Investigating Learning Transferability and Deployment for Neural NILM Strategies

Yacine Belguermi[1], Gilles Hermann[1], and Patrice Wira[1]

IRIMAS – Institut de Recherche en Informatique, Mathématiques, Automatique et Signal
IUT de Mulhouse, Université de Haute-Alsace
yacine.belguermi@uha.fr, gilles.hermann@uha.fr, patrice.wira@uha.fr

Abstract. This research focuses on Non-Intrusive Load Monitoring (NILM), a crucial component of energy management, enabling users to effectively monitor and reduce energy consumption. In a previous work, we developed a hybrid nerual network with combining 1D Convolutional Neural Networks (1D CNN) and Long Short-Term Memory networks (LSTM) for disaggregating 11 appliances using the AMPds2 dataset. Based on this work, we aim to create a generalized model capable of transferring knowledge gained from one building to others. Promising results have been achieved through fine-tuning techniques, indicating the model's adaptability and effectiveness in diverse settings. Notably, our research breaks new ground by employing transfer learning for the disaggregation of 10 appliances, surpassing previous work while maintaining a lower complexity. This study underscores the potential of NILM techniques in energy conservation and establishes a foundation for scalable, transferable models that can contribute to sustainable energy.

Keywords: Non-Intrusive Load Monitoring, energy disaggregation, Neural Networks, Transfer learning.

1 Introduction

Today's technologies are playing a pivotal role in addressing the global energy crisis by reshaping the way we design, construct, and manage buildings, giving rise to the concept of "smart buildings". In response to increasing energy demands, climate change concerns, and the need for sustainable practices, smart buildings leverage advanced technologies to optimize energy efficiency and resource utilization. These technologies enable buildings to dynamically adapt and monitor their heating, cooling, lighting, and power usage, thereby reducing energy consumption. Smart building technology has emerged as a key concept in modern urban infrastructure, prioritizing improvements in energy efficiency and sustainability [1]. A key facet of smart building operation involves the ongoing monitoring of electrical consumption by individual appliances, an aspect that has gained significant attention in recent years [2]. Through meticulous tracking and analysis of energy usage patterns within smart buildings, opportunities to address inefficiencies and implement precise energy-saving measures become evident. Authors in [3] reviewed existing research about the outcome of continuous and real-time monitoring of contained appliances within multiple buildings over multiple geographical areas. The results of the mentioned study demonstrate that appliance monitoring can yield energy savings of up to 23%. Nevertheless, it is imperative to acknowledge that implementing load monitoring in buildings necessitates the incorporation of additional hardware and resources [2, 4].

Actually, we can avoid the issue of installing more hardware to monitor appliances by using more sophisticated approaches. These methods involve enhancing the capabilities of the primary meter, enabling it to intelligently process total building energy consumption data for the purpose of monitoring individual appliances [2, 4, 5]. This methodology is commonly referred to as "Non-Intrusive Load Monitoring" (NILM) which has gained prominence in the field of energy management and smart building technology [2]. Recent research in NILM primarily focuses on using Machine Learning (ML) algorithms to address the challenge posed by the unpredictable nature of electrical load behavior [2]. In our previous study [6], we introduced a new neural network design to separate individual appliance signals from a main meter reading using a publicly available dataset. However, there is a common limitation with these solutions, generalizability was not verified across various buildings [5]. This issue arises because training ML models demands extensive datasets [5], which often means installing sub-meters for an extended period. Consequently, the main goal of our research is to test the applicability of our previous findings on a different dataset.

2 Non-Intrusive Load Monitoring (NILM) with Machine Learning

2.1 Overview and motivation

Non-Intrusive Load Monitoring (NILM), also known as energy disaggregation, is a technique used in the field of energy monitoring and management [4]. It involves the process of extracting detailed information about the individual appliances and devices within a building when there is only one meter measuring the total energy consumption for the entire building. This is modeled using equation (1) [4, 5]. In this scenario, NILM relies solely on the total power consumption signal, which is the aggregate power consumption of all appliances and devices operating within the building. By analyzing the unique electrical signatures and patterns associated with various appliances, NILM algorithms can identify when specific appliances are in use and estimate their power consumption. This enables users to gain insights into the energy usage of individual devices without the need for additional meters or sensors on each appliance, making it a non-intrusive and cost-effective method for monitoring and managing energy consumption in buildings. The total power consumption $P(t)$ is the sum of the unknown individual appliances' consumption $p_i(t)$ among N appliances and a measurement error $\varepsilon(t)$ [7]:

$$P(t) = \sum_{i=1}^{N} p_i(t) + \varepsilon(t) \tag{1}$$

Due to the inherent complexity of the problem, the inverse aggregation operation, particularly in cases where individual appliance consumptions are unknown and appliance behaviour is non-deterministic, poses a significant challenge. Traditional and straightforward algorithms may struggle to effectively disaggregate the aggregate signal under these conditions. Consequently, there has been a growing interest among scientists in employing ML algorithms to tackle this demanding task. ML offers a promising approach because it can adapt and learn from data, making it well-suited for capturing various appliances' diverse and often non-linear behaviours [2].

ML approaches for NILM are typically categorized into two primary domains: On/off detection (classification) and instantaneous consumption estimation (regression) [2, 8]. Our particular focus lies within the regression perspective. Among the array of ML algorithms proposed for this purpose, Markov models stand out as a noteworthy choice [2, 9].

Markov models possess the capability to autonomously discern and adapt to temporal dependencies within the data. Consequently, they hold significant promise in effectively addressing the challenges associated with load disaggregation [2, 9]. On the other hand, Artificial Neural Networks (ANNs) play a prominent role in NILM disaggregation [2, 5]. This involvement spans various types of ANNs, ranging from the foundational feedforward ANN to more advanced architectures like Convolutional Neural Networks (CNNs) and context-aware models such as Long Short-Term Memory (LSTM) networks [2]. We

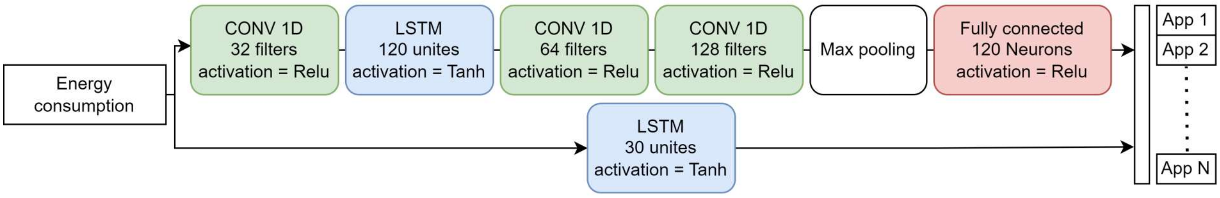

Fig. 1. Proposed multi-target disaggregation neural network [6]

have recently developed a new training scheme for disaggregating energy signals [6] using neural networks. Our approach consists of using a hybrid dual-channel deep neural network. We used 1D CNN along with LSTM layers to disaggregate 11 appliances in a Multi-Target Regression (MTR) framework with a sampling period of 60 seconds. The learning architecture is represented by Figure 1 [6]. Our method resulted in a disaggregation accuracy of 93.27% and a Normalized Root Mean Squared Error (NRMSE) of 0.19 [6]. Furthermore, our model performed better than previous works [6] using the AMPds2 dataset [10] which a reference low-frequency dataset. Nonetheless, a significant limitation of the majority of existing studies is their absence to assess the transferability and generalizability of the developed models to different households [5, 11]. In simpler terms, these studies often train and test their models using data from a single specific building, without considering how well the models might perform in other buildings.

2.2 NILM and transfer learning

In the context of this study, the authors in [5] introduced two distinct methodologies for addressing the concept of transfer learning on a mono-target regression basis for 5 appliances: Model-Agnostic Meta-Learning (MAML) and ensemble learning. In the MAML approach, a singular neural network undergoes pretraining on a meta-dataset and subsequently experiences fine-tuning when exposed to a new dataset. Conversely, in the ensemble approach detailed by the authors in [5], a network is structured as an ensemble of numerous sub-networks, each trained on distinct datasets. The work presented in [11] introduced a fine-tuning procedure using a pure 1D CNN within the framework of multi-target regression, utilizing four publicly available datasets. In these experiments, the number of disaggregated appliances is between 3 and 6. In their study detailed in [12], researchers examined the direct application of pre-trained models to alternate datasets without the need for fine-tuning or additional learning. This evaluation was conducted on three distinct appliances: A microwave, a dishwasher, and a fridge. The study introduced two distinct neural architectures for this purpose, a 1D CNN based architecture and another one consisting of a hybrid network incorporating both 1D CNN and bidirectional Gated Recurrent Units (GRU). Commonly, the mentioned studies [5, 11, 12] have presented intriguing results while encountering certain challenges during the transfer process. Additionally, the sampling period is the same (around 8 seconds). Lastly, a significant commonality among these studies is the transfer of knowledge to identical appliances, with each output corresponding to the same set of appliances.

Drawing inspiration from prior research and the concept of developing a versatile model, our objectives, using the best model obtained from [6], are as follows:

- To assess the adaptability of our trained model, as documented in [6], when applied to alternative datasets characterized by a notably low sampling frequency of 60 seconds.
- To expand the scope of appliance disaggregation to include a larger number of appliances, exceeding six in total.
- To introduce previously unconsidered (new) appliances into our model.

3 Investigation learning transferability for NILM

The intriguing aspect of transfer learning lies in its ability to enable a pre-trained model to generalize and perform a similar task on different data sources, requiring significantly less data and time [13]. Nevertheless, this necessitates pre-trained models that have been trained on very large multi-source datasets to ensure their optimal performance. To apply this concept to our specific task, we can easily fine-tune a pre-trained model to disaggregate appliance electrical consumption with just a few days of monitoring using plug meters.

In this section, we conduct our experiments and perform fine-tuning on the model originally obtained in [6], but this time we apply it to a different dataset. We utilize the same development environment as described in [6]. We recreate the identical network architecture from the previous work and import the parameters obtained in [6], subsequently initiating the fine-tuning process. Notably, we have switched the optimization algorithm from Adamax to Root Mean Square propagation (RMSprop) [14] which is an optimization algorithm that adapts the learning rates for each parameter during training to improve convergence.

3.1 DB description

Our initial research was conducted using the AMPds2 dataset, officially known as 'The Almanac of Minutely Power dataset 2 (AMPds2) [6, 10]. This dataset is publicly available and comprises two years of aggregated power consumption and load monitoring data collected from a household located in Canada [10]. It includes data from 20 different appliances, all sampled at a 60-second interval. We find this dataset particularly noteworthy due to its unique characteristic of low-frequency sampling and large number of monitored appliances [10]. In the other hand, the Electricity Consumption and Occupancy (ECO) dataset comprises appliance load measurements from six buildings over an eight-month period in Switzerland [15]. These measurements encompass current, voltage, and phase shift data from the three phases, sampled at a one-second frequency [15]. We preprocessed the dataset and calculated the real and reactive power, and then sub-sampled the whole set from 1 second to 60 seconds. Among the six houses we selected the second house because it contains more monitored loads than the others. Table 1 represents the most important characteristics of the used datasets. Note that we didn't consider the stove appliance from ECO dataset because of short monitoring duration. Figure 2 shows a heat map of appliances utilization over 244 days.

Table 1. Description and comparaison of major features for AMPds2 and ECO datasets

	AMPds2	ECO (house2)
Electrical distribution system	Single phase	Three Phases
Sampling period	T=60s	T=1s
Duration	2 years	244 days
Number of selected appliances	11/20	10/11
List of selected appliances	Basement, clothes washer, clothes dryer, dishwasher, fridge, furnace, heat pump, home office, wall oven, television, and hot water	Dishwasher (DWR), air exhaust (AXT), fridge (FRG), kettle (KTL), freezer (FRZ), television (TLV), tablet (TBL), entertainment (ENT), and lamp (LMP)
Selected features	Active and reactive powers	Active and reactive powers
Maximum active power (P)	11,706.0 watts	5,968.6 watts
Minimum active power (P)	0.0 Watts	0.0 Watts
Mean active power (P)	860.3 Watts	210.4 Watts
Standard deviation of active power (P)	826.7 watts	330.3 watts

3.2 Tests and results

We retrained the original model on the original dataset AMPds2 without considering the current (I) because ECO consists of a three-phase system. The results are almost the same with a tiny degradation of around 1%. Then we cloned the resulted model and fine tuned it using a learning rate of $5 \times 10^{-}5$ using 50 epochs. We considered randomly 20% of our data as training set and another 20% for testing. The Mean Absolute Error (MAE) in equation (2) was chosen to be both a metric to calculate the error and a loss function for the RMSprop optimizer. Along with MAE, we use the Desegregation Accuracy (DA) which indicates how well the model fit the ground truth (equation (3)). Note that N refers to the number of appliances, L is the length of the signal in samples, y_i^j represents the j^{th} sample that belongs to the i^{th} appliance for the real values. $\hat{y}$ refers to the estimated consumption with the same logic as for y.

$$MAE = \frac{1}{N \times L} \sum_{i=1}^{N} \sum_{j=0}^{L-1} \left| y_i^j - \hat{y}_i^j \right| \tag{2}$$

$$DA(\%) = 100 \times \left(1 - \frac{\sum_{i=1}^{N} \sum_{j=0}^{L-1} \left| y_i^j - \hat{y}_i^j \right|}{\sum_{i=1}^{N} \sum_{j=0}^{L-1} y_i^j} \right) \tag{3}$$

In order to investigate our results, we trained another model which has the same architecture but with randomly initialized parameters (weights and biases). Table 2 represents the results for the training of both models: Pre-trained model on AMPds2 (Fine-tuned) and the randomly initialized model (Trained). The columns titles represents the abbreviations of the chosen appliances and the last column is the overall performance. In Figure 3, we can visualize different disaggregation scenarios for three appliances from the test set: fridge, dishwasher, television, entertainment, lamp, and kettle. The area filled in grey

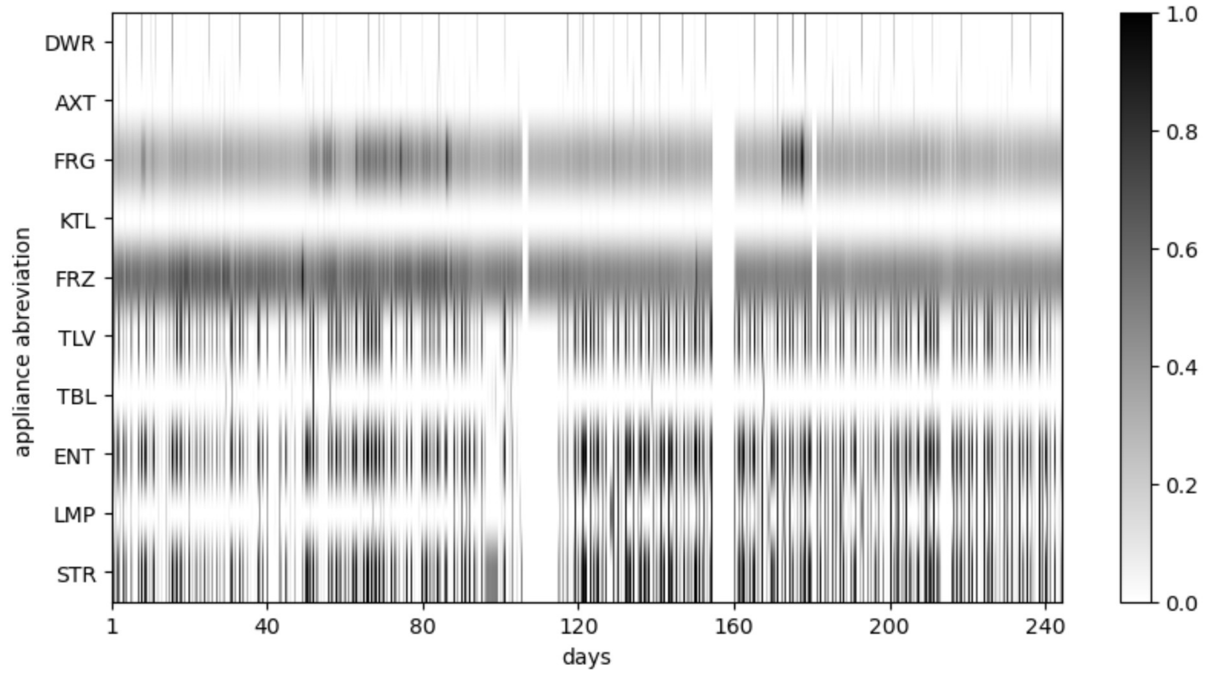

Fig. 2. Appliances' use rate over all monitored days

represents the total household consumption (main meter), the blue dashed line shows the actual consumption for these appliances (ground truth), the green line represents the disaggregation results from the fine-tuned model, and the red line shows the outcomes from the trained model.

4 Discussion and comparison with related works

In general, when we look at the results presented in Table 2 and Figure 3, we can see that the fine-tuned model performs much better than the model trained with random parameters. The fine-tuned model has an overall DA of 87.02% and a mean MAE of 5.56 Watts, whereas the randomly trained model has a DA of 82.58%. The most significant improvement is seen in the dishwasher appliance category. For the fine-tuned model, the DA for dishwashers is 79.67%, while the trained model only achieves a DA of 49.77%. This difference can be explained by the fact that the pre-trained model was exposed to the AMPds2 dataset, which contains more information about dishwasher usage compared

Table 2. Results of both models on the test set from ECO (fine-tuned trained on AMPds2 and fine tuned on ECO)

	model	DWR	AXT	FRG	KTL	FRZ	TLV	TBL	ENT	LMP	STR	ALL
MAE	Fine-tuned	**6.55**	**1.01**	**7.70**	6.25	**5.95**	**5.40**	1.32	**8.39**	**9.23**	**3.83**	**5.56**
	Trained	16.1	1.43	8.34	**6.18**	7.67	8.25	**1.13**	11.60	9.96	3.95	7.47
DA	Fine-tuned	**79.67**	**43.54**	**82.91**	47.09	**87.78**	**93.40**	40.35	**92.68**	**83.88**	**88.37**	**87.02**
	Trained	49.77	20.08	81.48	**48.48**	84.27	89.83	**49.20**	89.88	82.6	87.99	82.58

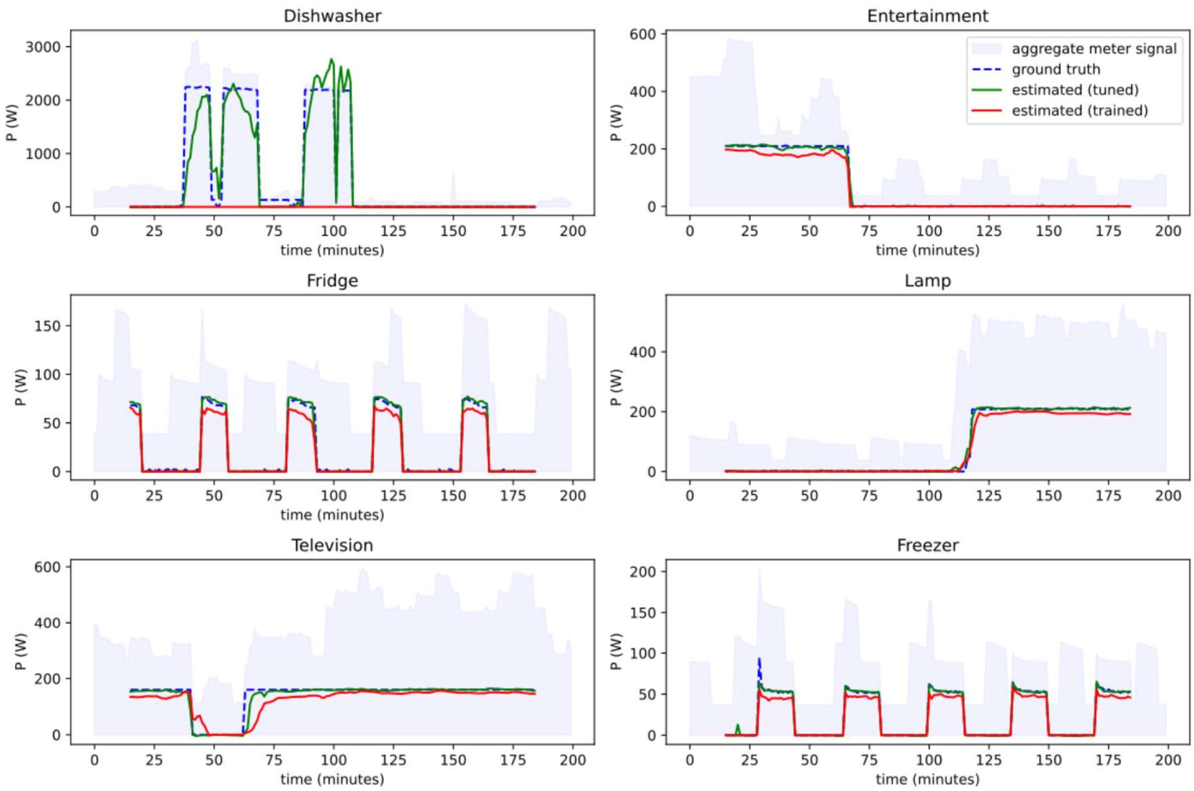

Fig. 3. Disaggregation performance of both models on the test set with respect to the aggregated signal and the real values (refereed to as ground truth)

to the ECO dataset. This exposure allowed the pre-trained model to generalize better, while the trained model struggled to accurately separate dishwasher usage from other appliances. Figure 2 proves our point of view on how this appliance was rarely used.

The datasets have other common appliances, namely the fridge and television. When comparing accuracy, the fine-tuned model performed better, especially for the television appliance, showing a notable increase of approximately 4% in accuracy. However, the fridge appliance's performance was quite similar for both models, with a slight advantage of 1.5% in favour of the fine-tuned model. This result can be attributed to the fact that the Fridge appliance is typically running continuously as we can see in Figure 2, allowing the trained model, even with limited data, to effectively disaggregate this appliance. We have observed intriguing outcomes regarding the entertainment appliance, where the fine-tuned model demonstrated superior disaggregation performance compared to the trained model. Interestingly, even though the exact appliance was absent in the AMPds2 dataset, the home office appliance consumption is much similar to the entertainment appliance. The freezer consumption signature is very similar to the fridge appliance which explains the good performance.

For the remaining appliances, the fine-tuned model consistently outperformed the trained model, except for two appliances: tablet and kettle. Regrettably, both models struggled to effectively disaggregate these two appliances. It is noteworthy that the tablet and kettle appliances were not included in the AMPds2 dataset and their use was rare (Figure 2), which consequently led to the pre-trained model's failure in accurately disaggregating them.

Table 3. Comparison of our work with related works

Work	Sampling period	Target	Number of appliances	Input width	Number of datasets	Transfer method
[5]	8s	mono-target	5	99	2	fine-tuning
[11]	6s	**multi-target**	3-7	599	**3**	fine-tuning
[12]	8s	mono-target	3	variant	**3**	**direct-test**
Our	**60s**	**multi-target**	**10**	**30**	2	fine-tuning

Let us now examine the output performances of both models, as depicted in Figure 3. In the left column, we observe the appliances common to both datasets. Notably, for the dishwasher appliance, the trained model exhibited complete inability to detect it, while the fine-tuned model achieved a partial fit to the actual consumption pattern, as reflected in the accuracy values reported in Table 2. Across the remaining examples, it becomes evident that the fine-tuned model consistently outperforms the trained version by providing a closer match to the ground truth a lower error. Finally, a comparative analysis is conducted to assess the intricacy of our research, the extent of disaggregated appliances considered, and whether the transfer necessitates fine-tuning or can be directly tested. The evaluation of complexity encompasses several critical facets, namely the width of the input window, the choice between multi or mono-target disaggregation, and the sampling period. Based on the information on Table 3, we notice that our method is less complex with a very high sampling period (60s) instead of 8s or 6s [5, 11, 12] with a window of 30 samples only. Our model is a multi-target meaning that we need only one model to disaggregate all appliances simultaneously. The limitation of our work is the number of used datasets where we trained our model on AMPds2 and fine-tuned it on ECO, the final limitation is that we still need fine-tuning meaning we still need some supervised data to apply our task.

5 Conclusion

In conclusion, our research has demonstrated the potential and effectiveness of transfer learning in the context of Non-Intrusive Load Monitoring (NILM). By training a model on the AMPds2 dataset and then transferring this knowledge to the ECO dataset, we achieved remarkable results. Notably, we successfully disaggregated 10 appliances more than any previous multi-target work, achieving an impressive 87.02% accuracy using just 20% of the dataset for training. Comparing our transfer learning approach to a model with randomly initialized parameters, which achieved an accuracy of 82.53%, underscores the significance of pre-trained knowledge in NILM tasks. While our results are promising, there remains room for improvement in the original model to enhance accuracy even further. Our overarching goal is to create a highly generalized model that requires minimal fine-tuning, making it adaptable to a wide range of scenarios. This entails harnessing larger and more diverse datasets to further advance the field of NILM and contribute to energy conservation efforts.

References

1. Alanne, K., Sierla, S.: An overview of machine learning applications for smart buildings. Sustainable Cities and Society **76** (2022) 103445

2. Kaselimi, M., Protopapadakis, E., Voulodimos, A., Doulamis, N., Doulamis, A.: Towards trustworthy energy disaggregation: A review of challenges, methods, and perspectives for non-intrusive load monitoring. Sensors **22**(15) (2022)
3. Zangheri, P., Serrenho, T., Bertoldi, P.: Energy savings from feedback systems: A meta-studies' review. Energies **12**(19) (2019)
4. Weißhaar, D., Held, P., Benyoucef, D., Ould Abdeslam, D., Wira, P., Merckl é, J. In: Smart Meters Improved by NILM. Springer International Publishing (2023) 29–54
5. Wang, L., Mao, S., Wilamowski, B.M., Nelms, R.M.: Pre-trained models for non-intrusive appliance load monitoring. IEEE Transactions on Green Communications and Networking **6**(1) (2022) 56–68
6. Belguermi, Y., Wira, P., Hermann, G.: A multi-output lstm-cnn learning scheme for power disaggregation within a nilm framework. In: 2023 IEEE 21st International Conference on Industrial Informatics (INDIN), Lemgo, Germany (2023) 1–6
7. Hart, G.W.: Nonintrusive appliance load monitoring. Proceedings of the IEEE **80**(12) (1992) 1870–1891
8. Cimen, H., Palacios-Garcia, E.J., Kolaek, M., Cetinkaya, N., Vasquez, J.C., Guerrero, J.M.: Smart-building applications: Deep learning-based, real-time load monitoring. IEEE Industrial Electronics Magazine **15**(2) (2021) 4–15
9. Wu, Z., Wang, C., Peng, W., Liu, W., Zhang, H.: Non-intrusive load monitoring using factorial hidden markov model based on adaptive density peak clustering. Energy and Buildings **244** (2021) 111025
10. Makonin, S.: AMPds2: The Almanac of Minutely Power dataset (Version 2) (2016)
11. 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. Applied Energy **329** (2023) 120223
12. Murray, D., Stankovic, L., Stankovic, V., Lulic, S., Sladojevic, S.: Transferability of neural network approaches for low-rate energy disaggregation. In: ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK (2019) 8330–8334
13. Vrbančič, G., Podgorelec, V.: Transfer learning with adaptive fine-tuning. IEEE Access **8** (2020) 196197–196211
14. Murphy, K.P.: Probabilistic Machine Learning: An introduction. MIT Press (2023)
15. Beckel, C., Kleiminger, W., Cicchetti, R., Staake, T., Santini, S.: The eco data set and the performance of non-intrusive load monitoring algorithms. In: Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings (BuildSys '14), New York, NY, USA (2014) 80–89