Performance Analysis of Similar Appliances Identification using NILM Technique under Different Data Sampling Rates

R. Gopinath
Energy Management System Division, CSIR-Central Scientific Instruments Organisation (CSIO), Chennai, India
rgopinath@ieee.org

Mukesh Kumar*
Energy Management System Division, CSIR-Central Scientific Instruments Organisation (CSIO), Chennai, India
kumarmukesh@csio.res.in

K. J. Lokesh
Energy Management System Division, CSIR-Central Scientific Instruments Organisation (CSIO), Chennai, India
lokeshkj006@gmail.com

Kota Srinivas†
Energy Management System Division, CSIR-Central Scientific Instruments Organisation (CSIO), Chennai, India
srinivaskota@csircmc.res.in

ABSTRACT

Similar power/similar loads identification using a non-intrusive load monitoring (NILM) system is a challenging task due to overlapping characteristics of similar appliances. Therefore, we need to identify robust features that are capable of distinguishing the events of similar appliances effectively. Further, the developed NILM system/methodology need to be examined for different sampling rates, since the performance evaluation of the algorithm with single sampling rate alone is not sufficient. In this work, we used four compact fluorescent lamps (CFL) which are having same specifications configured in different electrical network combinations. We explored locality constrained linear coding (LLC) based deep neural networks (DNN) to express the input feature vectors in terms of load independent basis vectors in a higher dimensional feature space to make the features robust for effective identification of similar loads in a NILM system. Further, we examined the effectiveness of LLC-DNN approach for the different sampling rates, 10 kHz, 25 kHz and 50 kHz respectively. From our study, it is observed that the performance of the LLC-DNN has been consistent and improves the similar appliances identification accuracy significantly with the increase in sampling rates.

CCS CONCEPTS

• **Computing methodologies** → **Neural networks**; **Support vector machines**; • **Hardware** → **Energy metering**.

KEYWORDS

Non-intrusive load monitoring (NILM), similar loads, energy management, locality constrained linear coding (LLC), deep neural networks (DNN)

*Also with Academy of Scientific and Innovative Research (AcSIR), Ghaziabad, India.
†Also with Academy of Scientific and Innovative Research (AcSIR), Ghaziabad, India.

NILM'20, November 18, 2020, Virtual Event, Japan

ACM ISBN 978-1-4503-8191-8/20/11…$15.00
https://doi.org/10.1145/3427771.3427858

ACM Reference Format:
R. Gopinath, Mukesh Kumar, K. J. Lokesh, and Kota Srinivas. 2020. Performance Analysis of Similar Appliances Identification using NILM Technique under Different Data Sampling Rates. In *The 5th International Workshop on Non-Intrusive Load Monitoring (NILM'20), November 18, 2020, Virtual Event, Japan.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3427771.3427858

1 INTRODUCTION

The recent advancements in smart metering infrastructure (AMI) has made significant developments in energy management of electrical utilities and renewable energy resources in buildings [3, 11]. Recently, smart meters are used in buildings to provide real-time energy consumption feedback to consumers continuously or in regular time intervals [1, 12]. Many studies revealed that the feedback of total energy consumption of building does not impact significantly on the energy usage behavior of consumers towards reducing the energy consumption or energy wastage. Subsequently, important loads/appliances are monitored through separate sensors for each appliances to get insights about the energy consumption of the appliances. However, this approach is expensive and not feasible in buildings due to the requirements of space, cables and sensors for every loads. To overcome the drawbacks in the intrusive approach, non-intrusive load monitoring (NILM) technique has been introduced for monitoring the events and energy consumption of the appliances from a single energy meter which is connected to the entry point of the main power supply in buildings [8].

NILM system performs appliance identification and energy disaggregation from the aggregated energy acquired through single energy meter. Though the early developments in NILM methodologies are progressed in a slower pace, the recent advancements in the smart metering infrastructure (AMI), smart grid and machine learning and deep learning algorithms have made significant improvements in the NILM system performance [6]. Despite the recent improvements in NILM methodologies [6, 15], there exists many challenges towards developing accurate NILM systems [13]. NILM system performance depends on the factors such as appliance type, number of similar power appliances and same appliances in the network, sampling rate, etc. Further, many real-time challenges need to be addressed when there is a change in the list of appliances (newly added/replaced with some other) and appliance performance degradation over its ageing in the monitoring environment.

Generally, use of multiple similar loads in commercial/residential buildings are common. Therefore, effective identification of similar loads is essential in order to develop more accurate NILM system [6]. Gupta et al., experimented with four same model compact fluorescent lamps (CFL) from the same manufacturer for similar appliance identification using electromagnetic interference (EMI) signals [7]. Gupta et al., reported that the magnitude of EMI signal changes as the location of lamp changes in the electrical network. Further, the spectra of four same lamps were not similar and there was a shift in the spectra of lamps, but the stable mean and variance of the Gaussian have been observed. Consequently, effective identification of similar appliances becomes challenging when they are located close to each other [2, 7]. Therefore, we need to identify robust features and methodologies for effective identification of similar appliances.

Sampling rate of the acquired aggregated energy plays an important role in the performance of NILM systems. Many researchers have adapted different sampling rates from the range of lower sample (in Hz) to higher sample rates (in kHz/MHz) for developing NILM system [1]. However, it is noted that most of the NILM algorithms in the literature are evaluated for single sampling rate only. Huang et al. suggested that reporting of the algorithm performance for a single sampling rate may not be sufficient for thorough performance evaluation of algorithms [9]. The study also reveals that the computation time increases as the sample rate increases, and appliance classification accuracy decreases non-linearly as the sample rate decreases [10, 14]. Therefore, we need to trade-off between the NILM system performance and computation time, and sample rates for developing effective hardware design configuration of NILM device. This necessitates to evaluate algorithm performance under different sampling rates for developing effective NILM system.

In this work, first we explore feature mapping based approach to make input features robust across similar appliances in NILM system. We use four CFL lamps having same technical specification that are from same manufacturer. The lamps are placed close to each other in the electrical network and they are turned on and off individually and simultaneously in the 16 possible configurations. Fundamental features are extracted and then expressed in terms of load independent basis vectors using a feature mapping technique, locality constrained linear coding (LLC) for effective similar appliances classification. Machine learning classifiers, support vector machine (SVM) and deep neural networks (DNN) are used for appliance detection. From the literature [4, 5, 18], it is noted that LLC has been experimented with linear SVM for its best performance in different applications due to the linear coding process. Further, it would be interesting to explore LLC features for capturing the signature of the similar appliances in NILM system with the DNN classifier that comprises of rectified linear unit (ReLU) in the initial layers and softmax activation function for the output layer. Therefore, SVM and DNN are chosen as the appropriate classifiers for experimenting with LLC features. Second, we analyse the reported feature mapping based approach for non-intrusive similar load identification under different sampling rates to check its effectiveness and therefore it would help in the hardware design of NILM system for product development. In the subsequent sections, data acquisition, feature extraction, methodology, and experiments and results are discussed. The last section concludes the paper.

2 DATA ACQUISITION AND FEATURE EXTRACTION

For similar load identification, we used four CFL lamps of having same technical specifications from the same manufacturer. Though the lamps are identical in specifications, it would be appropriate to term the loads as "similar" over identical since the lamps might be nanometrically distinct. The specifications of lamps are listed in Appendix. For data collection, four lamps are connected close to each other and acquired the aggregated voltage and current signals in sixteen possible electrical network configurations. In our test bench, we acquired the data at 10 kHz and 50 kHz respectively for evaluating the performance of NILM system. A single phase high-speed measurement setup for voltage and current has been designed using high speed ADC (ADS9224R), Magnelab 20 Amps current transformers and Raspberry Pi. Initially, 230 V has been stepped down to 2.3 V using resistor divider circuit, to fed into two channel ADC board for voltage and current measurement. Signals from the ADC has been transferred to Raspberry Pi using serial peripheral interface (SPI) communication to store in the database. The hardware setup/test bench for the experiments is shown in Figure 1.

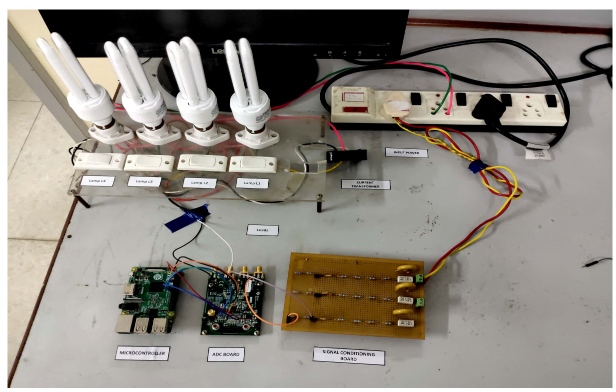

Figure 1: Hardware setup for non-intrusive load monitoring of similar appliances

The voltage and current signals are collected at five time intervals (0, 5, 10, 15, and 20 mins) to incorporate heating effects of lamps. For each time interval 2400 cycles of power signals are acquired. To complete the data collection, this process has been repeated for sixteen possible electrical network configurations by turning on and off the four lamps individually and simultaneously. Subsequently, feature extraction and appliance identification are carried out in Raspberry Pi for real-time monitoring of similar appliances in the NILM system. In this work, features such as V_{rms}, I_{rms}, real power (P), reactive power (Q), Power factor (PF), apparent power (S), I_{min}, I_{max}, I_{std}, $I_{skewness}$, V_{thd}, and I_{thd} are extracted for load identification. Further, we also use the similar dataset available in the public domain [16] to check the effectiveness of the NILM system. This dataset has been created by extracting fundamental features from the signals acquired at the sample rate of 25 kHz from the four CFL lamps having same technical specifications. The data collection procedure of this dataset is also same as discussed in our dataset. We choose this dataset for examining the NILM methodology for different sampling rates. The detailed information of this dataset can be found in this paper [16]. Further, this open source data set comprises of important fundamental features rather than the raw signals. This dataset has the features such as P, Q, S, PF, Irms, and first nine odd harmonic components for the analysis of similar load identification. In this paper, we develop feature mapping based approach for effective identification of similar appliances in the NILM system. Further, we examine the methodology under different sampling rates, 10 kHz, 25 kHz and 50 kHz to check its effectiveness/robustness for designing the better NILM system.

3 METHODOLOGY

Feature mapping/feature transformation techniques plays an important role in making the features robust across the appliances in NILM system. Generally, the extracted features exhibits non-linear characteristics in the input feature space due to the multiple number of similar power or similar appliances. Therefore, non-linear features may not fit effectively with available kernels. To overcome this problem, feature mapping approaches have been explored in different applications (e.g., image recognition) for learning the non-linear features. Sparse coding has been used widely in many applications to represent the input non-linear features in terms of linear combination of basis vectors. A deep sparse coding based approach has been experimented for NILM system [17]. However, many studies revealed that the sparse coding is not computationally efficient and it may select different basis for similar input feature vectors in order to favor sparsity in the feature coding process. This results in losing correlation between codes.

Wang et al., introduced locality constrained linear coding (LLC) to overcome the limitations in sparse coding [18]. LLC explicitly chooses k-nearest basis for similar feature vectors. Though LLC has been introduced in image recognition application, in this work, we explore the properties of LLC in non-intrusive load monitoring of similar appliances for improving the performance of NILM system. LLC learns basis vectors (dictionary) using Linde-Buzo-Gray (LBG) algorithm in a unsupervised manner. LLC represents the input feature vectors in terms of k-nearest basis vectors or codebooks. In this coding process, first codebooks are obtained and then features are represented in terms of k-nearest basis vectors. Moreover, LLC uses only few elements (sparse) to express the features. Figure 2 illustrates the graphical representation of LLC. The codebook $B = [b_1, b_2, ..., b_M] \in \mathbb{R}^{D \times M}$, is computed from the D-dimensional input feature vector $X = [x_1, x_2, ..., x_N] \in \mathbb{R}^{D \times N}$ using LBG algorithm. Consequently, input feature vectors are mapped into a M-dimensional code. The obtained codebook captures the inherent structures of input features. LLC can be expressed as:

$$\min_{\tilde{C}} \sum_{i=1}^{N} \|X_i - \tilde{c}_i B_i\|^2 \tag{1}$$

$$s.t\ 1^T \tilde{c}_i = 1,\ \forall i$$

where $C = [c_1, c_2, ..., c_M]$ represents codes. The shift invariant requirements in the LLC has been ensured by the constraint $1^T \tilde{c}_i = 1$. LLC has been solved using constrained least square fitting problem and thereby minimizing the error between mapped features and

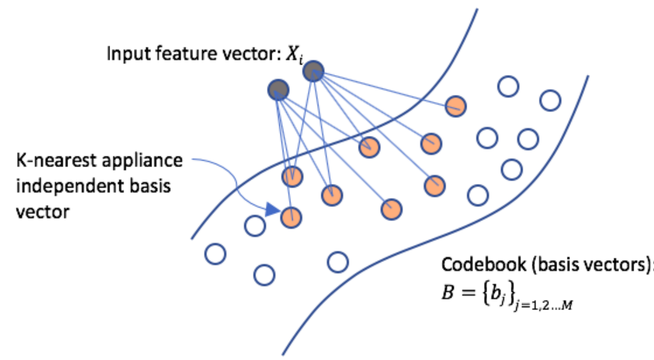

Figure 2: LLC: Represents features in terms of appliance independent basis vectors (codebooks)

input feature vectors. LLC chooses k-nearest local basis to represent the input features. Hence, LLC reduces the computation complexity to $O(M + k^2)$ from $O(M^2)$ (Sparse coding). In this paper, the input features extracted from the raw signals are expressed in terms of linear combination of k-nearest basis to obtain the appliance independent features that would help improve the identification of similar appliances in NILM systems.

4 EXPERIMENTS AND RESULTS

In this section, we first discuss the development of baseline system for similar load identification using SVM and DNN classifiers. Subsequently, performance improvement using a feature mapping technique, LLC with the back-end classifiers are discussed. All the experiments are conducted using the dataset acquired at different sampling rates, 10 kHz (dataset I), 25 kHz [16] (dataset II), and 50 kHz (dataset III) respectively to check the effectiveness of NILM system. Dataset acquired at 10 kHz and 50 kHz are created from our experimental setup. From the dataset I and dataset III, 12 features are extracted including the fundamental electrical and statistical features as discussed in Section 2. Similarly, dataset II has 14 fundamental electrical features [16].

4.1 Baseline system

For baseline system, input features extracted from the datasets are given as a input to the machine learning classifiers, SVM and DNN respectively. In SVM, experiments are carried out with different kernels, linear, radial basis function (RBF), and polynomial to classify similar appliances in NILM system. SVM constructs the hyperplane that separates the data samples either in original feature space using linear kernel or in the higher dimensional space using non-linear kernels. In this work, we have formulated as a 16 class problem since the data has been acquired from the four CFL lamps in 16 possible electrical network configurations by turning on and off individually and simultaneously. The baseline system performance with SVM kernels for three different sampling rates has been listed in Table 1. From the experiments, it is observed that the overall classification accuracy has been improved significantly for 50 kHz data samples when compared with 10 kHz and 25 kHz datasets. This trend has been consistent with the linear and RBF kernels.

However, the performance of polynomial kernel shows that there exists a non-linear relationship between the sampling rates and overall classification performance.

Table 1: Baseline system performance using input features under different sampling rates

Classifier	Parameters/ DNN layers	Overall Accuracy 10 kHz	25 kHz	50 kHz
SVM	Linear	41.63%	44.19%	**72.26**%
	RBF	46.40%	48.73%	62.23%
	Polynomial	43.56%	67.55%	60.60%
DNN	(512,256,64,16)	60.40%	69.32%	**80.37**%

Subsequently, experiments are carried out for DNN with the input features. DNN is capable of capturing the linear and non-linear relationships between the features and targets effectively. The fully-connected network has been used for multiple number of layers. Rectified linear unit (ReLU) and softmax activation function are used for initial layers and output layers respectively. Other parameters such as Adam optimizer, cross entropy loss function, and epochs (200 Nos.) are used for training the model. Backpropagation algorithm has been used for training DNN. The experiments are carried out with different parameters and selected the optimal parameters empirically for its performance. Table 1 lists the baseline system performance of DNN for similar appliance identification under different sampling rates. The experimental results show that the performance of DNN increases with the increase in sampling rates. DNN achieves overall accuracy of 80.37% for 50 kHz sampling rate. From the performance analysis of baseline system, it is noted that the similar appliance identification accuracy has been improved significantly with the increase in the sampling rates of the data. However, still the performance of the baseline system need to be improved for more accurate NILM systems.

4.2 Performance improvement using feature mapping based approach

The baseline system performance reveals that the lower sampling rate data is not much effective for similar appliance identification in NILM system. As the sampling rate increases, the performance of the baseline system increases, since the features extracted at higher resolution captures the signature effectively. Despite the higher sampling data, still the events of similar appliances are not accurately distinguishable due to the non-linear characteristics in the feature space. To overcome this problem, the input features are expressed in terms of appliance independent basis vectors (codebooks) using the feature mapping technique, LLC for making the features effectively distinguishable in new feature space. In this experiment, fundamental and statistical features extracted from the datasets are used to compute the appliance independent basis vectors. These basis vectors captures the inherent information about the similar appliance which helps to identify the events of multiple similar appliances. Codebook is computed from the training dataset, and then the input features are represented in terms of k-nearest appliance independent basis vectors. During this process, input features in the original space are mapped in to a new higher dimensional feature space, where the events of similar appliances are effectively classified. The computed codebook is common to test dataset. Therefore, the incoming test data samples will be represented using k-nearest basis vectors and then fed to the machine learning classifiers, SVM and DNN respectively. The experiments are performed for different codebook sizes and k-nearest neighbors to obtain the best classification performance. It is observed that the LLC represents the non-linear features in a higher dimensional new feature space, where the non-linear characteristics of similar appliances are classified effectively using linear kernel. Similarly, LLC mapped features helps rectified linear unit (ReLU) in DNN layers to learn the inherent patterns in the similar appliances effectively. From the experiments and results, it is observed that LLC based deep neural networks (LLC-DNN) improved the performance of similar appliance identification significantly when compared to the baseline, LLC-SVM and other approaches listed in [2]. The parameters for the SVM and DNN are chosen empirically for its best performance. The performance of similar appliances identification using LLC under different sampling rates are listed in Table 2. LLC-DNN achieves the overall accuracy of 96.57% for 50 kHz dataset with the DNN layers of (512, 256, 128, 64, 16) which is a fully connected network.

Table 2: Improved system performance using LLC features under different sampling rates

Classifier/ Parameters	Sample rate	Codebook size	kNN	Overall Accuracy
LLC-SVM	10 kHz	512	2	57.60 %
(Linear kernel)	25 kHz	512	28	80.05 %
	50 kHz	512	4	**93.16** %
LLC-DNN	10 kHz	512	60	72.90 %
(512,256,128,64,16)	25 kHz	512	118	84.47 %
	50 kHz	512	55	**96.57** %

5 CONCLUSIONS

In this work, we presented locality constrained linear coding (LLC) technique for non-intrusive load monitoring (NILM) application to represent non-linear features in terms of appliance independent basis vectors for making the events of similar appliances effectively distinguishable in a new higher dimensional feature space. The machine learning classifiers, support vector machine (SVM) and deep neural networks (DNN) are used for similar appliances identification. The experiments and results show that the LLC-DNN based approach outperforms the performance of the baseline and LLC-SVM systems. Further, we also examined the effectiveness of the NILM approaches under three different sampling rates, 10 kHz, 25 kHz, and 50 kHz respectively. The performance analysis of this study shows that the similar appliances identification accuracy increases as the sampling rate increases. Moreover, the performance of the reported NILM approach remains consistent and improved significantly with the increase in sampling rates. Therefore, the study of NILM approach for similar appliances identification under different sampling rates would help trade-off between the NILM system performance and sampling rates for selecting the optimal hardware specifications for product development of NILM device.

ACKNOWLEDGMENTS

The authors would like to thank Council of Scientific and Industrial Research (CSIR), India for the financial support through the grant under Niche High Science/High Technology theme (Grant No. MLP 0054).

APPENDIX

CFL Specifications	4 Lamps (ours)	4 Lamps ([16])
Active power	14 W	15 W
Rated voltage	230 V	127 V
Electric current	65 mA	190 mA
Power factor	0.85	≥0.5
Light efficiency	51.42 lm/W	56.3 lm/W
Maximum enclosure temp	70°C	85°C
Ambient temperature range	5 to 45°C	5 to 45°C
Nominal supply frequency	50/60 Hz	50/60 Hz
Time to reach luminous flux	40 s (for 60%)	60 s (for 80%)
Luminous flux	720 lm	844 lm
Average life	6000 h	6000 h

REFERENCES

[1] K Carrie Armel, Abhay Gupta, Gireesh Shrimali, and Adrian Albert. 2013. Is disaggregation the holy grail of energy efficiency? The case of electricity. *Energy Policy* 52 (2013), 213–234.

[2] Rayana Kristina Schneider Barcelos, Wanderley Cardoso Celeste, Luís Otávio Rigo Júnior, and Gisele de Lorena Diniz Chaves. 2019. Identification of Similar Loads for Electric Power Management in Smart Grid. *IEEE Latin America Transactions* 17, 08 (2019), 1318–1325.

[3] Nipun Batra, Amarjeet Singh, Pushpendra Singh, Haimonti Dutta, Venkatesh Sarangan, and Mani Srivastava. 2014. Data driven energy efficiency in buildings. *arXiv preprint arXiv:1404.7227* (2014).

[4] R Gopinath, C Santhosh Kumar, and KI Ramachandran. 2017. Fisher vector encoding for improving the performance of fault diagnosis in a synchronous generator. *Measurement* 111 (2017), 264–270.

[5] R Gopinath, C Santhosh Kumar, K Vishnuprasad, and KI Ramachandran. 2015. Feature mapping techniques for improving the performance of fault diagnosis of synchronous generator. *International journal of prognostics and health management* 6, 2 (2015), 12.

[6] R. Gopinath, Mukesh Kumar, C. Prakash Chandra Joshua, and Kota Srinivas. 2020. Energy management using non-intrusive load monitoring techniques - State-of-the-art and future research directions. *Sustainable Cities and Society* (2020), 102411. https://doi.org/10.1016/j.scs.2020.102411

[7] Sidhant Gupta, Matthew S Reynolds, and Shwetak N Patel. 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*. 139–148.

[8] George William Hart. 1992. Nonintrusive appliance load monitoring. *Proc. IEEE* 80, 12 (1992), 1870–1891.

[9] Bohao Huang, Mary Knox, Kyle Bradbury, Leslie M Collins, and Richard G Newell. 2017. Non-intrusive load monitoring system performance over a range of low frequency sampling rates. In *2017 IEEE 6th International Conference on Renewable Energy Research and Applications (ICRERA)*. IEEE, 505–509.

[10] Jana Huchtkoetter and Andreas Reinhardt. 2019. A study on the impact of data sampling rates on load signature event detection. *Energy Informatics* 2, 1 (2019), 24.

[11] Maysoun Ibrahim, Ali El-Zaart, and Carl Adams. 2018. Smart sustainable cities roadmap: Readiness for transformation towards urban sustainability. *Sustainable Cities and Society* 37 (2018), 530 – 540. https://doi.org/10.1016/j.scs.2017.10.008

[12] M Khazaei, L Stankovic, and V Stankovic. 2019. Trends and challenges in smart metering analytics. In *2019 MTMI International Conference on Emerging Issues in Business, Technology and Applied Sciences*. 111–117.

[13] Christoph Klemenjak, Andreas Reinhardt, Lucas Pereira, Stephen Makonin, Mario Bergés, and Wilfried Elmenreich. 2019. Electricity consumption data sets: Pitfalls and opportunities. In *Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*. 159–162.

[14] Henning Lange and Mario Bergés. 2016. Bolt: Energy disaggregation by online binary matrix factorization of current waveforms. In *Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments*. 11–20.

[15] Andreas Reinhardt and Christoph Klemenjak. 2020. How does Load Disaggregation Performance Depend on Data Characteristics? Insights from a Benchmarking Study. In *Proceedings of the Eleventh ACM International Conference on Future Energy Systems*. 167–177.

[16] Aloisio Paixao; Wanderley Celeste; Luis Rigo Jr.; Daniel Coura; Helder Rocha; Leonardo Silvestre; Silvia Rissino. 2019. Non-Intrusive Similar Loads Identification. https://doi.org/10.21227/mynq-a588

[17] Shikha Singh and Angshul Majumdar. 2017. Deep sparse coding for non–intrusive load monitoring. *IEEE Transactions on Smart Grid* 9, 5 (2017), 4669–4678.

[18] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong. 2010. Locality-constrained Linear Coding for image classification. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 3360–3367.