Multi-Target Energy Disaggregation using Convolutional Neural Networks

Mohammed Ayub[1], El-Sayed M. El-Alfy[2]

College of Computer Sciences and Engineering,

King Fahd University of Petroleum and Minerals,

Dhahran 31261, Saudi Arabia

Abstract—Non-Intrusive Load Monitoring (NILM) has become popular for smart meters in recent years due to its low cost installation and maintenance. However, it requires efficient and robust machine learning models to disaggregate the respective electrical appliance energy from the mains. This study investigated NILM in the form of direct point-to-point multiple and single target regression models using convolutional neural networks. Two model architectures have been utilized and compared using five different metrics on two benchmarking datasets (ENERTALK and REDD). The experimental results showed that multi-target disaggregation setting is more complex than single-target disaggregation. For multi-target setting of ENERTALK dataset, the highest individual F1-score is 95.37% and the overall average F1-score is 75.00%. Better results were obtained for the multi-target setting of the other dataset with higher overall average F1-score of 83.32%. Additionally, the robustness and knowledge transfer capability of the models through cross-appliance and cross-domain disaggregation was demonstrated by training for a specific appliance on a specific data, and testing for a different appliance, house and dataset. The proposed models can also disaggregate simultaneous operating appliances with higher F1-scores.

Keywords—*Energy disaggregation; smart meters; load monitoring; ENERTALK dataset; multi-target disaggregation; multi-target regression; NILM knowledge transfer*

I. INTRODUCTION

Using energy in an efficient manner has become one of the highest concerns for both utility and end-users nowadays, as the world is facing challenging problems including depletion of natural resources and emissions of environmentally hazardous gases. To be able to use electrical energy efficiently, first, both utility and consumers should know the amount of energy consumption of individual appliances. For this purpose, cost effective Non-Intrusive Load Monitoring (NILM) is considered a plausible alternative. NILM is a method that can deduce energy consumption of individual appliances from aggregated smart meter power data recorded at a single source. This process is based on software techniques and requires effective and efficient techniques for successful disaggregation. It can help understand consumer behavior and energy consumption of appliances, and hence provide feedback on how to save energy. According to [1], monitoring energy consumption can save up to 12% of electrical energy with positive impacts on natural resources and reduction of hazardous gas emission. Besides getting detailed insights of the energy usage, NILM is useful for better demand forecasting and tracing behavioral patterns of dwellers [2].

NILM can be realized using three major steps as shown in

Fig. 1: (1) Data acquisition which is the collection of data by installing hardware such as smart meters, (2) Feature extraction which is the derivation of features from the collected data, and (3) Learning and inferencing which is the deployment of models such as training machine learning to make prediction. NILM can be exploited for disaggregation using either classification or regression techniques. In classification methods, the detection or identification procedure is more sophisticated when unique signatures or fingerprints are needed to formulate for classification of appliances [3], [4]. In regression, the appliance's on/off state or classification is obtained by leveraging the disaggregation results based on the on-threshold value of an appliance [5].

There are a number of disaggregation studies that applied various machine learning techniques including Decision Trees (DT), K-Nearest Neighbors (KNN), Neural Networks (NN), Convolutional Neural Networks (CNN), Long Short Term Memory (LSTM), Denoising Auto Encoder (deAE), Sequence-to-sequence (seq2seq) and Sequence-to-point (seq2point) learning, Subtask Gated Network (SGU) and many others [6]–[13]. Sequence-to-sequence and sequence-to-point are recent NILM disaggregation paradigms that can predict direct energy consumption with promising results than the Hidden-Markov Model (HMM) and its variants. But in these techniques, choosing the receptive field (sliding window) is somewhat crucial as there are different types of appliances with the different activation cycles. Besides, there may be multiple predictions for a single time point in seq2seq which is redundant and calculation of mean is necessary for the final prediction. Considering all these points, this work formulates the disaggregation problem as a direct point to point regression problem on the motivation that it will retain the granularity of consumption information that will help the generalization capability and knowledge transfer in the disaggregation domain which is demonstrated in the experiment section. On the other hand, it will also reduce the burden of retaining contextual information and thereby releasing the computational and memory burden.

A disaggregation model can be trained either as a single-target [5], [14] or multi-target [15] regression problem, or as a single-label [2], [16]–[18], or multi-label [3], [7], [19]–[22] classification problem. Single- and multi-label classification and multi-class classification were explored in many works in the literature. To our best knowledge, multi-target regression models for disaggregation are still under-studied. As the principle application of NILM is to separate the individual consumption from aggregate reading, in real time scenario,
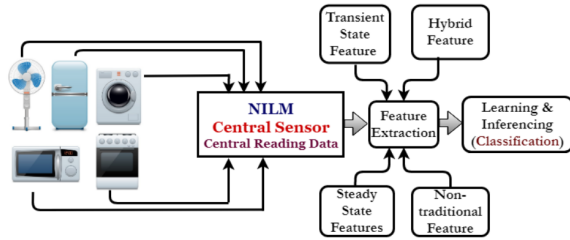
Fig. 1. Illustration of the Main Steps of Non-Intrusive Load Monitoring

direct disaggregation results are more preferable to the disaggregation results obtained after classification. Moreover, multi-target disaggregation models are more suitable on the reason that it can significantly reduce the training time by requiring low resources. Therefore, it is important to formulate NILM problem as a point-to-point multi-target regression model and compare its performance with single target regression models, especially for recent datasets such as ENERTALK. It is also emphasized that a multiple output classification or regression model poses challenges for disaggregation of multiple appliances of the same type [23], especially when the devices are multi-state devices and non-linear devices [24]. The works in [4], [25], [26] attempted the simultaneous detection of the load using different algorithms and features. However, their scopes were limited to classification of the load and there is still a need to explore this problem using regression models.

Keeping the above mentioned issues as main focus, this paper has the following contributions. Energy disaggregation is formulated as the direct point-to-point multi-target disaggregation learning that will help retain granular consumption meta information to foster generalization and knowledge transfer. The proposed multi-target disaggregation approach has highest individual and average F1-score of 95.37% and 75% respectively for ENERTALK dataset. For REDD dataset, the highest individual F1-score is 95.06% and overall average is 83.32%. For generalization, robustness and knowledge transferability of multiple and single target point-to-point disaggregation models, the highest F1-scores are 96.40% and 93.50% for single and multiple target, respectively. Furthermore, this study provides a base for NILM researchers to compare their results, especially for ENERTALK dataset which is a relatively recent dataset with large number of houses.

The remainder of this paper is organized as follows. Section II reviews related work and Section III describes the problem formulation and methodology. Section IV presents the details about experiments, results and discussions on ENERTALK and REDD. Finally, the paper summary and conclusions are presented in Section V.

II. RELATED WORKS

Shin and his team [13] proposed Subtask Gated Network (SGN) for non-intrusive load monitoring in 2019. Their model used a classification subnetwork and a regression subnetwork and then the output of classification is gated with that of regression. A concise study of appliance features used for NILM is presented in [27] using Random Forest (RF) and Recursive Feature Elimination (RFE). Mengistu et al. [28] described

an online cloud-based NILM system using HMM and Mean-Shift Clustering (MSC) algorithm in an unsupervised fashion. Unlike previous works, the study in [29] investigated the data reduction strategy in aggregated power signals by applying non-uniform subsampling (NUS).

A Graph-based semi-supervised multi-label classification model based on active/inactive (on/off) state of appliance was proposed in [19] using three graph-based algorithms (i.e. Local and Global Consistency (LGC), Gaussian Fields and Harmonic Functions(GFHF), and Manifold Regularization (MR)). Another multi-label classification of appliances using RAndom k-labELsets (RAkEL) with Decision Tree (DT) was explored in [7]. However, it was observed that the low-power consumption appliances were not correctly identified. Kim and Lee [3] investigated multi-label classification using audio signal processing techniques: Spectrogram, Mel-Frequency Cepstral Coefficient (MFCC) and Mel-spectrogram. The spectrogram based feature is proved to have the promising results. Inspired by the success of CNN, the authors in [16] proposed a novel appliance identification method using CNN for feature extraction, and Adaptive Linear Programming Boosting (ALPBoost) for classification. An accuracy of 95.4% for single appliance identification and 91.8% for multiple appliance identification were achieved. In the literature, the voltage-current (V-I) trajectories were found powerful for formulating appliance signature. In [21] and [22], appliances were classified using V-I trajectory converted to grey-scale and color coded image, respectively. Though the models in [21] were able to successfully detect a large number of appliances, the washing machine, fan, fridge and air conditioner were not identified with better score. Reference [22] used AlexNet transfer learning methodology. The authors in [17] used wavelet coefficients for identification of four appliances using Decision Tree (DT) and Nearest Neighbor (NN) classifiers on the setting of semi-supervised learning.

Jiang et al. in their work [5] investigated the on/off detection of appliances and energy disaggregation using CNN, RNN and Wavenet that used fast sequence-to-point learning. Kaselimi et al. [30] exploited the multi-channel CNN based architecture to include multiple input features in sequence-to-sequence (seq2seq) learning. However, the energies of multi-state devices were not correctly estimated. Schirmer and his team [14], like in Kaselimi's work, proposed a two-state disaggregation model using DNN and temporal contextual information (TCI). The performance for single- and multi-state device without power peak and non-linear devices are relatively low. How different combination of statistical and electrical features influence the disaggregation of various device types were studied in [6]. The authors in [9] studied the energy disaggregation using CNN, LSTM and CNN+LSTM on REDD dataset. In [31], a causal 1-D convolutional neural network based power disaggregation system (Wave-nilm) was proposed and the model was tested on AMPds2 dataset. It is noticed that the use of reactive power (Q) as input feature increases performance. In [32], Markov model was used to relate activity chain to each occupant of a household and then the energy usage per appliance was calculated and again these power usages were grouped under certain appliance categories. Appliance recognition and thereby disaggregation of energy using high frequency spectrogram feature (Short Time Fourier Transform (STFT)) was conducted in [33]. The authors in

[15] used a composite deep LSTM (CD-LSTM) to study the disaggregation using multi-target setting. This is the only work that used multi-target disaggregation. However, their work was based on sequence-to-sequence paradigm which has the drawbacks mentioned in the introduction section.

III. PROBLEM FORMULATION AND METHODOLOGY

The purpose of NILM is to deduce the individual appliance level power consumption from the total signal recorded through a smart meter. Inversely, the total aggregated energy of a residence can be estimated from all individuals' appliance consumption. This can be mathematically expressed as [24]:

$$p_t = \sum_{i=1}^{n} p_t^{(i)} + e_t \qquad (1)$$

where p_t is the aggregated power consumption at time t, $p_t^{(i)}$ is the individual appliance power consumption at time t, and e_t is line loss or error of measurement at time t. Based on Eq. 1, the energy disaggregation problem is defined as the direct point to point estimation $\hat{p}_t^{(i)}$ of the overall power p_t, where $\hat{p}_t^{(i)}$ is the estimated i^{th} appliance energy at time t, given the ground truth consumption $p_t^{(i)}$. In exact terms, this estimation can be achieved either by formulating the disaggregation as a classification or regression problem.

In this work, as described in the introduction section, the NILM is formulated as a regression problem which can have significant advantages over classification problems. Classification requires deployment of various algorithms to identify various working states of the household devices. This can become more challenging when the devices with the different level of energy demand are operating simultaneously, requiring high sampling data to create unique signatures to differentiate one from the other devices. Moreover, keeping track of on/off timestamp, duration of on/off, and calculation of average load consumption during specific 'on' periods makes the algorithms more computationally burdened. In sequence-to-sequence and sequence-to-point learning, training the model for long sequences of input to cover activation cycle requires a relatively large memory due to the need to keep track of contextual information. This sometimes jeopardizes the system in a low-resource environment. On the other hand, a regression model does not require all these operations and per-appliance disaggregation value is obtained directly from the results of regression output layer. Additionally, regression model provides detailed information about energy usage at every time point which cannot be the case with classification model.

A. Multi-Target Disaggregation Problem

Let $X = \{X_1, X_2, ...X_p\}$ and $Y = \{Y_1, Y_2, ...Y_q\}$ be the two sets of input and target variables, respectively; each set with features derived from aggregate consumption and individual appliance ground truth consumption at specific time points (steps). The training dataset is defined as,

$$\mathbb{D} = \{(\mathbf{x}_1, \mathbf{y}_1), ..., (\mathbf{x}_T, \mathbf{y}_T)\}, \qquad (2)$$

where T is the total number of time steps. Then each instance of $\mathbb{D}$ consists of input vector of p independent variables (predictors) and output vector of q target variables at time t, which respectively, are defined as,

$$\mathbf{x}_t = (x_t^{(1)}, x_t^{(2)}, ..., x_t^{(j)}, ..., x_t^{(p)}), \qquad (3)$$

and

$$\mathbf{y}_t = (y_t^{(1)}, y_t^{(2)}, ..., y_t^{(i)}, ..., y_t^{(q)}), \qquad (4)$$

where $j = 1, 2, ..., p$, $i = 1, 2, ..., q$ and $t = 1, 2, ..., T$.

Now, the task is to learn a multi-target disaggregation model $\mathbb{M}$ from instances of $\mathbb{D}$ such that a function f maps the vector $\mathbf{x}$ consisting of p aggregate feature values to a vector $\mathbf{y}$ consisting of q consumption values.

$$f : \mathbf{x} \mapsto \mathbf{y}, \qquad (5)$$

where $\mathbf{x} = (x^{(1)}, x^{(2)}, ..., x^{(p)})$ and $\mathbf{y} = (y^{(1)}, y^{(2)}, ..., y^{(q)})$. The trained model $\mathbb{M}$ can be used to predict the power consumption of all included appliances denoted by $\{\hat{\mathbf{y}}_{(T+1)}, \hat{\mathbf{y}}_{(T+2)}, ..., \hat{\mathbf{y}}_{(T')}\}$ using the new input instances $\{\mathbf{x}_{(T+1)}, \mathbf{x}_{(T+2)}, ..., \mathbf{x}_{(T')}\}$, i.e.,

$$\hat{\mathbf{y}}_{t'} = \mathbb{M}(\mathbf{x}_{t'}), t' = T + 1, T + 2, ..., T' \qquad (6)$$

For training loss, the Mean-Squared-Error (MSE) of single appliance disaggregation can be adopted for multiple devices as follows:

$$MSE = \frac{1}{q}\frac{1}{T} \sum_i \sum_t (y_t^{(i)} - \hat{y}_t^{(i)})^2 \qquad (7)$$

Setting $i = 1$ in Eq. (4) will transform a multi-target disaggregation problem to a single-target disaggregation problem.

B. CNN Multi-Target Regression Model

CNN was originally applied for image processing and computer vision where the input and each layer is multi-dimensional [34]. However, in this paper, NILM is formulated using 1-D CNN because the nature of data in load disaggregation is a uni-dimensional time series that keeps track of energy consumption of each appliance at a specific point. In line with this, every aggregate consumption feature at the specific time stamp is convolved with a kernel finding the relation between the features so that the part of aggregate consumption can be best mapped to the target appliance consumption value thereby finally producing optimum weights. Inspired by [35], two architectures of CNN are designed as follows. The first one will be referred to as CNN model-I (CNN1M), is built using two blocks of convolutional layers; the first block is followed with a max-pooling layer and the second block is followed with a global max-pooling layer. The fist convolutional block consists of three convolutional layers, each with 64 filters of size 2 and the second block also has three convolutional layers but each with 128 filters of size 2. The global max-pooling layer acts as a bottleneck layer to branch each appliance as a separate output with a dense layer of 512 neurons and a final output layer with one neuron. Another deeper model called CNN model-II was designed that has one more convolutional block of three layers each with 32 filters of size 2 and all the other parameters are the same as CNN Model-I. For CNN model-II, two variants were used: one for single target (CNN2S) and the other for multi-target (CNN2M), where S stands for Single and M for

Multiple. For both architectures, the 'stride' and 'learning rate' are set to 1 and 0.001, respectively. Also, 'ReLU' is used as activation for all hidden layers and 'linear' is used as activation for all output layers. The architectures are shown in Fig. 2 and Fig. 3.

IV. EXPERIMENTAL WORK

For all experimental works, a machine with NVIDIA GEFORCE GPU model GTX 950 which has 4GB dedicated graphics and 16GB RAM was used. The implementation scripts were written using Python 3.7.3 and for CNN model, Keras which run on Tensorflow back-end was deployed. For all experiments, data normalization used was L2-norm implemented in scikit-learn preprocessing.normalize(). The sketch of overall workflow is given in Fig. 4.

A. Energy Disaggregation based Metrics

Five performance measures have been reported to evaluate the proposed models. In the following formulae, $y_t^{(i)}$ represents the true energy consumed by appliance i at time point t and $\hat{y}_t^{(i)}$ represents the disaggregated power of appliance i at time point t.

1) Mean-Squared Error (MSE): MSE is defined as the average of the square of the difference between actual and predicted energy values as expressed by

$$MSE^{(i)} = \frac{1}{T} \sum_t (y_t^{(i)} - \hat{y}_t^{(i)})^2 \quad (8)$$

2) Normalized Disaggregation Error (NDE): It is the ratio between the sum of squared difference of estimated (disaggregated) energy and true energy, and sum of square of actual energy, and then taking square root of it. It is a slight variation of the metric used in [36]. Its mathematical formula is

$$NDE^{(i)} = \sqrt{\frac{\sum_t (y_t^{(i)} - \hat{y}_t^{(i)})^2}{\sum_t (y_t^{(i)})^2}} \quad (9)$$

3) Normalized Error in Assigned Power (NEAP): The sum of the absolute differences between the disaggregated power and the true energy consumption of appliance i in each time point t, divided by the total power consumption of appliance. In [37], the authors have shown that the error values greater than 1 is less representative towards the disaggregation performance and less explainable.

$$NEAP^{(i)} = \frac{\sum_t \left| y_t^{(i)} - \hat{y}_t^{(i)} \right|}{\sum_t y_t^{(i)}} \quad (10)$$

4) Energy-based Precision ($P^{(E)}$): It can be mathematically expressed as [38]:

$$P_i^{(E)} = \frac{\sum_t min(y_t^{(i)}, \hat{y}_t^{(i)})}{\sum_t \hat{y}_t^{(i)}} \quad (11)$$

5) Energy-based Recall ($R^{(E)}$): It can be mathematically expressed as [38]:

$$R_i^{(E)} = \frac{\sum_t min(y_t^{(i)}, \hat{y}_t^{(i)})}{\sum_t y_t^{(i)}} \quad (12)$$

6) Energy-based F1-measure ($F_1^{(E)}$): It is a geometric mean between the precision and recall and can be mathematically expressed as:

$$F_1^{(E)} = 2 \frac{P^{(E)} R^{(E)}}{P^{(E)} + R^{(E)}} \quad (13)$$

where,

$$P^{(E)} = \frac{1}{m} \sum_{i=1}^m P_i^{(E)}, R^{(E)} = \frac{1}{m} \sum_{i=1}^m R_i^{(E)} \quad (14)$$

where m is the number of appliances. Equation (13) can be used to calculate the F1 score of individual appliance using Equation (11) and (12).

B. Datasets and Data Preparation

Two public datasets were deployed for experimenting the proposed models in this work. The first one is ENERTALK dataset [39] that has consumption data collected from 22 residences in Korea using 15Hz sampling rate for both individual and total (aggregate) consumption. One of the challenging and crucial aspects of energy disaggregation of this dataset is the alignment of target appliance and aggregate meter reading. Keeping this in mind, the daily consumption of all appliances including aggregate reading and daily consumption were concatenated in horizontal and vertical fashion where the data were resampled to one second. To handle missing data, one-step backward filling was used and the remaining missing values were subsequently removed. The other dataset is REDD [40] that has data collected from six residential buildings in USA. For this work, the low frequency data of six houses were used; where the mains consumption are sampled at 1Hz and all appliances are sampled at $\frac{1}{3}$Hz. These different sampling rates were aligned by using horizontal concatenation and downsampling to 3 seconds. Afterwards, the preprocessing with L2-norm was applied to the resultant dataframe of both datasets to normalize feature values.

The active power (P) and reactive power (Q) are the two original features available in ENERTALK dataset. Based on these two features, some other features were extracted based on power triangle shown in Fig. 5. For REDD dataset, only active power for two mains are available and are used as features.

Active Power (P),

$$P = V * I * cos(\Phi) \quad (15)$$

where V is voltage, I is current and Φ is the phase angle.

Reactive Power (Q),

$$Q = V * I * sin(\Phi) \quad (16)$$

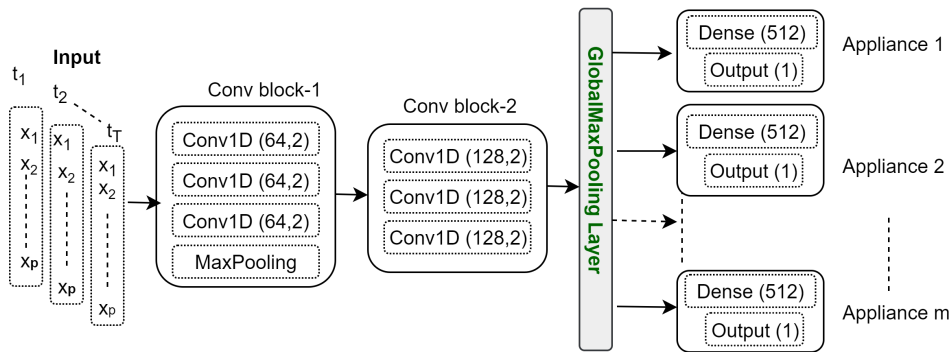Apparent Power (S),

$$S = V * I, S^2 = P^2 + Q^2 \quad (17)$$

Fig. 2. Multi-Target CNN Model-I Architecture; x_1 to x_p represent Energy Consumption (Features) at each time step t_1 to t_T, where p is the Total Number of Features, T is the Total Number of Time Steps and m is the Total Number of Appliances

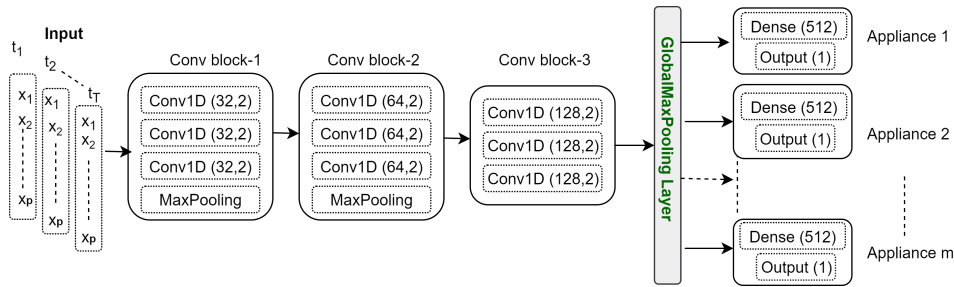
Fig. 3. Multi-Target CNN Model-II Architecture; x_1 to x_n represent Energy Consumption (Features) at each Time Step t_1 to t_T, where p is the Total Number of Features, T is the Total Number of Time Steps and m is the Total Number of Appliances
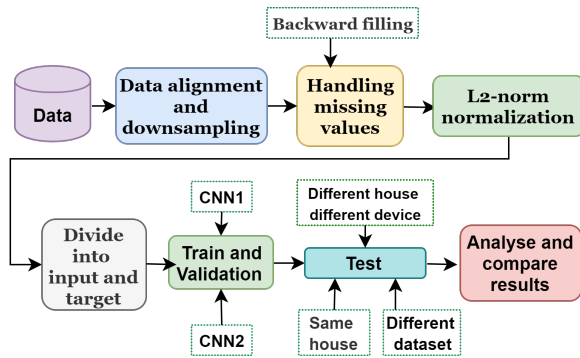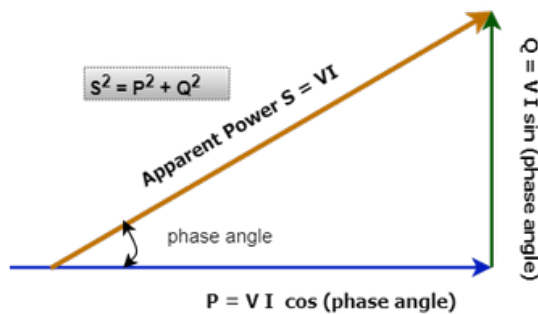

Fig. 4. Workflow Chart

Fig. 5. Power Triangle

where P, Q, S, V and I are as described earlier.

Power Factor (PF),

$$PF = cos(\Phi) = \frac{P}{S} \tag{18}$$

where P and S are as described earlier.

Besides the above features, six additional features are extracted: (1) difference power between S and P (DP_{sp}), (2) difference power between P and Q (DP_{pq}), (3) difference power between S and Q (DP_{sq}), (4) average of P, Q and S (P_{avg}), (5) sine of phase angle (Sin_Ph), and (6) tangent of phase angle (Tan_Ph), for ENERTALK dataset.

C. Results for ENERTALK Dataset

For all the experiments with this dataset, 20% of the total data is kept for testing whereas the rest 80% is divided again in 80:20 ratio for training and validation. The batch size of 512 and number of epochs of 20 are used for training of the models with the loss function of total Mean Squared Error (MSE) for all appliances.

1) Experiments on House00 Data: These experiments used 24-days data of House00 that consists of the recording of seven appliances: TV, Washing Machine (WM), Rice Cooker (RC), Refrigerator (R), Water Purifier (WP), Microwave (MW) and Kimchi Fridge (KR). A total of 2073580 instances were generated after applying the horizontal and vertical concatenation. After all the data preparation steps were applied as explained earlier, 1983102 instances were available. This experiment used S and PF as features besides P and Q. From the scores shown in Table I, it is observed that the highest F1-score of

TABLE I. PERFORMANCE SCORES OF ENERGY DISAGGREGATION OF CNN MODEL-II FOR ENERTALK-HOUSE00 (S= SINGLE, M= MULTIPLE)

Appliance	NDE	NEAP	P	R	F1
TV (S)	0.22661	0.38922	0.83034	0.98916	0.90282
TV (M)	0.74338	1.05110	0.50121	0.41428	**0.45362**
WM (S)	5.76306	6.33135	0.71930	0.75145	0.73503
WM (M)	19.4733	15.8424	0.39069	0.50867	0.44194
RC (S)	1.88011	1.96937	0.73848	0.91109	0.81576
RC (M)	10.1996	4.40733	0.42545	0.45535	0.43989
R(S)	0.15737	0.17893	0.93962	0.88391	0.91091
R(M)	0.46575	0.49511	0.79155	0.70009	0.74302
WP (S)	0.11281	0.13519	0.88762	0.99926	**0.94014**
WP (M)	0.39986	0.35087	0.86114	0.78504	**0.82133**
MW (S)	0.82996	1.02232	0.68390	0.78355	0.73034
MW (M)	1.44944	1.21941	0.62349	0.70693	0.66259
KR (S)	0.52759	0.64577	0.85249	0.89760	0.87446
KR (M)	2.82818	2.00645	0.39634	0.31392	0.35035
Average (S)	1.35671	1.52459	0.80739	0.88800	0.84577
Average (M)	5.07993	3.62467	0.56998	0.55489	0.56234

82.13% and 94.01% are achieved by Water Purifier (WP) in multi- and single target, respectively. The lowest achiever for this set of experiment is Kimchi Fridge (KR) in multi-target setting with F1-score of 35%. In the multi-target disaggregation, WP has also the highest performance in terms of NDE and NEAP.

2) Experiments on House12 Data: With this data of House12, the CNN model-II was validated using multi-state, continuously varying (nonlinear) consumption and always-on devices. It consists of 4 months aggregate and appliance level consumption data for two WMs, TV, RC, KR and R. A total of ten features mentioned in Section IV-B from 8639267 data samples were used for multi- and single target regression. All the experiments results are shown in Table II. As seen from the table, the refrigerator has the highest F1-score of 95.37% and 99.34% in multi- and single target setting, respectively. Both actual and predicted energy consumption of the refrigerator for multi-target setting are shown in Fig. 6. These results confirm that the CNN model-II is robust enough to successfully disaggregate energy with higher F1-score. It is noteworthy that the model successfully disaggregates energy of two of the three multi-state devices with the acceptable F1-scores along with other devices except WM2. For the multi-target setting, WM2 has the lowest F1-score of 43.47%, and for the single-target setting, WM1 has the lowest disaggregation score of 32.47%. Furthermore, if we analyze the results in terms of precision and recall for the multi-target setting, it is seen that the refrigerator has the highest scores of 97.59% and 93.24%, respectively. WM2 has the worst NDE and NEAP scores in the multi-target disaggregation setting.

D. Results for REDD Dataset

This experiment was executed on the total of 1099738 data samples resulted from combination of House1, House2 and House3 data of the REDD dataset for four common kitchen appliances (three multi-state devices and one continuous consumption device) such as Microwave (MW), Refrigerator (R), Dishwasher (DW) and Washer Dryer (WD). The CNN model-I was trained with 20 epochs and batch size of 512. The CNN model-II was modified according to the features and data used, i.e. a filter size of 1 and 16 epochs were used. Moreover, the MaxPooling layer after the first convolutional block was
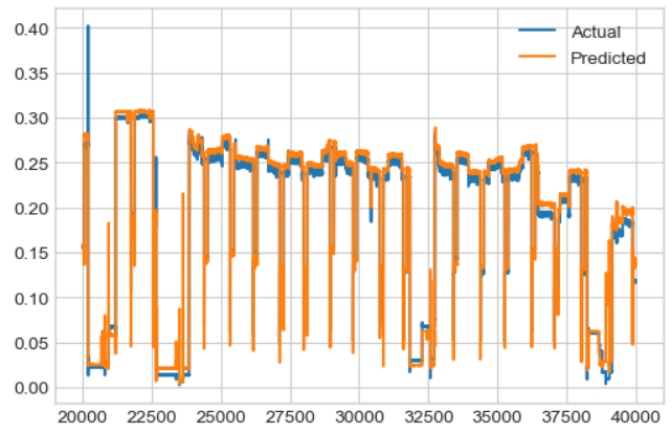

Fig. 6. Predicted Consumption of Refrigerator for Multi-Target CNN Model-II ((ENERTALK-House12)

TABLE II. PERFORMANCE SCORES OF ENERGY DISAGGREGATION FOR CNN MODEL-II ON ENERTALK-HOUSE12. (S= SINGLE, M= MULTIPLE)

| Appliance	NDE	NEAP	P	R	F1
WM1 (S)	1.81104	2.12090	0.31159	0.33915	0.32479
WM1 (M)	2.32865	1.84337	0.52510	0.74387	0.61563
TV (S)	0.17689	0.31679	0.99975	0.71681	0.83496
TV (M)	0.40574	0.38446	0.85595	0.77106	**0.81129**
RC (S)	3.98337	4.85974	0.39227	0.97747	0.55987
RC (M)	1.34695	0.83939	0.78732	0.74534	0.76576
KR (S)	0.04056	0.06156	0.99173	0.96386	0.97760
KR (M)	0.26382	0.24054	0.95253	0.84008	0.89278
WM2 (S)	0.63658	0.66677	0.90117	0.77531	0.83351
WM2 (M)	3.58233	2.33685	0.41711	0.45389	0.43472
R (S)	0.01768	0.01294	0.99705	0.98971	**0.99337**
R (M)	0.13047	0.09156	0.97591	0.93240	**0.95366**
Average (S)	1.11102	1.33978	0.76559	0.79372	**0.77862**
Average (M)	1.34299	0.95603	0.75232	0.74777	0.75004

removed to train the model for the combined data of REDD dataset. For the testing set, 30% of the original data is reserved and the remaining 70% is used for training and validation with 70:30 split ratio. The scores of disaggregation using CNN model-I (CNN1M) and CNN model-II (CNN2M) models are shown in Table III and the predicted disaggregation of MW and R are shown in Fig. 7. As seen from the table, the highest F1-score is achieved by R using CNN-I model. It should be emphasized that all the multi-state devices are disaggregated with acceptable F1-scores above 75% and highest F1-score of 88.15%. Another point to be noted is that for all devices, CNN model-I outperformed CNN model-II, though CNN model-II has a deeper architecture. This supports that a relatively shallower architecture can effectively disaggregate the low-sampling consumption data (REDD dataset was sampled at 3s whereas ENERTALK was sampled at 1s).

E. Discussion and Analysis

1) ENERTALK House00: For the analysis of energy disaggregation performance of the appliances of this house, if multi-target setting is considered, it has the average F1-score of 56.23% and the individual highest F1-score 82.13% for WP which is also the highest in the single output setting. As for the scores of single output setting, it is seen that WP has the highest F1-score of 94% and MW has the lowest F1-score of 73.03%. The performance of other appliances according
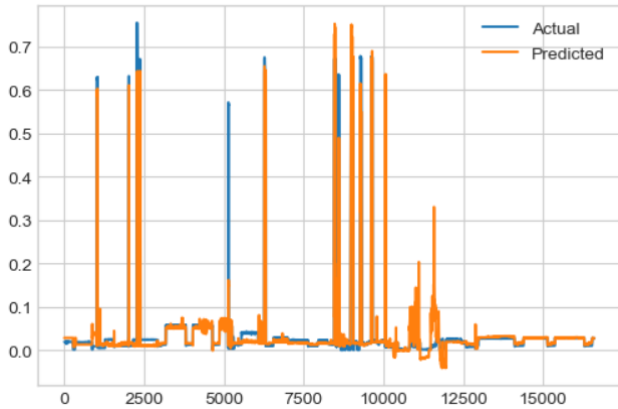
(a) Actual Vs. Estimated Consumption of MW for Multi-Target CNN Model-I on Combined Data)
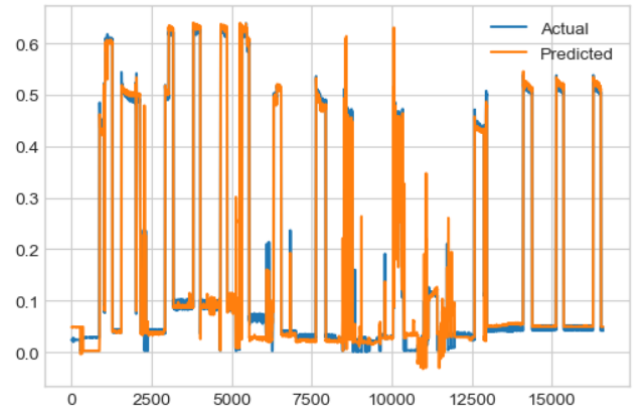
(b) Actual Vs. Estimated Consumption of R for Multi-Target CNN Model-I on Combined Data

Fig. 7. Actual Vs. Estimated Consumption for CNN Model-I Multi-Target on Combined Data (REDD:House1+House2+House3)
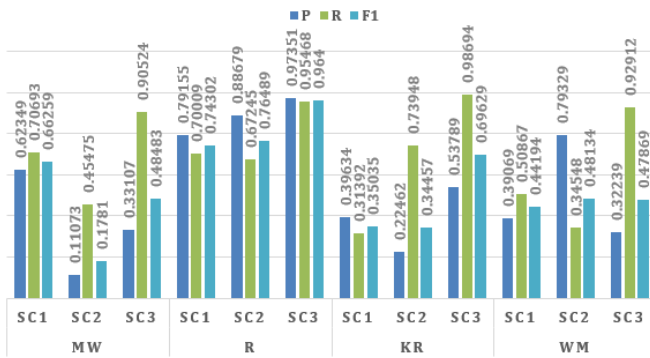

Fig. 8. Disaggregation Analysis of House00 (ENERTALK) Using the Models Trained on Different Appliance of Difference House of the Same Dataset and Same Appliance of Different Dataset. SC1, SC2 and SC3 stand for scenario 1, 2 and 3, respectively

.

TABLE III. PERFORMANCE SCORES OF CNN1M AND CNN2M ON REDD COMBINED DATA (REDD:HOUSE1+HOUSE2+HOUSE3)

Appliance	Model	P	R	F1
MW	CNN1M	0.87503	0.88806	**0.88150**
	CNN2M	0.83395	0.82906	0.83150
R	CNN1M	0.96422	0.93732	**0.95058**
	CNN2M	0.93352	0.94413	0.93880
DW	CNN1M	0.79036	0.72423	**0.75585**
	CNN2M	0.62219	0.45417	0.52506
WD	CNN1M	0.69692	0.78931	**0.76309**
	CNN2M	0.73521	0.64430	0.68676
Average (CNN1M)	CNN1M	0.83163	0.83473	**0.83318**
Average (CNN2M)	CNN2M	0.78122	0.71792	0.74824

to their F1-scores is 91.09%, 90.28%, 87.44%, 81.58% and 73.50% for R, TV, KR, RC and WM, respectively. In terms of all metrics scores, WP, R, TV, KR and MW have the highly acceptable disaggregation performance. WM and RC are not listed in this category because their respective NDE and NEAP have higher value even though they have better F1-score. Overall average performance F1-score is 84.58% which again confirms the robustness of point-to-point CNN model-II in single target setting for power disaggregation. For the

disaggregation performance of both setting, it is seen that the scores of multi-target model in this house are far lag behind those of the single target model (Table I).

2) ENERTALK House12: The main purpose of this experiment is to investigate how efficiently the CNN Model-II can disaggregate energy when there are simultaneously operating multiple mixed type of appliances. The experiment was conducted taking into consideration three multi-state devices (two WMs and one RC), one nonlinear device (TV) and two continuous consumption devices (R and KR). For this, the multi-target setting was considered and the results are shown in Table II. As seen from the table, the highest F1-score of 95.37% is for refrigerator and 75% is the average of all participating appliances. For the multi-state device (RC), F1-score is as high as 76.58%. This experiment also confirms that CNN mode-II can be used for future prototype for energy disaggregation of simultaneous operating of multiple multi-state devices. It should be also noted that our point-to-point multi-target CNN model can disaggregate the nonlinear device with F1-score of 81.13% in the presence of multiple multi-state devices which are also hard to disaggregate.

3) Disaggregation Performance Across ENERTALK Houses: For comparisons, please refer to Table I and Table II. It is seen that House00 has superior overall average performance than House12, for single target model. But in terms of multi-target setting, House12 has superior overall performance than House00. When compared the performance of appliances that are common (TV, KR, R, WM and RC) in all the two houses on the basis of appliance by appliance performance in single target setting, TV and RC have higher performance in House00 than House12, and WM1, KR and R have higher performance in House12 than House00. For multi-target setting, all appliances clearly outperform in House12 than House00. That is the reason that the overall performance of House12 is better than House00. Please note that the above analyses are based on F1-scores. To explore the differences, the data patterns of all appliances for all houses were analyzed. In that vein, we look into the data statistics in the houses and from the observation, it can generally be concluded that the usage patterns of individual appliance and

the duration of usage have an impact on the performance. This is based on the fact that the usage pattern and duration have the direct impact on the data distribution and range of data which in turn has remarkable influence on the performance. Neural network generally maps inputs to outputs based on certain mathematical operations, and penalizing the model for wrong mapping by adjusting the weights. This adjusting of weights become difficult for the model when data are sparsely distributed with most of the feature values with zero and some with very large values. This task becomes more difficult when the model has to extract small portion out from large values (blind-source separation) which is the case with the energy disaggregation. This point is particularly more applicable for the appliances that are less frequently used and disaggregation is done on event-less fashion. In addition, form the analyse of the average performance of multi-target regression models, it is observed that the highest F1-score of 75% is achieved by House12 and the lowest F1-score of 56.23% is achieved by House00. It is noteworthy that the reason of higher overall F1-score of House12. This may be due to the fact that when targets are correlated with one another, the multi-target produce better results. Again the correlation between the targets depend on data distribution which again is as mention above, depend on the usage pattern and duration. The study of correlation among the targets is beyond the scope this work. It is emphasized that in terms of NDE and NEAP also, the multi-target model in House12 performs better than both single and multi-target models in House00. Moreover, if we analyze the disaggregation performance of non-linear device (TV) in all two houses for multiple output regression model, the highest F1-score achieved is 81.13% for House12.

4) REDD Dataset: In combined data of House1, 2 and 3, the CNN1M has the highest individual F1-score of 95.06% with R and the overall average F1-score is 83.32%. In terms of individual scores CNN Model-I (CNN1M) has the best performance. This combined data experiment is further elaborated in IV-F.

F. Disaggregation Performance and Model Generalization

In general, according to our detail observation and analysis, the disaggregation performance of multi-target model is lower than single target model because in the single target setting, the weight is to adjust to multiple targets of different statistical differences using common features and parameter sharing. However, the multi-target models require less training resources and time than the single target model. On the other hand, the single target model can be customized for specific appliance but at the expense of extra training resources and time.

Generalization and robustness of single and multi-target disaggregation models are tested using CNN Model-II and CNN Model-I for ENERTALK and REDD dataset, respectively. To check the robustness of multi-target regression model, the model trained for four appliances (MW, R, DW, WD) on the combined data of House1+2+3 in REDD dataset was used to predict the energy of selected appliances of House 4, 5 and 6 of the same dataset. This is more challenging than the single output setting because there are ghost (unknown) appliance and miscellaneous outlet in House4 and House6. The scores and estimations are shown in Table IV, and Fig. 9.

TABLE IV. PERFORMANCE SCORES OF SELECTED APPLIANCES OF HOUSE4, 5 AND 6 DISAGGREGATED USING MODEL TRAINED ON HOUSE1+2+3 IN MULTIPLE OUTPUT SETTING (REDD)

House	Appliance	P	R	F1
House4	DW	0.39813	0.10865	0.17072
	WD	0.43354	0.05639	0.09980
	Unknown	0.17495	0.22044	0.19508
	Miscellaneous	0.30309	0.63820	**0.41099**
House5	MW	0.76739	0.55764	0.64591
	R	0.83057	0.66189	**0.73670**
	DW	0.89611	0.48298	**0.62767**
	WD	-0.01229	-0.55193	-0.02406
House6	R	0.94380	0.92640	**0.93503**
	DW	0.10150	0.98172	0.18398
	WD	0.897518	0.64567	**0.75104**
	Electronics	0.16903	0.82885	0.28080
	Unknown	0.52029	0.52725	**0.52734**

TABLE V. PERFORMANCE SCORES OF FOUR APPLIANCES OF HOUSE00 OF ENERTALK DISAGGREGATED USING CNN MODEL-I TRAINED ON COMBINED DATA OF REDD

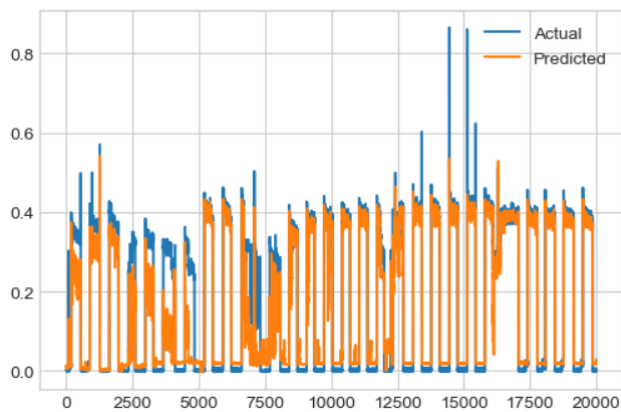| Appliance	P	R	F1
KR	0.22462	0.73948	0.34457
R	0.88679	0.67245	0.76489
WM	0.79329	0.34548	0.48134
MW	0.11073	0.45475	0.17810
Average	0.50386	0.55304	0.52731

It is seen that the model can disaggregate the unknown and miscellaneous outlets with F1-score of 52.73% and 41.1%, respectively. Moreover, it can disaggregate the multi-state device (WD) and permanent consumer device (R) with F1-score of 75.1% and 93.50%, respectively. Next, the CNN Model-I trained using combined data of REDD was tested on House00 data of ENERTALK to verify cross-domain disaggregation of the model. The model was trained on dataset and tested on another dataset of of a different geographical location. The tested appliances are KR, R, WM and MW. From the scores given in Table V, it is clear that the energy of R can be disaggregated with the highest precision of 88.68% and the highest F1-score of 76.49%. But in terms of recall score, KR has the highest disaggregation performance score of 73.95%. It should be emphasized that KR and WM were disaggregated using R and WD of combined data as there was no KR and WM in REDD combined data. As overall average of the four appliances, the scores of all three metrics are above 50%.

For single target regression, KR in House21 of ENERTALK dataset with total of 4215727 input instances is trained and the energy of all seven appliances (KR, R, TV, RC, WM, WP, MW) of House00 are disaggregated. In this scenario
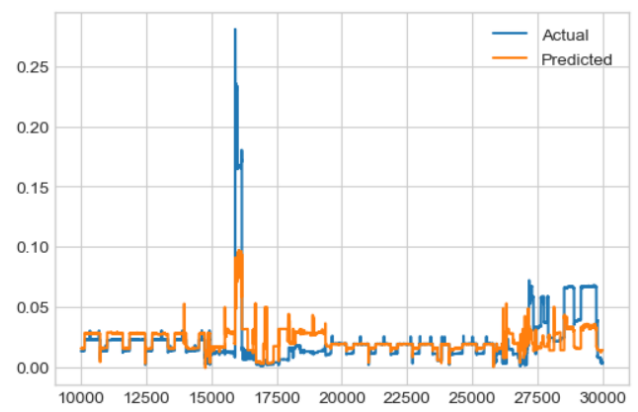
TABLE VI. PERFORMANCE SCORES OF SEVEN APPLIANCES OF HOUSE00 DISAGGREGATED USING CNN MODEL-II TRAINED ON KR OF HOUSE21 IN SINGLE OUTPUT SETTING (ENERTALK)

Appliance	P	R	F1
KR	0.53789	0.98694	0.69629
R	0.97351	0.95468	**0.96400**
TV	0.84321	0.93860	**0.88835**
RC	0.29038	0.96231	0.44614
WM	0.32239	0.92912	0.47869
WP	0.94289	0.97335	0.95788
MW	0.33107	0.90524	0.48483
Average	0.60591	0.95003	**0.73990**

(a) Energy of R of House6 Estimated by Multi-Target CNN Model-I Trained on House1+2+3

(b) Energy of MW of House5 Estimated by Multi-Target CNN Model-I Trained on House1+2+3

Fig. 9. Estimated Energy of Devices by the Model Trained on House1+2+3 (REDD)

of cross-appliance disaggregation, it is seen that the overall average F1-score is 73.99%. Cross-appliance disaggregation is using the model trained on one specific appliance for disaggregation of other different appliances. From Table VI, it is noticed that the model trained on KR of House21 data has successfully disaggregated energy of most of the appliances in House00, especially, nonlinear device TV with F1-score of 88.83% (see Table VI) and permanent consumer devices that have different operating states such as KR, R with F1-score of 69.63% and 96.40, respectively (see Table VI). But in terms of recall score, all seven devices have above 92% performance. This emphasizes that the direct point-to-point disaggregation models are powerful in learning granular information about consumption. Fig. 8 shows the disaggregation analysis of four selected appliances from House00 of ENERTALK. The scores are reported for three different scenarios: (1) The model trained and tested on House00 (CNN2M), (2) The model trained on REDD combined data (CNN1M) and tested on House00 (ENERTALK) and (3) The model trained on KR (CNN2S) of House21 (ENERTALK) and tested on House00 (ENERTALK). According to the scores in this figure, except MW, the other three appliances have higher disaggregation F1-scores in both scenario 2 and 3 than scenario 1. From previous analysis of generalization and robustness capability of single and multi-target point-to-point disaggregation, it can be emphasized that cross-appliance disaggregation and cross-domain disaggregation have very good performance.

V. CONCLUSION

This work investigated the NILM problem in terms of training disaggregation algorithms in multiple and single target regression setting, i.e. energy disaggregation of simultaneously operating multiple devices of same types. We used EN-ERTALK dataset, which is the latest publicly available energy dataset containing records of 22 Korean Houses. We also used REDD dataset, which is the first released energy dataset. We deployed state-of-the-art deep learning algorithms based on CNN and evaluated the disaggregation performance using five energy-based performance metrics. In NILM, aligning the consumption of the target appliances with the aggregate

consumption is more challenging. To achieve this, for EN-ERTALK, we used vertical and horizontal concatenation of appliance and dates, with mean resambling of data in one second, and for REDD, we used horizontal concatenation. For the experiments, dataframe was divided into training, validation and testing sets. To reflect better generalization capability of the models, the actual prediction is made on the data that was not seen during the training. In ENERTALK, CNN model-II has the highest superior disaggregation performance with F1-score of 95.37% for individual appliance and 75.00% for overall average performance across the dataset for multi-target disaggregation model. Moreover, in ENERTALK, the proposed models can disaggregate the energy of non-linear devices with higher F1-score of 81.13% in multi-target model. The generalization and robustness of the models were also validated by training the model for one appliance and testing it for different appliances in different houses, which shows the knowledge transferability in NILM using the proposed direct point-to-point disaggregation models. As an overall conclusion, the proposed point-to-point regression models have demonstrated computational efficiency and disaggregation effectiveness with superior scores. As future work, other techniques such as energy disaggregation using wavelets can be explored.

REFERENCES

[1] K. Ehrhardt-Martinez, K. A. Donnelly, S. Laitner *et al.*, "Advanced metering initiatives and residential feedback programs: a meta-review for household electricity-saving opportunities," no. E105. American Council for an Energy-Efficient Economy Washington, DC (ACEEE), 2010, pp. 1–140.

[2] M. A. Devlin and B. P. Hayes, "Non-intrusive load monitoring and classification of activities of daily living using residential smart meter data," *IEEE Transactions on Consumer Electronics*, vol. 65, no. 3, pp. 339–348, 2019.

[3] J.-G. Kim and B. Lee, "Appliance classification by power signal analysis based on multi-feature combination multi-layer lstm," *Energies*, vol. 12, no. 14, pp. 2804(1–24), 2019.

[4] A. Bouhouras, P. Gkaidatzis, K. Chatzisavvas, E. Panagiotou, N. Poulakis, and G. Christoforidis, "Load signature formulation for non-intrusive load monitoring based on current measurements," *Energies*, vol. 10, no. 4, pp. 538(1–21), 2017.

[5] J. Jiang, Q. Kong, M. Plumbley, and N. Gilbert, "Deep learning based energy disaggregation and on/off detection of household appliances," *arXiv preprint arXiv:1908.00941*, 2019.

[6] P. A. Schirmer and I. Mporas, "Statistical and electrical features evaluation for electrical appliances energy disaggregation," *Sustainability*, vol. 11, no. 11, pp. 3222(1–14), 2019.

[7] B. Buddhahai, W. Wongseree, and P. Rakkwamsuk, "A non-intrusive load monitoring system using multi-label classification approach," *Sustainable Cities and Society*, vol. 39, pp. 621–630, 2018.

[8] J. Kelly and W. Knottenbelt, "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*, 2015, pp. 55–64.

[9] İ. H. Çavdar and V. Faryad, "New design of a supervised energy disaggregation model based on the deep neural network for a smart grid," *Energies*, vol. 12, no. 7, pp. 1217(1–18), 2019.

[10] H. Rafiq, H. Zhang, H. Li, and M. K. Ochani, "Regularized lstm based deep learning model: First step towards real-time non-intrusive load monitoring," in *IEEE International Conference on Smart Energy Grid Engineering (SEGE)*, 2018, pp. 234–239.

[11] M. D'Incecco, S. Squartini, and M. Zhong, "Transfer learning for non-intrusive load monitoring," *IEEE Transactions on Smart Grid*, 2019.

[12] C. Zhang, M. Zhong, Z. Wang, N. Goddard, and C. Sutton, "Sequence-to-point learning with neural networks for non-intrusive load monitoring," in *Thirty-second AAAI conference on artificial intelligence*, 2018.

[13] C. Shin, S. Joo, J. Yim, H. Lee, T. Moon, and W. Rhee, "Subtask gated networks for non-intrusive load monitoring," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 1150–1157.

[14] P. A. Schirmer, I. Mporas, and A. Sheikh-Akbari, "Robust energy disaggregation using appliance-specific temporal contextual information," *EURASIP Journal on Advances in Signal Processing*, vol. 2020, no. 1, p. 3, 2020.

[15] M. Xia, K. Wang, W. Song, C. Chen, Y. Li *et al.*, "Non-intrusive load disaggregation based on composite deep long short-term memory network," *Expert Systems with Applications*, vol. 160, p. 113669, 2020.

[16] C. Min, G. Wen, Z. Yang, X. Li, and B. Li, "Non-intrusive load monitoring system based on convolution neural network and adaptive linear programming boosting," *Energies*, vol. 12, no. 15, pp. 2882(1–23), 2019.

[17] J. M. Gillis and W. G. Morsi, "Non-intrusive load monitoring using semi-supervised machine learning and wavelet design," *IEEE Transactions on Smart Grid*, vol. 8, no. 6, pp. 2648–2655, 2016.

[18] D. Jorde, T. Kriechbaumer, and H.-A. Jacobsen, "Electrical appliance classification using deep convolutional neural networks on high frequency current measurements," in *IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, 2018, pp. 1–6.

[19] D. Li and S. Dick, "Residential household non-intrusive load monitoring via graph-based multi-label semi-supervised learning," *IEEE Transactions on Smart Grid*, vol. 10, no. 4, pp. 4615–4627, 2018.

[20] V. Singhal, J. Maggu, and A. Majumdar, "Simultaneous detection of multiple appliances from smart-meter measurements via multi-label consistent deep dictionary learning and deep transform learning," *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 2969–2978, 2018.

[21] L. De Baets, J. Ruyssinck, C. Develder, T. Dhaene, and D. Deschrijver, "Appliance classification using vi trajectories and convolutional neural networks," *Energy and Buildings*, vol. 158, pp. 32–36, 2018.

[22] Y. Liu, X. Wang, and W. You, "Non-intrusive load monitoring by voltage-current trajectory enabled transfer learning," *IEEE Transactions on Smart Grid*, vol. 10, no. 5, pp. 5609–5619, 2018.

[23] J. Kim, T.-T.-H. Le, and H. Kim, "Nonintrusive load monitoring based on advanced deep learning and novel signature," *Computational intelligence and neuroscience*, vol. 2017, 2017.

[24] E. Aladesanmi and K. Folly, "Overview of non-intrusive load monitoring and identification techniques," *IFAC-PapersOnLine*, vol. 48, no. 30, pp. 415–420, 2015.

[25] W. A. de Souza, F. D. Garcia, F. P. Marafão, L. C. P. da Silva, and M. G. Simões, "Load disaggregation using microscopic power features and pattern recognition," *Energies*, vol. 12, no. 14, pp. 2641(1–18), 2019.

[26] P. Held, S. Mauch, A. Saleh, D. O. Abdeslam, and D. Benyoucef, "Frequency invariant transformation of periodic signals (fit-ps) for classification in nilm," *IEEE Transactions on Smart Grid*, vol. 10, no. 5, pp. 5556–5563, 2018.

[27] N. Sadeghianpourhamami, J. Ruyssinck, D. Deschrijver, T. Dhaene, and C. Develder, "Comprehensive feature selection for appliance classification in nilm," *Energy and Buildings*, vol. 151, pp. 98–106, 2017.

[28] M. A. Mengistu, A. A. Girmay, C. Camarda, A. Acquaviva, and E. Patti, "A cloud-based on-line disaggregation algorithm for home appliance loads," *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 3430–3439, 2018.

[29] M. Fagiani, R. Bonfigli, E. Principi, S. Squartini, and L. Mandolini, "A non-intrusive load monitoring algorithm based on non-uniform sampling of power data and deep neural networks," *Energies*, vol. 12, no. 7, pp. 1371(1–26), 2019.

[30] M. Kaselimi, E. Protopapadakis, A. Voulodimos, N. Doulamis, and A. Doulamis, "Multi-channel recurrent convolutional neural networks for energy disaggregation," *IEEE Access*, vol. 7, pp. 81 047–81 056, 2019.

[31] A. Harell, S. Makonin, and I. V. Bajić, "Wavenilm: A causal neural network for power disaggregation from the complex power signal," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 8335–8339.

[32] J. Holweger, M. Dorokhova, L. Bloch, C. Ballif, and N. Wyrsch, "Unsupervised algorithm for disaggregating low-sampling-rate electricity consumption of households," *Sustainable Energy, Grids and Networks*, vol. 19, p. 100244, 2019.

[33] Q. Wu and F. Wang, "Concatenate convolutional neural networks for non-intrusive load monitoring across complex background," *Energies*, vol. 12, no. 8, pp. 1572(1–17), 2019.

[34] Y. LeCun, Y. Bengio *et al.*, "Convolutional networks for images, speech, and time series," *The handbook of brain theory and neural networks*, vol. 3361, no. 10, p. 1995, 1995.

[35] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[36] J. Z. Kolter and T. Jaakkola, "Approximate inference in additive factorial hmms with application to energy disaggregation," in *Artificial intelligence and statistics*, 2012, pp. 1472–1482.

[37] E. T. Mayhorn, G. P. Sullivan, J. M. Petersen, R. S. Butner, and E. M. Johnson, "Load disaggregation technologies: real world and laboratory performance," Pacific Northwest National Lab.(PNNL), Richland, WA (United States), Tech. Rep., 2016.

[38] R. Bonfigli and S. Squartini, *Machine Learning Approaches to Non-Intrusive Load Monitoring*. Springer, 2019.

[39] C. Shin, E. Lee, J. Han, J. Yim, W. Rhee, and H. Lee, "The enertalk dataset, 15 hz electricity consumption data from 22 houses in korea," *Scientific data*, vol. 6, no. 1, pp. 1–13, 2019.

[40] J. Z. Kolter and M. J. Johnson, "Redd: A public data set for energy disaggregation research," in *Workshop on data mining applications in sustainability (SIGKDD), San Diego, CA*, vol. 25, no. Citeseer, 2011, pp. 59–62.