

RESEARCH

Open Access



# A feature fusion technique for improved non-intrusive load monitoring

Raghunath Reddy<sup>1\*</sup> , Vishal Garg<sup>1</sup> and Vikram Pudi<sup>2</sup>

\*Correspondence:

[raghunath.iiit@gmail.com](mailto:raghunath.iiit@gmail.com)

<sup>1</sup>Building Science Research Centre,  
IIIT-H, Gachibowli, Hyderabad,  
India

Full list of author information is  
available at the end of the article

## Abstract

Load identification is an essential step in Non-Intrusive Load Monitoring (NILM), a process of estimating the power consumption of individual appliances using only whole-house aggregate consumption. Such estimates can help consumers and utility companies improve load management and save power. Current state-of-the-art methods for load identification generally use either steady state or transient features for load identification. We hypothesize that these are complementary features and so a hybrid combination of them will result in an improved appliance signature. We propose a novel hybrid combination that has the advantage of being low-dimensional and can thus be easily integrated with existing classification models to improve load identification. Our improved hybrid features are then used for building appliance identification models using Naive Bayes, KNN, Decision Tree and Random Forest classifiers. The proposed NILM methodology is evaluated for robustness in changing environments. An automated data collection setup is established to capture 7 home appliances aggregate data under varying voltages. Experimental results show that our proposed feature fusion based algorithms are more robust and outperform steady state and transient feature-based algorithms by at least +9% and +15% respectively.

**Keywords:** Non-intrusive load monitoring, Feature learning, Appliance identification

## Introduction

The global energy demand is rising and is a major cause for global warming and climate change. The U.S. Energy Information Administration (EIA) projects that world energy consumption will grow by nearly 50% and the energy consumed in the buildings sector will increase by 65% by 2050 (EIA 2019). Improving energy efficiency and reducing energy consumption are two important sustainability measures. Detailed appliance specific energy usage feedback would enable consumers to reduce consumption by 5-15% (Darby and et al 2006). Load monitoring is a method of determining energy consumption and operating states of individual appliances. Intrusive Load Monitoring (ILM) monitors appliance consumption using a low-end energy sensor connected to an appliance. ILM can precisely monitor and control appliances but are not cost effective. NILM or load disaggregation is an approach to estimate individual appliance energy consumption using aggregate load measurement obtained from a single energy meter. NILM is economical

as it uses single energy meter for load monitoring. Hart (1992) in 1980 first introduced the research on NILM. NILM research is gaining importance due to advancements in the area of AI, IoT, Smart meters and Smart grids (Ruano et al. 2019).

The general process of load dis-aggregation involves four stages, namely data acquisition, event detection, feature extraction, and appliance identification (Ruano et al. 2019). In *data acquisition* stage, the aggregate power consumption data is acquired for load identification. In *event detection* stage, the state transitions of appliances are detected. The NILM approaches are either event-based or non-event-based depending on whether they rely on detecting events in the aggregate power signal. The accuracy of the event detection influences the performance of appliance classification. In *feature extraction* stage, *features* are extracted from the region around the neighborhood of an event. Load signature or features can be used for appliance identification. The performance of the NILM depends on the feature extraction step. The features used in NILM are mainly classified as steady state and transient features. Steady state features are extracted from the stable states of appliance operation. Transient state features are extracted from the short-term fluctuations in power or current during appliance state transitions. In *appliance identification* stage, the goal of load identification is to determine the operating state of an appliance. Load identification can be done using optimization or pattern recognition-based approaches. Pattern recognition-based approaches are classified as supervised, semi-supervised and unsupervised. Supervised approach requires training data to learn appliance identification models. Unsupervised approach does not require training but require one-time labelling of appliances. Unsupervised approach can build model with less training data. Some of the state of art load identification approaches are discussed below.

#### **Optimization based load identification**

NILM problem can be formulated as an optimization problem which finds an optimal combination of appliance consumption that minimises the residual sum between estimated consumption and actual aggregate consumption. There are different combinatorial search methods such as genetic algorithm, segmented integer quadratic constrained programming and mixed integer linear programming (Klemenjak and Goldsborough 2016). The optimization-based techniques are computationally intensive and are practically infeasible on large number of devices (Klemenjak and Goldsborough 2016).

#### **Pattern recognition based load identification**

Supervised and unsupervised techniques have been widely applied for load identification. Supervised approach employs machine learning, deep learning and some forms of Hidden Markov Model (HMM) techniques. Machine learning based load identification algorithms such as K. nearest neighbour (KNN), Support Vector Machine (SVM), Artificial Neural Network (ANN), Random Forest (RF), Naïve Bayes model and Decision Trees have been utilized for the load identification task (Zoha et al. 2012). In recent times, deep learning algorithms are preferred for load classification tasks. In Zhang et al. (2018), the authors propose sequence-to-point convolutional neural networks to train the model. The performance improvements in standard error measures were 84% and 92%. A convolutional neural network sequence to sequence model is proposed in Chen et al. (2018) and load dis-aggregation is performed on the Reference Energy Disaggregation Data Set (REDD) dataset. In De Baets et al. (2018) a convolutional neural network is used for

load identification task, here a V-I trajectory image is used as input to CNN and the output is the type of appliance. The authors in Kelly and Knottenbelt (2015), use three deep neural network architectures for energy dis-aggregation namely a long short-term memory (LSTM), denoising autoencoders, and a network which regresses the start time, end time and average power demand of each appliance activation. Deep learning methods are highly accurate compared to other methods but they require large amounts of data to build generic models (Liu et al. 2019). Unsupervised approach makes use of clustering techniques, HMM and several of its variants. HMM-based techniques have been widely studied for load dis-aggregation (Faustine et al. 2017). HMM are non-event based or state based NILM methods. State-based approaches are limited by the need for expert knowledge to set a-prior values for each appliance state. The key problem of HMM-based approaches is their high computational complexity. As the number of appliances to dis-aggregate increase, the time complexity increases exponentially (Kim et al. 2017). There are several NILM works in literature which are not covered here due to space limitation. A comprehensive qualitative and quantitative evaluation of some of recent algorithms can be found in Nalmpantis and Vrakas (2019).

### Contributions

We propose a supervised NILM algorithm using low frequency (10Hz) active power, reactive power, apparent power and impedance steady state and transient features. The main contribution of this paper is to develop a hybrid signature using a novel feature fusion technique that utilizes steady state and macroscopic transient features. Individually, steady state features or transient state feature represent partial appliance signature. Steady state features are helpful in identifying appliances with non-overlapping power values. Macroscopic transient features are helpful for differentiating appliances with overlapping power values. Therefore, a combination of numeric steady state features and time series transient state features will result in an improved appliance signature. We fuse features by converting transient features into numeric features and then merging with steady state features. Different distance measures such as dynamic time warping, Euclidean and Mahalanobis metric are used for transformation of transient features. Improved features are then used for building appliance identification models using Naïve Bayes, KNN, decision tree and random forest classifiers. An automated data collection setup is established to capture 7 home appliances aggregate data under varying voltages to validate the robustness of NILM algorithms in changing environments. Experimental results showed that the proposed feature fusion based algorithms are robust and outperforms active power steady state and transient feature-based algorithms. There are two existing works on similar lines to that of our approach. The authors in Chang et al. (2010) have shown that combined steady and transient features would improve load identification. The study by (Chang 2012) uses the wavelet transform (WT) to improve load transients. Our approach is different from these existing works as we develop a new hybrid signature from steady and transient features before training the model. We show that our approach is better than these existing works.

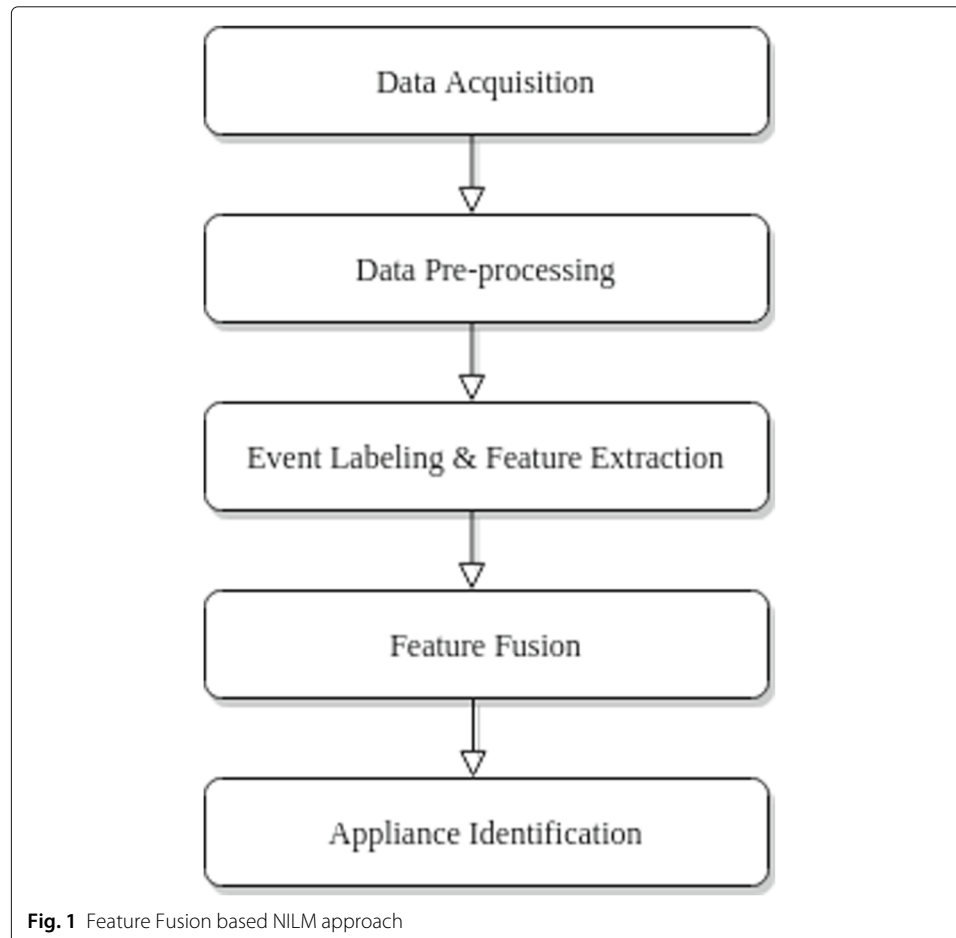
This paper is organized as follows: Next Section explains our proposed feature fusion based NILM methodology. Next “[Experimental results](#)” section discusses the details of experiments and analysis of results. Finally, we summarize the findings in the “[Conclusion](#)” section.

## Methodology

In this section, we explain our proposed NILM methodology. The objective of this study is to improve accuracy and generalization of standard machine learning NILM algorithms. The basic idea is to develop a novel feature by the fusion of steady state and transient state features. The sampling frequency is chosen in such a way that it captures both steady state and transient features. A very high frequency would usually acquire lot of data before reaching steady state. Very low frequency would miss transient information in the data. Therefore, we have chosen a sampling rate of 10 Hz in this study. The block diagram of our approach is as shown in Fig. 1. High frequency energy meters capture high resolution data at a frequency of more than 50 Hz. High frequency energy meters are expensive. The low frequency energy meter or ordinary energy meters capture low resolution data at a frequency of less than 1Hz (Basu et al. 2017). In India, supply voltages fluctuate and due to this the data distributions of appliance signatures will change affecting the performance of appliance identification techniques. Therefore, we test the robustness of NILM techniques on varying supply voltages.

## Data acquisition

The typical manner in which data is collected is cumbersome and time consuming as it requires manual intervention to generate appliance combinations. To overcome this, we

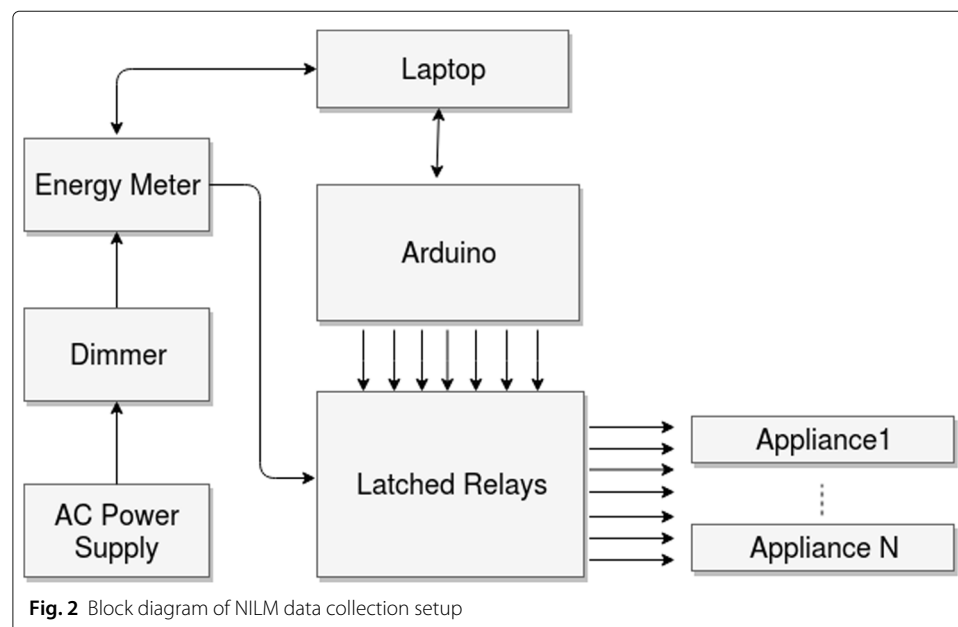


implement an automated setup to acquire NILM data. Figure 2 shows the block diagram of our setup. The Arduino micro-controller sends signals to the relays to power ON/OFF appliances. AC supply is regulated through a voltage regulator (dimmer) to vary the input source voltages. Yokogawa WT 310 digital power meter measures the aggregate NILM data. Appliances are connected to the power source through the energy meter. The aggregated data from energy meter is logged to a laptop. The setup provides aggregate NILM data of various appliance combinations in different appliance states. A seven-bit gray code sequence is fed to the micro-controller to restrict only a single device change state at a given time, usually referred as switch continuity principle (Hart 1992). This timely data generation makes it easier to label training data. The samples in a particular combination are collected for half a minute. This time duration is sufficient for home appliances to become steady after the transition.

The data of 7 Home Appliances namely Fan, Vacuum Cleaner, Geyser, Oven, Mixer, Air purifier, and Kettle was collected on 6 different source voltages 190, 200, 210, 220, 230 and 240. The data are collected at these voltages to analyze the effects of voltage variations on NILM techniques. We considered only some home appliances in the study so that it was feasible for us to generate events for different appliance combinations. We have excluded fridge and washing machine as they have long operating cycles which make it difficult to capture transient and steady state by our automated system. The features captured in the data are Voltage (V), Current (I), Active Power (P), Reactive Power (Q), and Apparent power (S). The total number of samples collected is 14,428. The number of events in the data is 702. The 127 all combinations for 6 different voltage variations result in 702 events. The data can be accessed from the following link <https://doi.org/10.6084/m9.figshare.11944932>.

### Event detection

Commonly used event-detection models are namely expert heuristic models, probabilistic models and matched-filters models. Heuristic models are based on change of standard



deviation, amplitude or cumulative sum of active power. Probabilistic models use generalized likelihood ratio or goodness of fit function for event detection. Template matching methods use Euclidean or dynamic time warping similarity measures to detect events (Anderson et al. 2012) Recently machine learning approaches are also been used for event detection (Kahl et al. 2019).

In our approach *event detection* is done manually by referring the control sequence of appliance ON/OFF used in the automated data setup. Such manually labelled event data do not miss any event and is suitable for studying the performance of appliance identification algorithms. This manual labeling cuts the dependency of learning algorithm on the event identification algorithms. Fourteen class labels are assigned to On/Off events of seven appliances. We improve the quality of data by data cleaning, feature selection and normalization. Resistance feature tends to be more stable even under fluctuations of voltage. A derived feature Resistance (R) is computed as a ratio of voltage and current. *Normalization* using reference voltage ( $V_{ref}$ ) minimizes the effect of voltage fluctuations (Hart 1992). We normalize the active power (P), apparent power (S) and reactive power (Q) using the equations given below.

$$P_{norm}(t) = (V_{ref}/V(t))^2 * P(t)$$

$$S_{norm}(t) = (V_{ref}/V(t))^2 * S(t)$$

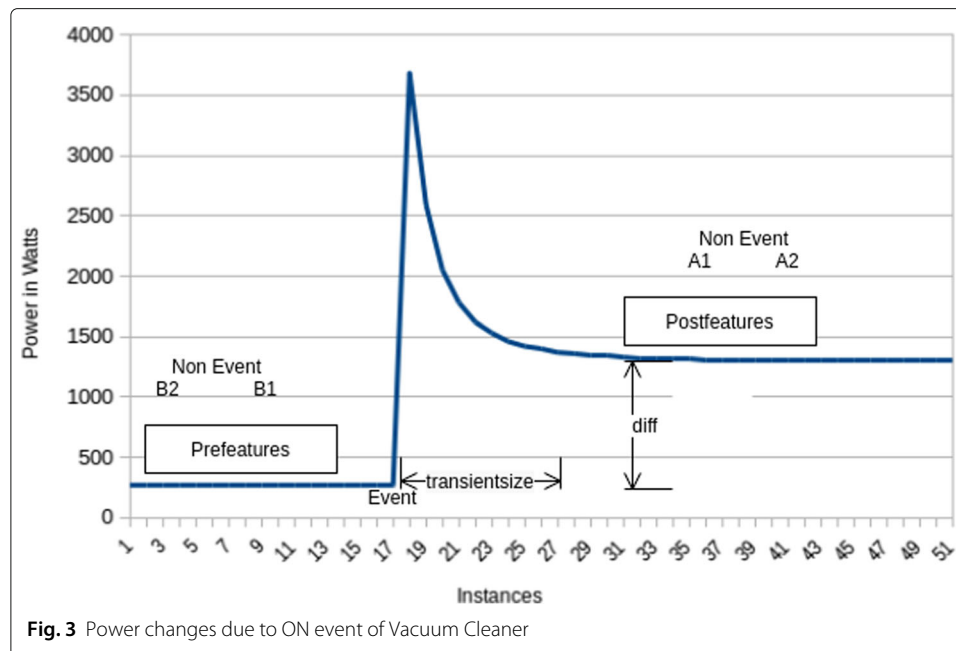
$$Q_{norm}(t) = (V_{ref}/V(t))^2 * Q(t)$$

### Feature extraction

Some of the commonly used low frequency steady state and transient features used in load identification are P-Q plane (Barsim et al. 2014), macroscopic transients, real power (Dinesh et al. 2016), current and voltage-based features. High frequency steady state and transient features used in load identification are spectral envelope, wavelets (Chang 2012; Su et al. 2011), shape features, raw wave-forms (Cole and Albicki 2000), VI trajectory (Hassan et al. 2013) etc. A comprehensive review of state of art feature extraction for NILM is discussed in Sadeghianpourhamami et al. (2017). Choosing an appropriate discriminating set of features is necessary for accurate appliance identification.

In our approach *appliance features* are extracted from the labelled event data. Steady state features are extracted by computing the difference between the feature values before and after events. Figure 3 shows the Power changes due to the ON event of the Vacuum Cleaner appliance. The difference *diff*, between A1, B1, and A2, B2 denote changes in power value after the Vacuum Cleaner ON transition. The transient of an appliance is of length *transientsize* begins at the event and ends up when the device stabilizes.

Algorithm 1 describes the process of extracting steady state and transient state features. Separate files containing appliance event data at different voltages are stored in an *all-voltsdata* folder. The *transientsize* is defined to specify the length of transient signature to be captured. Each data file is read one instance at a time and steady state features are extracted as shown in (lines 4 to 12 of Algorithm 1). The steady state difference in features is taken at different intervals before and after the event, shown as (*prefeatures* and *postfeatures*) in Fig. 3 so as to capture variations due to internal state transitions in appliances. The after event instances (*postfeatures*) are selected after the appliance becomes stable. The extracted features are written along with the appliance event label to an output *Steadystate* file corresponding to each input file. The (lines 13 to 22 of Algorithm 1) show



**Fig. 3** Power changes due to ON event of Vacuum Cleaner

how transient features are extracted. For every event, the transient features are the temporal first order difference of P, S, Q and R features in the region marked as *transientsize*. These transient features along with event label are written to a *Transientstate* file.

#### **Feature fusion algorithm**

Our feature fusion algorithm is explained in Algorithm 2. Figure 4 shows the process of feature fusion. The time-series transient data and steady state data obtained after feature extraction, A1, A2, ..., An is given as input to the feature fusion algorithm. The high dimensional time series data is transformed to equivalent low dimensional numeric features and is finally merged with steady state features. We capture the *discriminatory* features of the appliances from the transient data. We find discriminatory features by measuring the intra-class and inter-class distance of every instance in the transient data. We first compute representative sample or centroid of time series transient feature vectors of every appliance category C1, C2, ..., Cn. We then compute the distance of every instance of transient data to these centroids, for example: dA11C1 in Fig. 4 refers to the distance of transient instance 1 of appliance A1 to centroid C1, (line 6 of Algorithm 2). The time series transient features data of active power is of 702 X 46 dimensions, representing 702 ON/OFF events of 7 appliances. There are 14 classes and so 14 centroids are computed. There will be 14 distances of an instance to each of the centroids. These distances form the equivalent numeric feature vector of the transient data. Thus the 702 X 46 dimension transient feature data gets transformed into 702 X 14 dimension data. This transformed data is then merged with 702 X 1 steady state active power data to get the final 702 X 15 dimension feature fusion data representing an improved hybrid signature. (line 9 of Algorithm 2).

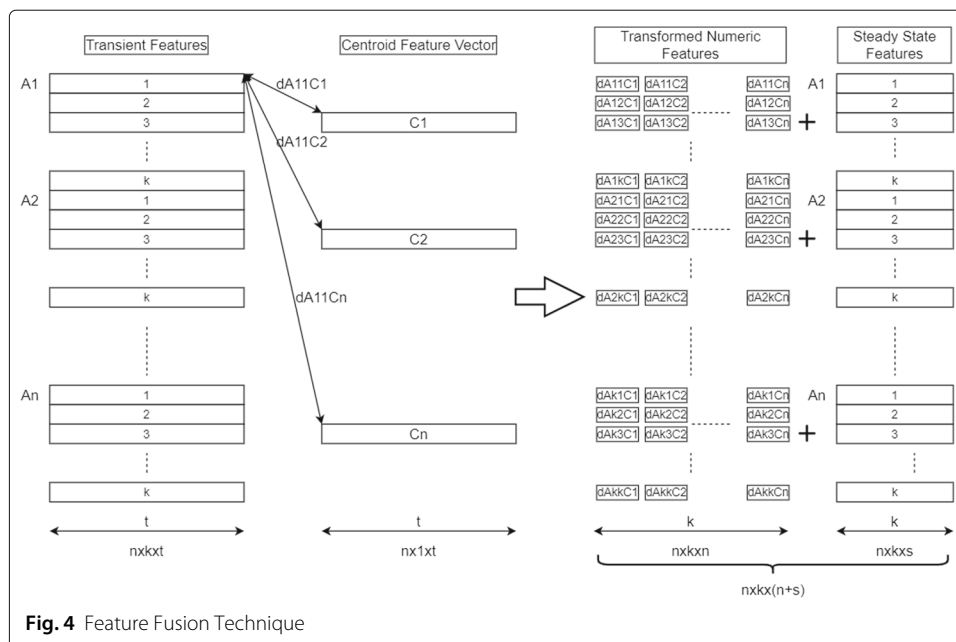
#### **Time-series distance computations for transient features**

The distance between two time series, sequences  $X = x_1, x_2, \dots, x_n$ , and  $Y = y_1, y_2, \dots, y_m$ , is computed using three different distance measures namely Euclidean,

**Algorithm 1** steady state and transient state feature extraction

```

1: Input: Appliance Event data allvoltsdata, transientsize
2: Output: Steadystate and Transientstate files
3: for file ∈ allvoltsdata do
4:   while not EOF file do // Extract steady state features
5:     Read an instance from file
6:     if is Event instance then
7:       prefeatures = P, S, Q and R features of some before event instances
8:       postfeatures = P, S, Q and R features of some after event steady state
       instances
9:       diff = prefeatures − postfeatures
10:      Write the diff along with the appliance event to steadystate file
11:    end if
12:  end while
13:  while not EOF file do // Extract transient features
14:    Read an instance from file
15:    if is Event instance then
16:      while transientsize > 0 do
17:        Write the temporal first order difference of P, S, Q and R features to
        Transientstate file
18:        transientsize = transientsize − 1
19:      end while
20:      Finally append the event label to the time series data
21:    end if
22:  end while
23: end for
    
```



**Fig. 4** Feature Fusion Technique



**Algorithm 2** Feature Fusion Algorithm

- 
- 1: **Input:** *Steadystate* and *Transientstate* files
  - 2: **Output:** Features fusion *Euclid*, *DTW* and *Maha* files
  - 3: Compute the centroid of each appliance category from the *Transientstate* file to represent the most representative sample of each appliance category
  - 4: **while** not EOF *Transientstate* file **do**
  - 5:     Read an instance from file
  - 6:     Compute distances of the read instance to each of the centroid
  - 7:     Corresponding to the distance metric used, write the transformed numeric features to the *Euclid*, *DTW* or *Maha* files respectively
  - 8: **end while**
  - 9: Merge the transformed transient features from *Euclid*, *DTW* and *Maha* files class-wise with features of *Steadystate* file
- 

Dynamic time warping and Mahalanobis. These distance metrics are analyzed for their suitability in appliance identification data. The Euclidean distance is computed as given below

$$dist(X, Y) = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2} \quad (1)$$

Dynamic Time Warping (DTW) is another widely used distance metric for comparing two time series. It has been used for analyzing temporal sequences of video, audio, graphics and also in load dis-aggregation (Liu et al. 2017). DTW tries to align two sequences in order to get the most representative distance measure. The recursive formulation for computing DTW is given by

$$\begin{aligned}
 dist(X, Y) &= D(n, m) \\
 D(i, j) &= dist(x_i, y_j) + \min\{D(i-1, j), D(i-1, j-1), D(i, j-1)\} \\
 dist(0, 0) &= 0, dist(0, j) = dist(i, 0) = \text{infinity}
 \end{aligned} \quad (2)$$

The Mahalanobis distance is based on the co-variance among variables in the feature vectors which are compared. The Mahalanobis distance groups means and variances for each variable and is scale invariant and takes into consideration correlation between features. Mahalanobis distance metric performs better than other distance metrics (Walters-Williams and Li 2010). The Mahalanobis distance between two sequences is defined as

$$dist(X, Y) = \sqrt{(X - Y)^T S^{-1} (X - Y)} \quad (3)$$

where S is the Inverse co-variance matrix of two sequences.

**Appliance identification**

The features constructed from the feature fusion algorithm are used for appliance identification. Here we use standard classification algorithms such as Naive Bayesian (NB), K-Nearest Neighbor (KNN), Decision Tree (J48) and Random Forests (RF) classifiers with 10-fold cross-validation.

## Experimental results

Experiments are designed to evaluate the effectiveness of our proposed feature fusion based NILM classifiers. We verify whether the performances improve due to feature fusion. We apply data pre-processing techniques such as normalization and data cleaning as explained in the earlier section and evaluate the effectiveness of normalization. We carry out experiments to see how feature fusion based classifiers performs compared to steady state or transient features based classifiers, using only P features and using P,S,Q and R features. In later experiments we evaluate which feature fusion signature works better with RF classifier using only P features and then with P,S,Q and R features. The transient signatures which are transformed using distance metrics such as Euclidean, DWT, Mahalanobis or discrete wavelets are used to obtain different feature fusion signatures.

## Evaluation metrics and implementation

The evaluation is done using the accuracy measure of a classifier (Hossin and Sulaiman 2015). The accuracy metric measures the sum of correct predictions divided by the total number of predictions. WEKA (Hall et al. 2009) experimenter tool is used for comparing the performance of the classifiers on the prepared data. The steady state and transient state feature extraction algorithm and feature fusion algorithm are implemented in python.

## Normalization results

We have normalized the data and the effect of normalization on steady state data and transient state data with P, S, Q and R features is shown in Fig. 5. There is good improvement in accuracy due to normalization on steady state data. The transient state data is not much affected by normalization.

## Fusion based NILM using only active power feature

In this experiment we have taken only active power (P) as the feature for training NILM classifiers. Most low end meters measure only active power. The performance results of using steady state based classifier, transient state based classifier and feature fusion based classifier using only active power (P) for NILM are compared in Fig. 6. The performance of feature fusion based classifier that uses Mahalanobis distance measures for transforming transient data, performs better as compared to separate steady state and transient state

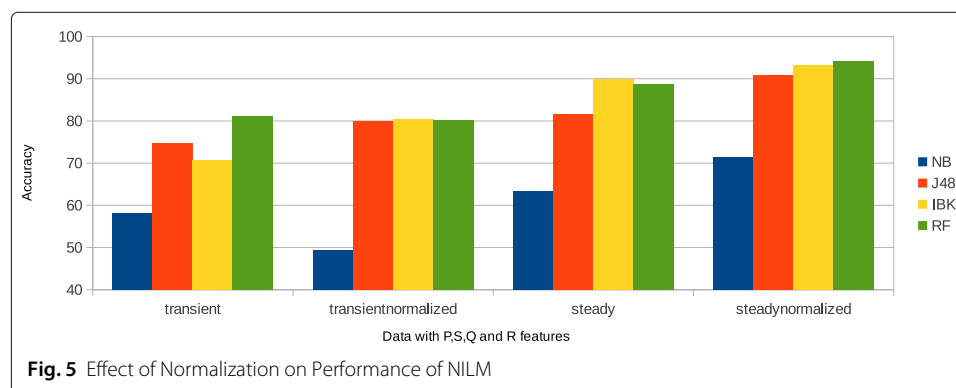
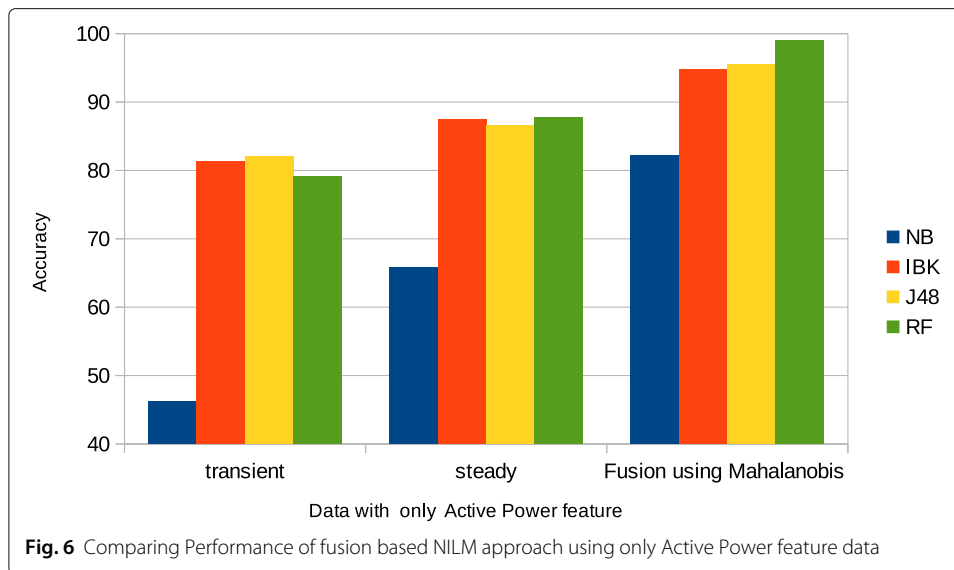


Fig. 5 Effect of Normalization on Performance of NILM



based classifiers. There is more than 8 percent improvement in the accuracy of appliance identification using fusion features in all classifiers. Notice that even when using only active power, our feature fusion technique is able to provide remarkable accuracy reaching up-to 98 percent using RF classifier.

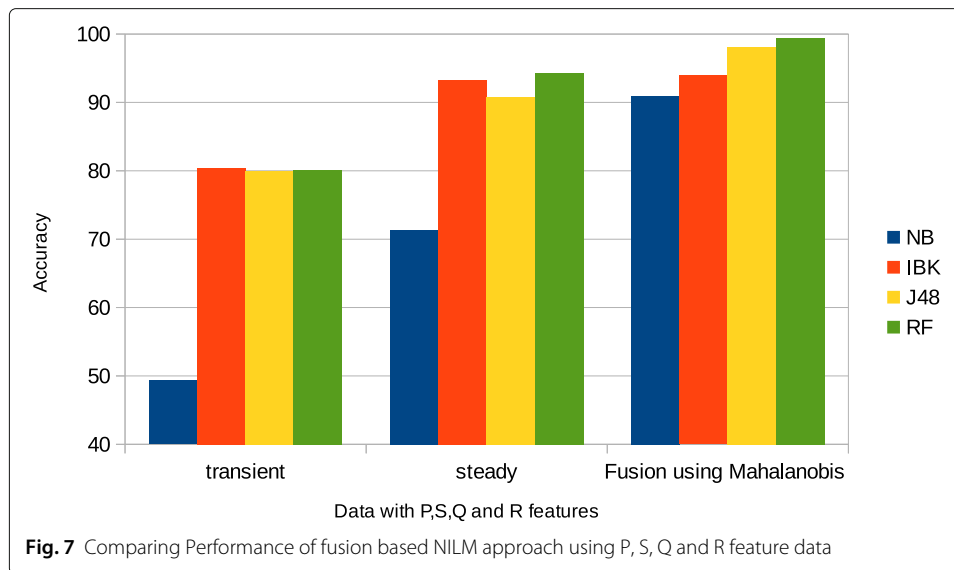
#### Fusion based NILM using P, S, Q and R features

In this experiment we have included active power (P), Apparent Power (S), Reactive Power (Q) and Resistance (R) as the features for training NILM classifiers. The performance results of steady state features based classifiers, transient features based classifiers and feature fusion based classifiers using P, S, Q and R features is compared in Figure 7. The performance of feature fusion based classifiers that uses Mahalanobis distance measures for transforming transient data, performs quite effectively as compared to separate steady state features based classifiers and transient features based classifiers. There is 4 percent improvement in accuracy of appliance identification using fusion feature approach in J48 and RF classifier. The overall improvement in accuracy with P, S, Q and R features is slightly more (1 to 2 percent) as compared to earlier experiment using only Active Power (P) feature.

#### Comparing the performances of different feature fusion based classifiers with transient feature based classifiers

In this experiment we compare different feature fusion based RF classifier and transient feature based RF classifier. The transient signatures which are transformed using distance metrics such as Eucliden, DWT, Mahalanobis or using discrete wavelets are used to obtain different feature fusion signatures. The data used in experiment are as listed below.

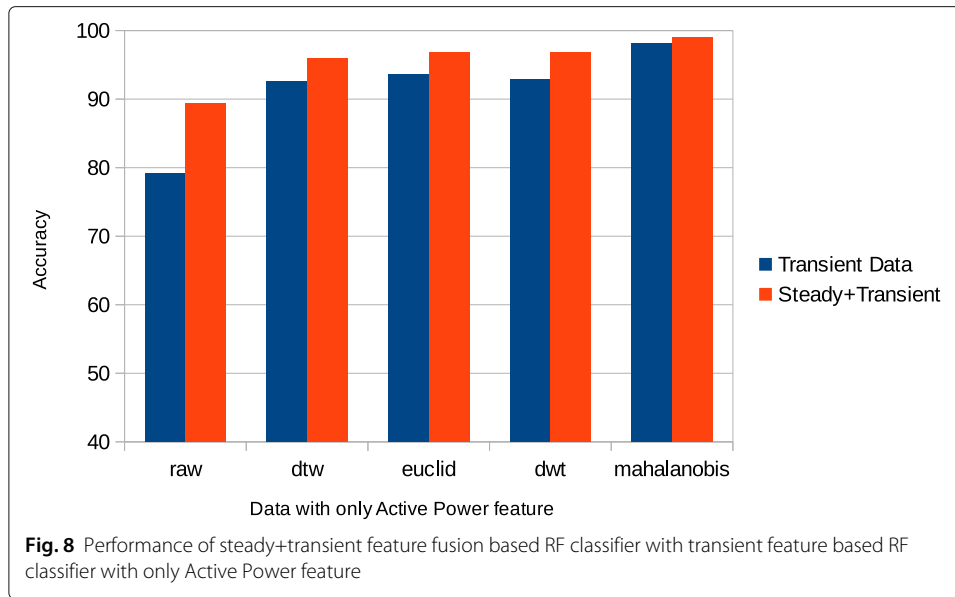
1. The raw data is the extracted transient data and steady state data, one of the existing work (Chang et al. 2010) have proposed such features.
2. The *dtw* data uses the dynamic time warping distance metric for transforming transient features



3. The *euclid* data uses the Euclidean distance metric for transforming transient features
4. The *Mahalanobis* data uses the Mahalanobis distance metric for transforming transient features
5. The *dwt* data uses single level discrete wavelet transform using Daubechies (db2) family wavelet followed by application of Principal Component Analysis (PCA). Existing work by Chang (2012) use this kind of feature extraction technique on transient signatures.

In Fig. 8, the first bar represents transient feature based RF classifier whereas the second represent feature fusion (steady+transient) based RF classifier. The feature fusion based RF classifier always performs better than the transient feature based RF classifier. With only active power (P) as the feature for training RF classifier the Fig. 8 shows the increasing order of accuracy on different signatures. The performance on *Mahalanobis* data is the best followed by *dwt* data. There is about 8 percent improvement in accuracy of appliance identification when compared with *raw*, Clearly indicating that our approach is better as compared to the existing approaches that use either *raw* features or *dwt* features. Using P,S,Q and R features for training, Fig. 9 shows the increasing order of accuracy of Random Forest classifier on different signatures. The performance on *Mahalanobis* data is the best followed by *dwt* data. There is 8 percent improvement in accuracy of appliance identification when compared with *raw*. Using all features slightly improve the accuracy about 1 to 2 percents as compared to earlier active power alone.

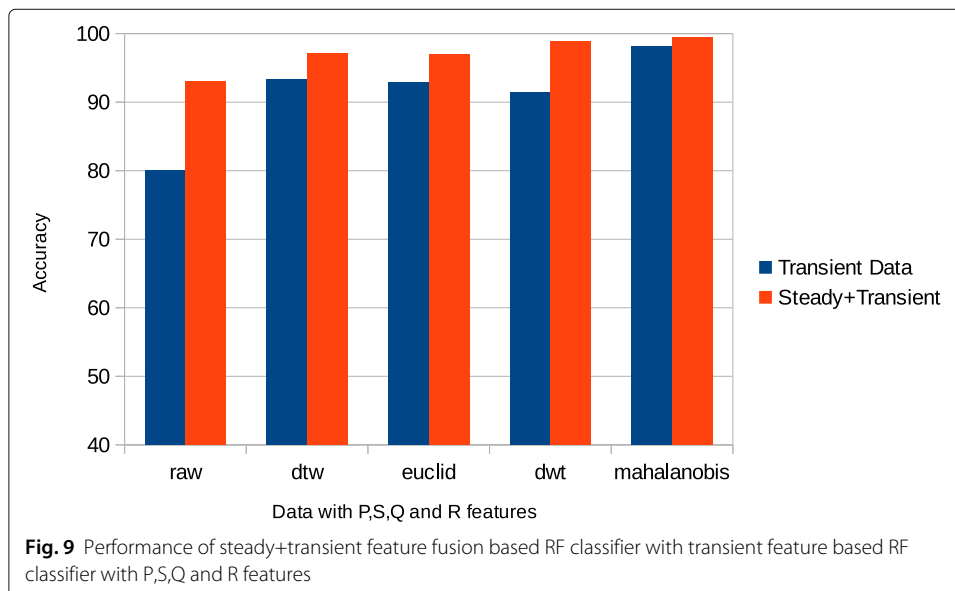
In all the above experiments the random forest (RF) and K-NN algorithms perform equally better and have higher accuracy. The Naive Bayesian classifier tends to perform poorly. Also the performance of NILM on only active power is sufficiently close to that using all features P, S, Q, and R data. Normalization helps in improving the accuracy. The performance on *Mahalanobis* data is the best followed by *dwt* data. The Mahalanobis distance performs best as it captures the co-relation among the features. There is almost 8 percent improvement in accuracy of appliance identification when compared with *raw* transients. The experimental results clearly establish feature fusion technique are better



than ordinary steady state or transient features in improving the performance of load identification.

**Conclusion**

There is a need for improved NILM algorithms which can accurately identify loads for effective energy management. Supervised NILM techniques are either using a steady state or transient state features. A novel fusion based Non-Intrusive load monitoring algorithm is proposed that combines the goodness of steady state feature and transient state features. Feature extraction and feature fusion algorithms are described. The data collection setup is established to automatically capture 7 home appliance aggregate data under varying voltages. This data helps in testing the robustness of NILM algorithms in changing



voltage environments. We provide the experimental results comparing the fusion based NILM with steady state feature NILM and transient state feature NILM. The experimental results validate that the novel fusion based algorithms outperforms the separate steady state and transient algorithms. It is also observed that the Mahalanobis distance used for transforming transient data is most effective for feature construction. We have shown that our approach performs better than two of the existing approaches that were some what similar to our work. Feature fusion technique can be applied to other supervised algorithms for improving NILM systems. There is a scope for comparing our algorithm on different data-sets and algorithms.

#### Abbreviations

NILM: Non-intrusive load monitoring; ILM: Intrusive load monitoring; SVM: Support vector machine; K-NN: K-Nearest neighbour; STFT: Short-time fourier transform; DWT: Discrete wavelet transform; DTW: Dynamic time warping; NB: Naive Bayesian; RF: Random forests; J48: Decision tree; PCA: Principal component analysis

#### Acknowledgements

We would like to acknowledge our thanks to Mr. Prabhakar, Senior Project Assistant at BSRC, IIIT-H for his help in establishing the NILM data collection setup.

#### Authors' contributions

RR, VG and VP contributed to the conception and experimental analysis of this research work. RR was the primary author of the manuscript under the guidance of VG and VP. All authors contributed equally to this work. All authors read and approved the final manuscript.

#### Funding

Not Applicable.

#### Availability of data and materials

The NILM dataset supporting the conclusions of this article is available in the repository at <https://doi.org/10.6084/m9.figshare.11944932>.

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>Building Science Research Centre, IIIT-H, Gacchibowli, Hyderabad, India. <sup>2</sup>Data Science and Analytics Centre, KCIS, IIIT-H, Gacchibowli, Hyderabad, India.

Received: 10 March 2020 Accepted: 6 August 2020

Published online: 08 September 2020

#### References

- Anderson KD, Bergés ME, Ocneanu A, Benitez D, Moura JM (2012) Event detection for non intrusive load monitoring. In: IECON 2012-38th Annual Conference on IEEE Industrial Electronics Society. IEEE, Montreal. pp 3312–3317
- Barsim KS, Streubel R, Yang B (2014) An approach for unsupervised non-intrusive load monitoring of residential appliances. In: Proceedings of the 2nd International Workshop on Non-Intrusive Load Monitoring 2014 Jun 3. University of Texas, Austin. pp 1–5
- Basu K, Debusschere V, Bacha S, Hably A, Van Delft D, Dirven GJ (2017) A generic data driven approach for low sampling load disaggregation. *Sustain Energy Grids Netw* 9:118–127
- Chang HH (2012) Non-intrusive demand monitoring and load identification for energy management systems based on transient feature analyses. *Energies* 5:4569–4589
- Chang HH, Lin CL, Lee JK (2010) Load identification in nonintrusive load monitoring using steady-state and turn-on transient energy algorithms. In: The 2010 14th International Conference on Computer Supported Cooperative Work in Design. IEEE, Shanghai. pp 27–32
- Chen K, Wang Q, He Z, Chen K, Hu J, He J (2018) Convolutional sequence to sequence non-intrusive load monitoring. *J Eng* 2018(17):1860–1864
- Cole A, Albicki A (2000) Nonintrusive identification of electrical loads in a three-phase environment based on harmonic content. In: Proceedings of the 17th IEEE Instrumentation and Measurement Technology Conference [Cat. No. 00CH37066]. IEEE, Baltimore. pp 24–29
- Darby S, et al (2006) The effectiveness of feedback on energy consumption. A Review for DEFRA of the Literature on Metering. *Billing Direct Displays* 486(2006):26
- De Baets L, Ruyssinck J, Devellder C, Dhaene T, Deschrijver D (2018) Appliance classification using vi trajectories and convolutional neural networks. *Energy Build* 158:32–36
- Dinesh C, Godaliyadda RI, Ekanayake MPB, Ekanayake J, Perera P (2016) Non-intrusive load monitoring based on low frequency active power measurements. *AIMS Energy* 4(3):414–443
- EIA (2019) International energy outlook. <https://www.eia.gov/outlooks/ieo/pdf/ieo2019.pdf>. Accessed 15 July 2020

- Faustine A, Mvungi NH, Kaijage S, Michael K (2017) A survey on non-intrusive load monitoring methodologies and techniques for energy disaggregation problem. arXiv preprint arXiv:170300785
- Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH (2009) The weka data mining software: An update. *SIGKDD Explor Newsl* 11:10–18
- Hart GW (1992) Nonintrusive appliance load monitoring. *Proc IEEE* 80:1870–1891
- Hassan T, Javed F, Arshad N (2013) An empirical investigation of vi trajectory based load signatures for non-intrusive load monitoring. *IEEE Trans Smart Grid* 5:870–878
- Hossin M, Sulaiman M (2015) A review on evaluation metrics for data classification evaluations. *Int J Data Min Knowl Manag Process* 5:1
- Kahl M, Kriechbaumer T, Jorde D, Ul Haq A, Jacobsen HA (2019) Appliance event detection—a multivariate, supervised classification approach. In: *Proceedings of the Tenth ACM International Conference on Future Energy Systems*. Association for Computing Machinery, New York. pp 373–375
- Kelly J, Knottenbelt W (2015) Neural nilm: Deep neural networks applied to energy disaggregation. In: *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*. Association for Computing Machinery, New York. pp 55–64
- Kim J, Le TTH, Kim H (2017) Nonintrusive load monitoring based on advanced deep learning and novel signature. *Comput Intell Neurosci*:2017. <https://doi.org/10.1155/2017/4216281>
- Klemenjak C, Goldsborough P (2016) Non-intrusive load monitoring: A review and outlook. arXiv preprint arXiv:161001191
- Liu B, Luan W, Yu Y (2017) Dynamic time warping based non-intrusive load transient identification. *Appl Energy* 195:634–645
- Liu Q, Kamoto KM, Liu X, Sun M, Linge N (2019) Low-complexity non-intrusive load monitoring using unsupervised learning and generalized appliance models. *IEEE Trans Consum Electron* 65(1):28–37
- Nalmpantis C, Vrakas D (2019) Machine learning approaches for non-intrusive load monitoring: from qualitative to quantitative comparison. *Artif Intell Rev* 52:217–243
- Ruano A, Hernandez A, Ureña J, Ruano M, Garcia J (2019) Nilm techniques for intelligent home energy management and ambient assisted living: A review. *Energies* 12:2203
- Sadeghianpourhamami N, Ruyssinck J, Deschrijver D, Dhaene T, Develder C (2017) Comprehensive feature selection for appliance classification in nilm. *Energy Build* 151:98–106
- Su YC, Lian KL, Chang HH (2011) Feature selection of non-intrusive load monitoring system using stft and wavelet transform. In: *2011 IEEE 8th International Conference on e-Business Engineering*. IEEE, Beijing. pp 293–298
- Walters-Williams J, Li Y (2010) Comparative study of distance functions for nearest neighbors. In: Elleithy K (ed). *Advanced Techniques in Computing Sciences and Software Engineering*. Springer, Dordrecht. pp 79–84
- Zhang C, Zhong M, Wang Z, Goddard N, Sutton C (2018) Sequence-to-point learning with neural networks for non-intrusive load monitoring. In: *Thirty-second AAAI conference on artificial intelligence*. Cornell University
- Zoha A, Gluhak A, Imran MA, Rajasegarar S (2012) Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey. *Sensors* 12(12):16,838–16,866

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)

---