Deep Learning Based Non-Intrusive Load Monitoring for Smart Energy Management in Households

Parvaneh Zavareh*, Warunee Soythong*, Basirah Noor†, Volker Hoffmann‡, and Huamin Ren†

*School of Economics, Innovation and Technology, Kristiania University College, Oslo, Norway
E-Mail: {paza002, waso001}@student.kristiania.no

†School of Economics, Innovation and Technology, Kristiania University College, Oslo, Norway
E-Mail: {Basirah.Noor, Huamin.Ren }@kristiania.no

‡SINTEF, Forskningsveien 1, 0373 Oslo, Norway
E-Mail: volker.hoffmann@sintef.no

***Abstract*—The increasing digitalization inside the energy sector is transforming the fact that how residential electricity consumption is analysed and optimized. This study aims to propose and evaluate a Deep learning-based model for Non-Intrusive Load Monitoring (NILM), also called energy disaggregation, designed for the real-world applications in smart homes energy management systems. The framework employs Convolutional Neural Network (CNN) based sequence-to-point (Seq2Point) and sequence-to-sequence (Seq2Seq) model to isolate appliance level consumption from the aggregate power, while addressing generalization challenges across diverse households and appliances type. The results suggest that model choice (Seq2Point vs Seq2Seq) effects the results, for example Seq2Point shows better results in detecting high power, short duration events,highlighting the importance of selecting architecture based on appliance-specific characteristics for optimal NILM performance.The practical implications suggest that the proposed study can enable real-time appliance monitoring, adaptive demand-side management, and sustainable energy optimization in future smart grid environments.**

Keywords-Non-Intrusive Load Monitoring (NILM); Convolutional Neural Networks; Seq2Seq; Seq2Point; Smart Home; Energy Management.

I. Introduction

Global electricity consumption grew by 2.2% in 2024, reflecting a faster-than-average rise in energy demand, according to International Energy Agency (IEA) [1]. Therefore, optimizing energy consumption has become a critical aspect now with the growing integration of smart households and smart grids [2]. The global market has been revolutionized when it comes to energy consumption because of energy aware technologies. Traditionally, aggregate consumption models were only used for load monitoring, which results in limited insights into the user behaviour [3]. As the energy demand is increasing globally, efficient energy monitoring across households can contribute more in optimizing energy usage, cost efficiency, and eventurally to the overall sustainability.

Non-Intrusive Load Monitoring (NILM) is a promising technique that can address this challenge by disaggregating the total household power consumption into appliance level consumption by using the data collected from single smart meter without the need of any device sensor [4]. The approach shows its application in energy conservation, demand-side management and smart home automation [5].

Recent advances in the field of Artificial Intelligence (AI) and Deep Learning (DL) have made NILM to perform even more better than before. Unlike traditional feature engineered models, CNN based approaches (Seq2Point and Seq2Seq) have this ability to automatically extract temporal and spatial patterns from raw energy data [6]. They can easily capture appliance signatures and therefore improves the appliance state detection and load disaggregation process.

Despite these models show promising results but previously many researchers have focused only on the energy disaggregation and not to the state detection (ON/OFF), which is an essential ingredient for automation and real time monitoring. This study will also adopt how preprocessing choices can affect the overall efficiency and generalization of the model across different appliances and households.

This study aims to achieve the following objectives to overcome the limitations of previous research.

- Systematically optimize the CNN based NILM models to improve both appliance state (ON/OFF) detection and energy disaggregation to individual load profile.
- Comparatively studying Seq2Seq and Seq2Point models for better load identification by keeping in view the variable household environments and different appliances.
- Analyzes how preprocessing parameters, such as learning rate, window size, normalization, etc. affect model accuracy and generalization.

The rest of the paper is organized in this way that Section II reviews the background and the related work on NILM and CNN based power disaggregation. Section III will present the proposed methodology including the dataset selection, data preprocessing, model training etc. The experimental results are presented and discussed in Section IV. Finally, Section V will provide the conclusion and future work.

 ISBN: 978-1-68558-357-6

II. Background and Recent Work

This section will define NILM and how Deep learning can help to improve its ability. Some recent work will also be presented to support the definition with evidence.

A. Non-Intrusive Load Monitoring

NILM is a promising technology to reduce the overall energy consumption. NILM researchers are continuously in this effort to introduce the best model to compliment the Hart's Theory [7]. CNNs are especially an effective technique to implement the NILM due to their capability to automatically extract time related information from the signal data and hence show better computational efficiency as compared to Recurrent Neural Networks (RNNs) or Long- and Short-Term Memory Networks (LSTMs). In this process the total power consumption at time t is calculated using the Hart's Signal Decomposition Theory [8].

$$P(t) = \sum_{i=1}^{n} a_i(t) P_i + e(t) \quad (1)$$

where:

- $P(t)$ is the total power consumption at time t,
- $a_i(t)$ is the ON/OFF state of the i^{th} appliance at time t,
- P_i is the power rating of appliance i,
- $e(t)$ is the error term accounting for power from appliances not considered, and
- n is the total number of appliances.

NILM has been framed historically both as classification and regression problems [9]. When regarded as classification problem, the ON/OFF state of each appliance is classified simultaneously and typically event-based methods focusing on detecting sudden changes in power signals have been proposed, while time-based methods aim to continuously monitor appliance state, such as in [10]. The authors in [11] leverages temporal patterns for their model Temporal Attention Bottleneck Variational Autoencoder (TAB-VAE) to capture long-term dependencies in appliance behaviour, thus offering more reliable time based continuous state monitoring. Whereas researchers from [12] used Conditional Random Fields (CRFs) with CNNs for predicting states by taking into account the sharp power changes typically associated with appliance state transitions, which are common in event-driven tasks.

B. Application of Deep Learning in NILM

Many CNN-based models have been proposed to enhance NILM performance. Authors of [13] introduced a Dual-CNN based model that independently learns state transitions and power consumption patterns, that helped to significantly improve classification accuracy. Similarly, [14] and [15] demonstrated that dilated convolutions and residual connections expand the receptive field without increasing computational complexity, allowing the models to effectively capture long-range dependencies. Several RNNs, particularly LSTM models, have also been employed in NILM to model temporal sequences effectively. A recent study [16] demonstrated that LSTM models trained on the PLAID dataset [17] achieved high levels of precision and recall across multiple appliances, particularly when integrated with IoT systems for near real-time classification. However, RNNs often face challenges when modelling long-duration appliance data, where convolutional models, especially CNN-based Seq2Seq and Seq2Point architectures, tend to perform better [18] and [19].

III. Methodology

This section will throw the light on the research design of the study. The goal of the study is to dis-aggregate the overall aggregate power of any household into appliance level power consumption using the CNN based Seq2Seq and Seq2Point models.

A. Data Acquisition

The data acquisition is the most important part of any model building. Two datasets were selected for this study. UK Domestic Appliance-Level Electricity (UK-DALE) is selected for model development and validation due to its acceptance in the NILM community, while the REFIT dataset is used as a secondary dataset for external testing to assess the generalization capability of the trained models.

Five household appliances were selected for further analysis -kettle, microwave, dishwasher, washer-dryer, and fridge-freezer - due to their nature of distinct operating pattern and high power consumption [20]. The required data has been collected from the UK-DALE dataset that comprises of the data from 5 households. Each household's data is organized into separate .dat files, where each file corresponds to single appliance level power consumption that were timestamped. REFIT provides data on household energy consumption from 20 homes in the UK and offers a wider range of appliances [21][22].

The data were split in two ways to improve the quality of the data. The UK-DALE set was first divided at the household level, where data from house 1,3,4 and 5 were used for training purposes and data from house 2 for testing purposes. Within each of the training houses, the data were further partitioned into an 80% training subset and a 20% validation subset. This internal split ensured that hyperparameter tuning and model selection were performed without contaminating the unseen test house. In the end data from REFIT was used to test the model and hence generalizability across the datsets.

B. Data Pre-Processing

After merging and aligning the dataset, the next step involved in the pipeline is the most important one as the results highly depends on these techniques. This study has utilized the following pre-processing techniques.

1) Cleaning and Resampling: This process involves cleaning of the data by handling missing entries. To address the issue of missing data gap-detection method is applied. The time difference between the two consecutive entries were calculated and if the difference is greater than 3 minutes, the gap would be filled with zero to show the inactivity. Whereas the gaps with smaller time differences were forward filled. Timestamps are converted to standard date-time format for easier manipulation.

After cleaning, Power readings were resampled to 30–60-second intervals using mean aggregation. This process reduced the high frequency attenuation. Furthermore, all NaN entries were dropped and the resultant dataset is been saved in a separate CSV files for each house.

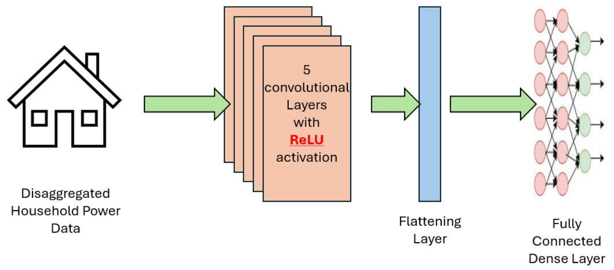

Figure 1. The Structure of Seq2point based CNN Model

2) Normalization: The process is used to standardize the dataset within the defined range. Several normalization techniques were applied including min-max method and z-score. Min-max normalization scales the data between the range (0 and 1), while z-score normalization scales the data so that it has a mean of 0 and a standard deviation of 1.

3) Labeling: Binary labels were generated to indicate the ON/OFF state. Different power criteria is applied to different appliances for example for kettle, it is considered ON if the power exceed from 1500 W while for fridge-freezer, it is considered ON, if power exceeds from 50 W. Minimum duration constraints were applied to avoid false activations from transient spikes.

4) Exploratory Data Analysis (EDA) and Visualization: The data in this step is analysed and visualized to better understand the power patterns. For the purpose, Python libraries, such as Matplotlib, Seaborn, and Plotly were used. the EDA helped us to find out that House 1 has the most complete data information over the period, that is why we will use it for training of the model. Houses 3, 4, and 5 offer moderate data and were also used for training. Where as house 2 had some gaps would be used for testing the model.

C. Model Implementation

Two CNN based models; Seq2Point and Seq2Seq were used. The following is the model configuration for CNN architecture that has 5 convolutional layers.

- Filter sizes: [30, 30, 40, 50, 50]
- Kernel sizes: [5, 4, 3, 3, 2]
- Activation function: ReLU for convolutional layers
- Dense layer: Size 256, ReLU activation
- Output layer: Linear activation, with size equal to the sliding window size
- Batch Size: 1024 (chosen to speed up training)

1) Seq2seq Based CNN Model: The seq2seq model uses aggregate energy readings and generates a corresponding output window of the same length. It maps an input sequence to a single output point. The model consists of 5 convolutional layer same as Seq2Point model based on ReLU activation, after that there is a flattening and Reshaping layer followed by dense layers for each output time step.

2) Seq2Point Based CNN Model: The seq2point model uses a window size of fixed length. It maps an input sequence to an entire output sequence. The model consists of 5 convolutional layer based on ReLU activation, after that there is a flattening layer followed by a fully connected dense layer as shown in Figure 1.

When we consider NILM based scenarios, the core difference between both models lies in their output structure, and hence the prediction would be different too. The Seq2seq model takes a sequence data (For example sequence of aggregate power readings) as input and outputs a corresponding sequence based data. The input and output sequences typically have the same length. Whereas for Seq2point, the input data is the same while it outputs only a single point data that shows the predicted power consumption at some defined, mostly the midpoint of the input window point for specific appliance. Therefore, Seq2point can be more helpful model where we need to capture sharp changes in appliance power like ON and OFF scenarios , while Seq2seq are better for representing continuous sequences and is well-suited for tasks where the entire output sequence is of interest.

The models are first trained as regression models that estimate continuous appliance power values, and the resulting predictions are then thresholded to derive binary ON/OFF states, which are evaluated using classification metrics as shown in the Figure 2

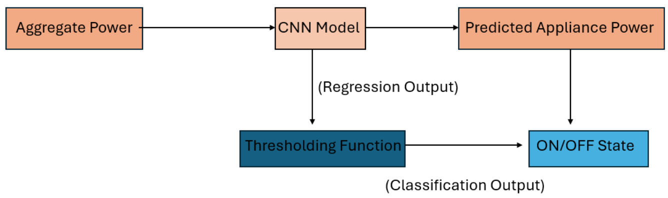

Figure 2. The Structure of Seq2point based CNN Model

D. Performance Evaluation

The study will use a combination of regression and classification metrics. The regression metrics like MAE and SAE can help to better access how the model can predict continuous appliance power consumption from the aggregated household signal, with lower values reflecting better disaggregation performance. Whereas classification metrics including Accuracy, Precision, Recall, and F1-score—are derived from the confusion matrix (TP, TN, FP, FN) can help access how good a model can predict from discrete information, for example on/off state of the appliances

IV. Results and Discussion

This section will throw light on the results. The results will be discussed on three different grounds based on the experiments conducted on the pre trained UK-DALE dataset and cross validation using REFIT dataset

A. Training Parameter Influence

This section outlines how different training parameters affected the performance of the Seq2Seq and Seq2Point models in NILM. A systematic parameter study was carried out to find the best configurations for higher accuracy and better generalization. Adjustments to parameters, such as window size, learning rate, and normalization technique had a clear influence on model behavior. For Seq2Point, using short windows, small batch sizes like for 128, the MAE is 12.80 for Kettle as compared to 13.82 for 1024 batch size. Similarly no dropout, and a learning rate of 0.001 consistently resulted in higher accuracy as MAE is less compared to larger windows and other larger layers. In Seq2Seq models, medium to long windows, larger Batch size large batch sizes like for 1024, the MAE is 13.65 for Kettle as compared to 14.09 for smaller 128 batch size. Similarly no dropout, and less learning rates produced clear gains for cyclic appliances like the fridge-freezer and dishwasher, improving sequence as shown in Table I. Overall, these findings show that the right parameter combinations can meaningfully increase model performance, reinforcing that targeted tuning yields stronger results than a uniform configuration. Among all the factors tested, window duration, learning rate, and normalization method had the greatest impact on performance, while dropout primarily helped reduce overfitting. These findings highlight the importance of careful hyperparameter tuning to achieve a good balance between model accuracy, training stability, and convergence speed.

B. Model Performance Across Different Appliance Load Patterns

For appliances that draw greater power and are used for small durations, such as the kettle and microwave, the Seq2Point model aligns more closely with the ground truth as shown in Figure 3. It captures the timing and magnitude of peaks with sharper transitions and a faster return to zero. In contrast, the Seq2Seq smooths these quick transitions and tends to underestimate peaks. Both models detect events accurately, but Seq2Point provides better precision hence more suitable for appliances with bursty on/off loads.

For appliances that show more regular and repeating behaviour, like fridge-freezer and dishwasher, the Seq2Point model is showing better results. As shown in the Figure 3 Seq2Seqmodel is still detecting most of the events, but its predictions are more fragmented lead to false positives occasionally.

The washer dryer, as clear from the Figure 3 , has a more complex pattern, where Seq2Seq is outperforming and producing continuous predictions across the operational period where as Seq2Points is shown struggling to keep accuracy throughout the operational period leading to certain false positives.

Therefore, the Seq2Point performs better with appliances that draws more power and are operational for small duration. For example the MAE for kettle is 13.25 for seq2seq and 11.12 for seq2point, showing a clear good results for seq2point as shown in Table II. While for wash dryer, Seq2Seq model perform better For example Seq2seq show an MAE of 8.43 while 13.22 for seq2point showing a victory for seq2seq model. These outcomes confirm the architectural strengths of each model: Seq2Seq is better for structured sequences hence maintains better continuity over longer operational periods whereas Seq2Point architecture helps it to excels in point-wise accuracy for detecting isolated events. This finding emphasizes the importance of appliance-specific model selection in NILM applications.

C. Comparison between Seq2Seq and Seq2Point

Threshold-based classification was used to convert the regression outputs into on/off states. Each appliance was assigned an activation threshold to avoid any spurious detections. Specifically, the thresholds were set as 1500W, 50W, 800W, 500W and 500W respectively for Kettle, Fridge-Freezer, Dishwasher, Microwave, Washer-Dryer. The experimental results indicate that the Seq2Point model generally outperforms Seq2Seq in detecting appliance on/off states, achieving higher F1-scores and better accuracy for devices with sharp, distinct power transitions as shown in Table II . Seq2Point excelled in identifying short, high-intensity bursts, such as those from the kettle (F1 = 0.787), fridge-freezer (F1 = 0.630), and dishwasher (F1 = 0.578), while also slightly outperforming Seq2Seq for the microwave (F1 = 0.274) despite the appliance's irregular usage pattern. However, for long and multi-phase operations like the washer-dryer, Seq2Seq performed better (F1 = 0.516) due to its ability to capture continuous temporal dependencies. Overall, Seq2Point proves more effective for appliances with short, bursty loads, whereas Seq2Seq remains advantageous for cyclic or extended activity patterns.

When comes to the regression metrics, Seq2Point typically achieves a 10–18% lower MAE and around 12–20% lower SAE, indicating more precise estimation of appliance power consumption.

D. Performance on the Holdout Set

Across all five appliances on the holdout set REFIT, Seq2Point models tends to outperform Seq2Seq models in most regression metrics as shown in Table III, although performance was notably reduced compared to evaluations conducted with the UK-DALE dataset. In terms of F1-score Seq2Point consistently achieves higher values for every appliance. For example, the kettle improves from 0.27 (Seq2Seq) to 0.37 (Seq2Point), and the washer-dryer shows a notable increase from 0.19 to 0.34. These gains are mainly driven by improvements in recall, indicating that Seq2Point is better at capturing true activation events. From a regression perspective,

ISBN: 978-1-68558-357-6

TABLE I. RESULTS OF MAE FOR SEQ2SEQ AND SEQ2POINT MODELS UNDER DIFFERENT PARAMETER SETTINGS.

	Window Duration		Batch Size		Dropout Ratio		Learning Ratio		Resampling	
	23 min	45 min	128	1024	0	0.25	0.0001	0.001	30s	45s
Seq2seq										
Kettle	14.14	13.33	14.09	13.65	13.59	14.08	14.14	13.59	13.41	13.50
Fridge-Freezer	26.55	26.30	26.57	26.17	25.81	27.39	25.36	25.81	26.87	26.54
Dish Washer	23.63	23.66	24.48	25.46	24.36	24.36	24.46	24.36	23.77	24.09
Microwave	9.62	10.26	10.33	11.00	11.41	11.27	10.80	11.41	9.56	8.56
Washer Dryer	20.97	19.09	18.24	17.00	17.91	17.81	15.15	17.91	12.69	14.49
Seq2point										
Kettle	13.52	13.93	12.80	13.82	13.44	13.90	14.36	13.44	13.11	13.30
Fridge-Freezer	27.11	25.64	24.63	25.05	25.68	25.14	25.93	25.68	26.66	26.61
Dish Washer	24.75	22.95	23.43	25.10	25.04	25.11	24.54	25.04	24.41	23.91
Microwave	10.44	11.91	10.97	9.96	12.07	11.50	14.15	12.07	9.42	10.81
Washer Dryer	23.20	18.67	17.77	18.89	15.60	18.01	20.40	15.60	13.45	13.68

TABLE II. RESULTS OF REGRESSION AND CLASSIFICATION METRICS FOR SEQ2SEQ AND SEQ2POINT MODELS APPLIED ON UK-DALE DATASET

Appliance	Seq2Seq						Seq2Point					
	Acc	Prec	F1	Rec	MAE	SAE	Acc	Prec	F1	Rec	MAE	SAE
Kettle	0.99	0.74	0.74	0.75	13.25	0.04	0.99	0.72	0.79	0.85	11.12	0.05
Fridge-Freezer	0.78	0.63	0.52	0.44	25.16	0.01	0.80	0.63	0.63	0.63	23.69	0.07
Dishwasher	0.98	0.57	0.52	0.51	23.68	0.15	0.98	0.56	0.57	0.59	23.49	0.01
Microwave	0.99	0.40	0.27	0.20	9.57	0.92	0.99	0.74	0.27	0.20	8.04	0.45
Washer-Dryer	0.99	0.51	0.52	0.52	8.43	0.18	0.99	0.21	0.33	0.73	13.22	0.81

TABLE III. COMBINED RESULTS OF REGRESSION AND CLASSIFICATION METRICS FOR SEQ2SEQ AND SEQ2POINT MODELS ON REFIT

Appliance	Seq2Seq						Seq2Point					
	Acc	Prec	F1	Rec	MAE	SAE	Acc	Prec	F1	Rec	MAE	SAE
Kettle	0.99	0.55	0.27	0.21	21.18	0.42	0.99	0.59	0.37	0.33	17.08	0.49
Fridge-Freezer	0.66	0.56	0.44	0.41	37.75	0.61	0.67	0.56	0.41	0.36	35.75	0.49
Dishwasher	0.98	0.59	0.24	0.18	39.17	0.68	0.98	0.57	0.32	0.26	39.01	0.63
Microwave	0.99	0.20	0.11	0.09	17.95	2.51	0.99	0.19	0.09	0.07	14.92	1.54
Washer-Dryer	0.98	0.26	0.19	0.17	29.22	0.80	0.99	0.31	0.34	0.41	29.52	0.94

Seq2Point again demonstrates stronger performance, achieving lower MAE and SAE scores across nearly all appliances. For instance, the MAE for the kettle drops from 21.18 W (Seq2Seq) to 17.08 W (Seq2Point), and the microwave improves from 17.95 W to 14.92 W. Similarly, SAE decreases for every appliance, showing that Seq2Point better captures total energy consumption over time. The decline in overall accuracy emphasizes the impact of dataset-specific variations in appliance usage patterns. Hence the results from REFIT revealed that the model did not perform as good as it did on UK-DALE, highlighting the challenges of cross-dataset generalization. The reduced performance on REFIT can be attributed to its lower sampling frequency, higher household and appliance variability, and increased aggregate noise compared to UK-DALE, which obscure short-duration appliance activations.

V. CONCLUSION AND FUTURE WORK

This study utilizes the CNN based Seq2Seq and Seq2Point architectures to effectively disaggregate the aggregate power of household to appliance level. The Seq2Point model outperformed at detecting short, high-power events, while Seq2Seq proved itself better in case of consistent loads. By optimizing key parameters, such as window duration, normalization, and learning rate, the research achieved substantial improvements in model accuracy. It is important to highlight this fact here

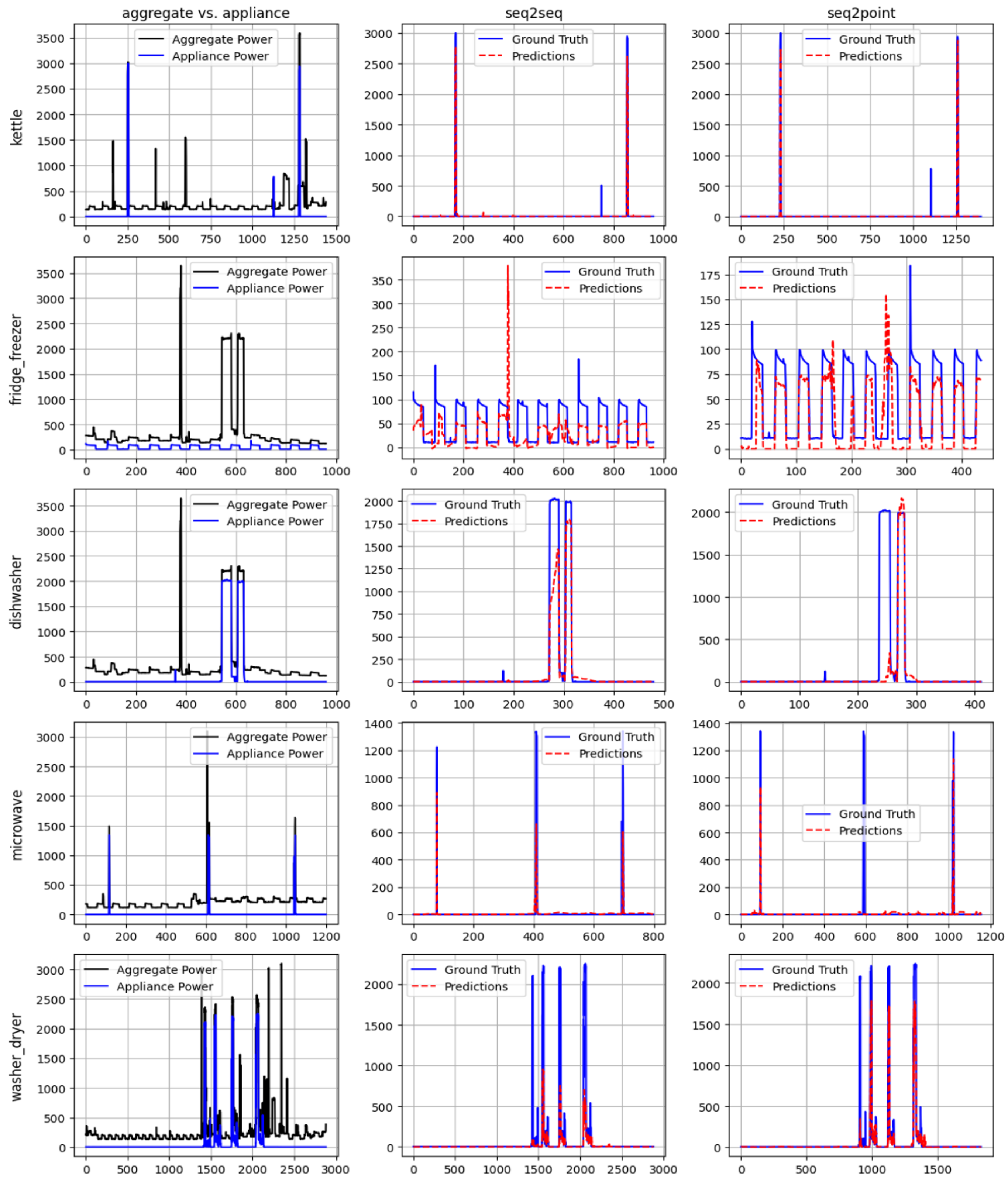

Figure 3. The disaggregation results for the appliances- kettle, fridge-freezer, dishwasher, microwave, and washer-dryer- across the five households of UK-DALE. Each row represents one appliance, with three columns, the first column represents the aggregate power versus the actual appliance consumption while the second and the third columns depicts the predicted power from Seq2Seq and Seq2Point models, respectively. The predicted values (Red) and ground truth (Blue) are compared simultaneously.

that although both Seq2Seq and Seq2Point models are trained purely as regression models that output continuous appliance power values, this study additionally evaluates their ability to perform appliance state detection. For the purpose, thresholding function is used to convert the continuous output to the ON/OFF states. While both models perform well on the UK-DALE dataset, this does not generalize uniformly on the REFIT dataset. Overall, Seq2Point outperforms Seq2Seq across both regression (MAE/SAE) and classification (F1/recall) metrics.

The proposed models and optimization framework advance the development of sustainable smart household power management systems. They also provide practical insights for real-world deployment in smart homes and IoT-enabled energy networks.

Despite these promising results, some limitations remain. Both models face challenges with low-power or overlapping appliance signals. Future work should explore hybrid CNN-transformer architectures, noise-robust preprocessing, and multi-dataset training to enhance generalization. Integrating optimized NILM models into edge devices and IoT systems will be a crucial step toward achieving real-time, scalable, and sustainable smart grid solutions.

References

[1] "Iea: Growth in global energy demand surged in 2024 to almost twice its recent average", Energy Connects, 2025, [Online]. Available: https://www.energyconnects.com/opinion/features/2025/march/iea-growth-in-global-energy-demand-surged-in-2024-to-almost-twice-its-recent-average/.

[2] "Load frequency regulation of interconnected multi-source multi-area power system with penetration of electric vehicles aggregator model", [Online]. Available: https://link.springer.com/article/10.1007/s00202-023-01923-2.

[3] S. Shams Amiri, M. Mueller, and S. Hoque, "Investigating the application of a commercial and residential energy consumption prediction model for urban planning scenarios with machine learning and shapley additive explanation methods", *Energy and Buildings*, vol. 287, p. 112 965, 2023. DOI: 10.1016/j.enbuild.2023.112965.

[4] A. Chatterjee and P. Heer, *Non-intrusive load monitoring (nilm) with very low-frequency data from smart meters in switzerland*, SSRN, 2025. DOI: 10.2139/ssrn.5113439.

[5] Y. Liu, Y. Wang, and J. Ma, "Non-intrusive load monitoring in smart grids: A comprehensive review", *arXiv*, Mar. 2024. DOI: 10.48550/arXiv.2403.06474.

[6] S. Zhang *et al.*, "Sequence-to-point learning based on spatio-temporal attention fusion network for non-intrusive load monitoring", *Complex Intelligent Systems*, vol. 11, no. 3, p. 171, Feb. 2025. DOI: 10.1007/s40747-025-01803-1.

[7] G. W. Hart, "Nonintrusive appliance load monitoring", *Proceedings of the IEEE*, 1992.

[8] G. Bucci, F. Ciancetta, E. Fiorucci, S. Mari, and A. Fioravanti, "State of art overview of non-intrusive load monitoring applications in smart grids", *Measurement: Sensors*, vol. 18, p. 100 145, Dec. 2021. DOI: 10.1016/j.measen.2021.100145.

[9] H. Ren, X. Su, R. Jenssen, J. Li, and S. N. Anfinsen, "Attention-guided temporal convolutional network for non-intrusive load monitoring", *2022 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, pp. 419–425, 2022. DOI: 10.1109/SmartGridComm52983.2022.9960976.

[10] P. G. Papageorgiou, G. C. Christoforidis, and A. S. Bouhouras, "Nilm in high frequency domain: A critical review on recent trends and practical challenges", *Renewable and Sustainable Energy Reviews*, vol. 213, p. 115 497, May 2025. DOI: 10.1016/j.rser.2025.115497.

[11] K. Oublal, "Temporal attention bottleneck is informative? interpretability through disentangled generative representations for energy time series disaggregation", *ICML 2023 Workshop on Deployment Challenges for Generative AI*, 2023.

[12] D. Precioso Garcelán and D. Gomez-Ullate, "Nilm as a regression versus classification problem: The importance of thresholding", *J Supercomput 79*, 2020.

[13] J. He *et al.*, "Msdc: Exploiting multi-state power consumption in non-intrusive load monitoring based on a dual-cnn model", in *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence (AAAI'23)*, vol. 37, 2023, pp. 5078–5086. DOI: 10.1609/aaai.v37i4.25636.

[14] Z. Jia, L. Yang, Z. Zhang, H. Liu, and F. Kong, "Sequence to point learning based on bidirectional dilated residual network for non-intrusive load monitoring", *International Journal of Electrical Power & Energy Systems*, vol. 129, p. 106 837, Jul. 2021. DOI: 10.1016/j.ijepes.2021.106837.

[15] G. Zhou *et al.*, *Sequence-to-sequence load disaggregation using multi-scale residual neural network*, 2020. DOI: 10.48550/arXiv.2009.12355.

[16] Z. Solatidehkordi, J. Ramesh, A. R. Al-Ali, A. Osman, and M. Shaaban, "An iot deep learning-based home appliances management and classification system", *Energy Reports*, vol. 9, pp. 503–509, May 2023. DOI: 10.1016/j.egyr.2023.01.071.

[17] "Time series classification website", [Online]. Available: https://www.timeseriesclassification.com/description.php?Dataset=PLAID.

[18] J. Jiang, Q. Kong, M. Plumbley, M. Hoogendoorn, and D. Roijers, "Deep learning-based energy disaggregation and on/off detection of household appliances", *ACM Transactions on Knowledge Discovery from Data*, vol. 15, pp. 1–21, Apr. 2021. DOI: 10.1145/3441300.

[19] J. Wang and K. A. Loparo, *Evolutionary deep nets for non-intrusive load monitoring*, 2023. DOI: 10.13140/RG.2.2.25983.28324.

[20] C. Zhang, M. Zhong, Z. Wang, N. Goddard, and C. Sutton, "Sequence-to-point learning with neural networks for non-intrusive load monitoring", in *Proc. AAAI Conf. Artif. Intell.*, vol. 32, Apr. 2018. DOI: 10.1609/aaai.v32i1.11873.

[21] J. Kelly, "Uk domestic appliance-level electricity (uk-dale) dataset", [Online]. Available: http://jack-kelly.com/data/.

[22] "Refit: Electrical load measurements (cleaned)", University of Strathclyde, [Online]. Available: https://pureportal.strath.ac.uk/en/datasets/refit-electrical-load-measurements-cleaned/.