifier to perform classification of appliances. Figure 1 flow chart shows an overview of our classification methodology implementation.

Preparation of data set

In this paper, the REDD data set for house #1 and house #3 is utilized (Kolter and Johnson 2011). Our approach is based on the event-based methodology for NIALM (Yang et al. 2015; Jazizadeh et al. 2014; Anderson et al. 2012; Jin et al. 2011). From the REDD data set data, the power measurement is obtained (Yang et al. 2015; Kolter and Johnson 2011). The REDD data set is utilized and the data used are labeled as house 1 which have 20 monitoring points that consist of whole home monitoring (two mains) and various other electrical appliances. The electrical appliances consist of electronics, lighting, refrigerator, disposal, dishwasher, furnace, washer dryer, smoke alarms, bathroom, kitchen outlets and microwave, as shown in Table 1. In house #3, there are 22 electrical outlets that were measured and this includes two main outlets. Eleven appliances are chosen based on the availability of sufficient data as shown in Table 2. The data set consists of total average power consumption (watts). In our work, all the data sets are manually labeled.

Next, the median filtering is performed to remove noise from the data (Yang et al. 2015; Arias-Castro and Donoho 2009). The median filtering is used to reduce the noise and to smoothen the signal for processing in

the subsequent stages. After median filtering, the event detection algorithm (Yang et al. 2015; Yang et al. 2014) is performed to identify the changes in the appliances’ power levels from the power consumption profile. The combination of ON-OFF based approach and goodness-of-fit (GOF) techniques are adopted to detect events in the total power consumption data. As the data set used is from the continuous power data, the data is separated into n-samples of frames. The frames can be identified as pre- and post-event frames. The event detection algorithm is to detect the switching ON or OFF of electrical appliances. The GOF method aims to establish that for some available probability distribution, a set of data could be derived from it. The GOF test utilizes the Chi-square test (Jin et al. 2011). Equation (1) is implemented using MATLAB.

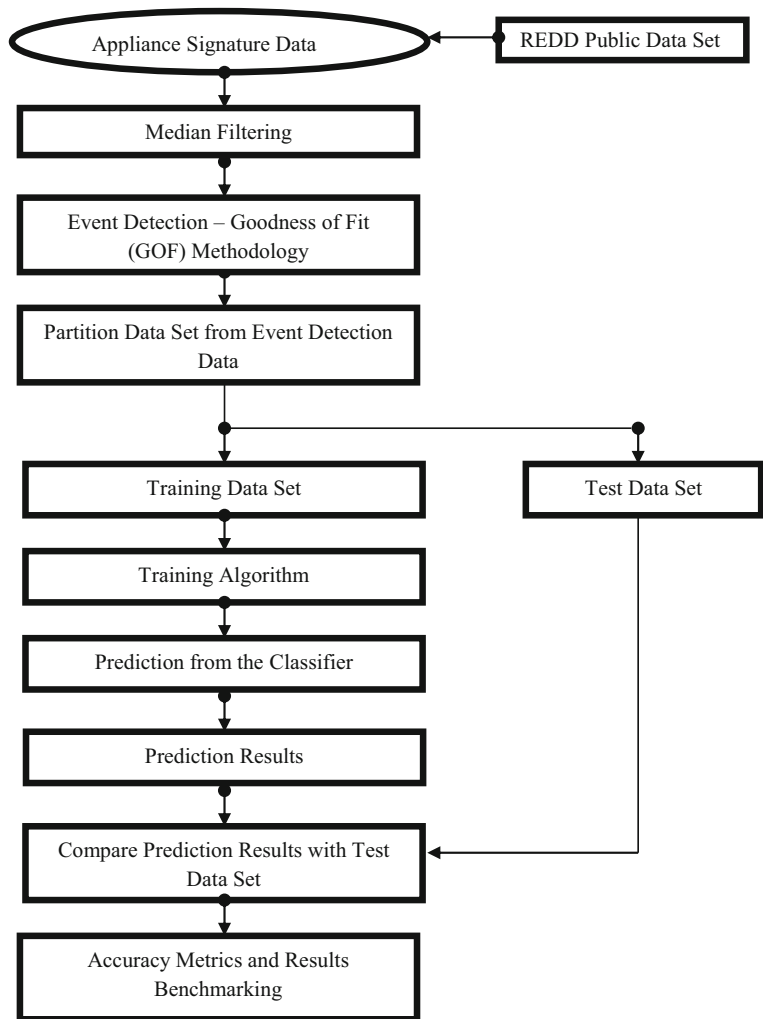
$$\ell_{GOF} = \sum_{i=1}^n \frac{(y_i - x_i)^2}{x_i} \quad (1)$$

For our calculation, we equate $\ell_{GOF} = \chi^2$. The data sample x_i is the sample from the pre-event window and y_i is the data sample from the post-event window. A total number of n data samples in the pre-event window are compared with a total number of n data samples in post-event window to determine the amount of deviation of both population of data sample. Equation (1) quantified this deviation. If an event takes place, the value of ℓ_{GOF} is expected to be high.

Partition data set from event detection data and training algorithm

The data set obtained from the event detection algorithm is manually labeled for each appliance. Next, the partition algorithm organizes the data according to the type of appliance and allocated 40% of the data set as test samples and the remaining 60 to 10% as the training samples. In our proposed method, the remaining 60 to 10% training sample data set is selected sequentially. This is because the ON/OFF event of an appliance must happen in succession. The minimum, maximum, median, and mean power consumption for each type of appliances are also listed in Tables 1 and 2 for House #1 and House #3, respectively.

Since we are interested in investigating the influence of training set size on the classification accuracy of the models, we design six different data sets assigned as

Fig. 1 Flow chart of the methodology implementation

data sets A, B, C, D, E and F for each of the appliances. The numbers of training samples for each of the appliances are as follows:

- Oven 1 (no. of training samples from 29 to 3)
- Oven 2 (no. of training samples from 25 to 4)
- Refrigerator (no. of training samples from 408 to 68)
- Dishwasher (no. of training samples from 110 to 18)
- Lighting 1 (no. of training samples from 41 to 7)
- Washer dryer 1 (no. of training samples from 28 to 5)
- Microwave (no. of training samples from 181 to 30)
- Bathroom gfi (no. of training samples from 34 to 6)
- Kitchen outlet 1 (no. of training samples from 24 to 4)
- Kitchen outlet 2 (no. of training samples from 16 to 3)
- Lighting 2 (no. of training samples from 18 to 3)
- Washer dryer 2 (no. of training samples from 136 to 23)

The test set contains the same test samples for all Data Set A through to F as shown in Table 1 (no. of test samples). This is to ensure that comparisons between the accuracy of different data sets can be made in a consistent manner using the same test samples for all training sets A through to F. Hence, there are a total of 705 test samples used for Data Set A, B, C, D, E and F.

In order to compare the effect of few training examples, a second set of test samples is used with the number of samples reduced approximately by half at $n/2 = 350$. The training samples will also be reduced

Table 1 REDD House #1—the power consumption values and number of samples for each type of appliances

Appliance type at house #1 (REDD data set)	No. of samples	No. of test samples	Power			
			Min	Max	Mean	Median
Oven 1	49	24	1446.30	4159.50	1673.50	1635.10
Oven 2	42	17	312.20	4595.80	2380.50	2439.00
Refrigerator (REF)	680	272	19.82	424.38	169.34	170.52
Dishwasher (DSW)	183	73	123.48	1140.90	383.07	214.58
Lighting 1 (LGT 1)	69	28	3.70	284.55	111.92	83.96
Washer dryer 1 (WSD 1)	47	19	436.92	1680.10	641.40	631.30
Microwave (MCW)	301	120	387.87	4038.70	1504.60	1497.20
Bathroom gfi (BTR)	57	23	1379.00	1728.10	1581.90	1580.60
Kitchen outlet 1 (KCO 1)	40	16	446.79	1141.70	1042.90	1061.10
Kitchen outlet 2 (KCO 2)	26	10	1310.20	1951.80	1512.00	1513.50
Lighting 2 (LGT 2)	30	12	0.19	153.30	68.47	67.12
Washer dryer 2 (WSD 2)	227	91	184.72	3135.70	2405.00	2419.90
Total samples	1751	705				

approximately by half. The $n/2 = 350$ sample size is used for the naive Bayes classifier.

Figures 2 and 3 show the distribution of power values for different types of appliances used in House #1 and House #3, respectively, of the REDD data set. These two figures also illustrate the extent of which the distribution of power values of appliances overlap with each other in a graphical manner. The overlapped power value distributions pose

difficulty in discriminating the appliances especially in cases where the power values fall in the ranges of two or more appliance categories.

Since we are interested in investigating the influence of training set size on the classification accuracy of the models between House #1 and House #3, similar treatment of allocation of data set in House#1 is being applied to House#3. Hence, there are a total of 1306 test samples used for House #3 as shown in Table 2.

Table 2 REDD House #3—the power consumption values and number of samples for each type of appliances

Appliance type at House #3 (REDD data set)	No. of samples	No. of test samples	Power			
			Min	Max	Mean	Median
Electronics (ELC)	67	27	2.18	1177.3	890.67	1098.2
Furnace (FUR)	190	76	62.8	743.88	367.64	393.60
Washer dryer 2 (WSD 2)	537	215	389.33	2687.5	2082.2	2110.4
Microwave (MCW)	80	32	48.8	2810.6	1503.4	1760.6
Bathroom gfi (BTR)	77	31	779.57	3410.1	1478.9	1684.6
Refrigerator (REF)	1382	553	2.34	1953.3	111.23	107.48
Dishwasher (DSW)	90	36	52	746	175.97	127
Lighting 1 (LGT 1)	74	30	34.17	358.56	172.28	172.27
Washer dryer 1 (WSD 1)	519	208	1567.3	2479.5	2277.0	2285.8
Lighting 2 (LGT 2)	173	69	30.43	383.73	142.59	140.99
Lighting 3 (LGT 3)	73	29	50.75	313.45	147.55	147.27
Total samples	3262	1306				

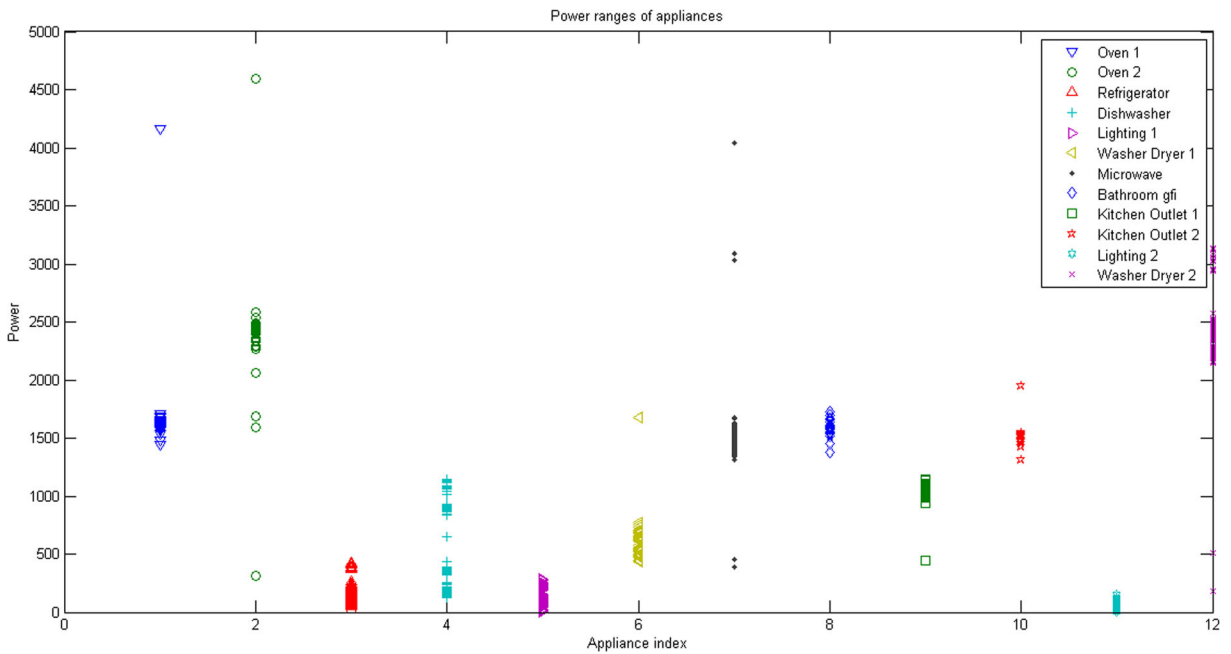


Fig. 2 House #1—power ranges of appliances

In order to compare the effect of few training examples, a second set of test samples is used with the number of samples reduced approximately by half at $n/2 = 650$. The training samples will also be reduced approximately by half. The $n/2 = 650$ sample size is used for the naive Bayes classifier.

Classifier’s algorithm

K-nearest neighbour (k-NN) classifier

The k-NN classifier finds the data’s nearest neighbours in training data in the determined feature space by

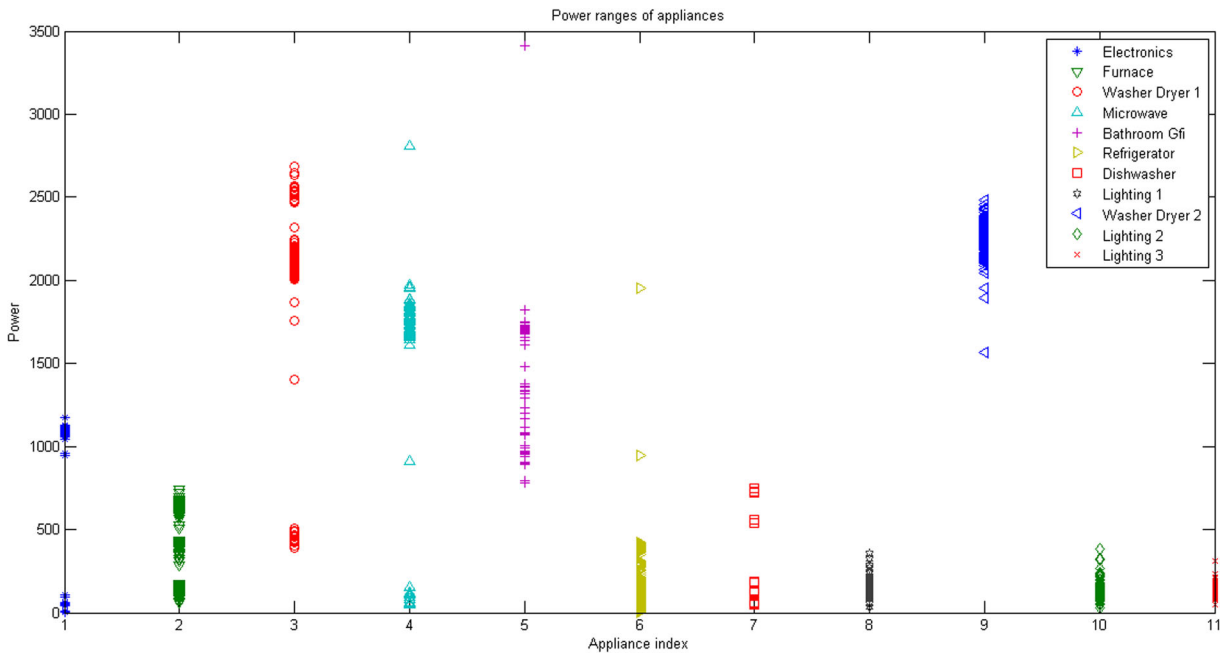


Fig. 3 House #3—power ranges of appliances

Table 3 State of the confusion matrix

	Actual ON	Actual OFF
Prediction ON	TP	FP
Prediction OFF	FN	TN

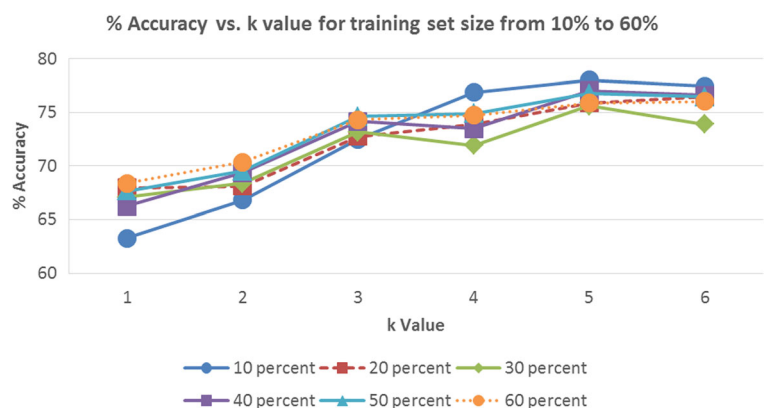
utilizing Euclidean distance as a distance metric. The majority choice of the nearest neighbours classes will be the class chosen for the data evaluated. The “*k*” value is the number of closest data points to the nearest neighbours. The kNN algorithm is chosen as one of the classifier in our work because we only utilize one feature to classify the appliance. This methodology was also utilized by Gupta et al. (2010).

In the k-NN classifier, the number of nearest neighbour, *k*, is varied, and the prediction accuracy for different *k* values is recorded. The highest percentage prediction accuracy with the corresponding training data set size, *k* value, is analyzed.

K-nearest-neighbour (k-NN) algorithm measures the distance between a query situation and a set of situation in the data set. For any new query situation, the attributes of the new query are compared with all the previously seen situation in the training data set, based on a distance measurement. The classification process is to assign the new query to the class of the majority of neighbouring query situation. For any instance *x_i* in a data set of size *m*, the distance *d* is based on the following Eq. (2):

$$d(x_i) = \min \left\{ d(x_i, x_1); d(x_i, x_2); \dots; d(x_i, x_m) \right\} \quad (2)$$

Fig. 4 House #1—k-NN result from training set size



In our work, Euclidean distance is used. This distance, where *m* is the total number of points, is calculated by the following Eq. (3)

$$d_E \left(\left\{ x_1, x_2, \dots, x_m \right\}; \left\{ y_1, y_2, \dots, y_m \right\} \right) = \sqrt{\sum_{i=1}^m (x_i - y_i)^2} \quad (3)$$

Naive Bayes classifier

In our proposed work, the naive Bayes classifier is also utilized. In our method, an individual appliance is the class and the feature is the power. This classifier is based on the Bayes’ theorem and assumes independence between each of the appliances’ state, as shown in Eq. (4).

$$p(c|d) = \frac{p(d|c) \cdot p(c)}{p(d)} \quad (4)$$

where

p(c | d) is the probability of instance *d* being in class *c*,

p(d | c) is the probability of producing instance *d* given class *c*,

p(c) is the probability of occurrence of class *c*,

p(d) is the probability of instance *d* occurring.

The advantage of using naive Bayes is that it only requires a small amount of training data to predict the appliances class. This fits into our method as we required training with few examples. It is because of this reason, we choose naive Bayes classifier. There are various researchers who carried out work using

Table 4 House #1—appliance confusion matrix for k-NN classifier, k = 5

		Actual												
		Oven 1	Oven 2	REF	DSW	LGT 1	WSD 1	MCW	BTR	KCO 1	KCO 2	LGT 2	WSD 2	
Predicted	Oven 1	10	1	0	0	0	0	2	6	0	0	0	0	
	Oven 2	0	0	0	0	0	0	0	0	0	0	0	0	
	REF	0	0	261	42	18	0	0	0	0	0	2	0	
	DSW	0	1	6	26	0	2	0	0	0	0	0	0	
	LGT 1	0	0	5	0	10	0	0	0	0	0	10	0	
	WSD 1	0	0	0	1	0	17	0	0	0	0	0	0	
	MCW	6	0	0	0	0	0	105	3	0	9	0	0	
	BTR	7	1	0	0	0	0	10	14	0	0	0	0	
	KCO 1	0	0	0	4	0	0	0	0	16	0	0	0	
	KCO 2	0	0	0	0	0	0	1	0	0	0	0	0	
	LGT 2	0	0	0	0	0	0	0	0	0	0	0	0	
	WSD 2	1	14	0	0	0	0	2	0	0	1	0	91	

○ Test samples of which the developed classifier algorithm has difficulty to discriminate between appliances

naive Bayes classifier and obtained good results (Marchiori and Han 2009; Meehan et al. 2014; Ashari et al. 2013).

With naive Bayes algorithm, the training data set is used to train the classifier, and the prediction accuracy is obtained by testing the classifier on unseen test set as per Eq. (5). In naive Bayes classifier algorithm, there is no parameter to vary unlike in the case of k-NN classifier.

Classifier’s performance

Confusion matrix is also constructed to help to identify the number of examples that are correctly classified as well as misclassified (Zoha et al. 2012; Alasalmi et al. 2012; Barker et al. 2014; Spiegel and Albayrak 2014; Ashari et al. 2013). In the analysis of

confusion matrix, the predicted appliances are mapped to the actual appliances. The classification accuracy of the model is calculated from the confusion matrix table based on the following Eq. (5).

$$Accuracy = [(TP + TN)/(TP + TN + FP + FN)] \quad (5)$$

where TP = number of true positive samples in test set, TN = number of true negative samples in the test set, FP = number of false positive samples in the test, FN = number of false negative sample in the test set.

Table 3 shows how the TP, TN, FP and FN parameters correspond to the entries of the confusion matrix (Powers 2011). From the confusion matrix, the individual appliance classification accuracy can be calculated. There are four types of conditions in

Fig. 5 House #1—% accuracy of k-NN classifier for each individual appliance

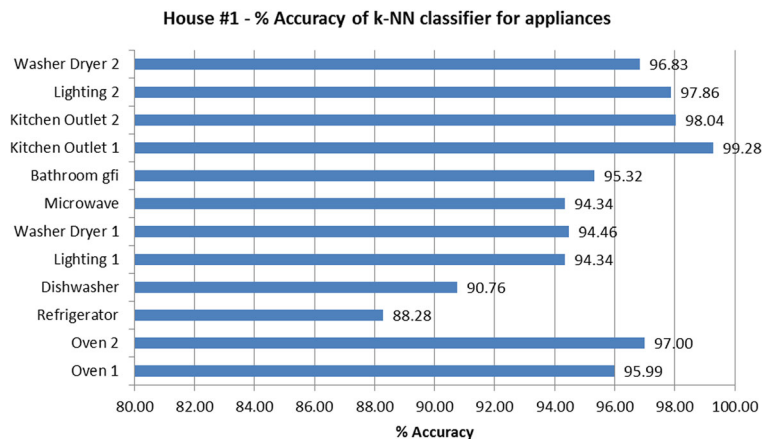
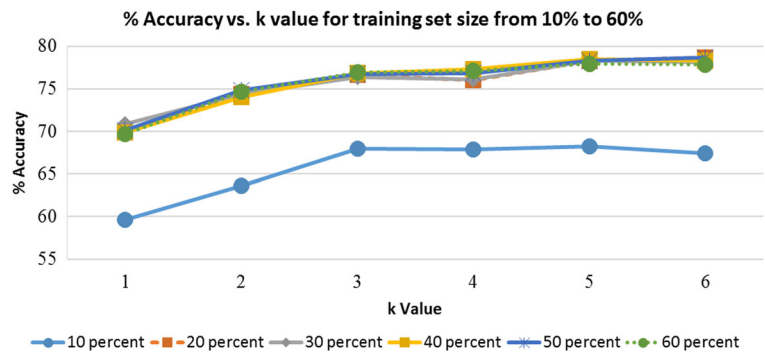


Fig. 6 House #3—k-NN result from training set size



the confusion matrix. The first type, true positives (TP), is examples correctly labeled as ON state. The second type, false positives (FP), corresponds to OFF state examples incorrectly labeled as ON state. The third type, true negatives (TN), refers to examples correctly labeled as OFF state. The fourth type, false negatives (FN), refers to ON state examples incorrectly labeled as OFF state. The confusion matrix constructed is for a two class classifier. The diagonal values represent the identification accuracy of the respective class. This means that the predicted ON/OFF corresponds to the actual ON/OFF. The row values in parallel illustrate that a particular class is wrongly predicted with other class.

Apart from the accuracy performance measurement discussed above, other researchers (Kolter and Jaakkola 2012; Li et al. 2014) have reported the precision and recall parameters as performance metrics. We also computed the precision and recall values to evaluate our classification task. Precision measures the segment of an estimated power of an appliance and

the higher the precision reflects the higher percentage of correct estimation (fewer false positive). Recall measures the segment of a given appliance power correctly classified and the higher recall indicates higher power level correctly estimated (fewer false negative) (Kim et al. 2011; Li et al. 2014; Davis and Goadrich 2006; Powers 2011). The equations for precision and recall are provided in the following eqs. (6) and (7), respectively.

$$\text{Precision} = \text{True Positive} / [\text{True Positive} + \text{False Positive}] \quad (6)$$

$$\text{Recall} = \text{True Positive} / [\text{True Positive} + \text{False Negative}] \quad (7)$$

Results and discussion

K-NN classifier

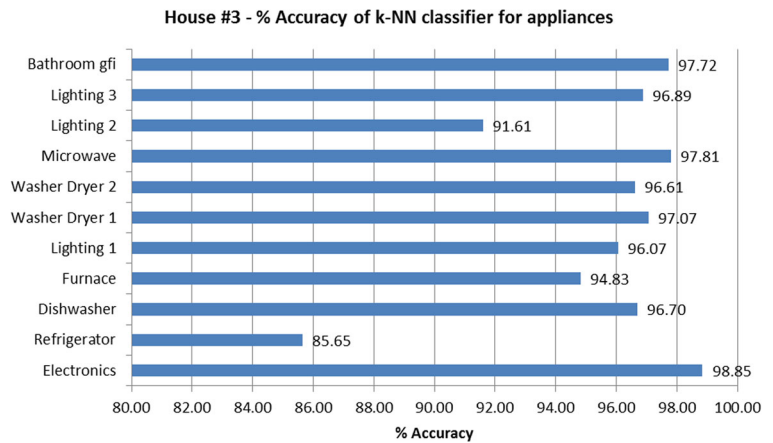
For House #1 of REDD data set, from the training set size of 60 to 10%, it is observed that the training set size of 10%, at the number of nearest neighbours, $k = 5$

Table 5 House #3—appliance confusion matrix for k-NN classifier, $k = 6$

		Actual										
		ELC	REF	DSW	FUR	LGT 1	WSD 1	WSD2	MCW	LGT 2	LGT 3	BTR
Predicted	ELC	17	0	0	0	0	0	0	0	0	0	2
	REF	9	514	18	29	18	0	1	3	42	13	0
	DSW	0	3	10	2	1	0	0	0	0	3	0
	FUR	0	10	4	39	0	0	3	0	0	0	2
	LGT 1	0	5	0	1	2	0	0	0	4	4	0
	WSD 1	0	0	0	0	0	188	11	0	0	0	0
	WSD 2	0	0	0	0	0	18	198	0	0	0	1
	MCW	0	0	0	0	0	2	1	14	0	0	2
	LGT 2	0	21	4	3	9	0	0	0	21	9	0
	LGT 3	0	0	0	2	0	0	0	0	2	0	0
	BTR	1	0	0	0	0	0	1	15	0	0	24

○ Test samples of which the developed classifier algorithm has difficulty to discriminate between appliances

Fig. 7 House #3—% accuracy of k-NN classifier for each individual appliance



produces the highest accuracy at 78.01%. Figure 4 shows the results. The accuracy is calculated using the following eq. (8).

$$\text{Accuracy} = (\text{Total correct prediction} / \text{Total Test Sample}) * 100\% \quad (8)$$

As the House #1 of REDD data set has the highest accuracy at 78.01% from the training set size of 10% and $k = 5$, an appliance confusion matrix table is built as shown in Table 4. The matrix table summarizes the number of samples that are correctly classified or misclassified. Each column of the confusion matrix represents the instances in an actual class, while each row represents the instances in a predicted class. The diagonal values represent the identification accuracy of the respective class. This means that the predicted ON/OFF corresponds to the actual ON/OFF. The row values in parallel illustrate that a particular class is wrongly predicted with other class.

From the confusion matrix in Table 4, it is observed that the k-NN classifier has difficulty in discriminating between refrigerator (REF) and dishwasher (DSW).

There are 42 power events that belong to the DSW but they are misclassified as REF by the k-NN classifier algorithm. There are also 6 power events that belong to REF but erroneously classified as DSW by the k-NN classifier algorithm. On top of that, from the confusion matrix in Table 4, it is observed that the k-NN classifier has difficulty in discriminating between refrigerator (REF) and lighting 1 (LGT 1). There are 18 power events that belong to the LGT 1, but they are misclassified as REF by the k-NN classifier algorithm. There are also 5 power events that belong to REF but erroneously classified as LGT 1 by the k-NN classifier algorithm. There is also 14 power events that belong to oven 2 but are misclassified as washer dryer 2 (WSD 2). In oven 2 and WSD 2, based on Table 1 and Fig. 2, it is observed that both appliances have similar power ranges. Overlapping of power value distributions from the two appliances creates difficulty in classifying power events. This difficulty has been reported in many studies (Hart 1992; Zeifman and Roth 2011; Zeifman 2012;

Fig. 8 House #1—naïve Bayes classifier result from training set size

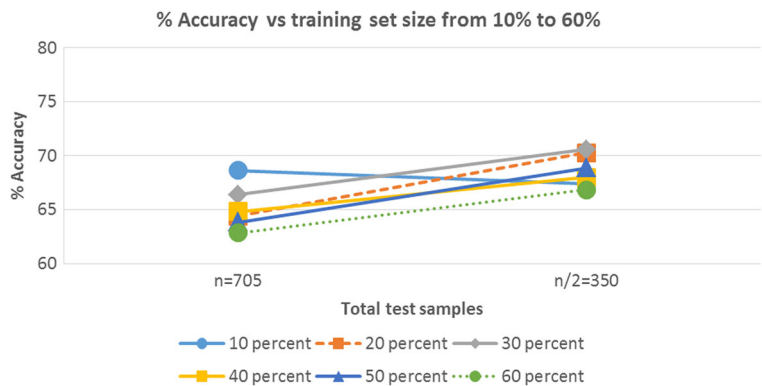


Table 6 House#1—appliance confusion matrix for naive Bayes classifier

		Actual												
		Oven 1	Oven 2	REF	DSW	LGT 1	WSD 1	MCW	BTR	KCO 1	KCO 2	LGT 2	WSD 2	
Predicted	Oven 1	7	0	0	0	0	0	1	1	0	0	0	0	
	Oven 2	0	1	0	0	0	0	0	0	0	0	0	30	
	REF	0	0	131	18	9	0	0	0	0	0	4	0	
	DSW	0	1	3	16	0	4	0	0	0	0	0	0	
	LGT 1	0	0	0	0	0	0	0	0	0	0	0	0	
	WSD 1	0	0	0	0	0	5	0	0	0	0	0	0	
	MCW	4	0	0	0	0	0	53	2	0	3	0	0	
	BTR	1	0	0	0	0	0	3	8	0	0	0	0	
	KCO 1	0	0	0	2	0	0	0	0	8	0	0	0	
	KCO 2	0	0	0	0	0	0	1	0	0	1	0	0	
	LGT 2	0	0	2	0	5	0	0	0	0	0	2	0	
	WSD 2	0	6	0	0	0	0	2	0	0	1	0	15	

○ Test samples of which the developed classifier algorithm has difficulty to discriminate between appliances

Goncalves et al. 2011; Zeifman et al. 2013). On top of that, refrigerator, dishwasher and lighting 1 are also having similar power range, which creates the difficulty in classifying these power events.

Figure 5 shows the accuracy of each individual appliance. The accuracy level is calculated using eq. (5) from the data obtained from Table 4 confusion matrix table. From Fig. 5, the appliances have classification accuracy of between 88 and 99%, which achieve very good performance.

Similar analysis is done for House #3 of REDD data set, from the training set size of 60 to 10%. Figure 6 illustrates the results for the training set size of 20%, at the number of nearest neighbours, $k = 6$ produces the highest accuracy at 78.64%.

The results in both Figs. 4 and 6 are compared, and it is observed that the percentage accuracy does not vary much in between 20 and 60% training set size. Despite big changes in the size of training set from 20 to 60%, the percentage accuracy is confined within the range of 68 up to 78%. The main difference is the performance of the k-NN classifier at training set size of 10%. As the House #3 of REDD data set has the highest accuracy at 78.64% from the training set size of 20% and $k = 6$, an appliance confusion matrix table is built as shown in Table 5.

From the confusion matrix in Table 5, it is observed that the k-NN classifier has difficulty in discriminating between refrigerator (REF) and lighting 2 (LGT 2). There are 21 power events that belong to the refrigerator

Fig. 9 House #1—% accuracy of naïve Bayes classifier for each individual appliance

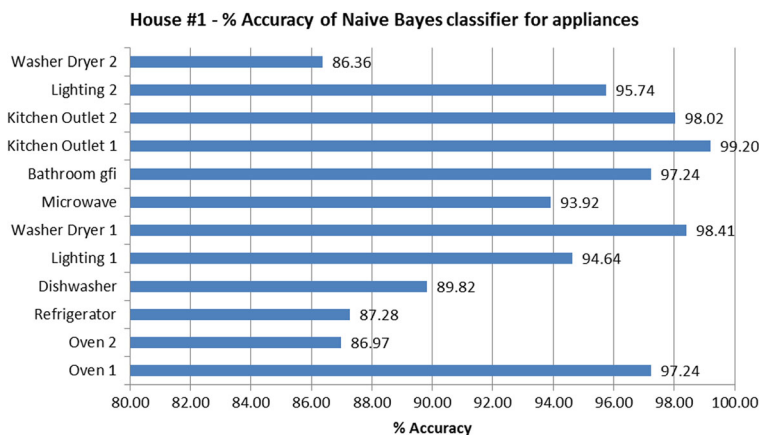
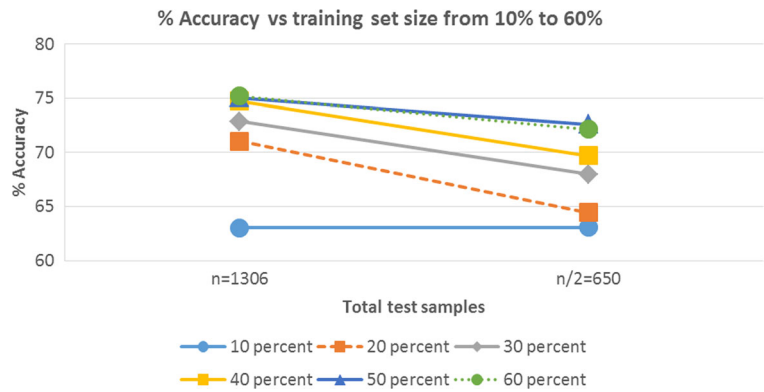


Fig. 10 House #3—result from naive Bayes classifier algorithm



(REF), but they are misclassified as LGT 2 by the k-NN classifier algorithm. There are also 42 power events that belong to LGT 2 but erroneously classified as REF by the k-NN classifier algorithm. On top of that, from the confusion matrix in Table 5, it is observed that the k-NN classifier has difficulty in discriminating between washer dryer 1 (WSD1) and washer dryer 2 (WSD 2). There are 18 power events that belong to the WSD 1 but they are misclassified as WSD 2 by the k-NN classifier algorithm. There are also 11 power events that belong to WSD 2 but erroneously classified as WSD by the k-NN classifier algorithm. Figure 7 shows the accuracy of each individual appliance. The accuracy level is calculated using Eq. (5) from the data obtained from Table 5 confusion matrix table. From Fig. 7, the appliances have classification accuracy of between 85 and 98%, which achieve very good performance.

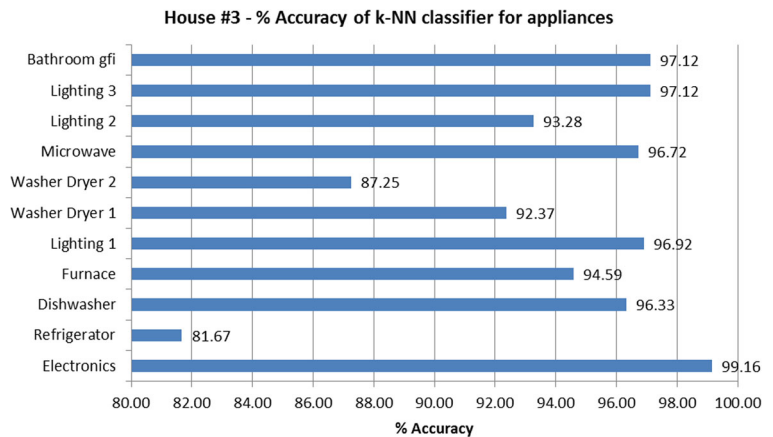
In summary, comparing both the Figs. 4 and 6, the k-NN classifier can achieve satisfactory accuracy for both House #1 and House #3 at 20% training data set size. At individual appliance level, the k-NN classifier is able to classify at satisfactory accuracy level ranging from 85 to 99% for both House #1 and House #3. Both house results also show that the performance of k-NN classifier not necessarily improves with more training examples. Even with 20% training data set size, the k-NN classifier achieves good accuracy performance. This is in line with the results obtained by Ashari et al. (Davis and Goadrich 2006). In their work (Davis and Goadrich 2006), the best performance of k-NN classifier is their set 1 (2827 data) and set 2 (4340 data) compared with their largest set 5 with 8630 data.

Table 7 Appliance confusion matrix for naive Bayes classifier

Predicted	Actual										
	ELC	REF	DSW	FUR	LGT 1	WSD 1	WSD2	MCW	LGT 2	LGT 3	BTR
ELC	10	0	0	0	0	0	0	0	0	0	1
REF	3	272	16	18	15	0	0	2	34	14	0
DSW	0	0	0	0	0	0	0	0	0	0	0
FUR	0	4	2	20	0	0	3	0	0	0	0
LGT 1	0	0	0	0	0	0	0	0	0	0	0
WSD 1	0	0	0	0	0	97	32	0	0	0	0
WSD 2	0	0	0	0	0	71	72	14	0	0	13
MCW	0	0	0	0	0	0	0	0	0	0	0
LGT 2	0	0	0	0	0	0	0	0	0	0	0
LGT 3	0	0	0	0	0	0	0	0	0	0	0
BTR	0	0	0	0	0	0	0	0	0	0	1

○ Test samples of which the developed classifier algorithm has difficulty to discriminate between appliances

Fig. 11 House #3—% accuracy of naïve Bayes classifier for each individual appliance



Naive Bayes classifier

From the training set size of 60 to 10%, the naïve Bayes classification algorithm is used. For House #1, two sample sizes are used. The first sample size is at $n = 705$ and the second sample size is at $n/2 = 350$. It is observed that at sample size of $n = 705$ and training set size of 10%, the classifier produces the highest accuracy at 68.65%. The accuracy is calculated using eq. (8). It is observed that at sample size of $n/2 = 350$ and training set size of 30%, the classifier produces the highest accuracy at 70.57%. Figure 8 shows the results.

From Fig. 8, it is shown that with the reduced number of total training sample size from $n = 705$ to $n/2 = 350$, the naïve Bayes classifier produces slight increase in accuracy level from 20 to 60% training set size. The exception is at 10% training which decrease in accuracy level. This might be because of the minimum number of training samples to be effective in producing satisfactory accuracy level. From the sample size of $n/2 = 350$ and training set size of 30%, a confusion matrix table is built. The training set size of 30% is chosen because in the naïve Bayes classifier algorithm, it produces the highest

accuracy at 70.57% as shown in Fig. 8. Table 6 shows the construction of an appliance labeled confusion matrix for the training set size of 30% and $n/2 = 350$.

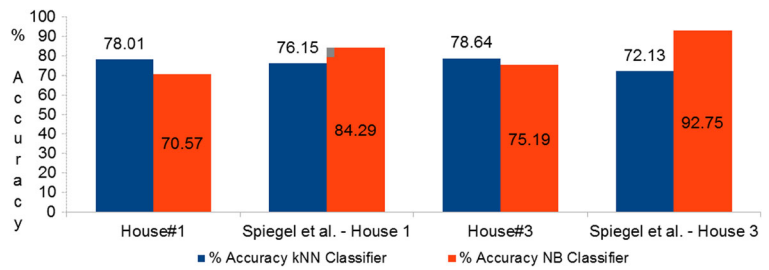
From the confusion matrix in Table 6, it is observed that the naïve Bayes classifier has difficulty in discriminating between oven 2 and washer dryer 2 (WSD 2). There are 6 power events that belonged to oven 2, but they are misclassified as washer dryer 2 (WSD 2) by the naïve Bayes classifier algorithm. There are also 30 power events that belonged to washer dryer 2 (WSD 2) but erroneously classified as oven 2 by the naïve Bayes classifier algorithm. As shown in Fig. 8, even with the reduced number of sample size by half, from $n = 705$ to $n/2 = 350$, the accuracy of the classifier is still within the 68 to 70% accuracy range. Figure 9 shows the accuracy of each individual appliance. The accuracy level is calculated using eq. (5) from the data obtained from Table 6 confusion matrix. From Fig. 9, the appliances have classification accuracy of between 86 and 98%, which achieve very good performance.

In order to compare the validity of the algorithm, REDD House #3 data set is also utilized in the naïve Bayes algorithm. As shown in Tables 1 and 2, there are

Table 8 % Accuracy of the proposed naïve Bayes (NB) classifier and k-NN classifier

House number	%Accuracy of kNN classifier, $k = 5$, training set size = 10%	%Accuracy of kNN classifier, $k = 6$, training set size = 20%	%Accuracy of NB classifier at n sample size, training set size of 10%	%Accuracy of NB classifier at n sample size, training set size of 60%	%Accuracy of NB classifier at $n/2$ sample size, training set size of 30%	%Accuracy of NB classifier at $n/2$ sample size, training set size of 50%
House #1	78.01	76.45	68.65	62.84	70.57	68.86
House #3	68.22	78.64	63.02	75.19	68.00	72.62

Fig. 12 Comparison of accuracy between the proposed method and the work of Spiegel and Albayrak (2014)



differences in the composition of electrical appliances between House #1 and House #3. In House # 3, two sample sizes are used. The first sample size is at $n = 1306$ and the second sample size is at $n/2 = 650$. It is observed that at sample size of $n = 1306$ and training set size of 60%, the classifier produces the highest accuracy at 75.19%. The accuracy is calculated using eq. (5). It is observed that at sample size of $n/2 = 650$ and training set size of 50%, the classifier produces the highest accuracy at 72.62%. Figure 10 shows the results. It can be noted that with the reduced sample size for training, the percentage accuracy reduces from 72 to 75%.

From the sample size of $n/2 = 650$ and training set size of 50%, a confusion matrix table is built. The training set size of 50% and sample size of $n/2 = 650$ was chosen because in the naive Bayes classifier algorithm, it produces the highest accuracy at 72.61% at that sample size value. This is to analyze that even if at fewer training samples and sample size, naive Bayes classifier produces comparative accuracy levels. Table 7 shows the construction of an appliance labeled confusion matrix for the training set size of 50% and $n/2 = 650$.

From the confusion matrix in Table 7, it is observed that the naive Bayes classifier has

difficulty in discriminating between washer dryer 1 (WSD 1) and washer dryer 2 (WSD 2). There are 7 power events that belonged to WSD 1 but they are misclassified as WSD 2 by the naive Bayes classifier algorithm. There are also 32 power events that belonged to WSD 2 but erroneously classified as WSD 1 by the naive Bayes classifier algorithm. As shown in Fig. 10, even with the reduced number of sample size by half, from $n = 1306$ to $n/2 = 650$, the highest accuracy level of the classier dropped slightly, from 75.19 to 72.62%. Figure 11 shows the accuracy of each individual appliance. The accuracy level is calculated using eq. (5) from the data obtained from Table 7 confusion matrix. Similarly, for House #3, the appliances have classification accuracy of between 81 and 98% which achieve very good performance as shown in Fig. 11.

Classifiers’ performance

Table 8 shows the percentage accuracy of our proposed naive Bayes and k-NN classifier. In the naive Bayes classifier, a small number of training set sizes do not affect the prediction accuracy of appliances as shown in Table 8. Even at $n/2$ sample size level, the

Fig. 13 Percentage classification accuracy of bathroom gfi, dishwasher and lighting

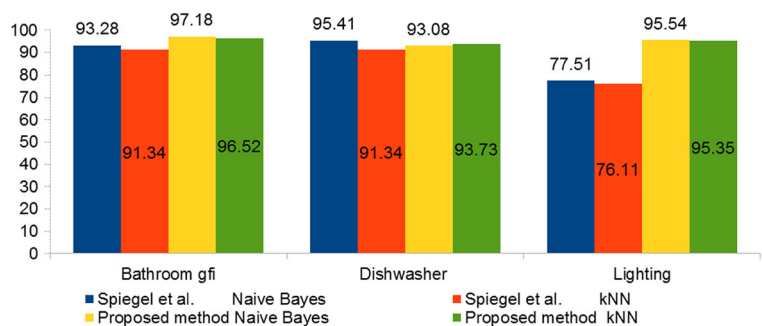
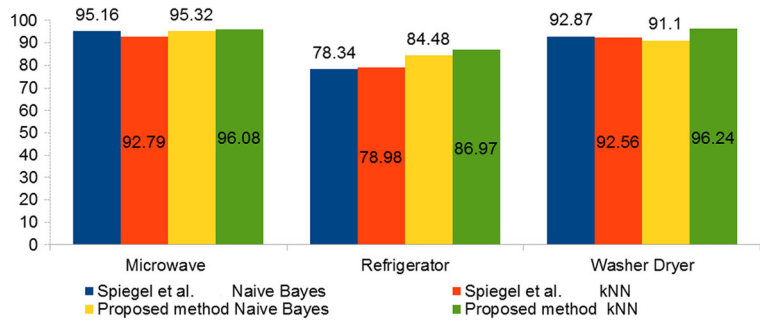


Fig. 14 Percentage classification accuracy of microwave, refrigerator and washer dryer



result shows comparable accuracy level with n sample size level. Both house results also show that the performance of naive Bayes and k-NN classifier not necessarily improves with more training examples. This result also agrees to the result done by Ashari et al. (Davis and Goadrich 2006). In their work (Davis and Goadrich 2006), the best performance of naive Bayes classifier is their set 2 with 4340 data compared to their largest set 5 with 8630 data.

The percentage accuracy of proposed k-NN classifier and naive Bayes classifier are also shown for comparison with the work done by Spiegel and Albayrak (2014). Figure 12 shows the comparison of accuracy. Based on comparison, it was noted that our proposed k-NN classifier has higher accuracy level compared to Spiegel and Albayrak (2014) results. For naive Bayes classifier, Spiegel and Albayrak (2014) work performed much better results compared to our results.

The percentage classification average accuracy of selected electrical appliances for the proposed k-NN classifier and naive Bayes classifier as compared to the work by Spiegel and Albayrak (2014) are shown

in Figs. 13 and 14. Based on comparison, both the proposed methods achieved comparable accuracy level for 4 types of appliances, namely bathroom gfi, dishwasher, microwave and washer dryer. For the appliance of lighting and refrigerator, both the proposed methods achieved a slightly higher average accuracy level. However, it should be noted that for our developed algorithm, average accuracy levels are the results of six appliances for House #1 and House #3. The average classification accuracy for the six selected appliances reported by Spiegel and Albayrak (2014) takes into the consideration the mean of an appliance from House #1 through House #6. The purpose of this bench marking is to gauge the difference in accuracy level from different methodology in the application of the naive Bayes classifier and k-NN classifier.

Figures 15 and 16 show precision and recall results for the proposed methods of k-NN and naive Bayes, respectively, for House #1. From the results shown in Fig. 15, it shows that four appliances performed very well in k-NN classifier, namely, washer dryer 2, refrigerator, washer dryer 1 and microwave. All these four

Fig. 15 Precision and recall results on the proposed methods with k-NN on. House #1 selected appliances

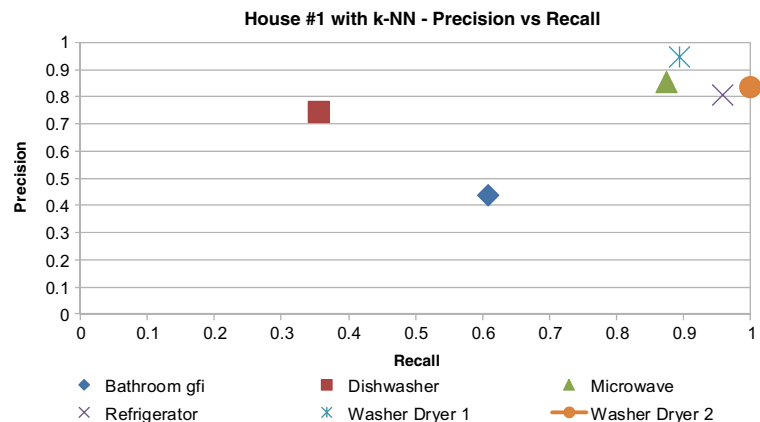
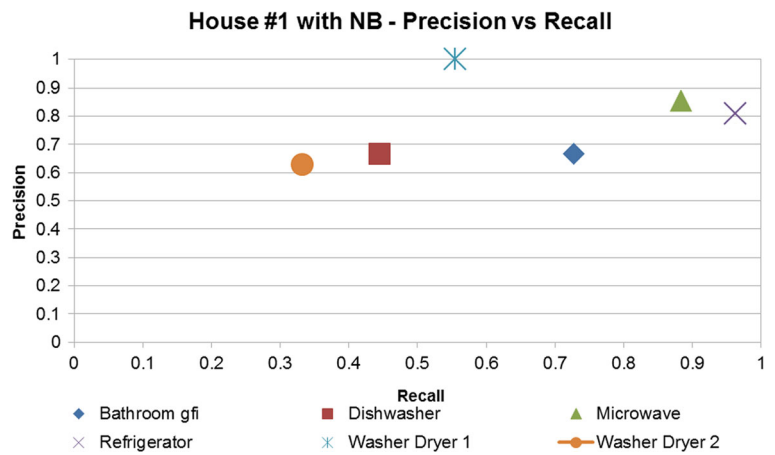


Fig. 16 Precision and recall results on the proposed methods with naive Bayes on House #1 selected appliances



appliances in House #1 have precision and recall values above 0.7 for the performance measure of k-NN.

From the results shown in Fig. 16, it shows that two appliances performed very well in naive Bayes classifier, namely, refrigerator and microwave. These two appliances in House #1 have precision and recall values above 0.7 for the performance measure of naive Bayes. For the appliance, washer dryer 1, it scores perfect result for precision (1.00) indicating all correct estimation. However, washer dryer 1 scores lower in recall indicating that lower power level is correctly estimated (more false negative).

Figures 17 and 18 show precision and recall results for the proposed methods of k-NN and Naive Bayes, respectively, for House #3. From the results shown in Fig. 17, it shows that three appliances performed very well in k-NN classifier, namely, washer dryer 2, refrigerator and washer dryer 1. All these three appliances in house #3 have precision and recall values above 0.7 for

the performance measure of k-NN. From the results shown in Fig. 18, it shows that two appliances performed very well in naive Bayes classifier, namely, refrigerator and washer dryer 1. For the appliance, dishwasher and microwave, both have all instances that were predicted as negative, therefore $TP + FP = 0$. The precision level is not applicable. For these two appliances, the recall value is zero meaning no power level is correctly estimated (all false negative). Overall, the precision and recall performance measurement produced better results from the k-NN classifier.

Conclusions

Our work in this paper only looked at approximately a month of data utilizing the REDD data sets. In future work, research could be done on other publicly available data sets. In the present work, classification accuracy

Fig. 17 Precision and recall results on the proposed methods with k-NN on House #3 selected appliances

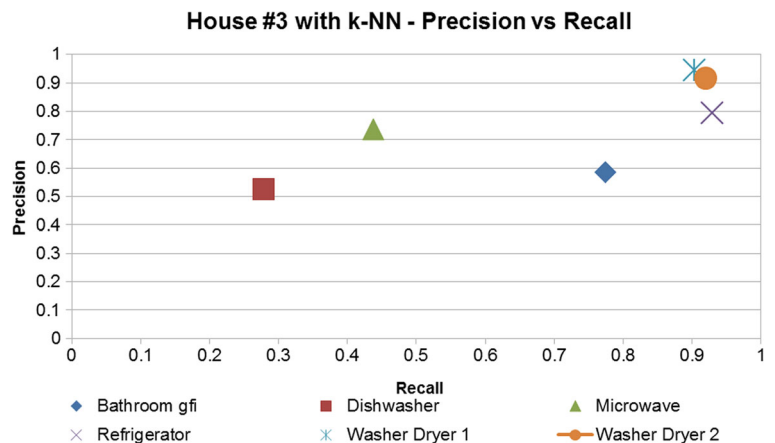
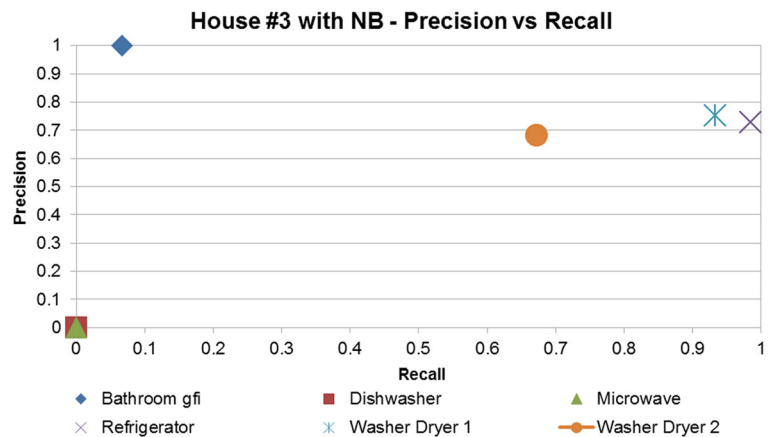


Fig. 18 Precision and recall results on the proposed methods with naive Bayes on House #3 selected appliances



metrics have been discussed. In the future work, as more data sets are considered, the estimation accuracy could be done (Makonin and Popowich 2015).

This paper reports a study carried out to classify the individual electrical appliances from a publicly available data set, REDD, utilizing NIALM with few training examples. There are two classification accuracy metrics obtained from the developed classifiers. This paper shows that despite the use of few training examples, the k-NN and naive Bayes classifiers can produce satisfactory level of accuracy. Increasing the size of training set does not seem to produce significant increase in the classification accuracy. In our classification of appliance feature presented in this paper, the classification performances of the k-NN and naive Bayes classifiers have saturated with the increase of the size of the training set and therefore unable to yield significant improvement in accuracy.

The result presented in this paper also shows that k-NN classifier performed better in the calculated accuracy performance measurement. This paper also shows that satisfactory accuracy can be achieved using few training examples with generic classifiers (k-NN and naive Bayes) and one feature (power), with the combination of ON-OFF based approach and goodness-of-fit (GOF) technique for event detection, to improve on appliance-level energy usage feedback for energy efficiency.

With the encouraging results obtained from k-NN for training with few examples, the proposed method is being applied in classifying individual electrical appliances through V-I trajectory-based load signature. The novelty of this accepted work (Yang et al. 2016) is to establish a systematic approach of load disaggregation through V-I trajectory-based load signature images for training with few examples by

utilizing a multi-stage classification algorithm methodology. Our contribution is in utilizing the “k-value”, the number of closest data points to the nearest neighbour, in the k-NN algorithm to be effective in classification of electrical appliances.

It is proposed that more studies need to be carried out to investigate the classification performances of both classifiers with more than one feature such as real power, reactive power and power factor and so on to further improve the feedback for energy efficiency.

Acknowledgements The work is funded by Ministry of Science, Technology and Innovation (MOSTI) Malaysia under the MOSTI Science fund project (No. 06-02-11-SF0162).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Alasalmi, T., Suutala, J., Rönning, J., (2012). Real-Time Non-Intrusive Appliance Load Monitor - Feedback System for Single-point per Appliance Electricity Usage. In Proceedings of SMARTGREENS 2012 - International Conference on Smart Grids and Green IT Systems, SciTePress, 2012, pp. 203–208.
- Anderson, K., Berges, M., Oceau, A., Benitez, D., Moura, J.M.F., (2012). Event detection for non intrusive load monitoring. In Proceedings of the 38th Annual Conference on IEEE Industrial Electronics Society (IECON), Montreal, Canada, 25–28, pp. 3312–3317.
- Arias-Castro, E., & Donoho, D. L. (2009). Does median filtering truly preserve edges better than linear filtering? *Annals of Statistics*, 37, 1172–1206.

- Ashari, A., Paryudi, I., & Tjao, A. M. (2013). Performance comparison between Naïve Bayes, decision tree and k-nearest neighbor in searching alternative design in an energy simulation tool. *International Journal of Advanced Computer Science and Applications*, 4(11), 33–39.
- Baets, L. D., Ruysinck, J., Develder, C., Dhaene, T., & Deschrijver, D. (2017). On the Bayesian optimization and robustness of event detection methods in NILM. *Energy and Buildings*, 145, 57–66.
- Barker, S., Musthag, M., Irwin, D., Shenoy, P. (2014). Non-Intrusive Load Identification for Smart Outlets, in: Proceedings of the 5th IEEE International Conference on Smart Grid Communications (SmartGridComm 2014), Venice, Italy, 3–6 November 2014.
- Basu, K. (2014). Classification techniques for non-intrusive load monitoring and prediction of residential loads. Electric power. Université de Grenoble. English. <NNT: 2014GRENT089><tel-01162610>.
- Basu, K., Debusschere, V., Bacha, S., Maulik, U., Bondyopadhyay, S. (2015a). Nonintrusive Load Monitoring: A Temporal Multilabel Classification Approach, *IEEE Transactions On Industrial Informatics*, Vol. 11, No. 1.
- Basu, K., Debusscherea, V., Douzal-Chouakria, A., & Bacha, S. (2015b). Time series distance-based methods for non-intrusive load monitoring in residential buildings. *Energy and Buildings*, 96, 109–117.
- Behavioural Insights Team. (2011). *Behaviour change and energy use*. London: Cabinet Office (Available at <https://www.gov.uk/government/publications/behaviour-change-and-energy-use-behavioural-insights-team-paper> [online]).
- Berges, M., Goldman, E., Matthews, H.S., Soibelman, L. (2009). Learning Systems for Electric Consumption of Buildings. In Proceedings of ASCE International Workshop on Computing in Civil Engineering, Austin, TX, USA, 24–27.
- Berges, M., Goldman, E., Matthews, H. S., & Soibelman, L. (2010). Enhancing electricity audits in residential buildings with nonintrusive load monitoring. *Journal of Industrial Ecology*, 14, 844–858.
- Berges, M., Goldman, E., Matthews, H. S., Soibelman, L., & Anderson, K. (2011). User-centered non-intrusive electricity load monitoring for residential buildings. *Journal of Computing in Civil Engineering*, 25, 471–480.
- Carrie Armel, K., Gupta, A., Shrimali, G., & Albert, A. (2013). Is disaggregation the holy grail of energy efficiency? The case of electricity. *Energy Policy*, 52, 213–234.
- Chahine, K., Drissi, K., Pasquier, C., Kerrroum, K., Faure, C., Jounnet, T., Michou, M. (2011). Electric load disaggregation in smart metering using a novel feature extraction method and supervised classification. In Proceedings of MEDGREEN 2011, Beirut, Lebanon, Energy Procedia, 6, pp. 627–632.
- Chang, H. H. (2012). Non-intrusive demand monitoring and load identification for energy management systems based on transient feature analyses. *Energies*, 5, 4569–4589.
- Chang, H.H., Yang, H.T., Lin, C.L. (2008). Load identification in neural networks for a non-intrusive monitoring of industrial electrical loads. In Computer Supported Cooperative Work in Design IV; Springer: Berlin, Germany, pp. 664–674.
- Chang, H.H., Lin, C.L., Lee, J.K. (2010). Load identification in nonintrusive load monitoring using steady-state and turn-on transient energy algorithms. In Proceedings of the 2010 14th International Conference on Computer Supported Cooperative Work in Design (CSCWD), Shanghai, China, 14–16, pp. 27–32.
- Chang, H. H., Lian, K. L., Su, Y. C., & Lee, W. J. (2014). Power Spectrum-based wavelet transform for non-intrusive demand monitoring and load identification. *IEEE Transactions on Industry Applications*, 50, 2081–2089.
- Cochran, W. G. (1952). The χ^2 test of goodness of fit. *The Annals of Mathematical Statistics*, 23, 315–415.
- Cover, T. M., & Hart, P. E. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13, 21–27.
- Darby, S. (2006). *The effectiveness of feedback on energy consumption: A review for DEFRA of the literature on metering, billing and direct displays*. Oxford: Environmental Change Institute, University of Oxford.
- Davis, J., Goadrich, M. (2006). The relationship between precision-recall and ROC curves. In Proceedings of the 23rd International Conference on Machine Learning, Pittsburgh, PA, USA, 25–29, pp. 233–240.
- Duda R.O., Hart P.E., Stork D.G. (2000). Pattern Classification, second ed., Wiley-Interscience, pp. 182–187.
- Ehrhardt-Martinez, K., Donnelly, K. A., & Laitner, J. A. (2010). *Advanced Metering Initiatives and Residential Feedback Programs: A Meta-Review for Household Electricity-Saving Opportunities. Technical Reports E105*. Washington, DC: American Council for an Energy-Efficient Economy.
- European Environment Agency (2013). 'Achieving energy efficiency through behaviour change: what does it take?', Technical report No 5/2013, (Available at <http://www.eea.europa.eu/publications/achieving-energy-efficiency-through-behaviour> [online]).
- Faruqui, A., Sergici, S., & Sharif, A. (2010). The impact of informational feedback on energy consumption—A survey of the experimental evidence. *Energy*, 35(4), 1598–1608.
- Figueiredo M., de Almeida A., Ribeiro B. (2011). An experimental study on electrical signature identification of Non-Intrusive Load Monitoring (NILM) systems In Adaptive and Natural Computing Algorithms, Springer, Berlin, Germany, Volume 6594, pp. 31–40.
- Fischer, C. (2008). Feedback on household electricity consumption: A tool for saving energy? *Energy Efficiency*, 1(1), 79–104.
- Forman, G., & Cohen, I. (2004). Learning from little: Comparison of classifiers given little training. *Knowledge Discovery in Databases: PKDD, 2004*, 161–172.
- Froehlich J. (2009). "Promoting Energy Efficient Behaviors in the Home through Feedback: The Role of Human-Computer Interaction," Human Computer Interaction Consortium 2009 Winter Workshop.
- Froehlich, J., Larson, E., Gupta, S., Cohn, G., Reynolds, M., & Patel, S. (2011). Disaggregated end-use energy sensing for the smart grid. *IEEE Pervasive Computing*, 10, 28–39.
- Giri, S., Berges, M., & Rowe, A. (2013a). Towards automated appliance recognition using an EMF sensor in NILM platforms. *Advanced Engineering Informatics*, 27, 477–485.
- Giri, S., Lai, P., Berges, M. (2013b). Novel Techniques For ON and OFF states detection of appliances for Power Estimation in Non-Intrusive Load Monitoring. In Proceedings of the 30th International Symposium on Automation and Robotics in

- Construction and Mining (ISARC), Montreal, Canada, pp. 522–530.
- Goncalves, H., Ocneanu, A., Berges, M., Fan, R.H. (2011). Unsupervised Disaggregation of Appliances Using Aggregated Consumption Data. In Proceedings of KDD 2011 Workshop on Data Mining Applications for Sustainability, San Diego, CA, USA, 21–24 August 2011.
- Gupta, S., Reynolds, M.S., Patel, S.N. (2010). ElectriSense: Single-point sensing using EMI for electrical event detection and classification in the home. In Proceedings of the 12th ACM International Conference on Ubiquitous Computing, Copenhagen, Denmark, 26–29, pp. 139–148.
- Hart, G. W. (1992). Nonintrusive appliance load monitoring. *Proceedings of the IEEE*, 80, 1870–1891.
- Jazizadeh, F., Becerik-Gerber, B., Berges, M., & Soibelman, L. (2014). An unsupervised hierarchical clustering based heuristic algorithm for facilitated training of electricity consumption disaggregation systems. *Advanced Engineering Informatics*, 28, 311–326.
- Jin, Y., Tebekaemi, E., Berges, M., Soibelman, L. (2011). Robust adaptive event detection in non-intrusive load monitoring for energy aware smart facilities. In Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, Prague, Czech Republic, 22–27, pp. 4340–4343.
- Johnson, M. J., & Willsky, A. S. (2013). Bayesian nonparametric hidden semi-Markov models. *Journal of Machine Learning Research*, 14(1), 673–701.
- Kato, T., Cho, H.S., Lee, D. (2009). Appliance recognition from electric current signals for information-energy integrated network in home environments. In Proceedings of the 7th International Conference on Smart Homes and Health Telematics, Tours, France, 1–3 July 2009, Volume 5597, pp. 150–157.
- Kelly J, Knottenbelt W, Neural NILM (2015). Deep Neural Networks Applied to Energy Disaggregation, Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments, November 04–05, 2015, Seoul, South Korea.
- Kim, H., Marwah, M., Arlitt, M., Lyon, G., Han, J. (2011). Unsupervised disaggregation of low frequency power measurements. In Proceedings of the 11th SIAM international conference on data mining, mesa, AZ, USA, 28–30, pp. 747–758.
- Kohavi, R., & Provost, F. (1998). Glossary of terms. *Machine Learning*, 30(2/3), 271–274.
- Kolter, J.Z., Jaakkola, T. (2012). Approximate Inference in Additive Factorial HMMs with Application to Energy Disaggregation. In Proceedings of the 15th International Conference on Artificial Intelligence and Statistics (AISTATS), pp. 1472–1482.
- Kolter, J.Z., Johnson, M.J. (2011). REDD: A Public Data Set for Energy Disaggregation Research. In Proceedings of the SustKDD Workshop on Data Mining Applications in Sustainability, San Diego, CA, USA, pp. 1–6.
- Kramer, O., Wilken, O., Beenken, P., Hein, A., Klingenberg, T., Meinecke, C., Raabe, T., Sonnenschein, M. (2012). On Ensemble Classifiers for Nonintrusive Appliance Load Monitoring. In Proceeding of the 7th international conference on Hybrid Artificial Intelligent Systems, Springer-Verlag, Berlin, Germany, Part I, pp. 322–331.
- Li, J., West, S., Platt, G. (2012). Power decomposition based on SVM regression. In Proceedings of International Conference on Modelling, Identification Control, Wuhan, China, 24–26, pp. 1195–1199.
- Li, Y., Peng, Z., Huang, J., Zhang, Z., Jae, H.S. (2014). Energy Disaggregation via Hierarchical Factorial HMM. In Proceedings of the 2nd International Workshop on Non-Intrusive Load Monitoring (NILM).
- Lin, G.Y., Lee, S.C., Hsu, Y.J., Jih, W.R. (2010a). Applying power meters for appliance recognition on the electric panel. In Proceedings of the 5th IEEE Conference on Industrial Electronics and Applications, Melbourne, Australia, 15–17, pp. 2254–2259.
- Lin, G.Y., Lee, S.C., Hsu, Y.J., Jih, W.R. (2010b). Applying power meters for appliance recognition on the electric panel. In Proceedings of the 5th IEEE Conference on Industrial Electronics and Applications (ICIEA), Taichung, Taiwan, 15–17, pp. 2254–2259.
- Makonin, S., & Popowich, F. (2015). Nonintrusive load monitoring (NILM) performance evaluation. *Energy Efficiency*, 8(4), 809–814.
- Makonin, S.; Popowich, F.; Bajic, I.V.; Gill, B.; Bartram, L. (2015). "Exploiting HMM sparsity to perform online real-time nonintrusive load monitoring," in Smart Grid, IEEE Transactions on, vol. PP, no.99, pp.1–11.
- Marceau, M. L., & Zmeureanu, R. (2000). Nonintrusive load disaggregation computer program to estimate the energy consumption of major end uses in residential buildings. *Energy Conversion and Management*, 41, 1389–1403.
- Marchiori, A., Han, Q. (2009). Using Circuit-Level Power Measurements in Household Energy Management Systems. In Proceedings of the First ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings, (BuildSys 09), ACM Press, pp. 7–12.
- Marchiori, A., Hakkarinen, D., Han, Q., & Earle, L. (2011). Circuit-level load monitoring for household energy management. *IEEE Pervasive Computing*, 10, 40–48.
- Meehan, P., McArdle, C., & Daniels, S. (2014). An efficient, scalable time-frequency method for tracking energy usage of domestic appliances using a two-step classification algorithm. *Energies*, 7, 7041–7066.
- Parson, O., Ghosh, S., Weal, M. J., & Rogers, A. (2014). An unsupervised training method for non-intrusive appliance load monitoring. *Artificial Intelligence*, 217, 1–19.
- Patel, S.N., Robertson, T., Kientz, J.A., Reynolds, M.S., Abowd, G.D. (2007). At the Flick of a switch: Detecting and classifying unique electrical events on the residential power line. In Proceedings of the 9th International Conference on Ubiquitous Computing, Innsbruck, Austria, 16–19, pp. 271–288.
- Powers, D. M. W. (2011). Evaluation: From precision, recall and F-measure to ROC, Informedness, Markedness and correlation. *Journal of Machine Learning Technologies*, 2, 37–63.
- Rahimi, S., Chan, A.D.C., Goubran, R.A. (2012). Nonintrusive load monitoring of electrical devices in health smart homes. In Proceedings of IEEE International Instrumentation and Measurement Technology Conference (I2MTC), Graz, 13–16, pp. 2313–2316.
- Roos, J., Lane, I., Botha, E., Hancke, G. (1994). Using neural networks for non-intrusive monitoring of industrial electrical loads. In Proceedings of the 10th IEEE Instrumentation and

- Measurement Technology Conference, Hamamatsu; IEEE: 10-12, pp. 1115–1118.
- Saitoh, T., Aota, Y., Osaki, T., Konishi, R., & Sugahara, K. (2010). Current sensor based home appliance and state of appliance recognition. *SICE Journal of Control, Measurement, and System Integration*, 3, 86–93.
- Salperwyck, C., Lemaire, V. (2011). Learning with few examples: An empirical study on leading classifiers. In Proceedings of The 2011 International Joint Conference on Neural Networks (IJCNN 2011), San Jose, California, USA, pp. 1010–1019.
- Shao, H., Marwah, M., Ramakrishnan, N. (2012). A Temporal Motif Mining Approach to Unsupervised Energy Disaggregation. In Proceedings of the 1st International Workshop on Non-Intrusive Load Monitoring, Pittsburgh, PA, USA, 7.
- Spiegel, S., Albayrak, S. (2014). Energy disaggregation meets heating control. In Proceedings of the 29th Annual ACM Symposium on Applied Computing, Gyeongju, Republic of Korea, 24–28.
- Srinivasan, D., Ng, W., & Liew, A. (2006). Neural-network-based signature recognition for harmonic source identification. *IEEE Transactions on Power Delivery*, 21, 398–405.
- Wang, Z., & Zheng, G. (2012). Residential appliances identification and monitoring by a nonintrusive method. *IEEE Transactions on Smart Grid*, 3, 80–92.
- Wong, Y.F., Ahmet Sekercioglu, Y., Drummond, T., Wong, V.S. (2013). Recent approaches to non-intrusive load monitoring techniques in residential settings. In Proceedings of the 2013 I.E. symposium on computational intelligence applications in smart grid (CIASG), Singapore, 16-19, pp. 73–79.
- Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng, A., Liu, B., Philip, S. Y., et al. (2008). Top 10 algorithms in data mining. *Knowledge and Information Systems*, 14, 1–37.
- Yang, C.C., Soh, C.S., Yap, V.V. (2014). Comparative Study of Event Detection Methods for Non-intrusive Appliance Load Monitoring. In Proceedings of the 6th International Conference on Applied Energy (ICAE2014), Taipei City, Taiwan, 30 May – 2 June 2014, Energy Procedia, pp. 1840–1843.
- Yang, C. C., Soh, C. S., & Yap, V. V. (2015). A systematic approach to ON-OFF event detection and clustering analysis for non-intrusive appliance load monitoring. *Frontiers in Energy*, *Frontiers in Energy*, 9(2), 231–237.
- Yang, C. C., Soh, C. S., & Yap, V. V. (2016). A Systematic Approach in Load Disaggregation Utilizing a Multi-Stage Classification Algorithm for Consumer Electrical Appliances Classification. *Frontiers in Energy*, Springer, (Accepted 12.October. 2016). doi:10.1007/s11708-017-0497-z.
- Yang, C. C., Soh, C. S., & Yap, V. V. (2017). A non-intrusive appliance load monitoring for efficient energy consumption based on naive Bayes classifier. *Sustainable Computing: Informatics and Systems*, 14, 34–42.
- Yoshimoto, K., Nakano, Y., Amano, Y., Kermanshahi, B. (2000). Non-intrusive appliances load monitoring system using neural networks. In ACEEE Summer Study on Energy Efficiency in Buildings, Pacific Grove, CA, USA, 20-25, pp. 183–194.
- Zeifman, M. (2012). Disaggregation of home energy display data using probabilistic approach. *IEEE Transactions on Consumer Electronics*, 58, 23–31.
- Zeifman, M., & Roth, K. (2011). Nonintrusive appliance load monitoring: Review and outlook. *IEEE Transactions on Consumer Electronics*, 57, 76–84.
- Zeifman, M., Roth, K., Stefan, J. (2013). Automatic recognition of major end-uses in disaggregation of home energy display data. In Proceedings of the IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, 11-14, pp. 104–105.
- Zhou, Z.H. , Zhan, D.C. , Yang, Q (2007). Semi-supervised learning with very few labeled training examples. Proceedings of the 22nd national conference on Artificial intelligence, p.675–680, July 22–26, 2007, Vancouver, British Columbia, Canada.
- Zoha, A., Gluhak, A., Imran, M. A., & Rajasegarar, S. (2012). Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey. *Sensors*, 12, 16838–16866.