Improving the Generalisability of Deep Learning NILM Algorithms using One-Shot Transfer Learning

Can one-shot transfer learning be leveraged to enhance the performance of a CNN-based NILM algorithm on unseen data?

Wim Verschuren

Supervisors: dr. ir. L. Cavalcante Siebert[1], ir. S. Kuilman[1]

[1]EEMCS, Delft University of Technology, The Netherlands

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
June 25, 2023

Name of the student: Wim Verschuren
Final project course: CSE3000 Research Project
Thesis committee: prof. dr. M.M. de Weerdt, dr. ir. L. Cavalcante Siebert, ir. S. Kuilman,

Abstract

Non-Intrusive Load Monitoring (NILM) is a technique used to disaggregate household power consumption data into individual appliance components without the need for dedicated meters for each appliance. This paper focuses on improving the generalizability of NILM algorithms to unseen households using Convolutional Neural Networks (CNNs) and one-shot transfer learning. The research investigates the effectiveness of one-shot transfer learning in fine-tuning a CNN model to accurately detect the ON/OFF state of appliances in households not seen during the training phase of the CNN. The study utilizes the Pecan Street dataset for training and evaluation, which includes detailed energy consumption records from various locations in the United States. The results suggest that one-shot transfer learning could enhance the performance of the NILM algorithm, particularly when multiple data samples are used for fine-tuning. However, the effectiveness of one-shot transfer learning varies strongly depending on the number of samples and the characteristics of the target household.

1 Introduction

Non-intrusive load monitoring (NILM) refers to the process of disaggregating a power time-series into individual components that constitute the aggregate signal. Employing NILM offers an effective means to measure electricity consumption of multiple appliances without the need to install a smart meter for each one. Such an approach holds various advantages, including providing electricity consumers with feedback regarding their appliance usage and establishing demand profiles [1].

The concept of NILM was initially introduced by G. Hart in the 1990s [2], after which extensive research has been conducted in this domain. Despite considerable research efforts, several critical issues still require thorough consideration to render NILM suitable for widespread deployment [3]. These issues encompass the following aspects:

1. Limited generalization of the algorithm to unseen households during training.
2. Inadequate detection of low-energy-consuming appliances.
3. Inability to identify appliances that switch on/off simultaneously.
4. High computational complexity.
5. Challenges associated with automatic detection of new appliances without necessitating algorithm retraining.

This research endeavor aims to address the first issue by attempting to enhance the generalization capability of the NILM algorithm, particularly concerning unseen households during training. The challenge of limited generalizability in NILM algorithms has been acknowledged in various studies, including [3], [4], and [5]. Solving this problem is essential to render NILM suitable for widespread deployment. Currently, achieving effective disaggregation performance in a building requires training the disaggregator using the appliance set specific to that building. This procedure proves inefficient and introduces intrusion [3] which would overthrow the entire purpose of non-intrusive load monitoring.

Previous attempts have been made to mitigate this challenge, such as those discussed in [6]. While prior studies have explored the transferability of houses within the same dataset, e.g., [7] and [8], as well as the application of one- or few-shot learning to improve NILM algorithms, e.g., [9], [10], and [11], none of them specifically address the utilization of one-shot learning to enhance the performance of a CNN on households not encountered during the training phase.

To enhance the generalization capability of deep learning NILM algorithms, we propose extending the deep learning algorithm with layers used for one-shot transfer learning. One-shot transfer learning refers to a machine learning paradigm where a model is trained to recognize or classify new objects or patterns based on only a single example or a very limited number of examples [12]. We will use a convolutional neural network as NILM algorithm. Data used for transfer learning could be gathered from consumers, in a non-intrusive way, through prompts (e.g., ”Are you currently using the dishwasher?”) on their smart thermostats. Consequently, the research question that arises is as follows: *”Can one-shot transfer learning be leveraged to enhance the performance of a CNN-based NILM algorithm on unseen households?”*

The main contribution of this study lies in demonstrating how one-shot transfer learning can be employed in conjunction with deep learning NILM algorithms to tailor them to specific households. Section 2 elaborates on frequently used topics in this paper. Section 3 will describe the methodology employed and the results will be presented in section 4. Ethical implications of this research will be discussed in section 5, followed by a discussion and conclusion in sections 6 and 7, respectively.

2 Background

This section aims to provide information regarding various concepts frequently used in this study. These encompass NILM algorithms, convolutional neural networks and one-shot transfer learning in sections 2.1, 2.2 and 2.3 respectively.

2.1 NILM algorithms

As is described in [4], NILM algorithms have evolved over time, transitioning from linear to nonlinear models and benefiting from advancements in deep learning techniques. In the early era of NILM (1995-2014), combinatorial optimization (CO) and unsupervised event detection were commonly employed. Support vector machines (SVMs), neural networks, decision trees (DTs), hybrid classification methods, and dynamic time warping (DTW) were utilized for event classification. Hidden Markov models (HMMs) and extensions were also explored, but they faced challenges in classifying unknown appliances and handling the complexity associated with increasing numbers of appliances. Linear decomposition and matrix factorization techniques were proposed as efficient approaches for estimating energy consumption per appliance.

Deep learning-based NILM algorithms gained traction from 2015 onwards, leveraging datasets from smart electric meters. Denoising autoencoders (DAEs) were used to reconstruct signals, while recurrent neural networks (RNNs), such as LSTM and GRU, proved effective for temporal dependencies in power signals. One-dimensional convolutional neural networks (CNNs), often combined with LSTM or recurrent convolutional networks, capture the temporal aspects of sequential time series data. Sequence-to-point CNN architectures and sequence-to-sequence models were also introduced.

Aside from the variety of techniques that are used, NILM algorithms also differ with regards to the task that they perform. Some algorithms exclusively perform appliance ON/OFF detection whilst others also attempt to predict their power consumption.

2.2 CNN

As said in [3], a CNN is a deep learning algorithm that performs exceptionally well on image-like (2D) data. This CNN architecture is also employed for sequence to sequence or sequence to point learning as in the case of NILM. Because of its exceptional success on 2D data NILM researchers have been motivated to apply this for the energy disaggregation problem as well.

The key feature of a CNN is its ability to automatically learn hierarchical representations of data through the application of convolutional filters. The network consists of multiple layers, including convolutional layers, pooling layers, and fully connected layers.

In the convolutional layer, the network applies a set of learnable filters, also known as kernels or feature detectors, to the input data. Each filter performs a convolution operation by sliding across the input, computing element-wise multiplications and summing the results to produce a feature map. These convolutional operations enable the network to capture local patterns within the data.

The pooling layer is typically inserted after the convolutional layer and helps reduce the spatial dimensions of the feature maps while preserving important features. Max pooling is a commonly used pooling technique, where the maximum value within a small region of the feature map is retained, discarding the remaining values.

The fully connected layers at the end of the CNN receive the output from the preceding layers and perform high-level feature extraction and classification. These layers are composed of neurons that are fully connected to the previous layer, similar to traditional neural networks. The final output layer typically uses an activation function such as sigmoid to produce the network's predictions.

2.3 One-shot transfer learning

One-shot learning is a machine learning approach that aims to enable models to recognize and classify objects or concepts with just a single example or a very limited amount of training data. Unlike traditional machine learning algorithms that typically require a large number of labeled samples to achieve high accuracy, one-shot learning leverages techniques that emphasize generalization and abstraction from a single or few instances. This approach is particularly useful when dealing with domains where acquiring abundant labeled data is challenging or impractical [12].

3 Methodology

This section outlines the methodologies employed in our research to improve the efficacy of Convolutional Neural Networks in the context of non-intrusive load monitoring through the utilization of one-shot transfer learning. We will conduct a comparative analysis between the performance of the CNN prior to the application of one-shot transfer learning and its performance subsequent to the implementation of one-shot transfer learning. The data from the dataset (section 3.1) is prepared as described in section 3.2. Consequently, our CNN as described in section 3.3, is trained as explained in section 3.4.

3.1 Dataset

The evaluation and training of our model will be conducted using data sourced from Pecan Street. The Pecan Street dataset[1] encompasses an extensive collection of comprehensive energy consumption records derived from numerous smart meters deployed across various locations within the United States. The data accessible as a university member includes static time-series data at three different temporal resolutions: 1-second energy, 1-minute energy, and 15-minute energy. The data consists of measurements from 73 homes, which are distributed across three research regions: New York, California, and Austin. Specifically, the New York dataset covers a period of six months and exhibits 100% completeness. In California, the dataset displays a 99% completeness rate for 23 homes, encompassing both 1-minute and 15-minute data. Finally, the dataset collected from the Austin, Texas area encompasses 99% completeness across all intervals for the 25 homes included in the study. The dataset encompasses a diverse range of building types, comprising single-family homes, townhomes, and apartments, thereby offering a broad representation of residential energy consumption patterns.

3.2 Data preparation

To train the CNN for ON/OFF detection, it is necessary to extract the appliance states (ON or OFF) from the power consumption data. The data is partitioned into 60-minute intervals, and during each interval, if the maximum power consumption surpasses a predetermined threshold, the corresponding device is labeled as ON.

To calculate the aggregate signal, the power consumption of all measured devices is aggregated. Although the original dataset includes a measurement of the total power consumption, we opted to construct our own grid signal by summing the power consumption of all devices. This decision stems from the disparities between the included grid measurement in the dataset and the actual sum of all measured devices. These disparities arise due to some devices in a household not being monitored by a smart meter and the presence of

[1] https://dataport.pecanstreet.org/

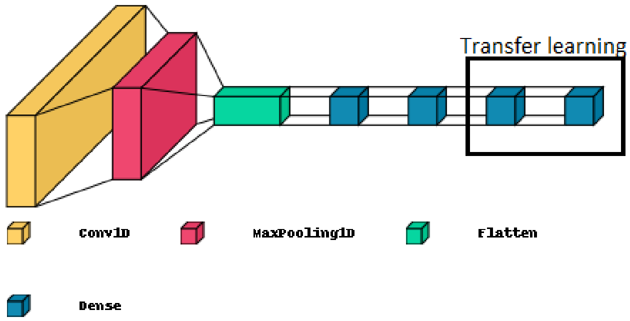

Figure 1: Graphic representation of the CNN architecture

solar panels, which can lead to negative power consumption values when excess energy is returned to the grid. In this research, we aim to eliminate the influence of solar panels and the discrepancy between the total grid measurement and the sum of all measured devices. Therefore, this data preparation step was implemented to ensure the integrity of the results.

3.3 Structure of the CNN

The convolutional neural network employed in this study comprises five distinct layers. Whereas other CNNs employ many more layers, the CNN is not the focus of this research, but rather the one-shot transfer learning extension. Therefore, it has been opted to keep the CNN simple and equip it with a mere five layers which are the exact amount to encompass a convolutional layer, max pooling layer, a flatten layer and a fully connected section (2 layers).

The initial layer is a convolutional layer, where multiple kernels traverse the one-dimensional input data. Subsequently, a Max Pooling 1D layer is applied to reduce the dimensionality of the feature map. This is followed by a layer that flattens the data, to consequently input it to two fully connected layers.

The CNN is extended with two layers to apply one-shot transfer learning. These are two fully connected layers. This is graphically represented in Figure 1.

3.4 Training

Subsequently, excluding the layers designed for one-shot transfer learning, the CNN is trained using the preprocessed data derived from the Pecan Street dataset. Once the training phase is complete, the layers are frozen, and two additional layers dedicated to one-shot transfer learning are added. Consequently, the network is trained using one to five data entries from the specific household for which the CNN is intended to be applied.

3.5 Evaluation

We will evaluate the complete model using data from the same household that we fine-tuned the model to using one-shot transfer learning. We will compare the performance of the model equipped with layers for one-shot learning to the performance of the model without layers for one-shot learning.

household id	n=0	n=1	n=2	n=3	n=4	n=5
661	0.0	16.8	41.7	33.7	41.7	33.7
1642	0.0	2.0	1.9	5.2	8.1	7.9
2335	0.7	16.4	25.3	21.6	34.3	29.4
2818	0.9	7.3	14.4	18.0	18.0	18.0
3039	0.0	0.7	19.1	19.0	18.6	18.6
3456	0.0	12.2	15.7	18.3	20.9	13.2
3538	3.2	11.8	7.5	12.7	18.9	16.6
4031	2.3	22.3	32.4	38.2	30.3	29.8
4373	3.3	2.8	14.1	10.2	12.6	15.0
4767	0.3	5.4	9.4	11.5	11.5	11.5
5746	0.0	7.6	20.3	17.3	19.2	20.4
6139	0.5	0.4	1.6	1.3	0.2	0.2
7536	0.9	2.5	17.7	9.1	18.2	20.9
7719	2.1	10.4	12.9	12.7	21.0	18.6
7800	0.0	8.7	9.8	10.0	9.7	9.7
7901	0.9	8.9	8.7	9.2	2.2	6.2
7951	1.5	6.6	12.0	7.2	11.1	10.5
8565	0.1	18.9	35.1	32.7	18.3	18.3
9019	0.3	5.6	4.7	5.3	5.6	5.3
9278	4.6	23.5	24.5	20.4	38.5	33.9
8156	0.7	0.5	1.4	2.1	0.6	0.6
8386	0.0	8.4	8.6	11.3	8.3	10.1
2361	0.7	10.3	27.7	42.5	42.5	42.5
9922	0.0	0.0	0.0	0.1	0.0	0.0
9160	0.0	0.0	0.0	10.7	13.6	20.4
mean	0.9	8.4	14.7	15.2	17.0	16.5

Table 1: The accuracy score in % on an Austin household after feeding n samples to the transfer learning layer.

4 Experimental Setup and Results

The CNN has been trained using the 1-minute data from the New York dataset[2]. After training it achieves an accuracy of 30.1% on the validation data. The CNN is acknowledged to be sub-optimal but performs, in terms of accuracy, far better than random guess since it does not merely give the prediction for one appliance, but for the entire appliance set in the household. Furthermore, achieving the highest accuracy is not the goal of this study. This CNN merely serves as a mechanism to construct upon the transfer learning layers, as the main goal of this study remains to show that one-shot transfer learning can be leveraged to enhance the generalizability of CNN-based NILM algorithms.

We want to apply this CNN for appliance ON/OFF detection for a specific household in the Austin dataset. To compare the results, we first evaluate the model without one-shot transfer learning using evaluation data from the household in Austin. Consequently, we apply one-shot learning and we re-evaluate the model to see whether the score has improved.

To evaluate the model we have taken various samples from the Austin dataset. Because the set of appliances differs per household it is interesting to try it out with different households and see how the model responds to this. Therefore, Table 1 contains the results from all households in the Austin

[2] https://gitlab.tudelft.nl/lcavalcantesie/flexibily_aggregation_smart_grids/-/tree/main_wim

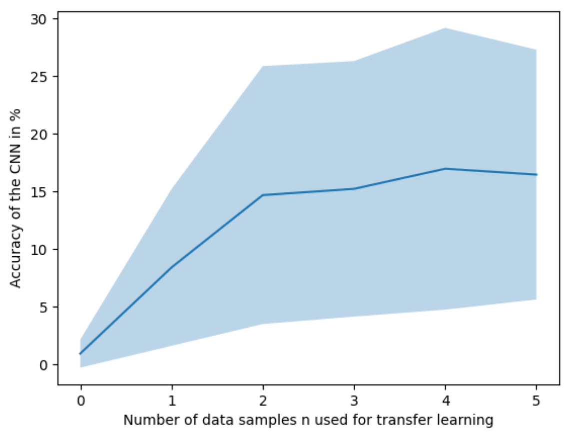

Figure 2: The mean accuracy values across all households, including standard deviation, after using n samples for transfer learning.

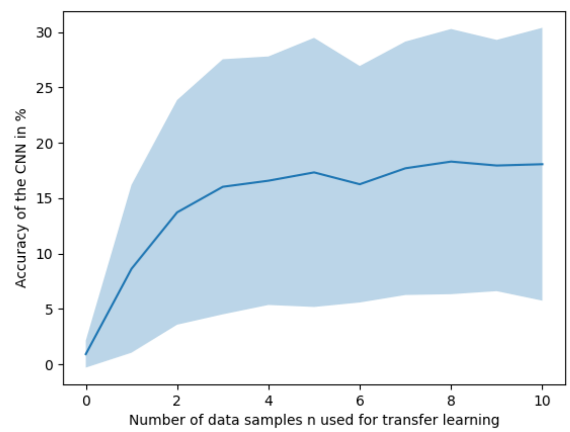

Figure 3: The mean accuracy values across all households, with standard deviation, extended to visualize 10 samples for transfer learning.

dataset. Additionally, it is interesting to see what happens if we take more than one sample to perform the transfer learning. This ranges from 1 to 5 samples. The mean improvements across all households, depending on the amount of samples used for one-shot transfer learning are plotted in Figure 2.

Table 1 suggests that one-shot transfer learning improves the performance of the NILM algorithm. This implies that one-shot learning could be an effective way for fine-tuning an algorithm to a specific household. However, as can be seen in Figure 2, the performance improvement is quite volatile, when differing the amount of samples used for one-shot transfer learning. Additionally, there is a significant variation in improvement across households i.e. different appliance sets. This explains the high standard deviation as shown in Figure 2.

As can be seen from Figure 2 at least two samples are required to obtain adequate results. The question remains how much improvement can be obtained from increasing n above 5. Our hypothesis would be that due to the shallowness of the transfer learning layer that it would easily over-fit on the training data. This phenomenon is described in [13].

After re-running and extending the number of samples used for transfer learning from 5 to 10 we obtain the results as in Figure 3. This plot implies that increasing the number of samples to train the transfer layers does not necessarily improve the performance for this implementation. Moreover, a flattening of the curve can be seen after 3 transfer learning samples which contradicts the hypothesis of over-fitting.

5 Responsible Research

While NILM holds tremendous potential in promoting energy efficiency and empowering consumers, it also raises significant ethical implications that need careful consideration.

The disaggregation of power consumption data can have serious implications on people's privacy. Whereas an aggregated signal does not give away much but the overall power usage, once it is disaggregated and the usage of the distinct appliances can be seen, this can give a detailed insight into individuals' habits. Even the TV channel somebody is watching can be predicted by looking at the aggregated energy consumption of all appliances in a house [14]. This raises questions about the extent to which individuals should have control over the collection and use of their data. It is important to strike a balance between data utility and privacy protection to prevent potential misuse or unauthorized access to personal information.

Furthermore, the implementation of NILM could lead to discriminatory practices. The gathering of more specific data about energy consumption could lead to differential pricing or targeted advertising. This may disproportionately affect vulnerable populations. It is therefore important that users are well-informed about the nature of data collection and how it will be used. Additionally, clear guidelines and regulations are necessary to ensure responsible data governance, including secure data store, data anonymization and restrictions on data sharing without the user's explicit consent.

It is important that researchers and companies that implement NILM are aware of the ethical challenges associated with data accuracy and reliability. Errors in appliance detection and classification could lead to misleading insight, impacting energy-saving decisions and potentially wasting resources. Therefore, manufactures and researchers should strive for transparency of NILM algorithms and mechanisms for users to verify and correct inaccuracies.

Since the potential of one-shot transfer learning to improve these NILM algorithms, they could make aforementioned aspects more relevant than ever. It is therefore important for NILM researchers to bear these ethical implications in mind. In this study, exclusively anonymized data has been used. Additionally, all code is publicly available to enable other parties to check the proposed methods.

6 Discussion

During this study on non-intrusive load monitoring, we have explored the use of convolutional neural networks and one-shot transfer learning to improve the performance of the NILM algorithm on unseen households. While the CNN employed in the research is acknowledged to be sub-optimal in terms of accuracy, the study suggests that one-shot learning could lead to an improvement in performance. However, we admit that the benefits of one-shot learning could be even greater if a better CNN architecture were initially deployed.

We acknowledge that each household possesses a unique set of appliances, and this variation can pose challenges for generalizing NILM algorithms to unseen data. This is also a possible explanation for the significant deviation in accuracy improvement across households as shown in Table 1 and depicted by the standard deviation in Figure 3. It would be interesting to investigate how different appliance sets and their ON/OFF distribution in the transfer learning training data, affect the effectiveness of one-shot learning and fine-tuning. Knowledge regarding this issue could benefit the selection of data used for transfer learning.

The research findings suggest that the actual improvement achieved through one-shot learning ($n = 1$) alone is relatively small. However, when multiple samples are used for one-shot transfer learning ($n \geq 2$), a more substantial improvement is observed. This highlights the potential benefits of collecting multiple data points from a household to fine-tune the NILM algorithm. Further research could explore the optimal number of samples required for effective one-shot learning and the trade-off between the number of samples and the resulting improvement in performance.

Lastly, the paper raises questions to what extent qualitative data can be obtained from a household using non-intrusive methods such as prompts on digital thermostats. The research suggests that gathering additional data from households can enhance the fine-tuning process and lead to improved performance. However, it is important to think about whether this data will be qualitatively sufficient since the labels (i.e. whether the device is ON or OFF) would be user generated. Further research would be needed to clarify this.

7 Conclusions and Future Work

In this research paper, we have explored the application of Convolutional Neural Networks (CNNs) and one-shot transfer learning to improve the performance of Non-Intrusive Load Monitoring (NILM) algorithms on unseen households. The main objective was to investigate if one-shot transfer learning can enhance the generalization of the CNN-based NILM algorithm to perform better on households not included in the training data.

Our findings indicate that one-shot transfer learning can lead to an improvement in the performance of the NILM algorithm. By fine-tuning the CNN model with a single data entry from the target household, we observed a noticeable enhancement in appliance ON/OFF detection accuracy. However, the degree of improvement varied depending on the number of data samples used for one-shot transfer learning and the specific household being analyzed. Therefore, further research has to be done to clarify the difference in accuracy in relation to the appliance set used in transfer learning data.

Furthermore, the research highlights the ethical implications associated with NILM, particularly concerning privacy, data accuracy, and potential discriminatory practices. It is essential to strike a balance between data utility and privacy protection, ensuring informed consent, secure data storage, and transparent data governance. Manufacturers and researchers should prioritize transparency and provide mechanisms for users to verify and correct inaccuracies in appliance detection and classification.

Overall, this study demonstrates the potential of one-shot transfer learning as a valuable approach to improve the generalization of CNN-based NILM algorithms to unseen households. It also underscores the importance of considering ethical considerations and the need for further research in optimizing CNN architectures and understanding the impact of different appliance sets on disaggregation accuracy.

By addressing these challenges, NILM can become a powerful tool for promoting energy efficiency, providing feedback to electricity consumers, and creating demand profiles. Further advancements in the field of NILM will contribute to a more sustainable and intelligent energy ecosystem.

References

[1] M. Devlin and B. P. Hayes, "Non-Intrusive Load Monitoring using Electricity Smart Meter Data: A Deep Learning Approach," *IEEE Power and Energy Society General Meeting*, vol. 2019-August, no. 8, 8 2019.

[2] G. W. Hart, "Nonintrusive Appliance Load Monitoring," *Proceedings of the IEEE*, vol. 80, no. 12, pp. 1870–1891, 1992.

[3] S. Dash and N. C. Sahoo, "Electric energy disaggregation via non-intrusive load monitoring: A state-of-the-art systematic review," *Electric Power Systems Research*, vol. 213, 12 2022.

[4] M. Kaselimi, E. Protopapadakis, A. Voulodimos, N. Doulamis, and A. Doulamis, "Towards Trustworthy Energy Disaggregation: A Review of Challenges, Methods, and Perspectives for Non-Intrusive Load Monitoring," *Sensors*, vol. 22, no. 15, p. 5872, 8 2022.

[5] Z. Zhou, Y. Xiang, H. Xu, Z. Yi, D. Shi, and Z. Wang, "A Novel Transfer Learning-Based Intelligent Nonintrusive Load-Monitoring with Limited Measurements," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–8, 2021.

[6] H. Rafiq, X. Shi, H. Zhang, H. Li, M. K. Ochani, and A. A. Shah, "Generalizability improvement of deep learning-based non-intrusive load monitoring system using data augmentation," *IEEE Transactions on Smart Grid*, vol. 12, no. 4, pp. 3265–3277, 7 2021.

[7] J. Kelly and W. Knottenbelt, "Neural NILM: Deep neural networks applied to energy disaggregation," in *BuildSys 2015 - Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built*. Association for Computing Machinery, Inc, 11 2015, pp. 55–64.

[8] D. Murray, L. Stankovic, V. Stankovic, S. Lulic, and S. Sladojevic, “Transferability of Neural Network Approaches for Low-rate Energy Disaggregation,” in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, vol. 2019-May. Institute of Electrical and Electronics Engineers Inc., 5 2019, pp. 8330–8334.

[9] L. De Baets, C. Develder, T. Dhaene, and D. Deschrijver, “Detection of unidentified appliances in non-intrusive load monitoring using siamese neural networks,” *Electrical Power and Energy Systems*, vol. 104, pp. 645–653, 2019.

[10] L. Wang, S. Mao, B. M. Wilamowski, and R. M. Nelms, “Pre-Trained Models for Non-Intrusive Appliance Load Monitoring,” *IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING*, vol. 6, no. 1, pp. 56–68, 2022.

[11] L. Matindife, Y. Sun, and Z. Wang, “Few-Shot Learning for Image-Based Nonintrusive Appliance Signal Recognition,” *Computational Intelligence and Neuroscience*, pp. 1–14, 2022.

[12] N. O’ Mahony, S. Campbell, A. Carvalho, L. Krpalkova, G. V. Hernandez, S. Harapanahalli, D. Riordan, and J. Walsh, “One-shot learning for custom identification tasks; a review,” in *Procedia Manufacturing*, vol. 38. Elsevier B.V., 2019, pp. 186–193.

[13] M. M. Bejani and M. Ghatee, “A systematic review on overfitting control in shallow and deep neural networks,” *Artificial Intelligence Review*, vol. 54, pp. 6391–6438, 123.

[14] P. A. Schirmer, I. Mporas, and A. Sheikh-Akbari, “Identification of TV channel watching from smart meter data using energy disaggregation,” *Energies*, vol. 14, no. 9, 5 2021.