

Received October 26, 2021, accepted December 2, 2021, date of publication December 6, 2021, date of current version December 17, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3133292

A Nonintrusive Load Monitoring Based on Multi-Target Regression Approach

BUNDIT BUDDHAHAI^{ID}¹ AND STEPHEN MAKONIN^{ID}², (Senior Member, IEEE)

¹Walailak University International College, Walailak University, Tha Sala, Nakhon Si Thammarat 80161, Thailand

²School of Engineering Science, Simon Fraser University, Burnaby, BC V5A 1S6, Canada

Corresponding author: Bundit Buddhahai (bundit.bu@mail.wu.ac.th)

This work was supported by the Research Grants from the Walailak University Research Fund under Contract WU64203.

ABSTRACT This paper proposes an experimental design process for the application of energy disaggregation using multi-target regression, a new data learning approach in this application area. The approach shows to be a suitable model for dealing with energy disaggregation problems in which the task is to predict multiple appliances usage from the aggregate data. The experiments were conducted by analyzing AMPDs2 and ECO public data sets for verifying the effectiveness of the approach. The data were analyzed through the machine learning process to select the optimal set of electrical features, learning algorithm, and model parameter so that the system resulting from the process could deliver the optimal performance for loads inference. Results of the data learning showed that the electrical features set of current (I), real power (P), reactive power (Q), and power factor (PF) for the aggregate data and Random Forest as the base regressor for multi-target regression model could provide the best disaggregation performance. The overall predictive performance of disaggregation accuracy and F-score outperformed the benchmarking Super State Hidden Markov Model (SSHMM) and Denoising Autoencoder (DAE) network approaches.

INDEX TERMS Home energy monitoring, machine learning process, multi-target regression, nonintrusive load monitoring.

I. INTRODUCTION

Non-Intrusive Load Monitoring (NILM) is the task of decomposing the whole energy information from a building into the information of energy used by appliances [1]. The details of energy feedback could provide users with decisive information such as the saving motivations [2], load deferral for high-power appliances. Studies showed that users could save energy costs up to 12% by having real-time feedback on appliance-level as against the conventional monthly billing feedback [3]. Grid utilities could also benefit in energy demand forecasting and balancing the energy supply policy [4]. The NILM operation relies on the machine learning process which consists of data collection, features engineering, and data learning and identification [5]. The key challenge of the system development is to design the learning algorithm to accurately infer the contributed loads operation

from the aggregate data for the proper action in the energy management scheme.

The NILM research directions were to develop a new or modified data learning algorithm for a better load predictive performance. Those works have been done by using different data learning frameworks mostly on Pattern Matching, Hidden Markov Model-based approach, Multi-Label Classification, and Deep Neural Network [5], [6]. Multi-output learning framework is an approach that has drawn more attention in many research areas on prediction of multiple output data [7]. In the NILM applications, the framework is compatible to represent the operating status of multiple appliances to its multiple-output data format. This work presents the use of Multi-Target Regression approach, a new data learning framework for the NILM. The framework could provide a direct estimation of power demand for the target appliances which is a benefit over the conventional multi-output classification-based approaches [8], [9]. We also propose the experimental design process based on the machine learning discipline to obtain the optimal predictive performance.

The associate editor coordinating the review of this manuscript and approving it for publication was Arup Kumar Goswami.

Thus, the key contributions of this work can be summarized as follows.

(1) We introduced Multi-Target Regression as the data learning framework and its suitability for energy disaggregation tasks.

(2) The experimental procedure was proposed to design a data learning system with optimal performance for appliance loads inference.

(3) We evaluated the proposed approach on two public data sets with performance benchmarking to verify the effectiveness of the approach.

The remainder of this paper is organized as follows: Section II reviews the development, advantages, disadvantages of the major existing approaches, and the research direction. Section III briefs the key descriptions of the proposed approach, system design process, and performance indexes. Section IV presents the experimental results and discussion, and Section V concludes this paper.

II. RELATED WORKS

The research in the NILM system development could be briefly classified into two categories of event-based and non-event-based approaches. The event-based approaches relied on detecting power state switching (ON \leftrightarrow OFF) which learned the steady-state features of changed power data (ΔP , ΔQ) [1], [10] or power ‘ON’ transient features for more accurate loads identification [11], [12]. The need of high-frequency sampling to detect the transient features made some limitations; the data storage problem and integration of the data modeling to the smart meter which normally operated at low-frequency sampling [5], [13].

The non-event-based approach referred to the method that identified the appliances operating status without detecting the switch events. Thus, the capability to handle low-frequency data was an advantage over the event-based approach. The well-known approaches were based on the Factorial Hidden Markov Model (FHMM) [14] which factorized the aggregate as a time series to predict the hidden power states of appliances. The load power data was inferred from the predicted state by the models [15], [16]. These approaches could well fit for modeling appliances with multiple power states but required high computational complexity through parameters and probabilistic modeling. Other non-event-based approaches used the conventional classification algorithms to classify the appliance power state using Neural Network [17], k-Nearest Neighbor (kNN) [18], Support Vector Machine [19]. The classification approach, however, modeled each appliance label independently which did not reflect the practical behavior of appliances correlation that could have existed. Recent NILM developments applied Deep Neural Network (DNN) as the data learning framework [20]. Some common network architectures were employed, for example, Convolutional Neural Network [21], Long-Short Term Memory [22], and Denoising Autoencoder [20]. The advantages of the framework were the automatic features

extraction and system scalability through transfer learning. However, the Deep Neural Network required a large number of training samples to obtain adequate accuracy performance and training the model involved the configuration of many parameters and the network structure adjustment.

To ascertain the applicability of low-frequency sampling data with less computational complexity, this work proposes Multi-Target Regression (MTR) framework [23], a new data learning system for the energy disaggregation task. The approach is a category of multi-output learning [24], [25] where the model can directly estimate the appliance power consumption data from the aggregate measurement with the incorporation of labels correlation in data modeling. Unlike the multi-label classification [8], [26] or HMM-based learning framework that relied on classification tasks in which the primary output was the discrete power state values. It then needs a further process to convert the power state into the estimated power data, which might induce deviation or prediction loss. The proposed approach takes advantage of this issue, in addition, the ability to incorporate multiple features of the aggregate data could enhance the overall predictive performance.

III. PROPOSED METHODS

The objective of a NILM system is to decompose the aggregate consumption data into the estimated power consumption of contributed appliances through the disaggregation algorithm as expressed in (1).

$$f(P(t) \mapsto \hat{p}_i(t)) : P(t) = (\hat{p}_1(t) + \hat{p}_2(t) + \dots + \hat{p}_m(t)) + \epsilon(t) \quad (1)$$

where P is the ground-truth aggregate power data which is the input given to the disaggregation model (f). The model is then used to predict and generate the output \hat{p}_i as the estimated power data of appliances i from the total m appliances at time t . The error term (ϵ) is the loss in AC line or measurement error. The task of predicting multiple appliance labels in which each label gives the estimated power consumption can then basically be represented as a problem of multi-output or multi-target regression [7], [23].

A. MULTI-TARGET REGRESSION

Multi-target regression is a type of supervised learning framework where the goal is to estimate the numeric value of the target variable from the observation data. It has been successfully utilized in many applications, for example, predicting gas levels in multiple tanks of a gas converter system [27], ecological modeling for a prediction of multiple variables that describe the quality of plantations [28], and estimation of biophysics parameters from multiple sensing images [29].

For the application of NILM, a data set \mathcal{D} of N instances consists of the observation X_1, \dots, X_k for k electrical features of the aggregate data and the ground-truth power data for m contributing appliances Y_1, \dots, Y_m . Since the observation

TABLE 1. Characteristics of some common-used methods for multi-target regression.

Method	General operation
<i>Problem Transformation Approach</i>	
Single-Target (ST) method	<ul style="list-style-type: none"> - Build an individual regressor for each output label [23]. - Use an off-the-shelf regression algorithm as the base regressor.
Regressor Chain (RC) method	<ul style="list-style-type: none"> - Select a random chain of target labels, the input set is transformed by cascading a label from the chain [31]. - Label dependency is incorporated through the sequence of a cascaded chain.
<i>Algorithm Adaptation Approach</i>	
Multi-target regression tree method	<ul style="list-style-type: none"> - Build a tree regressor for multiple outputs with an extension from the single output approach [32].
Rule induction method	<ul style="list-style-type: none"> - Build an ensemble of regression trees using a set of rules and select the best subset [32].

and power data for each instance can be characterized as a vector of input and output, the data set can be expressed in (2) as a term of the input-output pair.

$$\mathcal{D} = \{(x^1, y^1), \dots, (x^N, y^N)\} \quad (2)$$

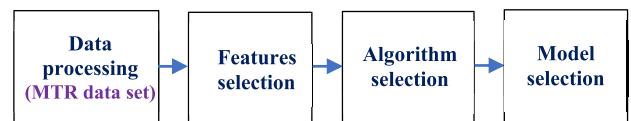
where x^j is the input vector of k features by the aggregate data for j^{th} instance (x_1^j, \dots, x_k^j)

y^j is the output vector of appliances power data of m appliances for j^{th} instance (y_1^j, \dots, y_m^j)

The multi-target regression model learns data by mapping the vector of aggregate data (x) to the vector of appliance power data (y) through a disaggregation function ($f: x \mapsto y$) or a multi-target regression algorithm. The model is then used for simultaneously predicting the appliance power consumption data ($\{\hat{y}_{T+1}, \dots, \hat{y}_{T'}\}$) from new aggregate instances ($\{x_{T+1}, \dots, x_{T'}\}$) where T' represents the period of new samples.

The data learning method for multi-target regression can be classified into 2 categories of Problem Transformation and Algorithm Adaptation approach. The first category transforms the multi-output regression problem (y data) into one or more sets of single-output problems using a multi-output regressor then applies an off-the-shelf regression algorithm or based regressor to build the regression models. The second category adapts a single-target regression algorithm to handle the multi-target data directly [23]. If the label correlation within the appliances exists, the Algorithm Adaptation approach could capture the dependency and provide better regression performance compare to the method from the Problem Transformation approach [30].

The characteristics of common-used multi-target regression methods for each approach can be summarized in Table 1.

**FIGURE 1.** The experimental design process for the proposed approach.

B. DESIGN PROCESS

The design process in this work utilizes the key procedure in machine learning design to optimize the performance of appliance loads prediction. The process includes (1) Data processing: To clean up, aggregate, and build a multi-target regression (MTR) data set. (2) Features selection: To choose the best combination of relevant electrical features that could deliver the best data estimation performance. (3) Algorithm selection: To choose the best multi-target regressor from common-used regression algorithms by comparing the predictive performance. (4) Model selection: To optimize the regression model through model parameter tuning. The process is then finalized by the performance comparison of the proposed approach to the existing ones. The experimental design process is summarized in Fig. 1.

C. DATA AND MULTI-TARGET DATA SET

The data sets publicly available for the NILM research have different characteristics in sampling frequency, electrical features to be monitored, and set of appliances. This work evaluated the AMPds2 data [33] and ECO data [34] since they both have a couple of electrical features for the mains meter (aggregate data) which were useful in the process of features selection and loads power data (P) were equipped for all sub-meters. For AMPds2, it is a collection of 20 load labels measurement with important electrical features, for example, current, power (active, reactive, apparent), and power factor from a house in Canada with a period of 2 years and 1-minute data sampling interval. The data set contains some appliance labels with a single appliance per label, for example, Clothes Washer, Heat Pump, Wall Oven. Other labels contain multiple appliances within a label, for example, Basement Room, Home Office, Entertainment Plug.

For ECO data, it is a collection of some appliance loads measurement from 6 houses in Switzerland for 8 months. The measurement consists of current, voltage, and their phase shift from three phases of the mains with 1 second of data sampling frequency. After cleaning and concatenating the aggregate data (x) to the set of output targets (y), i.e., the power consumption of each appliance label, a sample illustration of the multi-target data set is shown in Fig. 2. The data set has m appliance labels and N instances with the observation of I , P , Q , and PF features for the aggregate data.

D. EVALUATION TOOLS AND PERFORMANCE INDEX

This work adopted `sklearn.multioutput`, a method for learning and evaluating multiple-output problems, from Python 3.7 with scikit-learn package (ver. 0.24.2) [35]. Data learning methods available for multi-target regression problems are

I(A)	P(W)	Q(VAR)	PF	P _{L1}	P _{L2}	...	P _{Lm}
				p ₁₁	p ₂₁	...	p _{m1}
				p ₁₂	p ₂₂	...	p _{m2}
				p ₁₃	p ₂₃	...	p _{m3}
			
			
				P _{IN}	p _{2N}	...	p _{mN}

FIGURE 2. A sample illustration of multi-target regression data set.

from the problem transformation approach, consisting of *MultiOutputRegressor* for Single Target (ST) and *RegressorChain* for Regressor Chain (RC). Both methods are meta-estimators that require a base regressor in the model construction to extend single-output regressors to multi-output regressors. For the algorithm adaptation approach, we used the methods of multi-target regression tree and rule induction from CLUS [32]. It was a Java-based data learning API for learning the predictive clustering trees which were extensible for multi-task problems.

The performance evaluation of the regression algorithm is to measure how close the estimated values are as compared to the true value. There are 2 common performance indexes for energy disaggregation tasks as follows.

1) Disaggregation accuracy: This index defines the accuracy of power prediction by the one-complement of the difference between the predicted power to the true power over a given period [36]. We consider this index instead of Root Mean Square Error (RMSE) or Mean Square Error (MAE) since it provides the relative value rather than the absolute ones, so the error of load labels with different power consumption levels can be scaled and made a comparison. The overall disaggregation accuracy can be expressed as (3).

$$\text{Disaggregation accuracy} = 1 - \frac{\sum_{t=1}^T \sum_{i=1}^n |\hat{y}_t^i - y_t^i|}{2 * \sum_{t=1}^T \sum_{i=1}^n y_t^i} \quad (3)$$

where \hat{y}_t^i and y_t^i are the predicted and true power data for t^{th} appliance label within a period of T instances. For individual label accuracy, the equation can also be applied by just omitting the summation of n appliances calculation.

2) F-score: A performance measure for calculating the harmonic mean of precision (P) and recall (R). The F-score was evaluated by macro-average whereby, for a binary identification of power ‘On’ and ‘Off’ classes, the index calculates P and R for each class and does average as (4).

$$F\text{-score} = \frac{1}{k} \sum_{c=1}^k \frac{2 \times P_c \times R_c}{P_c + R_c} \quad (4)$$

where P_c and R_c are the precision and recall for each class of data, k is the number of appliance labels.

Since the primary output of the regression algorithm is the estimated power data for each appliance label, we need to

translate this data into the power state in binary form {0,1} to indicate if the appliance is being turned OFF or ON. A basic strategy is to use a threshold value to determine the power state from the ground-truth power data, then apply the same value to the estimated power data. The F-score value is then evaluated based on the true and predicted power states to obtain the precision and recall values as equation (4). The identification of the power state (p_{jk}) can be defined by (5).

For j^{th} data instance and k^{th} appliance label,

$$p_{jk} = \begin{cases} 1, & \text{if } P_{jk} \geq P_{Th(k)} \\ 0, & \text{if } P_{jk} < P_{Th(k)} \end{cases} \quad (5)$$

where $P_{Th(k)}$ is the threshold power data of appliance label k . It is derived from the current consumption data that corresponds to the power value, which differentiates the appliance operating state of ‘OFF’(‘0’) or ‘ON’(‘1’).

The F-score value is normally correlated to the disaggregation accuracy value since both indexes are generated by the estimated power data. In addition, a deviation of the selected threshold power might slightly affect the F-score due to some changes in the power status data (‘0’/‘1’) that could alter the F-score value.

IV. RESULTS AND DISCUSSION

This section described the experimental procedures and results which first started with the AMPds2 data set evaluation. We split the data for 3 months (129,600 samples) to evaluate the optimal components in data learning. The evaluation was conducted using 10-fold cross-validation ($K = 10$), the performance value thus composes of the mean score with its standard deviation values. Each procedure was presented in the following topics.

A. FEATURE SELECTION

The multi-target data set was created by using a set of electrical features that could indicate the presence of appliance operation, which were current (I), active power (P), reactive power (Q), and power factor (PF). The Single-Target method with Random Forest Regressor as the multi-output regressor and the based regressor respectively were used for the test purpose. To evaluate the performance by each set of features, the disaggregation accuracy was examined by using the forward selection technique based on the paired t-test [37]. It determined a candidate set of features that would be statistically significant of difference from the others. The result presented in Table 2 showed that the combination of all 4 electrical features could deliver the best disaggregation performance, the bold figure indicates the best value. On the other hand, this result shows an advantage of the multi-output learning framework, where more relevant features can be incorporated for higher predictive performance. Thus, these features set will be used for the next experiments.

B. LEARNING ALGORITHM SELECTION

This experiment aimed to choose the best regression algorithm for loads power estimation. A set of common single

TABLE 2. Disaggregation accuracy by different sets of features.

Features set	Disaggregation Accuracy
$P-Q$	$(83.25 \pm 0.17)\%$
$P-Q-I$	$(87.38 \pm 0.15)\%$
$P-Q-I-PF$	$(89.06 \pm 0.12)\%$

TABLE 3. Disaggregation accuracy by different multi-target regression algorithms.

	Regressor	Disaggregation Accuracy
Problem Transformation method	Ridge Regression	$(54.81 \pm 0.17)\%$
	Support Vector Regressor (SVR)	$(74.17 \pm 0.12)\%$
	Random Forest Regressor (RFR)	$(89.06 \pm 0.12)\%$
	Gradient Boosting Regressor (GBR)	$(83.81 \pm 0.15)\%$
	Regressors - GBR	$(83.92 \pm 0.12)\%$
	Chain - RFR	$(86.78 \pm 0.21)\%$
	Multi-Target Regression Tree	$(82.75 \pm 0.16)\%$
	Rule Induction	$(81.70 \pm 0.23)\%$

output regressors was used as the based regressor for the Problem Transformation approach. The candidate algorithms were ranged from the linear model of Ridge Regression, the non-linear model of Support Vector Machine, and ensemble algorithms of Random Forest Regressor and Gradient Boosting Regressor for both Single-Target and Regressor Chain methods. For the Algorithm Adaptation approach, two algorithms based on Regression Tree were deployed. Table 3 showed that the Single-Target method using Random Forest as the based regressor could provide the best disaggregation accuracy. Random Forest is an ensemble learning which proved to provide good performance by previous research [38], [39]. Although the regressor ignores label dependency, it could provide good performance. This could be described as the data set was containing 20 appliance labels which were a relatively high number. Thus, determining or revealing the correlation among appliances could be tough by the model. However, this behavior can be different for data sets with different characteristics such as the appliance power profiles or set of electrical features under measurement.

C. MODEL SELECTION

This experiment aims to tune up the model performance through a given range of model parameters using the

TABLE 4. Disaggregation accuracy after tuning up the model parameters.

Tuned parameter	Range	Optimal parameters	Disaggregation Accuracy
number of estimators ($n_estimator$)	[10, 20, 30, 50, 80, 100]	80	$(89.24 \pm 0.11)\%$
Maximum Tree Depth (max_depth)	[5, 10, 20, 30, 40, 50]	30	

TABLE 5. Execution time for data training and testing.

Training time (s)	Testing time (s)
73.1 ± 0.18	1.1 ± 0.06

Grid Search method in scikit-learn. The number of trees ($n_estimator$) and the maximum depth of tree branches (max_depth) were applied for the tuning process. These 2 parameters affect the way that the model creates decision trees for the Random Forest Regressor. Table 4 shows that the performance can be increased through a new set of model parameters.

D. EXECUTION TIME

Using the prior system configuration for data learning, this experiment aims to evaluate the execution time for data training and data testing. The 3-month samples (129,600 instances) were split 80% and 20% for training and testing, respectively. We used the Python method `time.time()` to check the starting and stopping time then evaluated the period for the training and testing process. This experiment was run on a 64-bit Windows 10 PC with an AMD RYZEN 5 (2.1 GHz processor) and an 8 GB memory. All evaluations were executed using a single thread setting. The figures of data training time and testing time presented in Table 5 were the mean and standard deviation values from 10 runs for fitting the model and predicting the test data.

E. TRUE AND PREDICTED POWER DATA

This experiment illustrates the plots of ground-truth (true) power against the predicted power data by some appliance labels. This is to visualize how well the predictions could track the true values in time series data as shown in Fig. 3. The data estimation could perform relatively well for high-power appliances (Clothes Dryer and Heat Pump) where the false positives had a low proportion. The low-power appliances, for example, the Kitchen Fridge had more proportion of falses in data estimation.

In a scenario where two or more appliances were turned ON or activated simultaneously, the true and predicted power together with the associated aggregate data for some of these

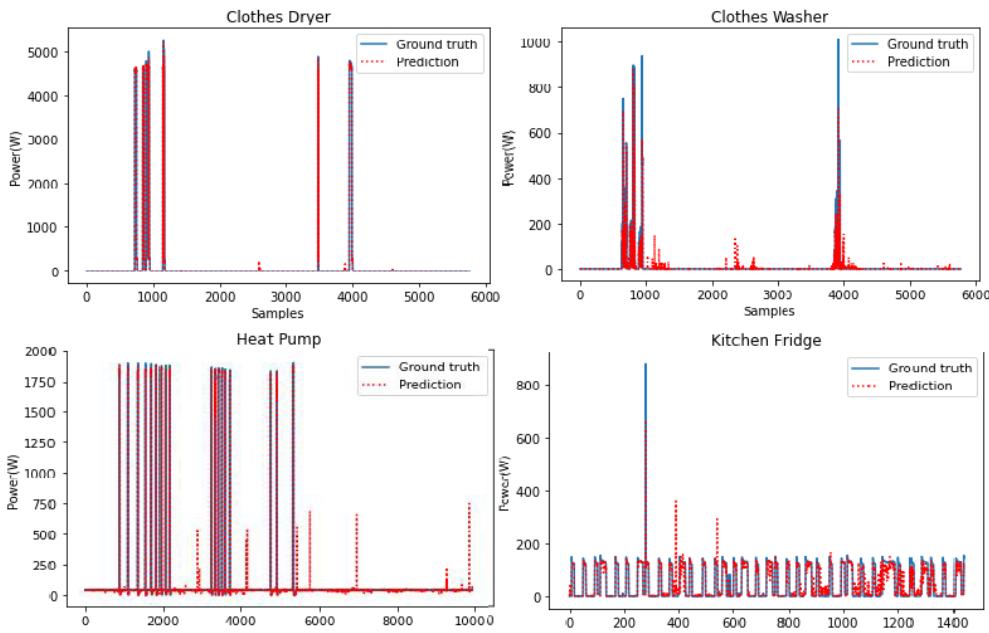


FIGURE 3. True and predicted power plots of individual appliances operation for AMPds2 data set.

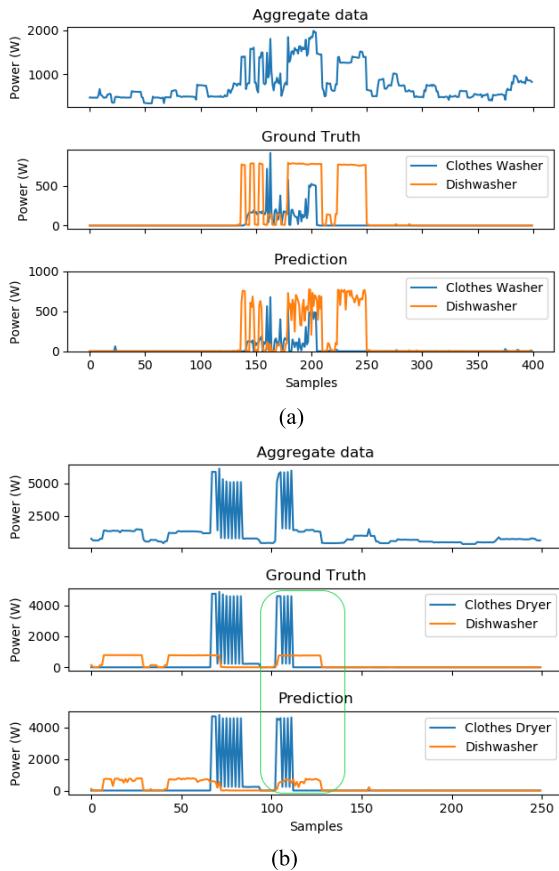


FIGURE 4. True and predicted power of appliances with simultaneous power ON activations for AMPds2 data set.

events were illustrated in Fig. 4 (a) and (b). Apart from the power estimation of an appliance at any given samples, this

result showed the capability of the proposed approach on disaggregating multiple appliances from the aggregate data at any specific time instance.

F. LABEL ACCURACY AND PERFORMANCE COMPARISON FOR AMPds2 DATA SET

This experiment evaluated the predictive performance for some appliance labels and the overall for the entire set of labels by the train/test split of 0.8/0.2. The benchmarking approaches were Super-State Hidden Markov Model (SSHMM) [2], [40] and Denoising Autoencoder (DAE) network [18]. The SSHMM was a variant of HMM-based which defined the power states from the combination of individual appliances status and the approach claimed to outperform some other HMM-based variants. The DAE approach used the network architecture presented in [18] which was shown to outperform the other network topologies. The implementation used the *neural-disaggregator* which was referenced from [41] and executed using Keras/Tensorflow and NILMTK [42]. The Denoising Autoencoder network aimed to reconstruct the clean power data of the target appliance from the aggregate measurement. The aggregate data was considered to be noisy because it also consists of data that was generated by other appliances.

We extended the range of AMPds2 data to 1 year (first year of the entire 2-year range) and the denoised data version was used. This was the data where the unmetered power data was subtracted from the actual aggregate measurement and it was for a fair comparative experiment. The evaluation results for disaggregation accuracy and F-score showed in Table 6 and Table 7, respectively.

Table 6 and Table 7 showed that this work outperformed the other two approaches by the overall and most of the

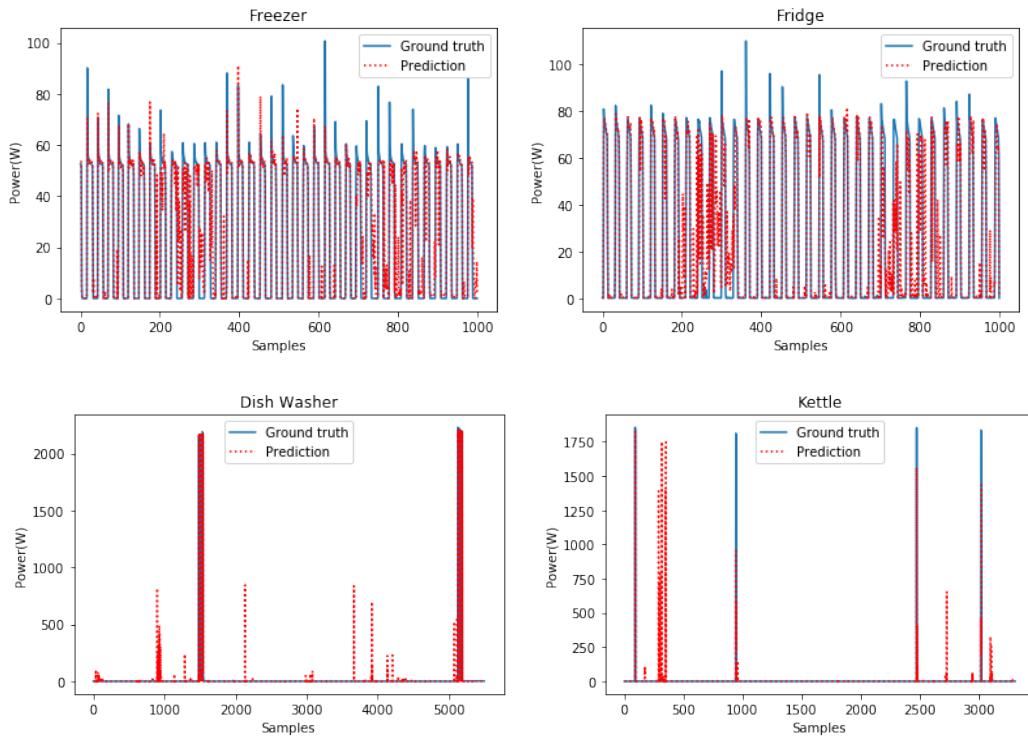


FIGURE 5. True and predicted power plots of individual appliances operation for ECO data set.

TABLE 6. Labels disaggregation accuracy and performance comparison for AMPds2 data set.

Load	Disaggregation accuracy		
	This work	SSHMM	DAE
Overall	92.2%	91.9%	85.0%
Basement	84.2%	83.1%	72.5%
Clothes Dryer	98.9%	96.2%	92.0%
Clothes Washer	66.3%	46.9%	65.2%
Dishwasher	62.2%	76.1%	60.5%
Kitchen Fridge	77.3%	82.5%	80.7%
HVAC/Furnace	98.6%	98.4%	82.2%
Heat Pump	98.1%	97.0%	90.8%
Home Office	78.4%	34.5%	60.2%
Wall Oven	73.2%	75.7%	65.5%

individual labels. The SHMM provided good performance to appliances of Kitchen Fridge and HVAC/Furnace. Whereas, the DAE performed well on high-power appliances with a large number of power activations like the Clothes Dryer and Heat Pump. Since the DAE network learned sets of sequential data when the appliance power was activated rather than learned by sample-based as the proposed approach. Thus, for the same amount of training samples by this experiment, the

TABLE 7. Labels F-score and performance comparison for AMPds2 data set.

Load	F-score		
	This work	SSHMM	DAE
Overall	91.9%	91.7%	82.5%
Basement	82.6%	85.3%	70.2%
Clothes Dryer	81.7%	79.0%	81.0%
Clothes Washer	60.1%	12.0%	55.5%
Dishwasher	75.3%	73.3%	60.5%
Kitchen Fridge	75.4%	84.9%	80.2%
HVAC/Furnace	97.4%	99.9%	85.5%
Heat Pump	98.7%	98.3%	92.5%
Home Office	32.1%	26.1%	25.2%
Wall Oven	65.8%	68.0%	55.8%

proposed MTR approach could provide better performance than that of the DAE network. Typically, to obtain a good performance by this network type, the training requires a large number of samples and empirical adjustment on the network parameters [18].

Overall, the performance values of disaggregation accuracy and F-score of high-power appliances (Clothes Dryer, HVAC/Furnace, and Heat Pump) were higher than that of

TABLE 8. Labels disaggregation accuracy and performance comparison for ECO data set.

Load	Disaggregation accuracy		
	This work	SSHMM	DAE
Overall	85.5%	83.1%	79.0%
Fridge	90.5%	88.6%	85.2%
Freezer	94.4%	90.6%	87.7%
Dish Washer	87.4%	80.2%	79.6%
Kettle	66.8%	70.1%	62.0%
Lamp	65.6%	63.3%	64.2%
TV	86.3%	82.7%	79.8%
Audio	62.2%	68.8%	55.0%

the lower power appliances (Kitchen Fridge, Clothes Washer) and the combination of appliances in a label (Basement, Home Office) which aligned to the results of previous studies [1], [13].

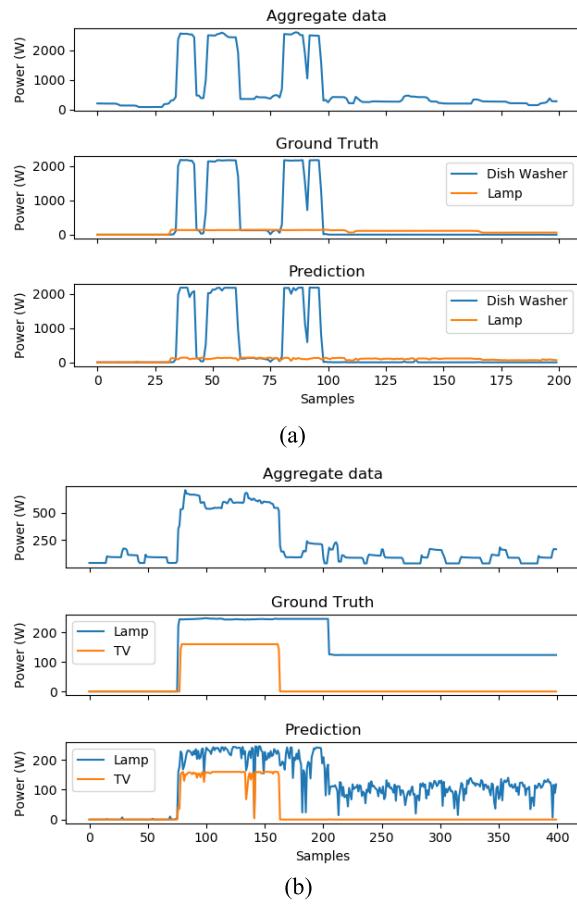
G. PERFORMANCE EVALUATION OF ECO DATA SET

Data set from House 2 was selected for evaluation since the proportion of unmetered power data was the lowest value among the six houses which was a benefit for data training and model performance. The data set was a noisy version (unmetered data was not excluded) and was resampled to a 1-minute interval and split for the train/test to 0.8/0.2. The same multi-target regression algorithm and model configuration were adopted for data learning. Some visualizations of the predictions and ground truth data for individual appliance operations and simultaneous power activations were shown in Fig. 5 and Fig. 6, respectively. Interestingly, the Lamp with low-power data could be disaggregated from the much more high-power data of the Dishwasher. In addition, the model could perform well for disaggregating the low-power appliances of Lamp and TV.

Table 8 and Table 9 presented the predictive performance values which showed to be comparable to the SSHMM and DAE approaches for some appliance labels, but they outperformed for the overall performance values. The performance values for Fridge and Freezer were quite high by the three approaches since these appliances had a continuous operation which could be the advantage for generating a sufficient number of training samples. The Kettle and Audio got lower performance than the other labels. For the Kettle, even if it was a high-power appliance but it was rarely activated in which its energy consumption contributed to 3.99% of the whole house [34]. Thus, the rare positive training samples could hardly create a model with high predictive performance. For Audio, this was the low-power appliance with normally contributed small changes to the aggregate data, thus making the model difficult to extract its presence.

TABLE 9. Labels F-score and performance comparison for ECO data set.

Load	F-score		
	This work	SSHMM	DAE
Overall	83.3%	79.0%	77.5%
Fridge	85.2%	87.9%	82.5%
Freezer	89.2%	91.2%	88.5%
Dish Washer	87.1%	74.8%	78.0%
Kettle	63.3%	68.8%	65.2%
Lamp	58.7%	55.0%	56.6%
TV	79.6%	69.9%	72.8%
Audio	51.3%	48.0%	47.0%

**FIGURE 6.** True and predicted power of appliances with simultaneous power ON activations for ECO data set.

According to the proposed Multi-Target Regression approach, we can summarize the key advantages against its peers as (1) The capability to incorporate several relevant features in the data training process. It could help the model learns data better, unlike the HMM-based or DAE approach that used just the power data (P) feature for model training. The direct data regression with lower data processing could also help reduce loss from estimation. (2) The approach is

more computationally efficient and requires fewer model configurations, just for the multi-output regressor and the base regressor model parameters. Some limitations of the approach, however, are that (1) It requires the full labeled data in training when there is a change in the set of target appliances, which is a property of the supervised learning. To ease the issue of intensive training, the semi-supervised concept which uses partially labeled data together with the unlabeled data in the training process [43] can be employed. Another issue is (2) the limited performance for the multi-power state appliances (like clothes washers or dishwashers). The multi-state operation can generate multiple levels of power data which makes the regression model harder to accurately estimate the power demand. Increasing more training samples that are associated with the relevant power level can help improve the predictive performance.

V. CONCLUSION

The multi-target regression approach shows the suitable capabilities for NILM application in which it can perform data learning tasks for estimating the power data of multiple appliance labels. The experimental results illustrate the evaluation procedure to obtain the optimal learning component for the best performance of loads disaggregation. Using the electrical features set of *I-P-Q-PF* and the multi-target regression model with the Random Forest as the base regressor, the predictive performance when evaluated AMPds2 data set could reach 92.2% and 91.9% for the overall disaggregation accuracy and F-score value, respectively. For the ECO data set, the overall disaggregation accuracy and F-score value reached 85.5% and 83.3%, respectively. The results outperformed the benchmarking approaches of the SSHMM and DAE network by the overall and most individual appliance labels.

The system design procedure can be applied for individual home energy monitoring. The proposed experimental process would be split into 2 phases. The first phase is the data acquisition and training data which is to do the data collection and perform data learning as the designated procedure. The second phase is applying the model for new or unknown aggregate data to obtain the power estimation and infer the status of appliances. The outperformance of loads identification by the proposed approach to the other benchmarks makes the application a more efficient system for home energy management.

REFERENCES

- [1] G. W. Hart, "Nonintrusive appliance load monitoring," *Proc. IEEE*, vol. 80, no. 12, pp. 1870–1891, Dec. 1992, doi: [10.1109/5.192069](https://doi.org/10.1109/5.192069).
- [2] S. Makonin, F. Popowich, I. V. Bajic, B. Gill, and L. Bartram, "Exploiting HMM sparsity to perform online real-time nonintrusive load monitoring," *IEEE Trans. Smart Grid*, vol. 7, no. 6, pp. 2575–2585, Nov. 2016, doi: [10.1109/TSG.2015.2494592](https://doi.org/10.1109/TSG.2015.2494592).
- [3] K. C. Armel, A. Gupta, G. Shrimali, and A. Albert, "Is disaggregation the holy grail of energy efficiency? The case of electricity," *Energy Policy*, vol. 52, pp. 213–234, Jan. 2013, doi: [10.1016/j.enpol.2012.08.062](https://doi.org/10.1016/j.enpol.2012.08.062).
- [4] L. Pereira and N. Nunes, "An empirical exploration of performance metrics for event detection algorithms in non-intrusive load monitoring," *Sustain. Cities Soc.*, vol. 62, pp. 1–16, Nov. 2020, doi: [10.1016/j.scs.2020.102399](https://doi.org/10.1016/j.scs.2020.102399).
- [5] A. Zoha, A. Gluhak, M. A. Imran, and S. Rajasegarar, "Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey," *Sensors*, vol. 12, no. 12, pp. 16838–16866, Dec. 2012, doi: [10.3390/s121216838](https://doi.org/10.3390/s121216838).
- [6] P. Huber, A. Calatroni, A. Rumsch, and A. Paice, "Review on deep neural networks applied to low-frequency NILM," *Energies*, vol. 14, no. 9, pp. 1–34, Apr. 2021, doi: [10.3390/en14092390](https://doi.org/10.3390/en14092390).
- [7] D. Xu, Y. Shi, I. W. Tsang, Y.-S. Ong, C. Gong, and X. Shen, "Survey on multi-output learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 7, pp. 2409–2429, Jul. 2020, doi: [10.1109/TNNLS.2019.2945133](https://doi.org/10.1109/TNNLS.2019.2945133).
- [8] K. Basu, V. Debusschere, S. Bacha, U. Maulik, and S. Bondyopadhyay, "Nonintrusive load monitoring: A temporal multilabel classification approach," *IEEE Trans. Ind. Informat.*, vol. 11, no. 1, pp. 262–270, Feb. 2015, doi: [10.1109/TII.2014.2361288](https://doi.org/10.1109/TII.2014.2361288).
- [9] B. Buddhabhai, W. Wongseree, and P. Rakwamsuk, "An energy prediction approach for a noninvasive load monitoring in home appliances," *IEEE Trans. Consum. Electron.*, vol. 66, no. 1, pp. 96–105, Feb. 2020, doi: [10.1109/TCE.2019.2956638](https://doi.org/10.1109/TCE.2019.2956638).
- [10] Z. Xu, W. Chen, and Q. Wang, "A new non-intrusive load monitoring algorithm based on event matching," *IEEE Access*, vol. 7, pp. 55966–55973, 2019, doi: [10.1109/ACCESS.2019.2913469](https://doi.org/10.1109/ACCESS.2019.2913469).
- [11] L. K. Norford and S. B. Leeb, "Non-intrusive electrical load monitoring in commercial buildings based on steady-state and transient load-detection algorithms," *Energy Buildings*, vol. 24, no. 1, pp. 51–64, May 1999, doi: [10.1016/0378-7788\(95\)00958-2](https://doi.org/10.1016/0378-7788(95)00958-2).
- [12] H.-H. Chang, "Non-intrusive demand monitoring and load identification for energy management systems based on transient feature analyses," *Energies*, vol. 5, no. 11, pp. 4569–4589, Nov. 2012, doi: [10.3390/en5114569](https://doi.org/10.3390/en5114569).
- [13] M. Zeifman and K. Roth, "Nonintrusive appliance load monitoring: Review and outlook," *IEEE Trans. Consum. Electron.*, vol. 57, no. 1, pp. 76–84, Feb. 2011, doi: [10.1109/TCE.2011.5735484](https://doi.org/10.1109/TCE.2011.5735484).
- [14] Z. Ghahramani and M. I. Jordan, "Factorial hidden Markov models," *Mach. Learn.*, vol. 29, pp. 245–273, Nov. 1997, doi: [10.1023/A:1007425814087](https://doi.org/10.1023/A:1007425814087).
- [15] H. Kim, M. Marwah, M. Arlitt, G. Lyon, and J. Han, "Unsupervised disaggregation of low frequency power measurements," in *Proc. SIAM Int. Conf. Data Mining*, Apr. 2011, pp. 747–758, doi: [10.1137/1.9781611972818.64](https://doi.org/10.1137/1.9781611972818.64).
- [16] J. Z. Kolter and T. Jaakkola, "Approximate inference in additive factorial HMMs with application to energy disaggregation," in *Proc. 15th Int. Conf. Artif. Intell. Stat.*, La Palma, Canary Islands, 2012, pp. 1472–1482.
- [17] A. G. Ruzzelli, C. Nicolas, A. Schoofs, and G. M. P. O'Hare, "Real-time recognition and profiling of appliances through a single electricity sensor," in *Proc. SECON*, Boston, MA, USA, Jun. 2010, pp. 1–9, doi: [10.1109/SECON.2010.5508244](https://doi.org/10.1109/SECON.2010.5508244).
- [18] M. Figueiredo, A. de Almeida, and B. Ribeiro, "Home electrical signal disaggregation for non-intrusive load monitoring (NILM) systems," *Neurocomputing*, vol. 96, pp. 66–73, Nov. 2012, doi: [10.1016/j.neucom.2011.10.037](https://doi.org/10.1016/j.neucom.2011.10.037).
- [19] T. Onoda, G. Ratsch, and K.-R. Müller, "Applying support vector machines and boosting to a non-intrusive monitoring system for household electric appliances with inverter," in *Proc. ICSC Neural Comput.*, Berlin, Germany, 2000, pp. 1–7.
- [20] J. Kelly and W. Knottenbelt, "Neural NILM: Deep neural networks applied to energy disaggregation," in *Proc. 2nd ACM Int. Conf. Embedded Syst. Energy-Efficient Built Environ. (BuildSys)*, Seoul, South Korea, Nov. 2015, pp. 55–64.
- [21] C. Zhang, M. Zhong, Z. Wang, N. Goddard, and C. Sutton, "Sequence-to-point learning with neural networks for nonintrusive load monitoring," 2018, *arXiv:1612.09106*.
- [22] X. Zhou, S. Li, C. Liu, H. Zhu, N. Dong, and T. Xiao, "Non-intrusive load monitoring using a CNN-LSTM-RF model considering label correlation and class-imbalance," *IEEE Access*, vol. 9, pp. 84306–84315, 2021, doi: [10.1109/ACCESS.2021.3087696](https://doi.org/10.1109/ACCESS.2021.3087696).
- [23] H. Borchani, G. Varando, C. Bielza, and P. Larrañaga, "A survey on multi-output regression," *WIREs Data Mining Knowl. Discovery*, vol. 5, no. 5, pp. 216–233, Sep./Oct. 2015, doi: [10.1002/widm.1157](https://doi.org/10.1002/widm.1157).
- [24] G. Tsoumakas and I. Katakis, "Multi-label classification: An overview," *J. Data Warehouse Mining*, vol. 3, no. 3, pp. 1–17, 2007.
- [25] G. Madjarov, D. Kocev, D. Gjorgjevikj, and S. Džeroski, "An extensive experimental comparison of methods for multi-label learning," *Pattern Recognit.*, vol. 45, no. 9, pp. 3084–3104, Sep. 2012, doi: [10.1016/j.patcog.2012.03.004](https://doi.org/10.1016/j.patcog.2012.03.004).

- [26] B. Buddhahai, W. Wongeree, and P. Rakkwamsuk, "A non-intrusive load monitoring system using multi-label classification approach," *Sustain. Cities Soc.*, vol. 39, pp. 621–630, May 2018, doi: [10.1016/j.scs.2018.02.002](https://doi.org/10.1016/j.scs.2018.02.002).
- [27] Z. Han, Y. Liu, J. Zhao, and W. Wang, "Real time prediction for converter gas tank levels based on multi-output least square support vector regressor," *Control Eng. Pract.*, vol. 20, no. 12, pp. 1400–1409, Dec. 2012, doi: [10.1016/j.conengprac.2012.08.006](https://doi.org/10.1016/j.conengprac.2012.08.006).
- [28] D. Kocev, S. Džeroski, M. D. White, G. R. Newell, and P. Griffioen, "Using single- and multi-target regression trees and ensembles to model a compound index of vegetation condition," *Ecol. Model.*, vol. 220, no. 8, pp. 1159–1168, Apr. 2009, doi: [10.1016/j.ecolmodel.2009.01.037](https://doi.org/10.1016/j.ecolmodel.2009.01.037).
- [29] D. Tuia, J. Verrelst, L. Alonso, F. Perez-Cruz, and G. Camps-Valls, "Multi-output support vector regression for remote sensing biophysical parameter estimation," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 4, pp. 804–808, Jul. 2011, doi: [10.1109/LGRS.2011.2109934](https://doi.org/10.1109/LGRS.2011.2109934).
- [30] G. Tsoumakas, I. Katakis, and I. Vlahavas, "Mining multi-label Data," in *Data Mining and Knowledge Discovery Handbook*, O. Maimon and L. Rokach, Eds. Boston, MA, USA: Springer, 2010, pp. 667–685.
- [31] E. Spyromitros-Xioufis, G. Tsoumakas, W. Groves, and I. Vlahavas, "Multi-target regression via input space expansion: Treating targets as inputs," 2012, *arXiv:1211.6581*.
- [32] J. Struyf and S. Džeroski, "Constraint based induction of multi-objective regression trees," in *Knowledge Discovery in Inductive Databases*. Berlin, Germany: Springer, 2006, pp. 222–233.
- [33] S. Makonin, B. Ellert, I. V. Bajić, and F. Popowich, "Electricity, water, and natural gas consumption of a residential house in Canada from 2012 to 2014," *Scientific Data*, vol. 3, no. 1, Jun. 2016, Art. no. 160037, doi: [10.1038/sdata.2016.37](https://doi.org/10.1038/sdata.2016.37).
- [34] C. Beckel, W. Kleiminger, R. Cicchetti, T. Staake, and S. Santini, "The ECO data set and the performance of non-intrusive load monitoring algorithms," in *Proc. 1st ACM Conf. Embedded Syst. Energy-Efficient Buildings*, Memphis, TN, USA, Nov. 2014, pp. 80–89, doi: [10.1145/2674061.2674064](https://doi.org/10.1145/2674061.2674064).
- [35] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, and D. Cournapeau, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Jan. 2011.
- [36] J. Z. Kolter and M. J. Johnson, "REDD: A public data set for energy disaggregation research," in *Proc. SIGKDD Workshop Data Mining Appl. Sustain.*, San Diego, CA, USA, 2011, pp. 59–62.
- [37] S. Derksen and H. J. Keselman, "Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables," *Brit. J. Math. Stat. Psychol.*, vol. 45, no. 2, pp. 265–282, Nov. 1992, doi: [10.1111/j.2044-8317.1992.tb00992.x](https://doi.org/10.1111/j.2044-8317.1992.tb00992.x).
- [38] P. R. Z. Taveira, C. H. V. De Moraes, and G. Lambert-Torres, "Non-intrusive identification of loads by random forest and firework optimization," *IEEE Access*, vol. 8, pp. 75060–75072, 2020, doi: [10.1109/ACCESS.2020.2988366](https://doi.org/10.1109/ACCESS.2020.2988366).
- [39] P. A. Schirmer, I. Mporas, and M. Paraskevas, "Evaluation of regression algorithms and features on the energy disaggregation task," in *Proc. 10th IISA*, Patras, Greece, Jul. 2019, pp. 1–4, doi: [10.1109/IISA.2019.8900695](https://doi.org/10.1109/IISA.2019.8900695).
- [40] S. Makonin and F. Popowich, "Nonintrusive load monitoring (NILM) performance evaluation," *Energy Efficiency*, vol. 8, pp. 809–814, Jul. 2015, doi: [10.1007/s12053-014-9306-2](https://doi.org/10.1007/s12053-014-9306-2).
- [41] O. Krystalakos, C. Nalmpantis, and D. Vrakas, "Sliding window approach for online energy disaggregation using artificial neural networks," in *Proc. 10th Hellenic Conf. Artif. Intell.*, Patras, Greece, Jul. 2018, pp. 1–6, doi: [10.1145/3200947.3201011](https://doi.org/10.1145/3200947.3201011).
- [42] N. Batra, J. Kelly, O. Parson, H. Dutta, W. Knottenbelt, A. Rogers, A. Singh, and M. Srivastava, "NILMTK: An open source toolkit for non-intrusive load monitoring," in *Proc. 5th Int. Conf. Future Energy Syst.*, Cambridge, U.K., Jun. 2014, pp. 265–276.
- [43] J. Levatić, D. Kocev, M. Ceci, and S. Džeroski, "Semi-supervised trees for multi-target regression," *Inf. Sci.*, vol. 450, pp. 109–127, Jun. 2018, doi: [10.1016/j.ins.2018.03.033](https://doi.org/10.1016/j.ins.2018.03.033).



BUNDIT BUDDHAHAI received the B.E. degree in electronics engineering from the King Mongkut's Institute of Technology Ladkrabang and the M.E. degree in electrical and information engineering and the Ph.D. degree in energy management technology from the King Mongkut's University of Technology Thonburi, Thailand. He is currently a Lecturer at Walailak University, Thailand. His research interests include energy monitoring, applied machine learning, and smart embedded systems.



STEPHEN MAKONIN (Senior Member, IEEE) is currently an Adjunct Professor in engineering science and the Principal Investigator of the Computational Sustainability Lab, Simon Fraser University. His research interests include computational sustainability and the understanding of socioeconomic issues. He is an active board member of IEEE communities and an editorial board of the *Nature Scientific Data journal*.

• • •