

Received May 19, 2019, accepted June 8, 2019, date of publication June 19, 2019, date of current version July 3, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2923742

Multi-Channel Recurrent Convolutional Neural Networks for Energy Disaggregation

MARIA KASELIMI¹, EFTYCHIOS PROTOPAPADAKIS¹, ATHANASIOS VOULODIMOS², (Member, IEEE),
NIKOLAOS DOULAMIS¹, (Member, IEEE), AND ANASTASIOS DOULAMIS¹, (Member, IEEE)

¹School of Rural and Surveying Engineering, National Technical University of Athens, 15773 Athens, Greece

²Department of Informatics and Computer Engineering, University of West Attica, 12243 Athens, Greece

Corresponding author: Maria Kaselimi (mkaselimi@mail.ntua.gr)

This work has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 768774.

ABSTRACT Power consumption signals of household appliances are characterized by randomly occurring events (e.g. switch-on events), making timeseries modeling a demanding process. In this paper, we propose a convolutional neural network (CNN)-based architecture with inputs and outputs formed as data sequences taking into consideration an appliance's previous states for better estimation of its current state. Furthermore, the proposed model endows CNN models with a recurrent property in order to better capture energy signal interdependencies. Using a multi-channel CNN architecture fed with additional variables related to power consumption (current, reactive, and apparent power), additionally to active power, overall performance, robustness to noise and convergence times are improved. The experimental results prove the proposed method's superiority compared to the current state of the art.

INDEX TERMS Convolutional neural network (CNN), deep learning, energy disaggregation, load monitoring, NILM, power, recurrent neural networks, tapped delay line, sequence to sequence modeling.

I. INTRODUCTION

Non-Intrusive Load Monitoring (NILM) estimates individual appliance power usage from aggregate measurements, thereby contributing to energy conservation through changing of consumers behavior, waste minimization, carbon footprint reduction, efficient network load handling and financial savings. The significance of the application has attracted the interest of an increasing number of researchers, leading to the proposal of a wide range of machine learning and signal processing techniques for energy disaggregation.

A number of recently proposed methods are based on deep learning, thus aiming to leverage the increased representational capabilities of models such as Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) networks and Stacked Denoising Autoencoders. In this work, we present a novel scalable CNN-based approach for energy disaggregation, called Multi-Channel Recurrent Tapped Delay Line CNN (MR-TDLCNN) which introduces

three innovative aspects to address respective challenges in NILM. Firstly, since disaggregation of total consumption includes inherently recurrent elements, our approach endows CNNs with a recurrent property that takes into consideration past estimation outputs. Secondly, the majority of proposed approaches only use active power as input measurement, thus disregarding potentially valuable information residing within other components, such as reactive power, apparent power and current. The proposed model is appropriately adapted to incorporate all four components; as is proven by the experimental results, the derived multi-channel model yields better performance compared to its single-channel counterparts. Finally, we propose a two-CNN architecture to increase robustness to noise. In particular, the architecture employs a second CNN, which utilizes the same inputs, along with the disaggregated signal of the first CNN. This approach allows for further noise reduction resulting in better overall disaggregation performance. The proposed approach goes beyond traditional event-based or state-based approaches. It provides refined outputs close to the actual appliance's patterns and exploits multiple channels of information than just the active power.

The associate editor coordinating the review of this manuscript and approving it for publication was Navanietha Krishnaraj Krishnaraj Rathinam.

The remainder of the paper is organized as follows: Section II provides a brief discussion on the taxonomy of applied disaggregation techniques. In Section III, we explain in detail the contribution of our approach compared to the related work. Section IV provides a detailed description of the proposed approach and of the related techniques, used for comparison purposes. Section V presents the experimental results, and Section VI concludes this work.

II. RELATED WORK AND PAPER CONTRIBUTION

Over the last few years deep learning has been applied to various application fields and scenarios, ranging from computer vision to natural language processing (NLP). In many cases deep learning outperformed previous work. The emergence of deep learning techniques in energy disaggregation dates back to 2015, when the first attempt to face NILM problem through deep learning techniques was made [1]. In that study, three neural network architectures on domestic loads are compared exploiting the most significant deep learning schemes: Convolutional Neural Networks (CNN), Long Short-term Memory (LSTM) networks, and Stacked Denoising Autoencoders. Since then, deep learning has gradually captured NILM researchers' interest [2], [3]. LSTM networks often outperform other deep learning schemes, due to their ability to handle lags of unknown duration between important events in a time series. Thus, LSTM networks have a comparative advantage among other deep learning methods in detecting changes in power consumption. Relevant studies have been carried out in the past ([4], [5]). Particularly, [4] proposes an LSTM model and additionally a novel signature to improve classification performance. Mauch and Yang use a generic two-layer bidirectional LSTM architecture [5]. In a previous work of ours we have also proposed a Bayesian optimized bidirectional LSTM regression model for NILM [6].

CNNs traditionally attain very good results in a variety of pattern recognition problems ([7], [8]), and they are popular for a wide variety of other applications as well. However, the usage of CNNs in the field of energy disaggregation is not straightforward mainly due to the difficulty of matching a traditional pattern recognition problem of 2D data with 1D data of power timeseries. Nonetheless, remarkable and worth mentioning studies are presented here, leading the way for a new era in energy disaggregation through deep learning techniques. Reference [9] presents a causal 1D CNN examining in parallel the effect of other variables that are related to the power, such as current, reactive power and apparent power. Reference [10] proposes a sequence to point CNN architecture underlying the importance of sliding windows to handle long-term timeseries. Barsim et al. developed a generic disaggregation model based on data-driven learning [11].

Apart from deep learning-based methods for NILM, there are various other machine learning approaches that are very popular in the field of energy disaggregation. These methods are either supervised or unsupervised. In cases of

supervised learning, problems can be further grouped into regression or classification problems.

Event-based NILM approaches leverage edge detection techniques for an optimal clustering of state transitions events [12]. Different classification tools are used, including Support Vector Machines (SVM) [13], Neural Networks, Decision Trees (DT) [14], and hybrid classification methods [12], [15]. Hart was the first to propose a method for disaggregating electrical loads based on combinatorial optimization (CO) through the clustering of similar events based on appliances characteristics [16]. Dynamic Time Warping (DTW), although with limited success in classifying multi-state appliances, nevertheless has been used for identifying unique load signatures for simpler appliance patterns [17]. Graph Signal Processing (GSP) [12] adopts a concept based on signal processing to correlate signals in time and space domain by embedding the structure of signals onto a graph. Recently, a Modified Cross-Entropy method for events' classification has been proposed [18], a method based on combinatorial optimization formulating NILM as a Cross-Entropy problem.

On the contrary, state-based NILM approaches require an a-priori knowledge or a large training dataset, to achieve good performance [19]. Hidden Markov models (HMM) and various extensions of this model were proposed to examine the different combinations of appliances' state sequences [20]–[23]. In this light, HMMs are state-based and so the studied appliances should have discrete states in their signatures [5]. As the number of appliances increases, the number of combinations of states sequences increases exponentially, thus increasing problem's complexity [5]. In addition to this, time complexity is also increased leading to the reduction of model's classification performance [4]. Makonin et al. proposed a super-state hidden Markov model and a sparse Viterbi algorithm in order to avoid unnecessary calculations and reduce complexity [24]. Another limitation is that this approach does not detect the presence of unknown appliances [5]. Rahimpour et al. proposed a matrix factorization technique for linear decomposition of the aggregated signal using as bases of this learned model the appliances' signatures resulting in an efficient estimation of the energy consumption per appliance [25].

Compared to the above approaches, our proposed methodology (Fig. 1) has a manifold contribution: a) provides detailed information on an appliance's consumption for a predefined time period (as a multi-output regression problem), b) exploits multiple input features (multi-channel CNN) and c) uses a secondary deep-learning approach for autocorrection and output refinement. In the following, we briefly summarize the standing point of our approach compared to existing state of the art.

A. REGRESSION VS CLASSIFICATION PROBLEM

NILM is often addressed in the literature as a classification problem, i.e., estimates the operational states of an appliance (e.g. ON/OFF or multi-state). However, such approaches have

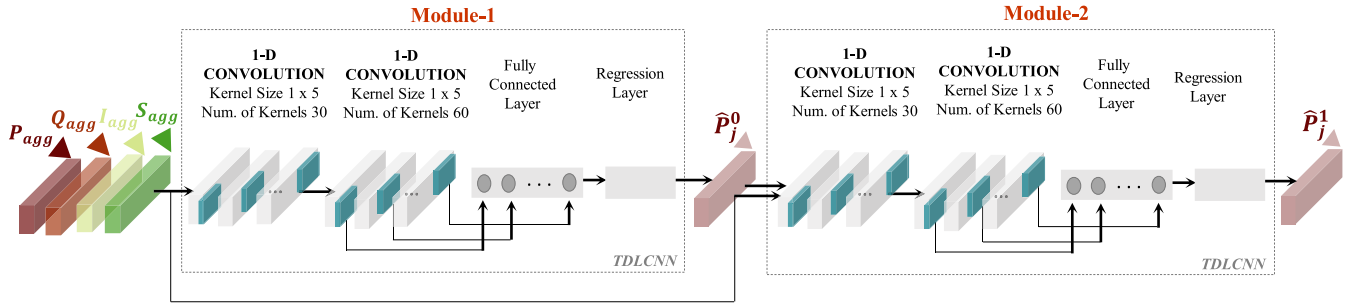


FIGURE 1. The proposed non-linear pipelined recurrent CNN structure.

the drawback that significant signature information regarding the electricity load is lost. Facing energy disaggregation as a regression problem helps us retain all the necessary knowledge regarding an appliance's signature.

B. CURRENT STATE VS HISTORY OF STATES

To reveal accurate appliance's operation state, knowledge gained from its previous operation is necessary, in order to detect time-dependent signature patterns. Tapped delay line model transfers relevant information, forming a sequence chain-like structure as input to feed the CNN network (see Section III.A).

C. SIMPLE VS RECURRENT CNN MODEL

Deep learning structures such as LSTM can be used to address NILM problem through a mechanism which passes the previous hidden state to the next step of the sequence and updates the new hidden state. On the contrary, our approach updates directly the output sequence, using a consecutive deep learning model. This approach is extremely beneficial: the first CNN focuses on the disaggregation part and the second CNN focuses on the correction of the signal. Taking into consideration previous regression output introduces a recurrent behavior to our proposed model (see Section III.B).

D. SINGLE-VARIABLE VS MULTI-VARIABLE APPROACH

Most of the aforementioned approaches employ active power as the sole electrical parameter for NILM. However, some approaches have adopted additional features, such as reactive power [26]. The use of reactive power has been employed initially by Hart [16]. Many NILM methods are based on active power and reactive power ([27]–[29]), while other approaches rely on harmonics [30], voltage and current waveform [31], or the voltage - current (V - I) trajectory analysis [32]. In Section III.C, we introduce a multi-channel approach to strengthen the model's performance.

III. THE PROPOSED MR-TDLCNN MODEL FOR NILM

The proposed methodology deals with the disaggregation problem by utilizing consecutive deep learning multi-input/multi-output regression models.

Let M be the number of household appliances and $p(t_n)$ be the measured aggregate active power over all appliances at a

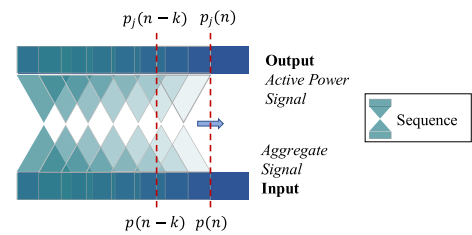


FIGURE 2. The convolutional sequence to sequence NILM model. Time windows progress incrementally in time. The output is derived as the average of the different overlapping output values.

TABLE 1. Notation used in this paper.

Notation	
$j = 1, \dots, M$	number of appliances
T	sampling interval
$t_n = nT = n$	time instance
p	active power
r	reactive power
q	apparent power
I	current
$p(n)$	aggregate signal
$p_j(n)$	j -th appliance signal
$\hat{p}_j(n)$	j -th appliance estimated signal
$e(n)$	noise
k	window of samples
$\hat{p}_j^0(n)$	output of Module-1 TDLCNN
$\hat{p}_j^1(n)$	output of Module-2 TDLCNN
$f(\cdot), g(\cdot)$	non-linear function

time instance t_n . Considering a discrete time sampling, we can express $p(t_n)$ as $p(t_n) = p(nT) = p(n)$, where $T = t_n - t_{(n-1)}$ is the sampling interval. Similarly, we denote $p_j(n)$ the active power load of the j -th appliance out of the M available. Then, the aggregate signal $p(n)$ can be given as [14]

$$p(n) = \sum_{j=1}^M p_j(n) + e(n) \quad (1)$$

where $e(n)$ denotes the additive noise of the measurements. In a NILM modeling framework, the measurements $p_j(n)$ are not available, since there are no smart plugs installed. Instead, only $p(n)$ is given. Therefore, the problem is to estimate $p_j(n)$ from $p(n)$. Let us note hereby that Table 1 includes the notations used in the paper.

A. TDLCNN MODEL: DESIGNING A BASELINE CNN FOR NILM

Each appliance has a unique spectral signature. This is the main principle we exploit to decompose the aggregate signal $p(n)$ into its components $p_j(n)$. Aggregate signal is actually derived as an integration of individual appliances' power consumption values over time. Thus, in order to get the estimates $\hat{p}_j(n)$ of $p_j(n)$, we need to assemble measurements of the aggregate signal $p(n)$ over a time window $[p(n)p(n-1) \dots p(n-k)]^T$. Variable k expresses the number of previous samples that should be considered for estimating the $p_j(n)$. The time window of the aggregate measurements (sequence) covers the mean appliance's operational duration in order to provide the full information regarding appliance's operational states for optimal feature maps selection later. Output sequences are created considering the same length T as in the input sequences, respectively (Fig. 2).

According to the already implemented approaches [6], sequence to sequence learning for NILM maps the input sequence of the aggregate signal to a same length output sequence of appliance's active power via LSTM networks. Here, we introduce an architecture based entirely on 1D convolutional neural networks that incorporates time series data. At instant n , the set of input to the module is a sequence of k most recent measurements of aggregate power values which can be represented as vector $\mathbf{p}(n)$ given by:

$$\mathbf{p}(n) = [p(n) p(n-1) \dots p(n-k)]^T \quad (2)$$

proposing a Tapped Delay Line CNN model (TDLCNN). The purpose of the convolutional layer is to apply non-linear transformations on the input data to maximize regression performance. A set of parameterizable filters (e.g., learnable kernels) is convolved with the input data selecting appropriate feature modalities and estimating kernel parameters, so that performance error on a labeled training set is minimized. The L feature maps, say f_1, f_2, \dots, f_L , optimally selected by the convolutional layer, will be used as input to the final regression layer. The output is a sequence of the j -th appliance active power data, formed as:

$$\hat{\mathbf{p}}_j^0(n) = [\hat{p}_j^0(n) \hat{p}_j^0(n-1) \dots \hat{p}_j^0(n-k)]^T \quad (3)$$

Therefore, we have that:

$$\hat{\mathbf{p}}_j^0(n) = g(\mathbf{p}(n)) + e(n) \quad (4)$$

where $g(\cdot)$ is a nonlinear relationship modeled by the learning process. As derived in this section, the first step to suitably decompose the aggregate signal to its components $p_j(n)$, is to consider several previous observations of the aggregate signal over a time window, in a way to maximize model's performance.

B. R-TDLCNN MODEL: INTRODUCING A RECURRENT CHARACTER TO CNN FOR NILM

It is intuitively clear that the active power signal observations per appliance are not independent over time. A widely

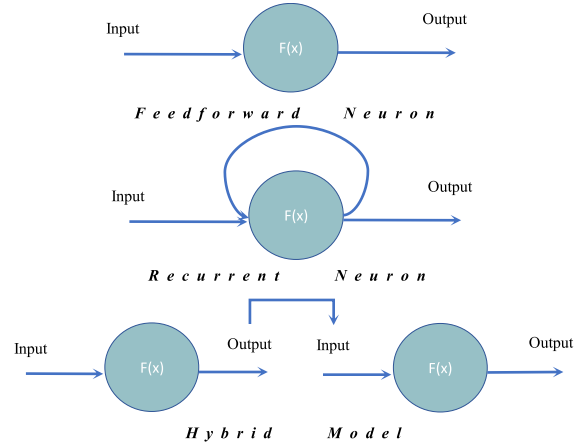


FIGURE 3. Schematic illustration of the proposed hybrid model in conjunction to a simple feedforward and a recurrent neuron.

accepted way to model this dependence and dealing with these inherently recursive data is through recurrent neural networks (RNN). RNNs can use the feedback connection to store information over time in form of activations, successfully handling sequential data and time series. On the other hand, CNNs capture patterns through non-linear relationships allowing weights to be dynamically updated in a complicated way. An approach based entirely on CNN is not adequate (see Section III.A), since significant variations in the disaggregated signal are observed that should be taken into consideration.

Thus, a hybrid CNN model is hereby introduced that incorporates RNN model's characteristics into the basic CNN structure, leading to a novel CNN model with a recurrent character, as illustrated in Fig.3. The difference is that in the case of RNN model the update occurs in the hidden state, whereas in our approach the update is carried out directly to the output. Except for matching sequence inputs with sequence outputs, as introduced in the previous step (Section IV.A), the model incorporates an a posteriori state estimation per appliance, at time n , given observations up to and including time n .

Thus, based on (4) and given (2) and (3), we form an updated non-linear framework:

$$\hat{\mathbf{p}}_j^1(n) = f(\mathbf{p}(n), \hat{\mathbf{p}}_j^0(n)) + e(n) \quad (5)$$

where:

$$\hat{\mathbf{p}}_j^1(n) = [\hat{p}_j^1(n) \hat{p}_j^1(n-1) \dots \hat{p}_j^1(n-k)]^T \quad (6)$$

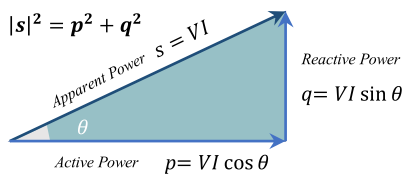
The main difficulty in (5) is that the non-linear relationship $f(\cdot)$ is actually unknown. To address the fact, machine learning methods can be applied to approximate $f(\cdot)$ in a way that minimizes error $e(n)$.

C. M-TDLCNN MODEL: INTRODUCING A MULTI-CHANNEL CNN ARCHITECTURE FOR NILM

An important benefit of using CNNs is that they can support multiple inputs. In the literature, the majority of the

TABLE 2. Performance metrics (MAE, RMSE, NRMS) for AMPDs dataset.

	MAE(W)	RMSE(W)	NRMS	MAE(W)	RMSE(W)	NRMS	MAE(W)	RMSE(W)	NRMS	MAE(W)	RMSE(W)	NRMS
	CDE			DWE			HPE			WOE		
TDLCNN	19.20	103.71	0.19	32.03	84.59	0.75	102.24	102.24	1.14	42.22	201.30	1.28
R-TDLCNN	11.86	95.79	0.18	16.00	74.76	0.66	49.31	181.02	0.90	20.38	139.18	0.88
M-TDLCNN	16.28	65.27	0.12	23.95	71.73	0.64	24.93	65.99	0.33	27.00	99.08	0.63
MR-TDLCNN	4.79	41.66	0.08	14.35	64.48	0.57	18.44	58.26	0.29	15.25	76.68	0.49
LSTM [5]	25.35	180.91	0.43	24.04	72.05	1.00	161.74	369.81	0.55	15.82	141.20	0.89
CNN-LSTM [1]	14.16	107.04	0.20	24.35	87.36	0.77	94.79	236.87	1.17	55.97	298.05	2.01
CO [34]	117.53	453.31	1.26	156.23	317.62	4.41	249.16	458.77	1.23	267.00	432.79	3.46
FHMM [34]	129.57	323.23	0.90	313.68	459.07	4.44	121.69	426.66	1.14	49.38	360.85	2.89

**FIGURE 4.** The power triangle.

proposed methods adopt a solution that uses only active power measurements. However, power utility companies are generally concerned about both active power (P) loads (W) and reactive power (Q) loads (VAR). In active power loads (e.g., an electric stove) dissipation of the performed worked takes place, whereas reactive power loads (e.g., a capacitor) store the power received from the grid, and release it back in the opposite direction later without dissipation. Loads can often be both active and reactive, as for example an air conditioning unit. Mathematically, active power results from in-phase voltage and current, whereas reactive power results from out of phase voltage and current [33]. Apparent power S , sometimes referred to as total momentary power, can also be a useful cue for disaggregation. Apparent power is conventionally expressed in volt-amperes (units in VA). These quantities are related as (Fig. 4):

$$s = IV; p = s \cos \theta; q = s \sin \theta \quad (7)$$

where I is current, V is voltage, and θ is the phase of voltage relative to current (i.e., the phase angle).

The s, p, q , and I are inserted in the CNN model as multi-variable timeseries in order to strengthen the model's reliability, resulting in a non-linear regression problem. In correspondence to Section IV.A, each variable (p, s, q, I) adopts the tapped delay line structure in order to be feed in the model. Each input sequence ($\{p\}, \{s\}, \{q\}, \{I\}$) is then, passed as a separate channel, in correspondence to different channels of an image (e.g. red, green and blue), forming a multi-channel CNN structure. As a result, a fused input that resembles a tensor is created. The tensorized input ensures that the model encapsulates all necessary information to produce the output, using a M-TDLCNN model.

D. MR-TDLCNN MODEL: LEVERAGING M-TDLCNN AND R-TDLCNN MODELS INTO A NOVEL NILM MODEL

The architecture, as shown in Fig. 1, consists of a pipelined recurrent structure. As shown in the figure, it is composed of two modules in parallel, TDLCNN Module-1 and TDLCNN Module-2, in such a manner that the output of Module-2 is used as an input to Module-1. MR-TDLCNN model combines simple recursive approach with CNN architecture in order to allow learning meaningful data-dependent weights. Furthermore, it exploits multiple input features succeeding high performance.

IV. EXPERIMENTAL EVALUATION

In this section, we will experimentally validate the superiority of the proposed MR-TDLCNN method in comparison to (i) the basic TDLCNN model, (ii) individual updates (R-TDLCNN and M-TDLCNN models) and, more importantly, (iii) other state of the art methods. Among them an LSTM network [5] and a hybrid CNN-LSTM method [1] are included. LSTM has been selected as a typical network for timeseries processing. Furthermore, an improved hybrid CNN-LSTM model is used with CNN layers for feature extraction on input data combined with LSTMs to support sequence prediction. We also compare the aforementioned results with the state-of-the-art NILM algorithms i.e., FHMM-based and CO-based methods from NILMTK [34], a Python-based extension which is widely used in energy disaggregation research. We evaluate the model's accuracy and convergence speed across different models and different appliances. RMSE error of training was selected as a frequently used measure of accuracy in order to keep track of the performance measure of our model during training.

A. DATASET DESCRIPTION AND EXPERIMENTAL SETUP

The evaluation of the proposed method is conducted on the public AMPDs dataset [24]. The AMPDs contains active, reactive and apparent power values as well as current measurements from a Canadian house, at one-minute interval over

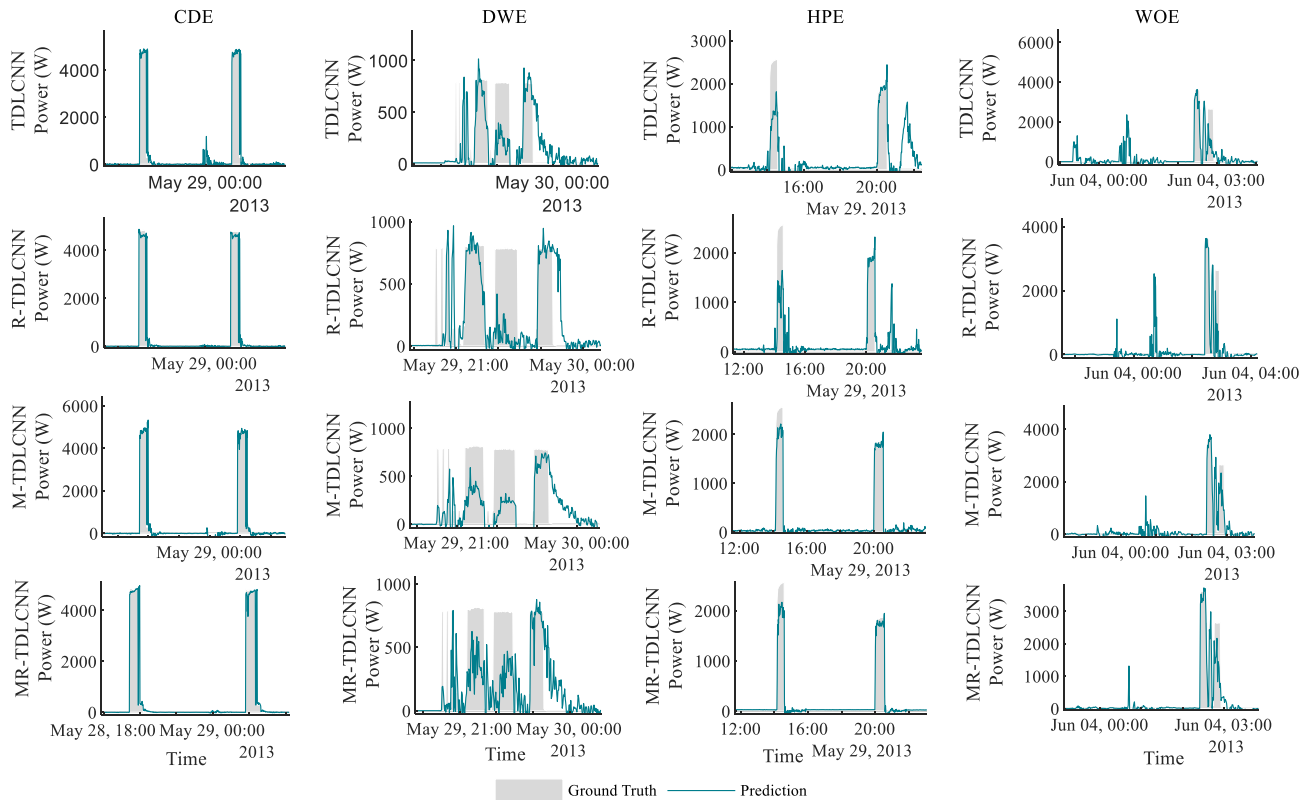


FIGURE 5. Comparison of the proposed method (green solid line) with ground truth (in gray) for selected appliances in AMPDs dataset.

a 2-year period. The selected appliances are of multi-state type, making difficult their respective power load estimation (see Fig. 5 with grey filled color).

B. PERFORMANCE EVALUATION AND COMPARISONS

We trained our models using adam (adaptive moment estimation) optimization with a learning rate of 1e-4. Model weights and coefficients are updated using a mini-batch size of 50 at each training iteration. The maximum number of epochs for training is selected to be 400. Training period starts at 18 August 2012 and ends at 13 April 2013; 30 days were used as test sample (17 May 2013-17 June 2013). This split is representative of the problem and in addition, the testing period is a transitional period, so we can evaluate the ability of the model to adapt to seasonal variations.

Deep learning performance is improved through data balance and normalization to 0-1 for each channel. The proposed MR-TDLCNN model, along with TDLCNN, M-TDLCNN, R-TDLCNN and LSTM are implemented using MATLAB software. CNN-LSTM algorithms have been trained and deployed using Python with Tensorflow and Keras libraries. CO and FHMM methods have implemented in NILMTK. Regarding the dataset, training and testing splits have already been pre-split and pre-normalized, to ensure that the conditions are the same and the results are comparable.

The proposed MR-TDLCNN regression model satisfies a set of crucial characteristics making it superior than the other

existing methods in literature, for NILM. Its modularity is one of its main advantages in comparison to FHMM and CO approaches, in which dimensionality is a major issue. In addition, the introduction of deep learning as part of the solution of NILM problem is also a comparative advantage. Furthermore, model’s performance strengthens with the use of all four components (current, active power, reactive power, and apparent power) available in AMPDs dataset, achieving faster convergence and higher performance than state-of-the-art results for the same dataset.

Table 2 presents the comparative results based on objective metrics of (i) Mean Absolute Error (MAE), (ii) Root Mean Square Error (RMSE) and iii) Normalized RMSE (NRMS), which are commonly used metrics for the evaluation of energy disaggregation. In this experimental setup, four appliances for AMPDs dataset are presented. Particularly, we have used clothes’ dryer (CDE), dishwasher (DWE), heat pump (HPE) and wall oven appliance (WOE) of single AMPDs house. Our proposed MR-TDLCNN method generally performs best mainly due to its capability to effectively model time dependencies and its ability to incorporate different data observations (p, s, q, I) strengthen model’s performance. R-TDLCNN and M-TDLCNN models proved to have better performance compared to basic TDLCNN model. Here, it is worth mentioning that, CNN-LSTM model’s performance is quite high and is a good alternative as a proposed solution to solve NILM problem. It should be mentioned that

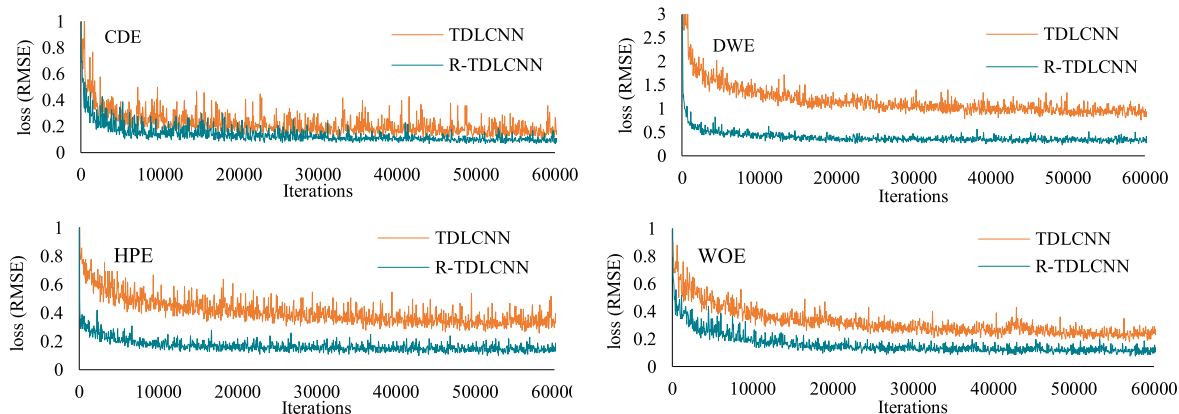


FIGURE 6. Comparison of loss curves between the basic TDLCNN (orange line) and R-TDLCNN (green line) models.

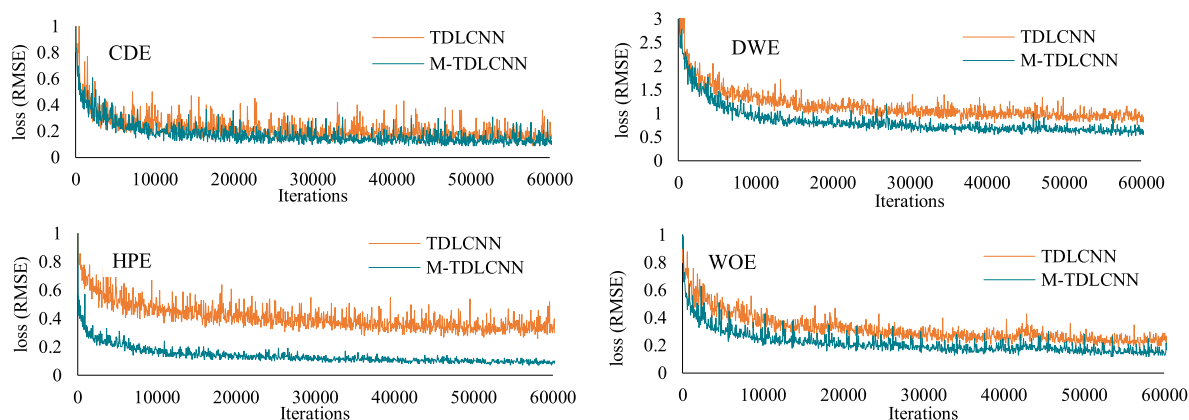


FIGURE 7. Comparison of loss curves between the basic TDLCNN (orange line) and M-TDLCNN (green line) models.

the model for detecting the HPE appliance (AMPds) is not so accurate mainly due to seasonal signal’s changes caused by external contextual conditions / parameters. It should be mentioned that for all scenarios the metrics have been calculated over all the examined time period, in which the appliances can be either in operation or not.

Fig. 5 shows the comparison among the predicted signal and the ground truth for clothes dryer (CDE), heat pump (HPE) and wall oven appliance (WOE) of single AMPds House. Fig. 5 is representative of the MR-TDLCNN method’s superiority against the remaining TDLCNN, M-TDLCNN and R-TDLCNN methods that have been presented before. Thus, the integrated MR-TDLCNN model succeeds better performance in comparison to the results that the basic model architecture succeeds. Particularly, the baseline TDLCNN model presents the worst results among others for all the presented appliances. Also, we can notice the existence of false detections, especially in HPE and WOE appliance. M-TDLCNN and R-TDLCNN models’ have adequate performance, while MR-TDLCNN presents high-levels of performance and additionally, false detections have been eliminated.

A way to get insight into the model’s learning behavior is through evaluation on the training dataset. Thus, a model’s

learning rate can be described using performance/epochs diagram. Fig. 6, 7, 8 illustrate loss curves for the four aforementioned models, namely are TDLCNN, R-TDLCNN, M-TDLCNN and MR-TDLCNN, in pairs.

Fig. 6 shows TDLCNN and R-TDLCNN models’ loss progress during training using RMSE error and considering 60000 iterations per appliance for each of the presented CNN based models. In general, the loss function is being minimized during training. As observed, R-TDLCNN model’s performance is slightly better than TDLCNN model’s, as the former has a lower loss than the latter. It is worth mentioning that the training loss for HPE and DWE appliances present considerably lower values in R-TDLCNN than the values deriving from TDLCNN model. The loss curve, as illustrated with green for R-TDLCNN model, has lower starting point value, decreases with a smaller rate and is smoother than TDLCNN model’s loss curve (with orange line).

Fig. 7 shows loss progress for basic TDLCNN and M-TDLCNN models during training. In general, the loss function is being minimized during training. In most cases, M-TDLCNN losses are lower compared to TDLCNNs ones. DWE appliance presents intense slope with high starting point values (3 for TDLCNN and 1.5 for M-TDLCNN) that converge at values <1. Finally, the loss curve for M-TDLCNN

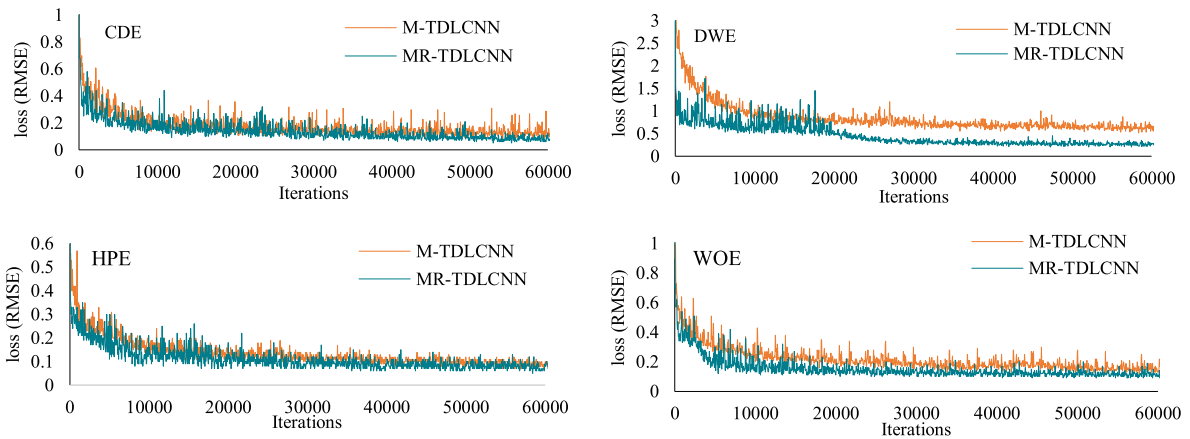


FIGURE 8. Comparison of loss curves between the basic M-TDLCNN (orange line) and MR-TDLCNN (green line) models.

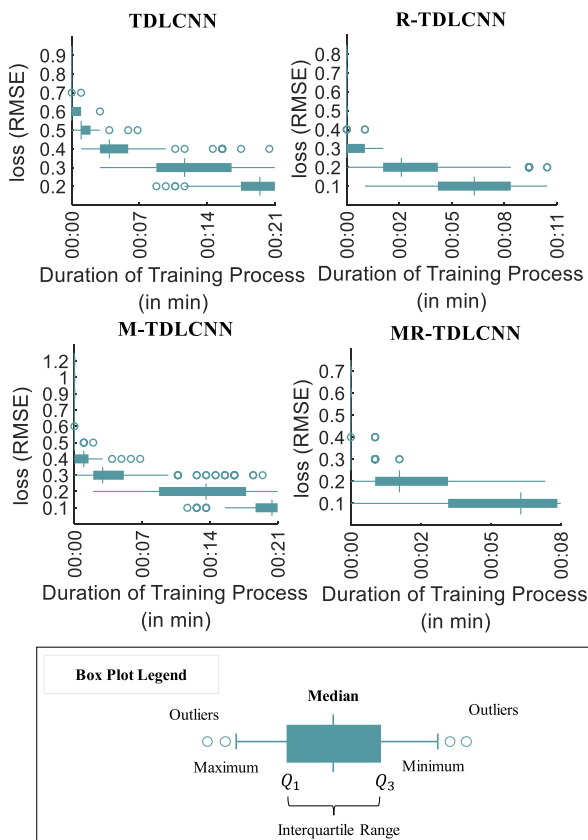


FIGURE 9. Elapsed convergence time for TDLCNN, R-TDLCNN, M-TDLCNN and MR-TDLCNN models of WOE appliance.

and MR-TDLCNN models is presented with orange and green color line, respectively, as Fig. 8 shows.

Taking as example WOE appliance, the box plot has been used to display the distribution of RMSE error reduction during training in conjunction to elapsed time, based on the five-number summary: minimum, first quartile, median, third quartile, and maximum. Furthermore, surprisingly high and low values called outliers, are illustrated by dots. The central rectangle spans the first quartile to the third quartile.

A segment inside the rectangle shows the median and “whiskers” above and below the box show the locations of the minimum and maximum. Comparing TDLCNN and R-TDLCNN in Fig.9, we notice an increase in convergence speed leading to the reduction of needed time for training in comparison to the time needed for TDLCNN, even though the initial RMSE error (0.9) is greater than RMSE error of the simple CNN single channel model (0.8). Furthermore, TDLCNN succeeds RMSE error 0.2 while R-TDLCNN reaches the value of 0.1. Also, the presence of RMSE error value 0.1 starts early (the first minute of training time), even though as an “outlier”. As regards the other two models, M-TDLCNN model reaches the value of 0.1 RMSE error, while MR-TDLCNN reaches the same value earlier and the training phase stops 10 minutes earlier.

V. CONCLUSION

In this paper, we introduce a novel deep learning based method for energy disaggregation. The proposed recurrent deep-learning multi-input/multi-output regression model based on CNN leverages the recurrent property to effectively model the temporal interdependencies of the power signals. Moreover, the incorporation of multiple channels, each for a different signal (active, reactive, apparent power and current), offers additional streams of information resulting in a more accurate model. Experimental results suggest higher performance and faster convergence times compared to state of the art approaches. As future work, we will consider Bayesian optimization techniques for hyperparameter fine-tuning, as well as investigating the applicability of transfer learning for improving generalization in other power consumption scenarios.

REFERENCES

- [1] 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.*, Nov. 2015, pp. 55–64.
- [2] W. He and Y. Chai, “An empirical study on energy disaggregation via deep learning,” presented at the 2nd Int. Conf. Artif. Intell. Ind. Eng. (AIIE), 2016.

- [3] L. Mauch and B. Yang, "A novel DNN-HMM-based approach for extracting single loads from aggregate power signals," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2016, pp. 2384–2388.
- [4] J. Kim, T.-T.-H. Le, and H. Kim, "Nonintrusive load monitoring based on advanced deep learning and novel signature," *Comput. Intell. Neurosci.*, vol. 2017, Aug. 2017. Accessed: Sep. 16, 2018, Art. no. 4216281. [Online]. Available: <https://www.hindawi.com/journals/cin/2017/4216281/>
- [5] L. Mauch and B. Yang, "A new approach for supervised power disaggregation by using a deep recurrent LSTM network," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Dec. 2015, pp. 63–67.
- [6] M. Kaselimi, N. Doulamis, A. Doulamis, A. Voulodimos, and E. Protopapadakis, "Bayesian-optimized bidirectional LSTM regression model for non-intrusive load monitoring," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 2747–2751.
- [7] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
- [8] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep learning for computer vision: A brief review," *Comput. Intell. Neurosci.*, vol. 2018, Feb. 2018, Art. no. 7068349.
- [9] A. Harell, S. Makonin, and I. V. Bajić, "Wavenilm: A causal neural network for power disaggregation from the complex power signal," Feb. 2019, *arXiv:1902.08736*. [Online]. Available: <https://arxiv.org/abs/1902.08736>
- [10] C. Zhang, M. Zhong, Z. Wang, N. Goddard, and C. Sutton, "Sequence-to-point learning with neural networks for nonintrusive load monitoring," Dec. 2016, *arXiv:1612.09106*. [Online]. Available: <https://arxiv.org/abs/1612.09106>
- [11] K. S. Barsim and B. Yang, "On the feasibility of generic deep disaggregation for single-load extraction," Feb. 2018, *arXiv:1802.02139*. [Online]. Available: <https://arxiv.org/abs/1802.02139>
- [12] K. He, L. Stankovic, J. Liao, and V. Stankovic, "Non-intrusive load disaggregation using graph signal processing," *IEEE Trans. Smart Grid*, vol. 9, no. 3, pp. 1739–1747, May 2018.
- [13] H. Altrabalsi, V. Stankovic, J. Liao, and L. Stankovic, "Low-complexity energy disaggregation using appliance load modelling," *AIMS Energy*, vol. 4, no. 1, pp. 884–905, Jan. 2016.
- [14] J. Liao, G. Elafoudi, L. Stankovic, and V. Stankovic, "Non-intrusive appliance load monitoring using low-resolution smart meter data," in *Proc. IEEE Int. Conf. Smart Grid Commun. (SmartGridComm)*, Nov. 2014, pp. 535–540.
- [15] Y.-H. Lin and M.-S. Tsai, "Non-intrusive load monitoring by novel neuro-fuzzy classification considering uncertainties," *IEEE Trans. Smart Grid*, vol. 5, no. 5, pp. 2376–2384, Sep. 2014.
- [16] G. W. Hart, "Nonintrusive appliance load monitoring," *Proc. IEEE*, vol. 80, no. 12, pp. 1870–1891, Dec. 1992.
- [17] G. Elafoudi, L. Stankovic, and V. Stankovic, "Power disaggregation of domestic smart meter readings using dynamic time warping," in *Proc. 6th Int. Symp. Commun., Control Signal Process. (ISCCSP)*, May 2014, pp. 36–39.
- [18] R. Machlev, Y. Levron, and Y. Beck, "Modified cross-entropy method for classification of events in NILM systems," *IEEE Trans. Smart Grid*, to be published.
- [19] B. Zhao, K. He, L. Stankovic, and V. Stankovic, "Improving event-based non-intrusive load monitoring using graph signal processing," *IEEE Access*, vol. 6, pp. 53944–53959, 2018.
- [20] J. Z. Kolter and T. S. Jaakkola, "Approximate inference in additive factorial HMMs with application to energy disaggregation," in *Proc. AISTATS*, Mar. 2012, pp. 1472–1482.
- [21] R. Bonfigli, E. Principi, M. Fagiani, M. Severini, S. Squartini, and F. Piazza, "Non-intrusive load monitoring by using active and reactive power in additive factorial hidden Markov models," *Appl. Energy*, vol. 208, pp. 1590–1607, Dec. 2017.
- [22] W. Kong, Z. Y. Dong, D. J. Hill, J. Ma, J. H. Zhao, and F. J. Luo, "A hierarchical hidden Markov model framework for home appliance modeling," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3079–3090, Jul. 2018.
- [23] D. Bajović, K. He, L. Stanković, D. Vukobratović, and V. Stanković, "Optimal detection and error exponents for hidden semi-Markov models," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 5, pp. 1077–1092, Oct. 2018.
- [24] S. Makonin, F. Popowich, I. V. Bajić, 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.
- [25] A. Rahimpour, H. Qi, D. Fugate, and T. Kuruganti, "Non-intrusive energy disaggregation using non-negative matrix factorization with sum-to-k constraint," *IEEE Trans. Power Syst.*, vol. 32, no. 6, pp. 4430–4441, Nov. 2017.
- [26] B. Qi, L. Liu, and X. Wu, "Low-rate non-intrusive load disaggregation with graph shift quadratic form constraint," *Appl. Sci. Switz.*, vol. 8, no. 4, p. 554, Apr. 2018.
- [27] C. Dinesh, B. W. Nettasinghe, R. I. Godaliyadda, M. P. B. Ekanayake, J. Ekanayake, and J. V. Wijayakulasooriya, "Residential appliance identification based on spectral information of low frequency smart meter measurements," *IEEE Trans. Smart Grid*, vol. 7, no. 6, pp. 2781–2792, Nov. 2016.
- [28] M. Figueiredo, B. Ribeiro, and A. de Almeida, "Electrical signal source separation via nonnegative tensor factorization using on site measurements in a smart home," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 2, pp. 364–373, Feb. 2014.
- [29] Y. Xu and J. V. Milanović, "Artificial-intelligence-based methodology for load disaggregation at bulk supply point," *IEEE Trans. Power Syst.*, vol. 30, no. 2, pp. 795–803, Mar. 2015.
- [30] W. Wichakool, A. T. Avestruz, R. W. Cox, and S. B. Leeb, "Modeling and estimating current harmonics of variable electronic loads," *IEEE Trans. Power Electron.*, vol. 24, no. 12, pp. 2803–2811, Dec. 2009.
- [31] W. Wichakool, Z. Remscrib, U. A. Orji, and S. B. Leeb, "Smart metering of variable power loads," *IEEE Trans. Smart Grid*, vol. 6, no. 1, pp. 189–198, Jan. 2015.
- [32] T. Hassan, F. Javed, and N. Arshad, "An empirical investigation of V-I trajectory based load signatures for non-intrusive load monitoring," *IEEE Trans. Smart Grid*, vol. 5, no. 2, pp. 870–878, Mar. 2014.
- [33] A. Harell, S. Makonin, and I. V. Bajić, "Wavenilm: A causal neural network for power disaggregation from the complex power signal," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Feb. 2019, pp. 8335–8339.
- [34] 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.*, Jun. 2014, pp. 265–276.



MARIA KASELIMI received the Diploma degree from the National Technical University of Athens (NTUA), Greece, in 2015, and the M.Sc. degree from NTUA in geoinformatics in 2017, where she is currently pursuing the Ph.D. degree and was a Junior Researcher in several European projects focusing on machine learning, signal processing techniques, data analysis, and modeling with applications in the fields of energy, geosciences, and environment.



EFTYCHIOS PROTOPAPADAKIS received the degree in production engineering and management, the M.S. degree in management and business administration, and the Ph.D. degree in decision systems from the Technical University of Crete. He has been an Engineer in European (4D-CH-World, BENEFFICE, eVACUATE, ROBO-SPECT, Terpsichore, and WaterSpy) and Intereg (e-Park and Poseidon) projects, since 2010. He has coauthored more than 40 publications. His research interest includes machine learning applications. He has explored the applicability of semi-supervised techniques in maritime surveillance, energy applications, elder people support, industrial workflow monitoring, structural assessment of tunnel infrastructures, and cultural heritage applications.



ATHANASIOS VOULODIMOS received the Dipl.Ing., M.Sc., and Ph.D. degrees (Hons.) from the School of Electrical and Computer Engineering, National Technical University of Athens (NTUA). He is currently an Assistant Professor with the Department of Informatics and Computer Engineering, University of West Attica. He has been involved in several European research projects, as a Senior Researcher and as a Technical Manager. He has coauthored more than 90 papers in international journals, conference proceedings and books. His research interests include machine learning and signal processing. He has received awards for his academic performance and scientific achievements.



ANASTASIOS DOULAMIS received the Diploma and Ph.D. degrees (Hons.) in electrical and computer engineering from the National Technical University of Athens (NTUA). Until 2014, he was an Associate Professor with the Technical University of Crete. He is currently an Assistant Professor with NTUA. He has received several awards in his studies, including the Best Greek Student Engineer and the Best Graduate Thesis Award. He has also served as program committee in several major conferences of IEEE and ACM. He has authored more than 350 papers in leading journals and conferences receiving more than 3800 citations.

• • •



NIKOLAOS DOULAMIS received the Diploma and Ph.D. degrees (Hons.) in electrical and computer engineering from the National Technical University of Athens (NTUA), where he is currently an Associate Professor. He has authored more than 75 (240) journals (conference) papers in signal processing and machine learning. He received more than 3900 citations. He is involved in large scale European projects, such as H2020 Stop-It and H2020 Benefice. He has received many awards (e.g., Best Student among all Engineers, best paper awards). He has served as an Organizer and/or a TPC in major IEEE conferences.