

Article

Adversarial and Hierarchical Distribution Alignment Network for Nonintrusive Load Monitoring

Haozhe Xiong, Daojun Tan, Yuxuan Hu, Xuan Cai and Pan Hu *

State Grid Hubei Electric Power Research Institute, Wuhan 430077, China

* Correspondence: hupaninwh@whu.edu.cn

Abstract

Nonintrusive Load Monitoring (NILM) models often suffer from significant performance degradation when deployed across different households and datasets, primarily because of distribution discrepancies. To address this challenge, this study proposes an adversarial hierarchical distribution alignment unsupervised domain adaptation network for non-intrusive load disaggregation. The network aims to reduce the distribution divergence between the source and target domains in both the feature and label spaces, enabling effective adaptation to transfer learning scenarios in which the source domain has limited labeled data and the target domain has abundant unlabeled data. The proposed method integrates adversarial training with a hierarchical distribution alignment strategy that uses Correlation Alignment (CORAL) to align global marginal distributions. It employs Multi-Kernel Maximum Mean Discrepancy (MK-MMD) to constrain the conditional distributions of individual appliances, thereby enhancing cross-domain generalization. Extensive experiments on three public datasets demonstrate that, in both in-domain and cross-domain settings, the proposed method consistently reduces Mean Absolute Error (MAE) and Signal Aggregation Error (SAE), outperforming baseline approaches in cross-domain generalization.

Keywords: Nonintrusive Load Monitoring (NILM); load disaggregation; transfer learning; Domain Adaptation (DA); Temporal Convolutional Network (TCN); Multi-Kernel Maximum Mean Discrepancy (MK-MMD)

1. Introduction

To improve energy efficiency and promote environmental sustainability, accurate appliance-level electricity usage information is essential. This would increase people's awareness of their energy consumption, enabling more effective interaction between end-users and energy suppliers [1]. With the development of smart grids, power systems are becoming increasingly informative, digitalized, and intelligent, increasing the demand for fine-grained perception and management of customer electricity consumption. Against this backdrop, load disaggregation, often called Nonintrusive Load Monitoring (NILM), plays a pivotal role in Home Energy Management Systems (HEMS), demand response programs, and the operational optimization of smart grids [2,3]. Based on monitoring methods, load disaggregation can be categorized into intrusive and nonintrusive approaches. Intrusive load monitoring requires installation of separate sensors for each electrical device. Although it provides high-accuracy monitoring data, its time-consuming

Academic Editor: José Matas

Received: 22 December 2025

Revised: 28 January 2026

Accepted: 28 January 2026

Published: 2 February 2026

Copyright: © 2026 by the authors.
Licensee MDPI, Basel, Switzerland.
This article is an open access article
distributed under the terms and
conditions of the [Creative Commons
Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

installation and high maintenance costs, along with potential privacy risks, have limited large-scale deployment. In contrast, NILM, first proposed in 1992, estimates the electricity consumption of individual appliances using only aggregate measurements at the main meter. Its advantages, including its low cost, ease of use, and privacy-friendly nature, have granted it greater application prospects [4,5].

With advances in artificial intelligence, NILM is commonly formulated as a blind source separation problem. The core objective is to learn parametric mapping from the aggregate load sequence to appliance-level power consumption. Early studies primarily employed probabilistic machine learning methods based on hidden Markov models [6]. However, mainstream approaches have shifted toward deep learning architectures, such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and encoder–decoder models [7–9]. Although these methods perform well on in-domain or similar datasets, they generally assume that training and testing data follow similar distributions [10]. In practice, power-load distributions vary significantly across households and datasets owing to differences in appliance composition, user electricity usage habits, and metering conditions. This severely limits models' generalizability across domains.

To address this issue, researchers have proposed improvements at both data and model levels. Generative models are widely used to augment the training data. For example, TraceGAN [11] and Wasserstein deep convolutional Generative Adversarial Networks (GANs) [12] synthesized appliance power sequences, whereas NILM-Synth [13] created synthetic NILM datasets by superimposing reference load profiles. On the model side, transfer learning strategies such as fine-tuning and direct transfer of pretrained load models have been employed to improve cross-domain adaptability [14,15]. Although these methods can improve cross-domain performance to some extent, they often struggle to align the feature distributions of complex appliance loads when the source domain has limited labeled data and the target domain is rich in unlabeled data. This mismatch can result in insufficient or negative transfer.

Domain Adaptation (DA), a key branch of transfer learning, aims to reduce the distribution discrepancy between source and target domains [16,17]. When the target domain has no labeled data, this problem falls within the scope of Unsupervised Domain Adaptation (UDA), where the model must achieve cross-domain adaptation without label supervision from the target domain. Owing to the significant distribution differences between the source and target domains in both aggregate load and appliance-level load patterns, a single distribution constraint may be insufficient to mitigate the shift. Several UDA studies have been conducted in this context. Examples include careful electrical feature selection to enhance cross-domain generalization [18], combining a Temporal Convolutional Network (TCN) with a domain-adaptation loss to jointly optimize disaggregation and domain alignment [19], and FL-WGAN, which integrates federated learning, Wasserstein GANs, and self-attention for domain transfer [20]. However, because appliance features at different levels exhibit complex multilayered shifts, a more systematic alignment strategy is required to simultaneously alleviate global and appliance-level distribution discrepancies. Such a strategy would improve the robustness and accuracy of NILM models in cross-domain scenarios.

Therefore, this study proposes a UDA network with hierarchical distribution alignment, which combines Correlation Alignment (CORAL) and Multi-Kernel Maximum Mean Discrepancy (MK-MMD) to perform joint alignment of global and appliance-level distributions. The goal is to reduce the model's reliance on labeled data and effectively mitigate distribution shifts at different levels. The main contributions of this study are as follows.

1. We propose a hierarchical distribution alignment UDA model to alleviate both global and appliance-level distribution discrepancies, thereby improving the generalization capability of NILM in cross-domain scenarios.
2. The model was designed for application to both in-domain and cross-domain transfer tasks, treating different users or datasets as distinct domains. Through extensive experimental scenarios and comparative studies, we verified the adaptability and stability of the model under various transfer conditions, demonstrating the advantages of the joint application of CORAL and MK-MMD in enhancing the cross-domain performance of NILM.

2. Model Framework

Focusing on the issue of poor model generalization caused by data distribution shifts in cross-domain NILM, this study proposes a DA network-based NILM model to enhance decomposition accuracy. The framework of the proposed model is illustrated in Figure 1.

The proposed model primarily consists of a feature extractor, domain discriminator, and load disaggregator, along with MK-MMD and CORAL modules. A feature extractor, based on a TCN, was designed to effectively capture both the global and local features of the load sequences. The domain discriminator, CORAL, and MK-MMD modules jointly facilitated feature alignment. They operate from the perspectives of marginal and conditional distributions. As a result, the distribution discrepancy between the source and target domains was reduced.

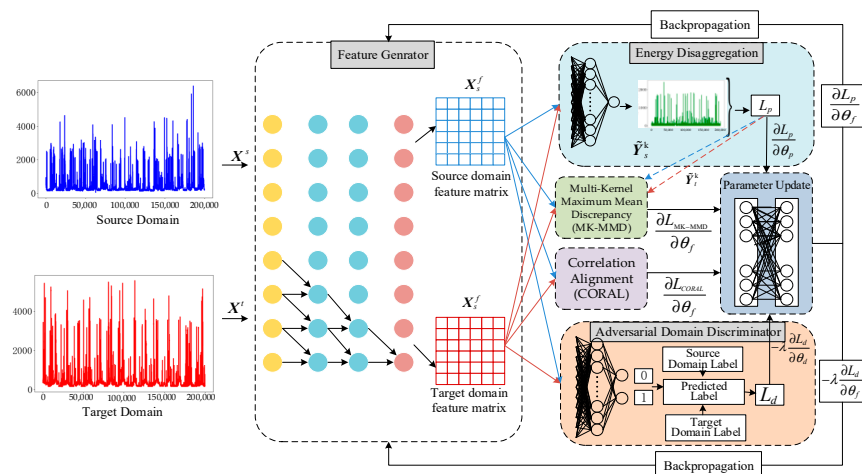


Figure 1. Nonintrusive Load Monitoring (NILM) model based on domain adversarial transfer network.

2.1. Domain-Invariant Feature Extractor Based on TCN

In the load disaggregation task, the initial step was feature extraction from the aggregated power data. This study employs a TCN as the feature extractor, the core components of which are causal convolutions and dilated convolutions. The overall architecture consisted of an input layer, multiple convolutional modules, and an output layer. Causal convolution ensures strict temporal causality in temporal modeling. This introduces a one-sided zero-padding strategy. This strategy constrains the feature output at each time step to depend only on current and previous input values.

Let the aggregated power time-series input be $\mathbf{X} = (x_1, x_2, x_3, \dots, x_t, \dots, x_T)$ and the convolution kernel be $f: \{0, 1, \dots, k-1\}$. The operation of causal convolution can be expressed as follows:

$$G(t) = \sum_{i=0}^{k-1} f_i x_{t-i} \quad (1)$$

To expand the receptive field of the model without increasing the number of parameters or the computational complexity, TCN incorporates a dilated convolution. This method inserts zeros between kernel elements, enabling the capture of longer-range dependencies without increasing the kernel size. Denoting the dilation factor as d , the dilated convolution operation with a causal constraint is formulated as follows:

$$G(t) = \sum_{i=0}^{k-1} f_i x_{t-d \cdot i} \quad (2)$$

where d is the dilation factor, and $t - d \cdot i$ represents the past time steps. This structure facilitates the multiscale extraction from local features to global dependencies within the aggregated power sequence. Its structural diagram is shown in Figure 2(a).

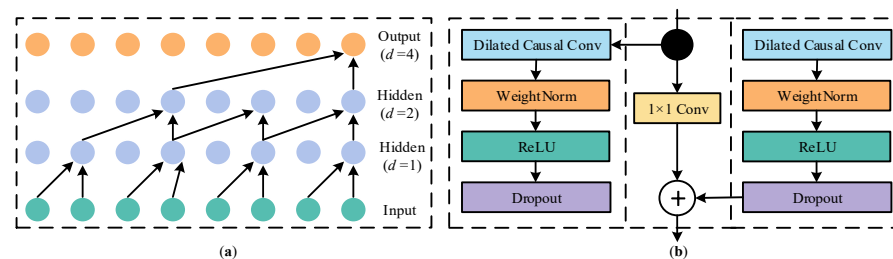


Figure 2. The TCN architecture: (a) dilated causal convolution and (b) A TCN residual block.

TCN adopts residual modules as its fundamental building blocks to further enhance the training efficiency of deep networks. Its structural diagram is shown in Figure 2(b). Each module consists of multiple convolutional layers, normalization layers, activation functions, and shortcut connections. This design mitigates the vanishing gradient problem and improves the generalization capability of the model. A schematic diagram of the module is shown in the figure, and its computational process can be expressed as follows:

$$\mathbf{X}^{(h)} = \delta \left(F \left(\mathbf{X}^{(h-1)} \right) + \mathbf{X}^{(h-1)} \right) \quad (3)$$

where $F(\bullet)$ represents the transformation operation, and $\delta(\bullet)$ is an activation operation.

2.2. Adversarial Strategy Based on MK-MMD and CORAL Modules

A domain discriminator is a core component of the domain-adversarial learning framework. It extracts domain-invariant features through adversarial training by using a feature extractor. However, relying solely on a domain adversary often leads to sub-optimal disaggregation performance. This study introduces MK-MMD and CORAL to enhance the transfer learning capability. MK-MMD was adopted for feature alignment. This aims to reduce the discrepancy in the joint distribution of the features and labels. It also addresses the inconsistency in appliance power consumption patterns across different domains. MK-MMD measures the distribution difference between data from different domains. It serves as a multikernel variant of the Maximum Mean Discrepancy (MMD). MMD quantifies the difference between the source-domain and target-domain feature distributions. This is achieved by computing the distance between their mean embeddings in a Reproducing Kernel Hilbert Space (RKHS). The calculation formula is as follows:

$$D_{\text{MMD}} = \text{MMD}(X^s, X^t) = \left\| \frac{1}{n^s} \sum_{i=1}^{n^s} \Phi(x_i^s) - \frac{1}{n^t} \sum_{j=1}^{n^t} \Phi(x_j^t) \right\|_H^2 \quad (4)$$

where $X^s = \{x_i^s\}_{i=1, \dots, n^s}$ and $X^t = \{x_j^t\}_{j=1, \dots, n^t}$ denote the feature vectors of the source and target domains, respectively.

Compared to conventional MMD, MK-MMD incorporates a linear combination of multiple kernel functions. This design enables the capture of distribution discrepancies at different scales. This also allows the model to capture diverse distribution patterns. The MK-MMD captures the different characteristics and scales of the data distributions within the RKHS. The kernel function is formulated as follows:

$$D_{\text{MK-MMD}} = \sum_{m=1}^M \beta_m \text{MMD}_m^2(X^s, X^t) \quad (5)$$

Because this study employed MK-MMD for joint distribution alignment, the predicted power from the load disaggregator was incorporated into the calculation. The loss function for MK-MMD on the joint distribution is formulated as:

$$L_{\text{MK-MMD}} = \sum_{m=1}^M \beta_m \left[\frac{1}{n_s^2} \sum_{i,j} k_m(z_i^s, z_j^s) + \frac{1}{n_t^2} \sum_{i,j} k_m(z_i^t, z_j^t) - \frac{2}{n_s n_t} \sum_{i,j} k_m(z_i^s, z_j^t) \right] \quad (6)$$

$$z_i = [X^f, \tilde{Y}^k] \quad (7)$$

CORAL achieves cross-domain distribution adaptation by aligning the Frobenius norms of the feature matrices of the source and target domains. Its loss definition is

$$L_{\text{CORAL}} = \frac{1}{4d^2} \|C_s - C_t\|_F^2 \quad (8)$$

where $\|\cdot\|_F^2$ is the Frobenius matrix, d denotes the feature dimension, C_s and C_t represent the covariance matrices of the source and target domains, calculated as follows:

$$C_s = \frac{1}{n^s - 1} \left[D_s^T D^s - \frac{1}{n^s} (\mathbf{1}^T D^s)^T (\mathbf{1}^T D^s) \right] \quad (9)$$

$$C_t = \frac{1}{n^t - 1} \left[D_t^T D^t - \frac{1}{n^t} (\mathbf{1}^T D^t)^T (\mathbf{1}^T D^t) \right] \quad (10)$$

where $\mathbf{1}$ is a column vector with all elements equal to 1; D^s and D^t are the feature matrices of the source and target domains, respectively; and n^t and n^s are the numbers of features in the source and target domains, respectively.

In the framework of this study, CORAL and MK-MMD perform feature alignment from the aspects of marginal distribution and conditional distribution, respectively, reducing the distribution difference between the source and target domains and enhancing the generalization capability of the model.

2.3. Parameter Settings

The training and prediction performance of the model are highly dependent on the hyperparameter configuration. A stepwise hyperparameter-tuning strategy was employed to balance model complexity and generalization capability. First, reasonable default values were set as the base parameters. Subsequently, grid search optimization was performed on the key parameters. The finalized parameter configurations are listed in Table 1.

Table 1. Model parameter settings.

Type	Parameter Name	Parameter Value
Global Parameters	Epoch	150
	Optimizer	Adam
	Batch	32
	Learning Rate	0.001
Feature Generator	Number of Encoder Layers	6
	Hidden Dimension	128
	Number of Residual Blocks	6
	Kernel Size	3
	Dilation Factor	2^i
Energy Disaggregator	Number of Decoder Layers	3
	Hidden Layer	[256,512,256]
Adversarial Domain Discriminator	Number of Network Layers	3
	Number of Neurons per Layer	[128,64,2]
Multi-Kernel Maximum Mean Discrepancy (MK-MMD)	Number of Kernel Functions	3

2.4. Working Process

During the training phase, the aggregated power sequences of the source domain D^s and target domain D^t were fed into the shared feature extractor to capture the temporal features of X_s^f and X_t^f from the total load data of different households. For the source domain samples, the data include the real power consumption Y_s^k of each electrical appliance, which can be used for supervised training. In contrast, only aggregated power sequences are provided for the target domain data to support adversarial learning. Subsequently, the features extracted by the feature extractor are input into the load disaggregator and the domain discriminator. In the domain discriminator module, adversarial training and adversarial loss L_d constraints are introduced to reverse promote the feature extractor to learn domain-invariant feature representations. The features processed by the load disaggregator were mapped to the predicted power consumption $\tilde{Y}_s^k$ (for the source domain) and $\tilde{Y}_t^k$ (for the target domain). The predicted power consumption is generated from the source domain features through the load disaggregator. This output was subsequently utilized to calculate the prediction loss. By contrast, the predictions derived from the target domain features did not participate in the prediction loss computation. Instead, these predictions serve as pseudo-labels for the subsequent distribution alignment. The calculation process is as follows:

$$L_p = \frac{1}{n^s} \sum_{i=1}^{n^s} F_p(\tilde{Y}_s^k, Y_s^k) \quad (11)$$

$$L_d = \frac{1}{n^s + n^t} \sum_{i=1}^{n^s + n^t} F_d(D(G_f(X_i)), d_i) \quad (12)$$

where n^s and n^t denote the number of training samples in the source and target domains, respectively; Y_s^k and $\tilde{Y}_s^k$ represent the actual and predicted power of the k_{th} appliance in the source domain; $F_p(\bullet)$ is the loss function of the load decomposer; d_i is the domain label indicating whether a sample originates from the source or target domain; and $F_d(\bullet)$ is the loss function of the discriminator.

In the model training phase, a collaborative adversarial mechanism was established. This mechanism was formed between the designed feature extraction module and the domain discrimination module. Specifically, it is analogous to the adversarial interaction

between a generator and discriminator in GANs. Consequently, the feature extractor was trained to produce domain-invariant representations through this adversarial process. The two modules are connected through a Gradient Reversal Layer (GRL) to achieve parameter transfer and adversarial optimization. In this structure, the domain discriminator distinguishes the domain origins of the input features. This task was conducted to enhance the discriminative ability of the model. Simultaneously, a GRL was utilized. During forward propagation, features are transmitted directly through the GRL. However, the gradient direction was reversed during backpropagation. Consequently, the feature extractor is enabled to update the parameters in a direction that reduces discriminative accuracy. Through this adversarial process, the model gradually learns consistent cross-domain feature representations, thereby improving the domain invariance of the feature space. The total loss L of the framework is formulated as

$$L = L_p - L_d + \lambda L_{\text{MK-MMD}} + L_{\text{CORAL}} \quad (13)$$

where L_p is the predicted loss, L_d is the domain adversarial loss, $L_{\text{MK-MMD}}$ is the multi-core maximum mean deviation loss, and L_{CORAL} is the correlation alignment loss.

A dynamic weighting mechanism was incorporated into the pseudo-label-driven MK-MMD alignment process. This mechanism was designed to alleviate the potential confirmation bias introduced during conditional distribution alignment. Specifically, pseudo-labels were generated for the target domain. However, these labels may suffer from considerable uncertainty during the early training stages. If a strong conditional distribution alignment is enforced based on such unreliable labels, the early prediction errors may be amplified. Consequently, a self-reinforcing error accumulation process may be triggered within the model. To alleviate this issue, dynamic weight λ_i is designed to increase progressively during the training procedure, allowing the model to gradually exploit target-domain pseudo-labels only after the feature extractor and load disaggregator have reached a relatively stable state.

$$p_i = \frac{i}{E} \quad (14)$$

$$\lambda_i = \frac{2}{1 + \exp(-10 * p_i)} - 1 \quad (15)$$

where i denotes the current iteration step, and E represents the total number of training iterations. This design ensures that $\lambda_i \approx 0$ at the early stage of training, resulting in a weak conditional distribution alignment, such that the model primarily relies on the real source-domain labels L_p to learn discriminative features. As training progressed, the source-domain and target-domain feature representations gradually stabilized, and λ_i increased accordingly, leading to a progressively strengthened MK-MMD alignment and gradual release of the influence of pseudo-labels, thereby enabling a safe and progressive alignment process. In addition, global constraints are imposed on the proposed framework. These constraints were implemented through CORAL-based marginal distribution alignment and adversarial training. Consequently, the dominance of pseudo-label errors was effectively suppressed during the overall training process. Furthermore, this strategy mitigated the risk of confirmation bias. As a result, the robustness of the cross-domain feature alignment was significantly enhanced.

During the training phase, the domain discriminator, MK-MMD, and CORAL were used as auxiliary constraint tools. However, these components did not participate in the model prediction phase. Consequently, subsequent prediction steps were executed independently of the auxiliary modules. The trained model directly invokes the time-series

feature output using the feature extractor through the load disaggregator, thereby achieving an accurate estimation of the electrical appliance power load.

3. Experimental Setup

3.1. Dataset and Data Preprocessing

(1) In this study, three open datasets were used for validation: two real-world datasets, REDD (**R**) [21] and UK-DALE (**U**) [22], and a synthetic dataset, SynD (**S**) [23].

The REDD dataset provides electricity readings from six households in the United States. It consists of aggregate power measurements and sub-metered appliance-level data. Aggregate-level data sampling takes place at a rate of 1 s, and appliance-level data sampling takes place at a rate of 3 s. The REDD datasets from different households span 3–19 days. By contrast, the UK-DALE dataset includes five households in the UK, with aggregate data sampling at 1 s and individual appliance data sampling at 6 s, covering 39–600 days across households. The SynD dataset provides 180 days' worth of simulated household electricity consumption data for 21 appliance types generated by domain experts, based on typical Austrian household load profiles. Compared with other datasets, SynD offers a longer monitoring period and is therefore more suitable for experimental studies.

Five representative appliances were selected as research objects: kettle (KT), microwave (MV), dishwasher (DW), washing machine (WM), and fridge (FG). The UK-DALE and SynD datasets included all five appliance types, whereas the REDD dataset contained only four.

To facilitate this research, we simplified the representation of the different datasets and houses. Here, **H** denotes the house ID in a dataset, where (**R**, **U**, **S**) corresponds to the REDD, UK-DALE, and SynD datasets, respectively. The notation " $\rightarrow$ " indicates the mapping from a source domain to a target domain; for example, **U1** $\rightarrow$ **R1** indicates the mapping from House 1 in the UK-DALE dataset to House 1 in the REDD dataset. Additionally, **HA** represents all remaining houses in the target domain dataset, excluding the target house itself.

(2) Data preprocessing: To improve the training efficiency and generalization performance of the model, raw data were preprocessed, including data cleaning, alignment, sample construction, and class balancing with normalization.

Data cleaning: To ensure data integrity and accuracy, the raw datasets were cleaned by identifying and imputing missing values. The original 20 s time-series segments were further divided into sub-sequences to eliminate short invalid time periods caused by device malfunction or sampling errors.

Data alignment: To guarantee the consistency of the aggregate power and temporal alignment of individual appliance operating times, the data were aligned. An upsampling method was applied to resample all appliance power data at a uniform time interval of 1 s. This uniform interval was selected for use throughout this study. The resampling process eliminates the time offsets caused by inconsistent logging or sampling intervals across different devices.

Sample construction: To convert time-series data into model-processable samples, a sliding window approach was used to generate input samples. A fixed window length for $T = 600$ was adopted, and the window was slid across a continuous time-series with different step sizes to produce the required samples. The same sliding step sizes were used for the UK-DALE and SynD datasets: 16 for kettle (KT), 24 for microwave (MV), 32 for dishwasher (DW), 8 for washing machine (WM), and 68 for refrigerator (FG). Owing to its smaller data volume, the REDD dataset used a step size of 2 for all appliances to ensure an adequate number of samples.

Class balancing and normalization: Given the significant imbalance in the electricity consumption data, OFF and ON state samples were randomly selected from the training set. This sampling strategy prevents the model from being dominated by the majority class during training and maintains an OFF-to-ON sample ratio of 0.2. The data were then normalized using the formula $x = (x - \mu) / \sigma$, where μ and σ are the mean and standard deviation of aggregate power, respectively.

3.2. Evaluation Indicators

In this study, the Mean Absolute Error (MAE) and Signal Aggregation Error (SAE) were selected as evaluation metrics using the following formulas:

$$\text{MAE} = \frac{1}{T_m} \sum_{t=1}^{T_m} |\tilde{y}_t - y_t| \quad (16)$$

$$\text{SAE} = \frac{1}{T_e} \left| \sum_{t=1}^{T_e} \tilde{y}_t - \sum_{t=1}^{T_e} y_t \right| \quad (17)$$

where T_m is the total length of the concatenated sequence. $\tilde{y}_t$ and y_t represent the predicted and real power values of the appliance, respectively, at time t . $\sum_{t=1}^{T_e} \tilde{y}_t$ and $\sum_{t=1}^{T_e} y_t$ denote the cumulative predicted and real power values, respectively, over the time interval T_e . In this work, $T_e = 3600$. Thus, SAE represents the absolute value of the cumulative difference between the predicted and true values within one hour, reflecting the average power deviation of the model over an hour. In contrast, MAE is the average of the absolute power errors across all sampling points in T_m . When the sampling rate is 1 s, it represents the average power deviation of the model at each sampling moment within the interval.

3.3. Feature Extractor Performance Experiment

Experiments were conducted on the UK-DALE dataset by using the same DA strategy to verify the performance of the feature extractor in the proposed model. The source and target domains were set to U1→U2. We compared the proposed TC-Net with five state-of-the-art feature extraction neural networks, including a Multilayer Perceptron (MLP), CNN, LSTM, Informer, and Transformer. The results are summarized in Table 2.

Table 2. Comparison of load disaggregation results with different feature extractors.

Methods	MAE (Watt)					SAE (Watt)				
	KT	MV	DW	WM	FG	KT	MV	DW	WM	FG
MLP	30.12	18.14	30.25	28.64	20.52	21.21	22.86	28.56	34.05	20.55
CNN	25.22	15.21	25.22	22.43	15.22	16.83	18.22	22.88	28.14	16.89
LSTM	13.45	13.14	22.11	23.11	15.31	10.67	17.13	21.44	24.13	16.24
Informer	12.76	11.82	22.37	21.53	15.13	9.58	10.29	21.09	20.07	12.55
Transformer	7.53	9.57	21.94	19.84	13.87	7.55	9.18	21.53	18.17	9.88
TCN	8.22	10.14	22.14	20.13	14.23	7.84	9.44	21.65	18.66	10.21

As shown in Table 2, the model using TCN as the feature extractor achieved superior overall performance compared to other architectures. Based on the MAE and SAE results in Table 2, the following conclusions can be drawn.

(1) MLP fails to capture temporal dependencies in time-series data, which limits its effectiveness in feature extraction and results in high MAE and SAE values. In contrast, TCN leverages a combination of causal convolutions and dilated convolutions to capture long-range temporal dependencies, enabling more accurate global feature extraction.

(2) CNN primarily relies on fixed-size local receptive fields for feature extraction. Although it can capture local spatial patterns, processing a long time-series requires an expanded receptive field. This is typically achieved by increasing the number of layers and the kernel sizes. Consequently, the model parameter count and computational cost increase. In contrast, TCN uses dilated convolutions to achieve a larger receptive field with fewer layers and parameters, leading to a lower MAE and SAE.

(3) LSTM effectively models long-term dependencies through gated memory units. However, its recursive computation mechanism leads to slower training when processing long sequences. In comparison, TCN, with its causal and dilated convolutions, typically converges faster and exhibits more reliable training on long sequences.

(4) Transformer is a sequence model based on the self-attention mechanism. Informer is an optimized variant designed for long-sequence time-series prediction. Although Informer's sparse attention mechanism improves the modeling efficiency for long sequences, it lacks the ability to capture short-term dependencies and sudden fluctuations. In contrast, Transformer's global attention mechanism balances long-term and short-term dependencies more effectively, resulting in stronger sequence feature extraction. Both models achieved lower MAE and SAE than TCN, with Transformer demonstrating overall better performance. However, both Informer and Transformer rely heavily on large datasets and tend to perform poorly on small datasets or in high-noise scenarios. Additionally, their self-attention layers involve a large number of parameters. This results in high computational complexity, making them unsuitable for low-latency and edge deployment scenarios in practical NILM applications.

(5) In summary, TCN employs causal and dilated convolutions to balance parameter efficiency and computational performance. This architecture effectively captures both the short- and long-term dependencies. As a result, TCN achieves lower MAE and SAE values, making it more suitable for NILM scenarios with limited data and enabling superior overall performance compared with other feature extractors.

3.4. Effectiveness Analysis of Domain-Adaptive Modules

To verify the impact of the MK-MMD and CORAL modules on transfer learning performance, we used t-SNE visualization to compare and analyze the feature distributions. An in-domain transfer task was constructed using the UK-DALE dataset. The source and target domains were set to $U1 \rightarrow U2$. The feature distributions and performance differences were evaluated for the three configurations. The configurations included adversarial training only, adversarial training with the CORAL module, and adversarial training with both the CORAL and MK-MMD modules. The feature distributions of the source and target domains are shown in Figure 3, where FG denotes the feature distributions of the source and target domains.

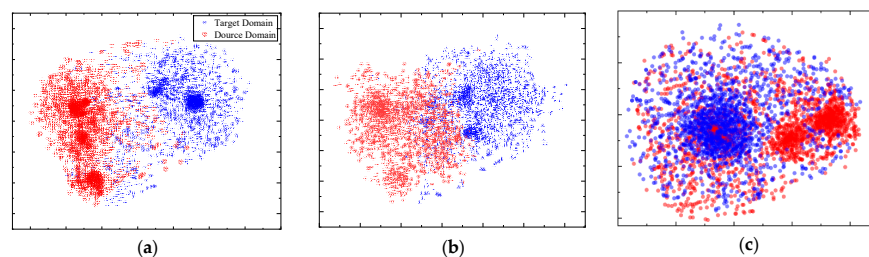


Figure 3. t-SNE visualization: (a) adversarial training only; (b) adversarial training with the CORAL module added; (c) adversarial training with both the CORAL and MK-MMD modules added.

As shown in Figure 3, a scenario with only adversarial training was considered. The FG features of the source and target domains form two separate but overlapping clusters. This phenomenon suggests that the model captures a certain degree of domain invariance. Subsequently, the CORAL module was incorporated. The distance between the two FG clusters was significantly reduced by aligning the second-order statistics of the global features. This improvement demonstrated the effectiveness of the proposed method for marginal distribution alignment. Finally, a complete model using the MK-MMD module was evaluated. In this configuration, the FG features of the two domains were highly mixed. Consequently, precise in-domain adaptation was achieved.

As shown in Figure 4, with an increase in the number of MK-MMD kernel functions, both the MAE and SAE metrics of the model for KT and FG appliances exhibit a decreasing trend. This indicates that increasing the number of kernel functions effectively enhanced the DA capability of the model. By introducing multiple Gaussian kernels, MK-MMD captures the domain distribution discrepancies at different scales, enabling a more comprehensive alignment of the feature distributions. On this basis, the model further introduces CORAL loss to align second-order statistics, resulting in a synergistic enhancement of DA performance. The experimental results indicate that the combination of MK-MMD and CORAL significantly strengthened domain alignment. Consequently, the distribution mismatch between the source and target domains was effectively mitigated. Moreover, the overall accuracy of load disaggregation was substantially improved.

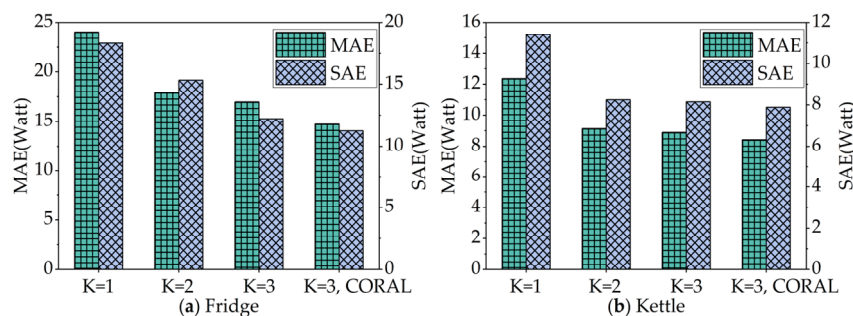


Figure 4. Comparison of results with different kernel function configurations: (a) fridge; (b) kettle.

3.5. Domain Migration Experiment

To verify the performance of each module in the proposed framework under both in-domain and cross-domain NILM scenarios, the following four models were designed for comparative analysis with the proposed model.

(1) Baseline: The model with all DA components removed from the proposed framework retains only the feature extractor and the load disaggregator. Its structure is similar to that of Seq2Point [24], and it was used to verify the disaggregation performance of the proposed model without DA components.

(2) CTL (Cross-Domain Transfer Learning) [25]: Regarding the cross-domain transfer learning approach, the model was initialized with source domain training and further fine-tuned using a small number of labeled target domain samples to facilitate better adaptation to the target domain distribution. The samples were used to verify the UDA performance of the proposed model.

(3) DTCN: The Baseline model is augmented with an adversarial domain discriminator, which enables the feature extractor to learn domain-invariant features through adversarial training. This was used to verify the disaggregation performance of the proposed model without MK-MMD and CORAL modules.

(4) CMMD: The Baseline model with additional CORAL and MK-MMD losses added to the original loss function. This was used to verify the performance of the CORAL and MK-MMD modules in the proposed model.

3.5.1. In-Domain Transfer Experiment

First, we verified the nonintrusive load disaggregation performance of the proposed model under in-domain transfer scenarios. Five models—Baseline, CTL, DTCN, CMMD, and the proposed model—were selected for comparative analysis using two real-world datasets: UK-DALE (U) and REDD (R). We evaluated the disaggregation performance of the five models under both single-source and multisource domain settings. The MAE and SAE results for each appliance across various datasets are summarized in Table 3. Specifically, the term “improved” was used to denote the relative reduction in error. This metric represents the performance gain achieved by the proposed model compared to the Baseline.

Table 3. Results of intradomain adaptation.

App.	Methods	UK-DALE(U)						REDD(R)					
		MAE (Watt)			SAE (Watt)			MAE (Watt)			SAE (Watt)		
		U1→U3	U1→U2	UA→U2	U1→U3	U1→U2	UA→U2	R1→R4	R1→R2	RA→R2	R1→R4	R1→R2	RA→R2
KT	Baseline	28.02	15.55	11.83	25.57	16.35	8.65	-	-	-	-	-	-
	CTL	35.22	14.00	13.65	33.01	14.71	7.78	-	-	-	-	-	-
	DTCN	21.01	21.66	28.87	29.18	22.26	16.49	-	-	-	-	-	-
	CMMD	26.81	9.33	20.10	15.34	19.81	6.19	-	-	-	-	-	-
	Ours	14.76	13.32	11.14	15.12	10.83	5.63	-	-	-	-	-	-
	Improved	47.32%	14.35%	5.82%	40.87%	33.76%	34.91%	-	-	-	-	-	-
MV	Baseline	92.42	20.62	25.07	89.89	17.21	23.88	30.48	5.58	24.40	23.77	5.18	16.57
	CTL	83.18	18.56	22.56	80.90	25.49	21.49	27.44	10.02	21.96	21.40	6.65	14.91
	DTCN	91.31	15.46	28.80	87.42	22.91	17.91	32.86	6.18	18.30	27.83	7.89	12.43
	CMMD	55.45	22.37	25.04	53.93	20.32	17.33	28.29	8.35	20.64	24.26	5.11	9.94
	Ours	65.82	13.64	17.82	52.35	11.76	16.72	22.89	5.31	13.92	19.71	4.31	15.92
	Improved	28.78%	33.84%	28.92%	41.76%	31.66%	29.99%	24.91%	4.81%	42.96%	17.09%	16.82%	3.91%
DW	Baseline	74.47	32.99	31.67	59.24	33.18	15.10	58.71	47.03	35.11	51.24	33.10	28.81
	CTL	67.03	29.69	38.50	53.32	29.86	23.59	52.84	42.33	31.60	46.11	29.79	25.93
	DTCN	55.85	34.74	43.75	34.43	28.88	31.33	44.04	35.27	26.33	38.43	24.83	21.61
	CMMD	44.68	29.80	33.00	35.55	29.91	19.06	45.23	28.22	24.07	30.74	19.86	23.29
	Ours	50.82	28.73	30.46	36.82	26.28	14.84	26.41	14.64	23.83	20.12	12.97	18.41
	Improved	31.76%	12.92%	3.81%	37.85%	20.79%	1.73%	55.02%	68.87%	32.13%	60.73%	60.82%	36.09%
WM	Baseline	37.01	50.68	73.00	30.85	54.88	33.79	63.39	29.79	74.89	54.30	26.45	65.75
	CTL	33.31	45.62	65.70	27.76	49.39	30.41	57.05	26.81	67.40	48.87	23.80	59.18
	DTCN	27.76	38.01	54.75	33.13	41.16	35.34	47.54	22.34	56.17	40.72	19.84	49.31
	CMMD	32.21	30.41	63.80	28.51	32.93	40.27	58.03	17.87	64.93	52.58	15.87	49.45
	Ours	27.49	25.93	54.61	24.72	25.88	27.13	43.14	13.75	53.26	37.45	13.57	43.89
	Improved	25.73%	48.84%	25.19%	19.86%	52.84%	19.71%	31.94%	53.84%	28.88%	31.03%	48.69%	33.25%
FG	Baseline	47.49	37.20	29.85	11.21	28.87	8.41	62.79	29.07	28.23	38.51	11.96	17.14
	CTL	49.74	43.48	26.86	10.09	15.98	17.57	56.51	26.16	35.40	34.66	14.77	15.42
	DTCN	35.61	37.90	22.38	18.41	3.65	10.31	67.09	29.80	28.17	38.88	18.97	12.85
	CMMD	38.49	35.32	17.91	26.73	17.32	9.05	67.67	27.44	26.94	33.11	17.18	14.28
	Ours	31.93	32.45	27.81	11.03	4.35	7.24	60.72	24.91	27.85	32.89	14.83	16.31
	Improved	32.76%	12.76%	6.82%	1.64%	84.93%	13.91%	3.29%	14.31%	1.33%	14.59%	-23.98%	4.82%

The bold entities denote the best performing algorithm for the appliance in each case and in different cases, respectively.

As shown in Table 3, the proposed model achieved a notable performance improvement in most in-domain transfer scenarios. Overall, the proposed framework outperforms the other four comparative models in terms of both MAE and SAE. Consequently, the

effectiveness of this framework was demonstrated for transferring tasks across different houses within the same domain. These results further confirmed the robustness of the model in in-domain scenarios.

Specifically, the improvement was most pronounced for power-intensive impulse-type appliances, such as MV and DW. For MV, under the U1→U3 scenario, the Baseline achieved an MAE of 92.42 Watt, whereas Ours reached 65.82 Watt, representing a 28.78% relative reduction. For DW under the same scenario, the Baseline SAE was 59.24 Watt, while Ours achieved 36.82 Watt, a 37.85% improvement. These results indicate that the proposed method can effectively reduce MAE and SAE when transferring between different houses within the same domain.

However, the performance of different models varies across appliances and scenarios. For example, with KT under the U1→U3 scenario, the Baseline achieved an MAE of 28.02 Watt, whereas CTL/DTCN/CMMD exhibited unstable performance. In contrast, our model reduced the MAE to 14.76 Watt, a 47.32% improvement. Simple fine-tuning or single-alignment strategies, such as KT, often struggle to disaggregate short-duration impulse loads. This difficulty arises because these methods fail to capture transient features effectively. In contrast, the proposed model implements a joint-alignment strategy. Consequently, transient features are captured more robustly, leading to an improved disaggregation accuracy. Additionally, it is observed from Table 3 that DTCN yields higher MAE and SAE than the Baseline in certain scenarios. This phenomenon suggests that the improvements from the adversarial domain discriminator alone are insufficient. In particular, discrepancies in both the conditional and label distributions cannot be fully addressed by this single-alignment mechanism.

Furthermore, different appliances exhibited varying responses to increased source domain data within the same domain. When the source data were sufficient and the appliance's operating mode was stable, adding more source data generally improved the model performance, reducing MAE and SAE. However, for appliances with complex variable operating modes (e.g., WM), excessive source data introduce additional distribution noise, impairing the disaggregation performance. In summary, under in-domain transfer scenarios, the proposed adversarial joint DA network demonstrates significant advantages for nonintrusive load disaggregation.

3.5.2. Cross-Domain Transfer Experiment

Next, we verified the nonintrusive load disaggregation performance of the proposed model under cross-domain transfer. Five models—Baseline, CTL, DTCN, CMMD, and the proposed model—were compared using two real-world datasets (UK-DALE (U) and REDD (R)) and one synthetic dataset (SynD (S)). We evaluated the cross-domain disaggregation performance of the five models using SA as the source domain. The MAE and SAE results for each appliance across the different target domains are presented in Table 4.

Table 4. Results of interdomain adaptation.

Metrics	Methods	UK-DALE (SA→U1)					REDD (SA→R1)				
		KT	MV	DW	WM	FG	KT	MV	DW	WM	FG
MAE (Watt)	Baseline	16.37	23.80	92.69	38.42	60.64	-	21.20	26.12	50.53	35.05
	CTL	14.73	21.42	83.42	44.57	34.58	-	19.08	23.50	55.47	25.54
	DTCN	12.27	17.85	69.52	38.81	45.48	-	15.90	19.59	47.90	26.29
	CMMD	19.82	24.28	65.61	43.05	46.38	-	12.72	15.67	50.32	31.03
	Ours	11.94	12.89	56.78	36.56	42.83	-	8.93	13.76	49.32	23.84
	Improved	27.04%	45.83%	38.74%	4.83%	29.37%	-	57.87%	47.31%	2.39%	31.98%
	Baseline	13.25	17.99	76.92	35.61	36.89	-	16.61	22.17	44.21	27.29

SAE (Watt)	CTL	11.93	16.19	69.23	42.05	33.20	-	14.95	19.95	39.79	16.56
	DTCN	9.94	13.49	57.69	36.71	47.67	-	12.46	16.63	43.16	20.47
	CMMD	17.95	20.79	56.15	41.37	42.13	-	9.97	13.30	36.53	26.38
	Ours	10.78	14.51	47.86	33.86	38.63	-	7.32	12.64	33.57	18.93
	Improved	18.65%	19.35%	37.78%	4.91%	-4.72%	-	55.93%	42.98%	1.45%	30.64%

The bold entities denote the best performing algorithm for the appliance in each case and in different cases, respectively.

As shown in Table 4, the MAE and SAE values of all models were generally higher in the cross-domain transfer scenarios than in the in-domain transfer scenarios. However, the proposed model significantly outperformed the other models, demonstrating its strong disaggregation performance in cross-domain scenarios.

Specifically, in the $SA \rightarrow U1$ scenario, the proposed method achieved notable advantages for appliances, such as MV, DW, and FG. The MAE of MV decreased from 23.80 Watt (Baseline) to 12.89 Watt (Ours), representing a 45.83% relative reduction. The SAE of DW dropped from 76.92 Watt (Baseline) to 47.86 Watt (Ours), a 37.78% improvement. Under the $SA \rightarrow R1$ scenario, the improvements were even more pronounced. The MAE of MV decreased from 21.20 Watt to 8.93 Watt (57.86% reduction), and the SAE of DW also decreased by 42.98%. These results indicate that the proposed method can maintain effective disaggregation performance in cross-domain scenarios.

Furthermore, Table 4 reveals that the proposed method does not always yield positive improvements for certain steady-state appliances (e.g., FG). For example, under the $SA \rightarrow U1$ scenario, the SAE of FG showed no improvement. This suggests that while MAE may decrease for long time-series, SAE can sometimes fail to improve, owing to persistent prediction biases or systematic errors over the time interval.

To further visualize the disaggregation results from a temporal perspective, the disaggregated load sequence curves of all the models are presented in Figure 5. These results corresponded to the $SA \rightarrow R1$ scenario. According to this figure, the largest deviation between the disaggregated curves and ground truth is exhibited by the Baseline model. This is attributed to the fact that the Baseline model completely ignores cross-domain distribution discrepancies. These results indicate that direct transfer learning is almost entirely ineffective in cross-domain scenarios with significant shifts in distribution.

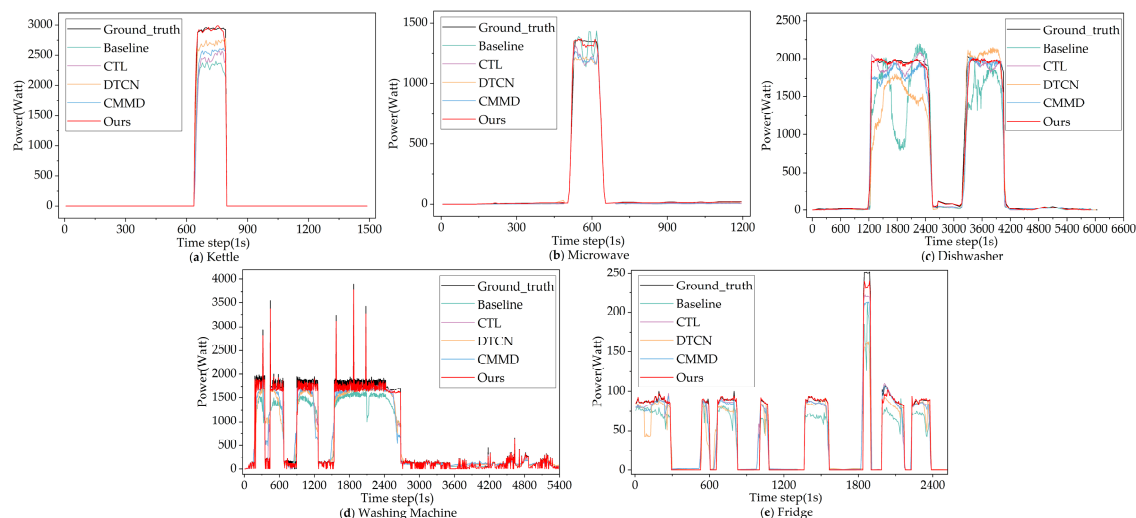


Figure 5. Disaggregated load sequence curves of all models under $SA \rightarrow R1$ scenario: (a) kettle; (b) microwave; (c) dishwasher; (d) washing machine; (e) fridge.

In comparison, CTL leveraged a small number of target domain labels for fine-tuning. Therefore, better performance was achieved by CTL than by the Baseline model. In addition, CMMD yields favorable disaggregation results through its feature alignment mechanism being an unsupervised DA method. In contrast, multilayer feature alignment is implemented in the proposed model. Consequently, the characteristics of multistate appliances were captured more effectively than those of the other models. Thus, the accuracy of load disaggregation in cross-domain scenarios was significantly improved.

4. Conclusions

In NILM, the cross-domain data distribution shift and limited availability of labeled source domain data are critical challenges. To address these issues, an UDA network was proposed in this paper. Specifically, this network integrates adversarial training with a hierarchical distribution alignment. In the proposed method, CORAL and MK-MMD were jointly employed. Through this integration, global marginal distributions and appliance-level conditional distributions were aligned. Moreover, discrepancies in both the feature and label spaces were simultaneously mitigated.

The experimental results on the REDD, UK-DALE, and SynD datasets demonstrate that the proposed model achieves superior performance. This superiority was observed in both in-domain and cross-domain scenarios. In particular, for appliances with complex operating patterns or pronounced power fluctuations, the proposed method yields significant reductions in MAE and SAE. Consequently, it consistently outperformed the baseline methods and representative transfer learning models such as CTL. For appliances with relatively simple operating characteristics, the proposed approach achieves a performance comparable to that of the state-of-the-art models.

Furthermore, scenarios with limited labeled data were investigated. It is observed that merely increasing the diversity of the training samples does not necessarily improve DA performance. In contrast, the hierarchical distribution alignment strategy enhances cross-domain generalization more effectively. These findings indicate that the proposed framework provides robust technical support for fine-grained demand-side energy management in smart grids.

Author Contributions: Methodology, H.X. and D.T.; Software, X.C., P.H. and D.T.; Validation, Y.H. and D.T.; Writing—Original Draft, P.H.; Writing—Review and Editing, H.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by State Grid Hubei Electric Power Co., Ltd. Technology Project (521532240024) and the National Natural Science Foundation of China (52407118).

Data Availability Statement: The original contributions presented in the study are included in the article, and further inquiries can be directed to the corresponding author.

Conflicts of Interest: All authors were employed by the company State Grid Hubei Electric Power Research Institute, Wuhan, China. The authors declare that this study received funding from the State Grid Hubei Electric Power Co., Ltd. Technology Project (521532240024) and the National Natural Science Foundation of China (52407118). The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

Abbreviations

NILM	Nonintrusive Load Monitoring
DA	Domain Adaptation
UDA	Unsupervised Domain Adaptation
TCN	Temporal Convolutional Network

MK-MMD	Multi-Kernel Maximum Mean Discrepancy
MMD	Maximum Mean Discrepancy
CORAL	Correlation Alignment
GRL	Gradient Reversal Layer
GAN	Generative Adversarial Network
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
MLP	Multilayer Perceptron
CTL	Cross-Domain Transfer Learning
MAE	Mean Absolute Error
SAE	Signal Aggregation Error
KT	Kettle
MV	Microwave
DW	Dishwasher
WM	Washing Machine
FG	Fridge
RKHS	Reproducing Kernel Hilbert Space
HEMS	Home Energy Management Systems

References

- Athanasiadis, C.L.; Papadopoulos, T.A.; Kryonidis, G.C.; Doukas, D.I. A holistic and personalized home energy management system with non-intrusive load monitoring. *IEEE Trans. Consum. Electron.* **2024**, *70*, 6725–6737.
- Donciu, C.; Serea, E.; Temneanu, M.C. Residential Electricity Consumption Behaviors in Eastern Romania: A Non-Invasive Survey-Based Assessment of Consumer Patterns. *Energies* **2025**, *18*, 4883.
- Luo, Q.; Yu, T.; Liang, M.; Pan, Z.; Guo, W.; Hu, X. Review of advances in scaling non-intrusive load monitoring for real-world applications. *Appl. Energy* **2025**, *398*, 126462.
- Athanasiadis, C.L.; Papadopoulos, A.; Kryonidis, G.C.; Doukas, D.I. Multi-objective data-driven framework to support network operation via residential flexibility. *Sustain. Energy Grids Netw.* **2025**, *44*, 102042.
- Alwaz, N.; Bashir, M.M.; Rehman, A.U.; Ullah, I.; Galea, M. Sustainable Optimization of Residential Electricity Consumption Using Predictive Modeling and Non-Intrusive Load Monitoring. *Sustainability* **2025**, *17*, 11193.
- Ghasrodashti, E.K.; Adibi, P.; Karshenas, H.; Kashani, H.B.; Chanussot, J. Multimodal Image Classification Based on Convolutional Network and Attention-Based Hidden Markov Random Field. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5511114.
- da Silva Nolasco, L.; Lazzaretti, A.E.; Mulinari, B.M. DeepDFML-NILM: A new CNN-based architecture for detection, feature extraction and multi-label classification in NILM signals. *IEEE Sens. J.* **2021**, *22*, 501–509.
- Lee, M.H.; Moon, H.J. Nonintrusive load monitoring using recurrent neural networks with occupants location information in residential buildings. *Energies* **2023**, *16*, 3688.
- Chu, X.; Pang, Y.; Ma, Y.; Li, S.; Qu, Y.; Wang, Y. Data-driven recommendation model based on meta-learning for non-intrusive load monitoring. *IEEE Trans. Consum. Electron.* **2023**, *70*, 3562–3572.
- Hu, L.; Kan, M.; Shan, S.; Chen, X. Unsupervised domain adaptation with hierarchical gradient synchronization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 4043–4052.
- Harell, A.; Jones, R.; Makonin, S.; Bajić, I.V. TraceGAN: Synthesizing appliance power signatures using generative adversarial networks. *IEEE Trans. Smart Grid* **2021**, *12*, 4553–4563.
- Li, J.; Chen, Z.; Cheng, L.; Liu, X. Energy data generation with wasserstein deep convolutional generative adversarial networks. *Energy* **2022**, *257*, 124694.
- Henriet, S.; Simsekli, U.; Richard, G.; Fuentes, B. Synthetic dataset generation for non-intrusive load monitoring in commercial buildings. In Proceedings of the 4th ACM International Conference on Systems for Energy-Efficient Built Environments, Delft, The Netherlands, 8–9 November 2017; pp. 1–2.
- Bao, G.; Huang, Y. Non-intrusive load monitoring based on ResNeXt network and transfer learning. *Autom. Electr. Power Syst.* **2023**, *47*, 110–120.
- Liu, H.; Liu, C.; Zhao, H. Noninvasive Decomposition of Electric-Gas Load in Public Buildings Based on Deep Learning. *Power Syst. Technol.* **2023**, *47*, 1188–1197.
- Su, H.; Hou, K.; Gao, J. Research on Non-Intrusive Load Monitoring Based on Domain Adaptive Learning. *Gansu Sci. Technol.* **2021**, *37*, 20–26+66.

17. Li; Meng, L.; Zhang, K. Review of Studies on Domain Adaptation. *Comput. Eng. Appl.* **2021**, *47*, 1–13.
18. Houidi, S.; Fourer, D.; Auger, F.; Sethom, H.B.A.; Miègeville, L. Comparative evaluation of non-intrusive load monitoring methods using relevant features and transfer learning. *Energies* **2021**, *14*, 2726.
19. Lin, J.; Ma, J.; Zhu, J.; Liang, H. Deep domain adaptation for non-intrusive load monitoring based on a knowledge transfer learning network. *IEEE Trans. Smart Grid* **2021**, *13*, 280–292.
20. Li, D.; Li, J.; Zeng, X.; Stankovic, V.; Stankovic, L.; Xiao, C.; Shi, Q. Transfer learning for multi-objective non-intrusive load monitoring in smart building. *Appl. Energy* **2023**, *329*, 120223.
21. Johnson, M.J.; Kolter, J.Z. A public data set for energy disaggregation research. *Data Min. Appl. Sustain.* **2011** *25*, 59–62.
22. Kelly, J.; Knottenbelt, W. The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes. *Sci. Data* **2015**, *2*, 1–14.
23. Klemenjak, C.; Kovatsch, C.; Herold, M.; Elmenreich, W. A synthetic energy dataset for non-intrusive load monitoring in households. *Sci. Data* **2015**, *7*, 108.
24. Zhang, C.; Zhong, M.; Wang, Z.; Goddard, N.; Sutton, C. Sequence-to-point learning with neural networks for non-intrusive load monitoring. In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018; Volume 32.
25. D’Incecco, M.; Squartini, S.; Zhong, M. Transfer learning for non-intrusive load monitoring. *IEEE Trans. Smart Grid* **2019**, *11*, 1419–1429.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.