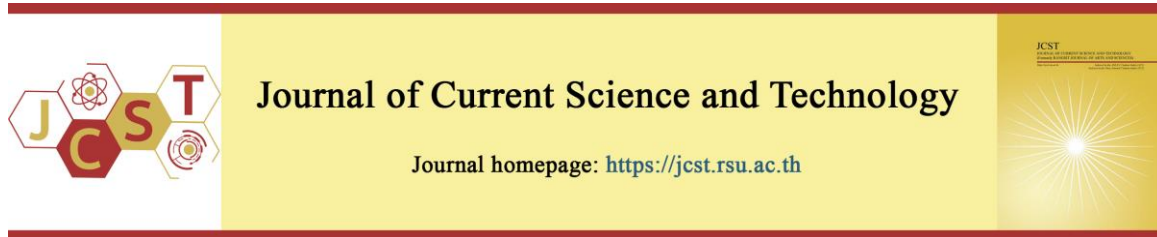


Cite this article: Chanhom, P., & Tangchittichariya, J. (2026). Performance evaluation of S2P-CNN for multi-load disaggregation of EV and PV generation under varying data sampling rates. *Journal of Current Science and Technology*, 16(4), Article 219. <https://doi.org/10.59796/jcst.V16N4.2026.219>



Performance Evaluation of S2P-CNN for Multi-Load Disaggregation of EV and PV Generation under Varying Data Sampling Rates

Peerapon Chanhom^{1,*}, 
Jiranat Tangchittichariya², 

¹Division of Electrical Engineering, Faculty of Engineering and Architecture, Rajamangala University of Technology
Suvarnabhumi, Nonthaburi 11000, Thailand

²Smart Grid Division, System Planning Department Provincial Electricity Authority of Thailand, Bangkok 10900, Thailand

*Corresponding author; E-mail: peerapon.c@rmutsb.ac.th

Received 30 December 2025; Revised 22 March 2026; Accepted 16 June 2026; Published 15 September 2026

Abstract

This study evaluates the performance of a multi-output regression Sequence-to-Point Convolutional Neural Network (S2P-CNN) for Non-Intrusive Load Monitoring (NILM), with a specific focus on disaggregating Electric Vehicle (EV) charging load and Photovoltaic (PV) generation under varying data sampling rates. While the growing integration of Distributed Energy Resources (DERs) has increased the complexity of household energy patterns, most utility smart meters still record data at low resolutions (e.g., 15 minutes), presenting a significant challenge for traditional algorithms. To address this issue, experiments were conducted utilizing the real-world UK-DALE dataset solely as a background load, onto which synthetic EV and PV profiles were superimposed at resolutions of 1, 5, 10, and 15 minutes. To ensure statistical robustness, all evaluations were validated across multiple random seeds and compared against a standard Long Short-Term Memory (LSTM) sequence-modeling baseline. The quantitative results reveal a clear engineering trade-off: while lowering the sampling rate degrades point-wise signal shape accuracy due to the loss of high-frequency transients, the model successfully preserves the total energy mass. Crucially, multi-seed validation highlights a specific sensitivity at certain resolutions (e.g., 10 minutes), causing feature extraction to be highly dependent on initial weights. However, at the primary target resolution of 15 minutes, the model regains stability and demonstrates highly consistent behavior, maintaining Signal Aggregate Error (SAE) values safely below the 10% acceptable threshold. Furthermore, baseline evaluations indicate that the S2P-CNN architecture offers greater robustness and computational efficiency under coarse sampling conditions compared to the LSTM. These findings demonstrate that the S2P-CNN is a viable solution for aggregate-level energy estimation using existing Advanced Metering Infrastructure (AMI), offering a practical and cost-effective deployment pathway for smart grid management without necessitating immediate hardware upgrades.

Keywords: *Non-Intrusive Load Monitoring (NILM), S2P-CNN, Electric Vehicle (EV), Photovoltaic (PV), sampling rate, smart meter*

1. Introduction

Rising energy demand and growing environmental awareness have accelerated the transition of traditional power systems into Smart Grids. A key element of this transformation is the intelligent integration of Distributed Energy Resources (DERs) into the power network (Al-Shetwi, 2022; Fang et al., 2012). Within the context of Home Energy

Management Systems (HEMS), the widespread adoption of Electric Vehicles (EVs) and Photovoltaic (PV) systems has introduced significant complexity and fluctuation to household load profiles. To effectively manage these dynamics and maintain system balance, access to granular, device-level energy consumption data is essential (Himeur et al., 2021; Zoha et al., 2012).

Non-Intrusive Load Monitoring (NILM) provides an efficient method for identifying individual appliance usage based on aggregated signals from a single metering point (Hart, 1992). The landscape of NILM has evolved significantly from traditional statistical models, such as Hidden Markov Models (HMMs) (Zeifman & Roth, 2011), to advanced deep learning architectures. As comprehensively reviewed by Huber et al. (2021), while Recurrent Neural Networks (RNNs) and LSTMs were initially popular for time-series data, they struggle with high computational costs and vanishing gradients when processing long sequences. In contrast, Convolutional Neural Networks (CNNs), particularly the Sequence-to-Point (S2P-CNN) architecture proposed by Zhang et al. (2018), have emerged as a robust benchmark due to their strong balance of feature extraction capabilities and computational efficiency (Jaime et al., 2026). Recently, attention-based architectures and Transformers have gained significant traction in NILM research (Sykiotis et al., 2022; Dash & Sahoo, 2023). While Transformers excel at capturing long-range dependencies, their self-attention mechanisms require quadratic computational complexity and massive datasets. For HEMS operating on local, computationally constrained smart grid infrastructure, deploying massive Transformer models remains impractical. Therefore, S2P-CNN is selected in this study as it provides a practical trade-off, offering robust local temporal pattern recognition with significantly lower computational overhead, making it highly suitable for sliding-window inference on edge devices.

Recently, deep learning techniques, particularly CNNs, have demonstrated remarkable capabilities in NILM for complex high-power loads. However, a significant gap exists between academic research and utility-scale implementation. Most existing studies utilizing S2P-CNN evaluate performance strictly on high-frequency data (e.g., 1 Hz to 1-minute intervals) to maximize accuracy. The assumption that these architectures will seamlessly generalize to low-resolution data is highly problematic. Currently, large-scale deployment of Advanced Metering Infrastructure (AMI) typically operates at a 15-minute data resolution. This interval is technically defined and governed by smart grid communication standards such as IEEE2030.5-2018 (IEEE Standards Association, 2018) and IEC 62056-1-0 (International Electrotechnical Commission (IEC), 2020), primarily due to practical constraints in communication bandwidth and data storage (Völker et al., 2021; Yin et al., 2023).

DERs under such low-frequency conditions presents unique challenges as it severely limits the available shape features. Recent empirical studies have explicitly demonstrated that deep learning architectures suffer severe performance degradation at these coarser resolutions (Todic et al., 2023). The foundational work by Vavouris et al. (2022) explicitly tackled the low-frequency NILM problem for EVs and PVs. They demonstrated that while low-frequency disaggregation (e.g., 15 minutes) is highly challenging due to the loss of high-frequency switching transients, it is physically possible to extract load trends. However, Vavouris et al. (2022) primarily focused on training individual, separate models for each specific appliance. Furthermore, while prior multi-appliance NILM literature has explored multi-task learning, it typically focuses on predicting appliance ON/OFF states rather than continuous power estimation. Simultaneous multi-output continuous regression of EV charging and PV generation under overlapping operating conditions from a single aggregate meter is rarely addressed. This study critically extends the foundational baseline set by Vavouris et al. (2022). Instead of single-target models like those used by Nizami et al. (2020), we propose a simultaneous multi-output S2P-CNN. This approach explicitly models the complex interactions and reverse power flows between EV charging and PV generation simultaneously, offering a more holistic and system-level approach to low-frequency AMI data analysis.

The key contributions of this study are explicitly summarized as follows:

- 1) Proposed a multi-output S2P-CNN framework capable of simultaneously disaggregating both EV charging load and PV generation from a single low-resolution aggregate meter.
- 2) Conducted a systematic performance evaluation across multiple sampling rates with multi-seed statistical validation.
- 3) Empirically demonstrated that a 15-minute AMI data resolution achieves an SAE of less than 10%, supporting cost-effective deployment.
- 4) Highlighted the computational efficiency of the proposed S2P-CNN, demonstrating significantly faster training convergence compared to the LSTM baseline, which is highly advantageous for practical applications

2. Objectives

The primary objective of this study is to evaluate the performance of a simultaneous multi-output S2P-CNN framework across various sampling rates (1, 5, 10, and 15 minutes). Specifically, this research aims to determine the feasibility of achieving a Signal Aggregate Error (SAE) below 10% for both EV charging load and PV generation at a 15-minute resolution.

It is hypothesized that decreasing the temporal resolution below 10 minutes introduces diminishing returns in shape reconstruction accuracy due to the loss of high-frequency load signatures, whereas the energy estimation accuracy (SAE) remains relatively stable until a critical 15-minute threshold.

3. Methods

This study presents a CNN-based framework for multi-load disaggregation, with a particular emphasis on evaluating performance across varying temporal resolutions. As illustrated in the research workflow, the methodology comprises four distinct stages: (1) Data Preparation and Synthetic Load Generation, (2) Multi-Resolution Data Processing, (3) Proposed S2P-CNN Architecture, and (4) Performance Evaluation Metrics. The specific details of each stage are outlined below.

3.1 Data Preparation and Synthetic Load Generation

This study employs real-world data from House 1 of the UK-DALE dataset (Kelly & Knottenbelt, 2015). This specific household was deliberately selected for its extensive duration and high data quality, serving as a controlled baseline to isolate the effects of data sampling rates without the confounding factor of inter-household variability. To rigorously evaluate the S2P-CNN model under challenging Low Signal-to-Noise Ratio (SNR) conditions, the background base load was amplified by a factor of 2. This specific multiplier was chosen to intentionally degrade the SNR, creating a "worst-case" stress-test scenario. It pushes the background power magnitude closer to the target EV and PV peaks without completely drowning the target signals. This ensures the model is tested against rigorous, noise-heavy backgrounds rather than idealized low-consumption conditions. Following this base load preparation, target EV and PV load profiles were synthetically generated and superimposed onto the background.

1) EV Charging Model: This model is designed to simulate realistic residential charging patterns. To

rigorously test the model's sensitivity, the peak power is limited to 1,750 W (approximately 8A), representing a Level 1 slow or portable charging scenario. While higher-power Level 2 charging (3.3–7.2 kW) is increasingly prevalent, Level 1 was deliberately selected because its lower power signature is easily obscured by common background household appliances. Successfully disaggregating a Level 1 signature demonstrates the algorithm's lower-bound detection capability, implying that the model possesses the inherent robustness to capture the more prominent and massive power spikes of Level 2 charging. Instead of assuming a constant rectangular load, the charging profile incorporates an exponential decay phase toward the end of the session to accurately reflect real-world battery charging characteristics. Furthermore, charging sessions are randomly scheduled between 23:00 and 05:00 (aligning with typical Time-of-Use off-peak periods) with durations ranging from 4 to 6 hours.

2) PV Generation Model: The solar generation profile is simulated using a Gaussian function, producing a bell curve that corresponds to active daylight hours (09:00 to 17:00). The peak power output is capped at 1,000 W. This synthetic generation pattern aligns with established performance models for small-scale residential rooftop systems, consistent with the profiles generated by the National Renewable Energy Laboratory (NREL) PVWatts calculator (Dobos, 2014). Additionally, to evaluate the model's robustness against weather-induced variability, a simulated cloud cover effect is introduced by injecting random noise fluctuations into the ideal generation curve.

Finally, the synthetic target signals are combined with the amplified base load to generate the Augmented Aggregate Load ($P_{agg}(t)$) for each temporal resolution. Residential PV generation reduces net power drawn from the grid. To reflect real-world net-metering dynamics, $P_{PV}(t)$ is subtracted from total consumption. Additionally, a constant baseline offset C is applied to prevent negative aggregate loads. This avoids numerical instability during data normalization and model training by ensuring a strictly non-negative input domain. This process is formally defined in (1), and a representative example is illustrated in Figure 1.

$$P_{agg}(t) = P_{base}(t) + P_{EV}(t) - P_{PV}(t) + C \quad (1)$$

Where:

$P_{agg}(t)$ denotes the augmented aggregate load power at time t .

$P_{base}(t)$ denotes the scaled base load power derived from the dataset.

$P_{EV}(t)$ denotes the EV charging load power.

$P_{PV}(t)$ denotes the PV power generation.

C denotes the constant baseline offset applied to maintain a non-negative aggregate load.

t denotes the time-instance considered.

3.2 Multi-Resolution Data Processing

To ensure the data is suitable for Deep Learning and supports the comparative resolution analysis, the raw data undergoes the following preprocessing stages. Initially, missing data points within the raw dataset were addressed using linear interpolation to ensure temporal continuity before any further transformations.

1) Data Resampling: To represent various metering standards, the original time-series data is down-sampled using an averaging method. This process generates four distinct datasets with sampling intervals of 1, 5, 10, and 15 minutes.

2) Normalization: Given the high variance and wide range of electrical power values (x_t), Min-Max Scaling is applied to normalize the data into a $[0, 1]$ range. This step facilitates faster model convergence and improves training stability, as defined in (2). Crucially, to prevent data leakage, the x_{max} and x_{min} parameters are computed strictly from the training subset.

$$x'_t = \frac{x_t - x_{min}}{x_{max} - x_{min}} \quad (2)$$

3) Sliding Window Segmentation: The time-series data is segmented using sliding windows of a fixed length of 60 data points across all sampling resolutions (1, 5, 10, and 15 minutes). Although this causes the physical time span to vary (from 1 hour at a 1-minute resolution to 15 hours at a 15-minute resolution), it is necessary to prevent model complexity from confounding the results. In CNN-based architectures like the S2P framework (Zhang et al., 2018), altering the input sequence length fundamentally changes the network's architecture and the total number of trainable parameters. Specifically, while convolutional layers can process variable-length sequences, a longer input sequence produces a larger flattened vector, which directly expands the

dimensions of the weight matrix and increases the number of trainable parameters. For instance, a fixed 15-hour span would force the 1-minute model to process 900 points while the 15-minute model processes only 60, creating an unfair comparison between models of vastly different sizes. Furthermore, this approach fits the physical nature of the signals, where high-resolution data contains sharp details requiring only a short time frame, whereas coarse data relies on broader trends that need a longer context. Thus, maintaining a constant 60-point window naturally adapts to each resolution's information density, ensuring a fair evaluation by keeping the model structure completely identical.

3.3 Proposed S2P-CNN Architecture

This study proposes a modified S2P-CNN architecture tailored for Multi-Output Regression, as shown in Figure 2. The model takes a time-windowed sequence of the aggregate load (x'_t) as input and simultaneously predicts the EV charging load and PV generation at the window's midpoint. The architecture comprises three core processing stages:

1) Feature Extraction Layers: A stack of five convolutional layers is used to hierarchically extract temporal features, ranging from low-level patterns (e.g., edges, on/off transitions) to high-level representations (e.g., charging curves and solar profiles). The network structure, specifically the filter depth sequence (30, 30, 40, 50, 50), adopts the benchmark S2P-CNN architecture (Zhang et al., 2018). However, unlike the original variable kernel sizes, we standardized the kernel size to 5 across all layers. This modification was empirically selected to optimize the receptive field for the lower-resolution data (1–15 minutes) used in this study. Specifically, a kernel size of 5 provides an optimal balance; it ensures the effective capture of appliance state transitions without over-smoothing, while simultaneously maintaining computational efficiency for potential real-time edge-device deployments. A ReLU (Rectified Linear Unit) activation function follows each convolution to model non-linear relationships.

2) Fully Connected and Regularization Layers: The extracted feature maps are flattened into a 1D vector and fed into a dense layer containing 1,024 units. While this exceeds the original S2P-CNN configuration (Zhang et al., 2018), this capacity was empirically increased during the preliminary tuning phase. This higher dimensionality was found necessary to sufficiently capture the complex, overlapping features of the simultaneous multi-output task (EV

and PV), which requires a larger representational space than single-appliance disaggregation. To prevent potential overfitting, particularly when processing high-resolution (1-minute) data, a dropout layer with a rate of 0.5 was implemented immediately after the dense layer. This ensures the model maintains robust generalization capabilities across all sampling rates investigated.

3) Multi-Output Layer: To independently capture the distinct characteristics of the two target loads, the network branches into two separate output layers. The first layer predicts EV power consumption, while the second predicts PV generation. Both layers use a linear activation function to provide continuous power value predictions, allowing the model to handle overlapping signals and reverse power flows simultaneously.

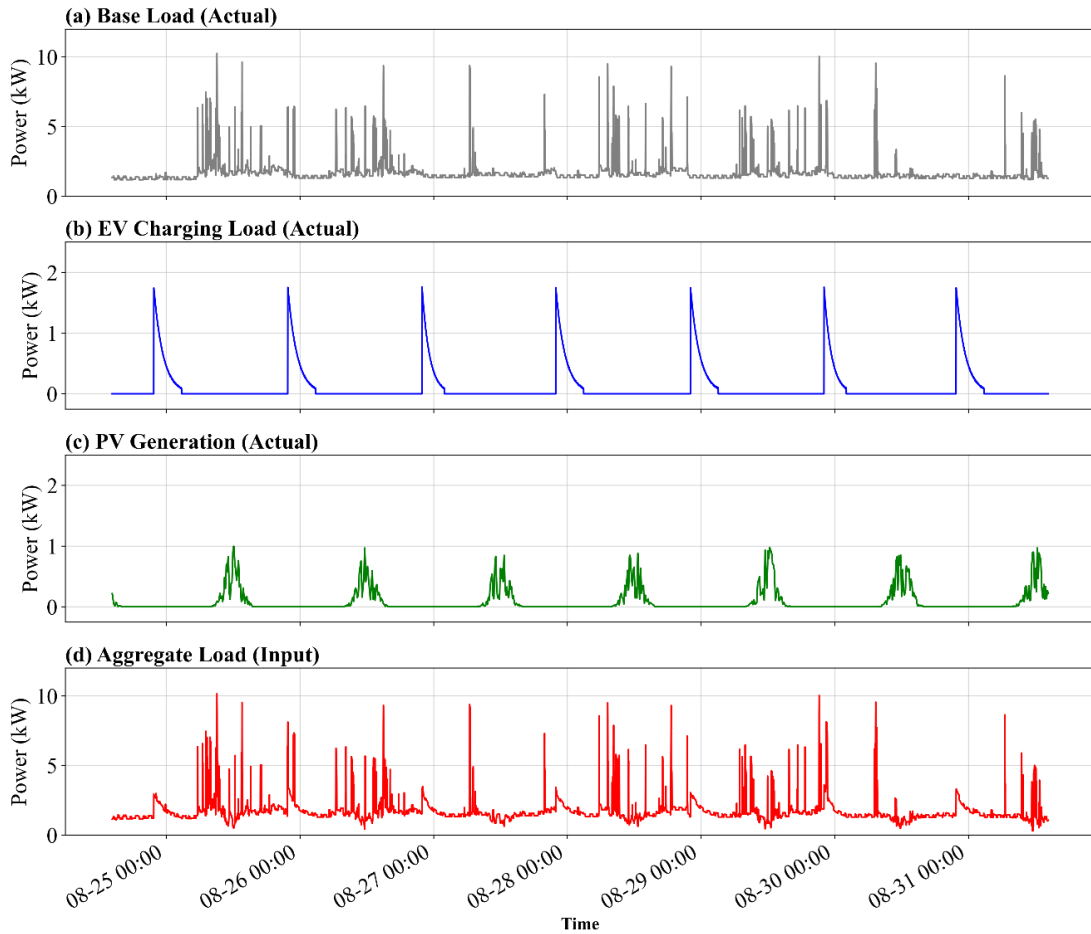


Figure 1 Augmented aggregate power profile with superimposed EV charging load and PV generation

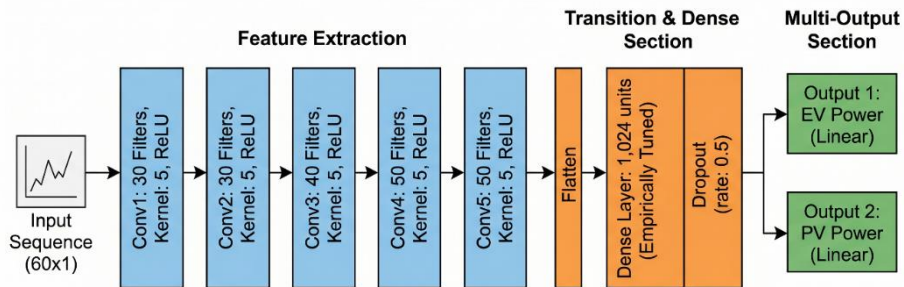


Figure 2 S2P-CNN model architecture diagram

3.4 Performance Evaluation Metrics

To comprehensively evaluate the disaggregation performance, model predictions are compared against the ground truth utilizing three quantitative metrics. These metrics are specifically selected to assess both point-wise shape reconstruction accuracy (Mean Absolute Error and Root Mean Square Error) and overall energy estimation capability (Signal Aggregation Error). It is important to note that the reported performance metrics are derived from the model checkpoint achieving the lowest validation loss in each respective run. Furthermore, to address the inherent stochastic variance of deep learning training processes and ensure robust statistical reliability, all experiments were rigorously conducted across five independent training runs utilizing different random weight initializations (seeds). Consequently, the quantitative results presented in this study are strictly reported as the Mean $\pm$ Standard Deviation (SD).

1) Mean Absolute Error (MAE): This metric quantifies the average magnitude of power estimation errors (in Watts), as defined in (3).

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (3)$$

2) Root Mean Square Error (RMSE): This metric calculates the square root of the average squared errors (in Watts). It is particularly useful for penalizing large deviations, such as peak prediction errors, as described in (4).

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (4)$$

3) Signal Aggregate Error (SAE): This metric measures the relative error of the total accumulated energy over a given period. As defined in (5), SAE is critical for assessing the model's feasibility for energy billing and cumulative monitoring applications.

$$\text{SAE} = \frac{|\hat{E} - E|}{E} \quad (5)$$

Where:

y_i denotes the ground truth power at time i .

$\hat{y}_i$ denotes the predicted power at time i .

E denotes the total ground truth energy.

$\hat{E}$ denotes the total predicted energy.

N denotes the total number of samples.

4. Results

This section evaluates the performance of the proposed S2P-CNN architecture, specifically focusing on how temporal resolution impacts the disaggregation accuracy of EV charging load and PV generation. The analysis is organized into three main subsections. First, the Experimental Setup details the training parameters and computational environment. Next, the Impact of Sampling Rates presents a quantitative comparison of results across resolutions of 1, 5, 10, and 15 minutes. Finally, the Feasibility Analysis examines the viability of using standard 15-minute metering data for real-world energy management applications.

4.1 Experimental Setup

All experiments were implemented in a local computing environment featuring an Intel Core i5-8350U CPU @ 1.70 GHz (8 CPUs) with 8 GB of RAM, operating on Windows 10 Pro 64-bit. The software stack utilized Jupyter Notebook, Python, TensorFlow, and Keras. To provide a standardized measure of computational cost across different data resolutions, the training speed is reported in milliseconds per step. On average, the model training required approximately 100 ms/step. To rigorously evaluate the model's stability, reproducibility, and inherent variance, a comprehensive multi-seed validation approach was employed. All experiments were systematically executed using five independent random seeds (42, 123, 415, 707, 999) to initialize weights across all relevant libraries (Random, NumPy, TensorFlow). While deep learning variance analysis in NILM is often limited to single-seed reporting, utilizing five independent runs provides a statistically robust baseline to characterize the model's true performance distribution. This multi-seed approach is strategically designed to isolate the variance caused by weight initialization instability from the variance caused by the inherent sparsity and high-power volatility of the physical load profiles (e.g., infrequent EV charging events at coarser resolutions). Furthermore, to provide a rigorous benchmark, a baseline LSTM architecture was implemented and evaluated under identical conditions (data splits, window size, and random seeds) to directly compare disaggregation accuracy and computational efficiency against the proposed S2P-CNN.

For data partitioning, the dataset was split into a Training Set (80%) and a Testing Set (20%). Within the training phase, 20% of the training data was

further reserved as a Validation Set to monitor performance. The model was trained using the Adam optimizer with Mean Squared Error (MSE) as the loss function. Additionally, an Early Stopping strategy with a patience of 5 epochs was employed to monitor validation loss. This mechanism automatically halts training when performance plateaus, effectively preventing overfitting and ensuring the retention of the optimal model weights. Further parameters details are provided in Table 1.

Table 1 Parameter configuration

Parameter	Value
Window Size	60
Batch Size	256
Max Epochs	75
Early Stopping Patience	5
Optimizer	Adam (Learning Rate 0.001)
Loss Function	MSE

Following the training phase, the optimal model from each independent run is evaluated on the Testing Set using standard statistical metrics (MAE and RMSE) to assess the accuracy of load signal reconstruction. Crucially, the study also incorporates the SAE metric to quantify total energy estimation error. To accurately reflect the multi-seed validation, the final quantitative results are reported as the average and standard deviation (Mean $\pm$ SD). This rigorous evaluation metric is essential for analyzing the feasibility of implementing varying data resolutions in real-world utility metering systems.

4.2 Experimental Results and Comparative Analysis

This section evaluates the S2P-CNN disaggregation performance for EV charging load and PV generation across four distinct sampling rates: 1, 5, 10, and 15 minutes. The analysis is divided into two parts: a qualitative visual inspection of signal

reconstruction and a quantitative evaluation that includes rigorous statistical significance testing across multiple random seeds.

4.2.1 Quantitative Disaggregation Results and Multi-Seed Variance Analysis: Table 2 and the corresponding multi-seed variance distributions present the aggregated performance metrics (MAE, RMSE, and SAE) for EV charging load and PV generation. Reporting the variance across five independent random seeds provides a more comprehensive perspective on how temporal resolution and weight initialization interact. To objectively validate these comparative trends, a rigorous non-parametric statistical analysis was conducted, with significant differences denoted by superscript letters in Table 2. For EV charging, the error metrics exhibit divergent trends depending on the evaluation criteria. Specifically, across the 1, 5, 10, and 15-minute resolutions, the shape reconstruction errors (MAE and RMSE) show a generally linear increase; the MAE rises from 14.60 ± 1.76 W to 16.44 ± 3.24 W, 17.79 ± 3.24 W, and 18.89 ± 3.12 W, respectively. The RMSE follows a similar trajectory, increasing from 47.18 ± 1.79 W at 1-minute to 60.50 ± 6.95 W at 15-minutes. Conversely, the total energy estimation error (SAE) appears to be inversely affected, often remaining relatively stable or even decreasing at coarser resolutions (resulting in $4.35 \pm 1.60\%$, $3.53 \pm 3.01\%$, $3.64 \pm 3.56\%$, and $2.37 \pm 1.53\%$, respectively). Crucially, the statistical evaluation corroborates this divergence: while the structural degradation observed in the EV RMSE is statistically significant (transitioning from superscript 'a' at 1-minute to 'b' at 15-minutes), the SAE remains statistically indistinguishable across all resolutions (sharing the superscript 'a'). The rise in MAE and RMSE is likely due to the progressive blurring of the EV's sharp, step-like transient edges caused by the down-sampling effect.

Table 2 Quantitative disaggregation results for EV charging load and PV generation under varying sampling rates (Mean $\pm$ SD over 5 independent seeds)

Sampling Rate	EV MAE (W)	EV RMSE (W)	EV SAE (%)	PV MAE (W)	PV RMSE (W)	PV SAE (%)
1min	14.60 ± 1.76^a	47.18 ± 1.79^a	4.35 ± 1.60^a	29.38 ± 1.36^a	65.51 ± 0.69^a	3.87 ± 2.94^a
5min	16.44 ± 3.24^a	52.85 ± 4.16^{ab}	3.53 ± 3.01^a	31.65 ± 3.18^a	71.60 ± 2.67^b	5.78 ± 3.93^a
10min	17.79 ± 3.24^a	56.40 ± 6.31^b	3.64 ± 3.56^a	33.74 ± 6.38^a	70.61 ± 3.18^b	9.63 ± 9.61^a
15min	18.89 ± 3.12^a	60.50 ± 6.95^b	2.37 ± 1.53^a	29.38 ± 1.75^a	69.34 ± 2.81^{ab}	1.92 ± 2.01^a

Note: Values within the same column sharing the same superscript letter are not significantly different, whereas different letters indicate statistically significant differences (Overall effect assessed via Kruskal-Wallis test; pairwise comparisons conducted via Mann-Whitney U test with Bonferroni correction at $p < 0.0083$).

However, this heavy averaging process inherently conserves the overall area under the power curve. Consequently, while the point-by-point shape matching deteriorates, the model's ability to estimate the integrated total energy (SAE) tends to improve or stabilize, suggesting that coarser resolutions might be sufficient if overall energy metering is the primary objective rather than precise profile reconstruction.

Conversely, the PV generation metrics display a non-linear, inverted U-shaped behavior. For instance, the PV SAE progresses from $3.87 \pm 2.94\%$ (1-minute) and $5.78 \pm 3.93\%$ (5-minute) to a peak of $9.63 \pm 9.61\%$ (10-minute), before dropping significantly to $1.92 \pm 2.01\%$ (15-minute). The MAE reflects a peak at the 10-minute resolution (33.74 ± 6.38 W), while the RMSE reaches its maximum at the 5-minute resolution (71.60 ± 2.67 W) and remains comparably high at 10 minutes (70.61 ± 3.18 W), before recovering at 15-minutes. Notably, as the high standard deviation at 10-minutes extends the upper error limit to approximately 19.24%, this resolution cannot reliably satisfy the acceptable 10% SAE threshold for PV. This trend may be related to the fixed input window size of 60 data points employed across all sampling rates. At the 1-minute resolution, the window covers merely one hour, allowing the model to process localized linear segments with relatively low error. As the resolution decreases to 5 and 10 minutes, the time span expands to 5 and 10 hours, respectively. These windows encompass the steep transitional edges of the solar curve during sunrise and sunset. The high intra-window variance during these phases appears to introduce instability for the midpoint prediction, causing the shape error metrics to rise and directly accounting for the SAE threshold failure. When the resolution is further reduced to 15 minutes, the 60-point window spans 15 hours, covering the entire daytime generation cycle. At this macroscopic scale, the low-pass filtering effect of the 15-minute down-sampling may help smooth out high-frequency fluctuations, potentially allowing the network to capture the underlying Gaussian trend more effectively, causing the average error to drop again. Despite the threshold failure caused by the variance spike at 10-minutes, the overall statistical grouping (superscript 'a') for PV SAE across all sampling rates further underscores that the fundamental energy conservation property of the model remains robust.

As illustrated in Figure 3, while the average SAE for EV loads decrease linearly (as summarized in Table 2), the multi-seed variance of the SAE exhibits an

inverted U-shaped pattern for both EV charging load and PV generation. This variance behavior appears to be connected to the fixed input window size of 60 data points. At the 1-minute resolution, the window covers a narrow time span of 1 hour, providing clearer gradients that generally allow the seeds to converge to similar solutions with lower variance. When the resolution drops to 5 and 10 minutes, the window expands to cover 5 and 10 hours, which may include blurred transitional edges. This ambiguity can create a more complex optimization landscape where different seeds might settle into different local minima, potentially causing the SAE variance to peak. At the 15-minute resolution, the 15-hour window acts as a strong low-pass filter, flattening the data into a simplified macro trend. With fewer complex features left, the seeds tend to converge to a similar average, which could explain the variance drop that completes the inverted U shape.

Further observation at the 10-minute resolution reveals differing metric behaviors depending on the load type. For PV generation, the evaluation metrics align reasonably well. However, EV charging presents a distinct pattern where MAE generally trends similarly to SAE, but RMSE exhibits an inverse relationship. Seed 415 exemplifies this by achieving a lower RMSE yet a substantially higher SAE. This divergence might be explained by the temporal misalignment inherent in sequence-to-point modeling. Most seeds likely preserve the overall energy mass, keeping SAE and MAE relatively lower, but shift the activation peaks slightly in time. This temporal lag creates large point-by-point errors that are heavily penalized by the squared nature of RMSE. Conversely, Seed 415 appears to align the temporal phases more accurately but underestimates the charging peak amplitude. By avoiding completely misaligned peaks, it reduces large outlier errors and keeps the RMSE low. However, this amplitude reduction fails to capture the concentrated energy of the EV load, resulting in a noticeable loss of total energy and a higher SAE.

This training variance in Figure 4 is also reflected in the model's convergence dynamics as illustrated by the early stopping epochs. At high resolutions (1 and 5 minutes), the optimization landscape appears relatively well-defined, allowing the seeds to converge consistently within 10 to 23 epochs. At lower resolutions (10 and 15 minutes), the training behavior diverges; some seeds halt prematurely, while others require extensive epochs to find a generalized solution, such as Seed 999 at the 10-minute resolution requiring 36 epochs.

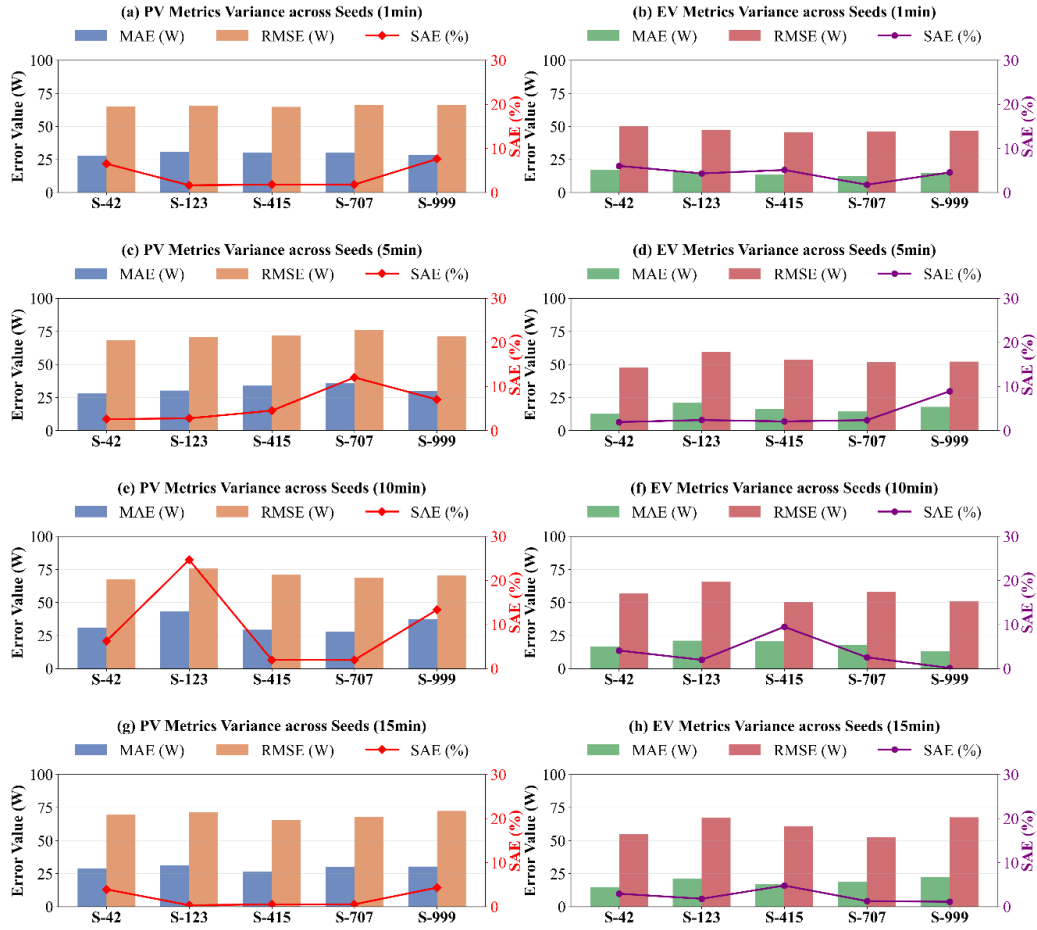


Figure 3 Quantitative performance variance (MAE, RMSE, and SAE) across five independent random seeds

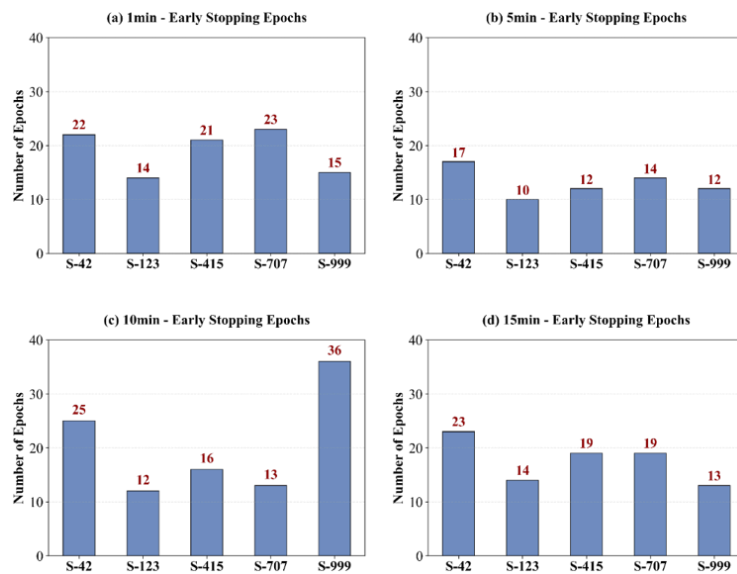


Figure 4 Number of epochs required for model convergence (triggered by early stopping) across five independent random seeds at different sampling rates

This indicates that at lower sampling rates, the loss of essential features in low-resolution data increases the model's sensitivity to weight initialization, making the training process highly dependent on the initial state of the network.

4.2.2 Qualitative Visual Inspection and Statistical Reliability: As observed in the multi-seed variance

analysis (Figure 3), Seed 42 consistently demonstrates stable evaluation metrics across all sampling rates. Thus, it serves as a reliable representative for qualitative visual inspection. Figures 5 and 6 present the actual versus predicted profiles, which generally corroborate the quantitative metrics. For EV charging, the S2P-CNN captures sharp transients at the 1-minute resolution.

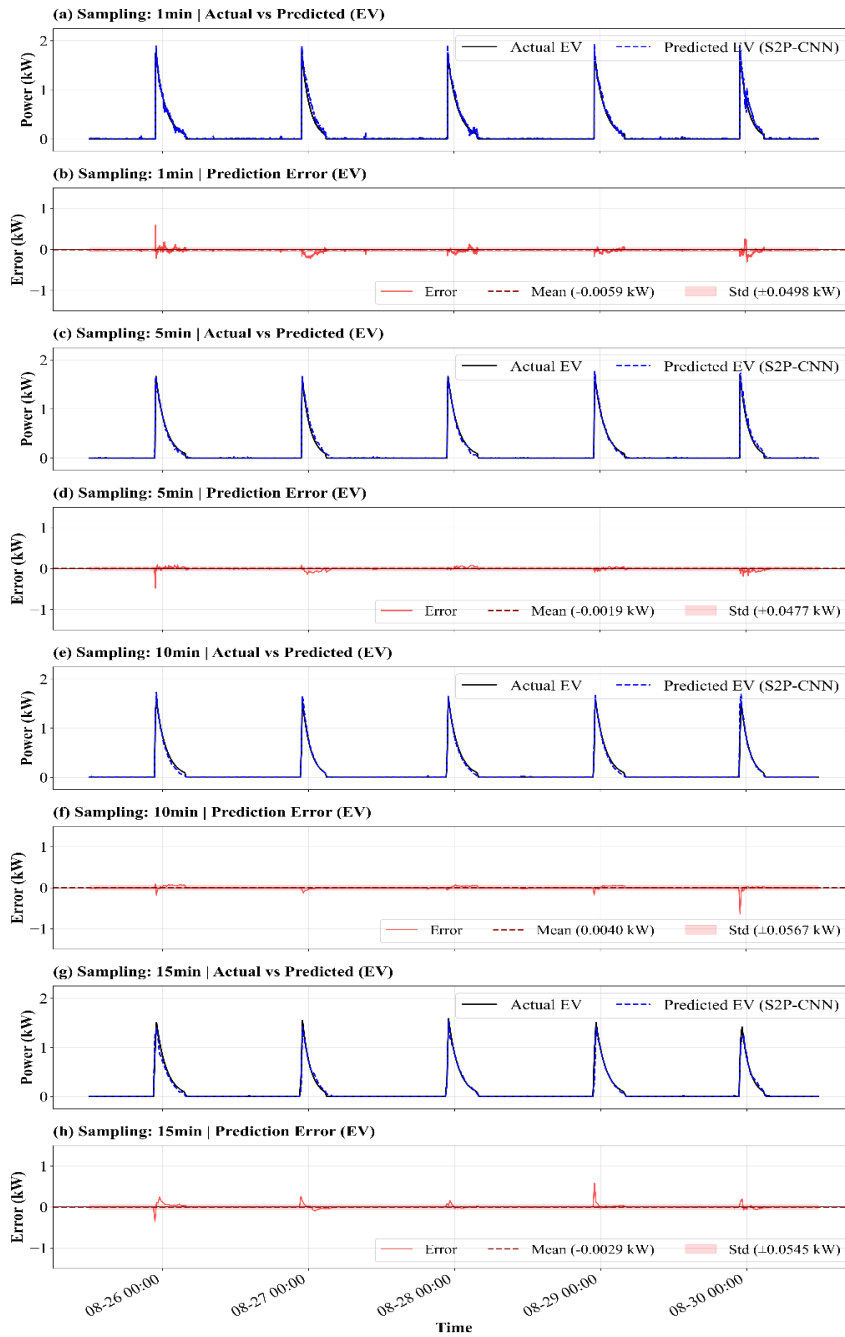


Figure 5 Qualitative disaggregation results of EV load across different sampling rates (Seed 42)

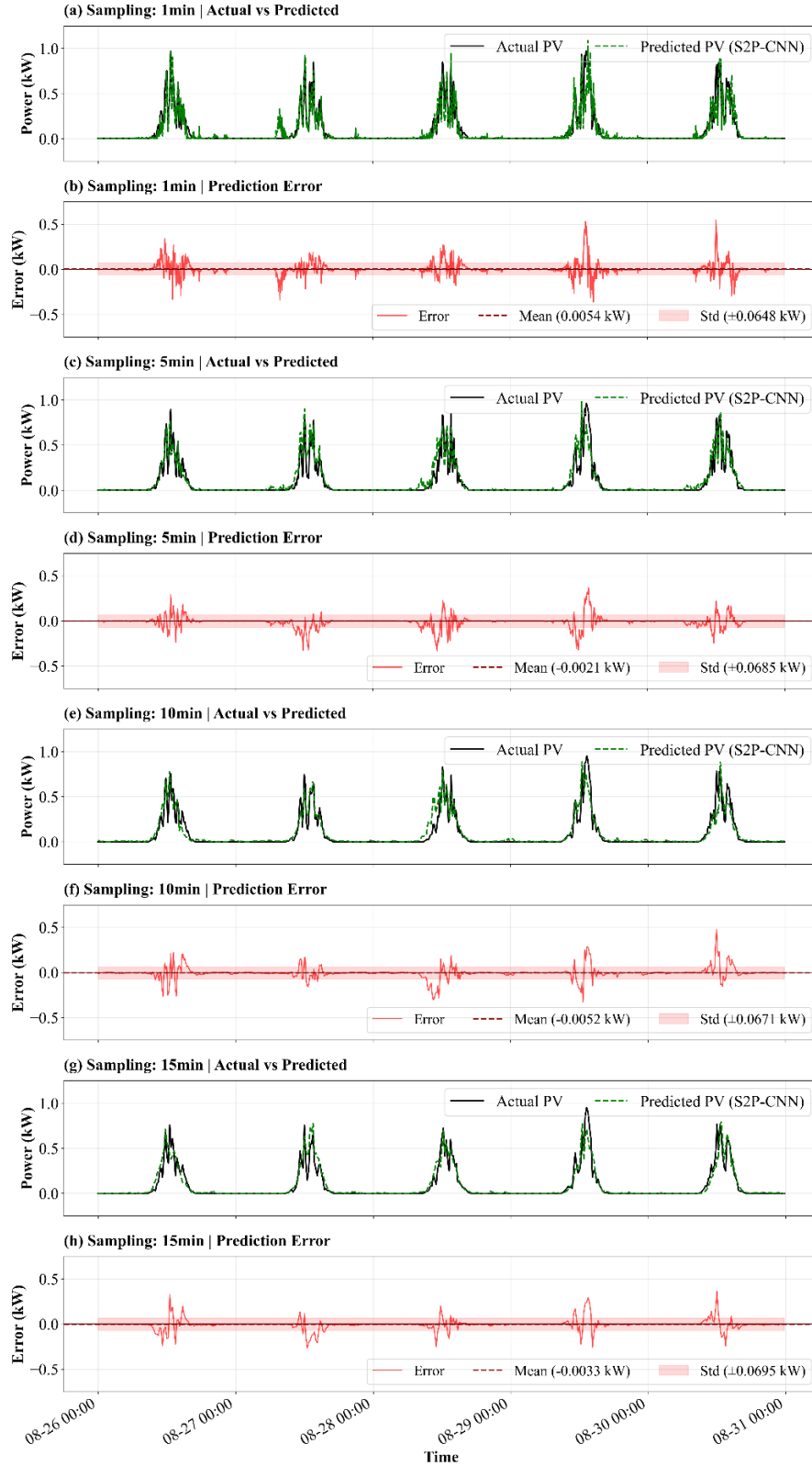


Figure 6 Qualitative disaggregation results of PV generation across different sampling rates (Seed 42)

As the resolution decreases to 15 minutes, the predicted peaks appear visually smoothed, reflecting the linear increase in MAE and RMSE. This smoothing is considered an inherent characteristic of many AMI systems that compute average power from accumulated energy. Importantly, while the peaks are flattened, the activation base tends to expand, which helps conserve the total area under the curve. Regarding PV generation, the evaluation metrics are tightly clustered across all resolutions, reflecting the multi-seed variance previously shown in Figure 3. Consequently, the prediction behavior of Seed 42 does not differ significantly when the sampling rate changes. Although high-resolution data captures stochastic weather-induced micro-fluctuations, the 15-minute temporal window acts as a natural low-pass filter that mitigates high-frequency environmental noise. As a result, the model consistently traces the macroscopic Gaussian bell curve of the daily solar generation. This visual consistency aligns with the stable evaluation metrics, indicating that the model's prediction performance remains relatively stable across different temporal resolutions. For a statistical comparison between the actual and predicted distributions under these sampling constraints, Paired Samples T-tests were conducted separately for EV and PV profiles at the 15-minute resolution ($n = 12,515$). Due to the inherently large sample size, the p-values approached zero ($p < 0.05$), a common hypersensitivity artifact in large datasets. Therefore, the Cohen's d effect size was evaluated to assess practical reliability. The analysis yielded trivial effect sizes of ($d = 0.05$) for EV and ($d = 0.04$) for PV, corresponding to average point-by-point differences of approximately 2.88 W and 3.29 W, respectively. In

a practical context, these sub-4-Watt deviations translate to acceptable SAE values of 2.84% for EV and 3.81% for PV, falling well below the target 10% threshold. This statistical evidence suggests that the S2P-CNN model can preserve the foundational energy mass without introducing substantial systematic bias.

4.2.3 Baseline Comparison: To evaluate the robustness and efficiency of the proposed multi-output S2P-CNN, a standard multi-output Long Short-Term Memory (LSTM) network was implemented as a sequence-modeling baseline. Regarding architectural advancements, the proposed framework extends the functional capability of the original S2P-CNN (Zhang et al., 2018) by transitioning from a single-output to a multi-output design. This expansion inherently optimizes computational resources through parallel processing, enabling the simultaneous disaggregation of both EV charging load and PV generation in a single inference pass.

Figure 7 illustrates the performance and computational efficiency comparison between the two models across varying temporal resolutions. In this analysis, MAE is utilized as the primary representative metric to quantify point-wise reconstruction fidelity. This metric provides a clear measure of structural tracking capability and average deviation. At the 1-minute resolution, the LSTM model appears competitive, achieving an EV MAE of 15.47 ± 1.96 W and a PV MAE of 28.72 ± 1.70 W, which is comparable to the S2P-CNN's 14.60 ± 1.76 W and 29.38 ± 1.36 W, respectively. However, a performance divergence emerges as the sampling rate decreases.

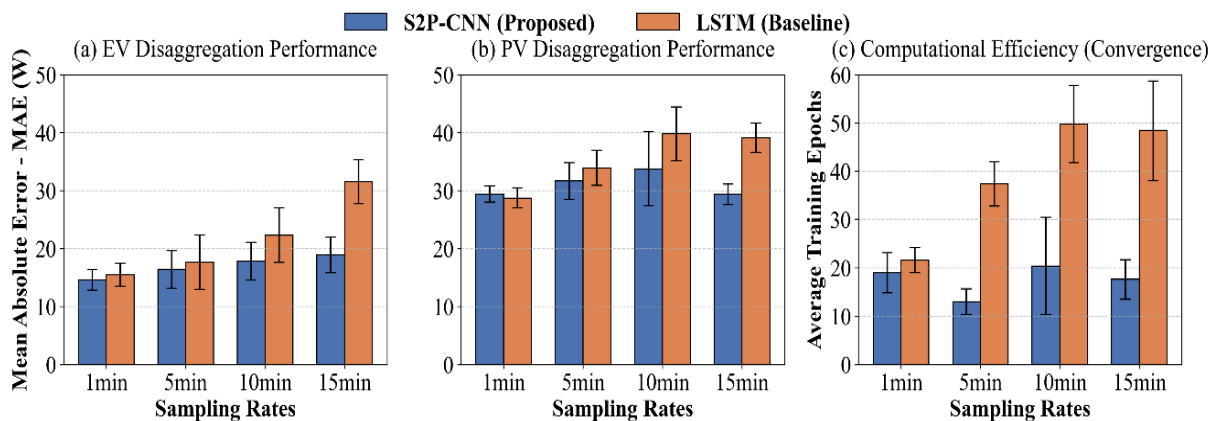


Figure 7 Performance and computational efficiency comparison between the proposed S2P-CNN and the baseline LSTM model across varying sampling rates

The S2P-CNN demonstrates more robustness across the 5, 10, and 15-minute resolutions (maintaining its EV MAE between 16.44 W and 18.89 W, and PV MAE between 29.38 W and 33.74 W), while the LSTM displays a progressive increase in MAE, reaching up to 31.53 ± 3.84 W for EV and 39.11 ± 2.57 W for PV at the 15-minute resolution. This quantitative trend suggests that sequence modeling may be more sensitive to sampling rates. Furthermore, the training dynamics depicted in Figure 7 reveal a significant shift in computational efficiency. While the number of training epochs remains comparable at the 1-minute resolution (requiring 21.6 ± 2.6 epochs for the LSTM versus 19.0 ± 4.2 epochs for the S2P-CNN), computational divergence becomes significant as the sampling rate decreases. At lower resolutions (5 to 15 minutes), the LSTM requires more than double the number of epochs to reach convergence (ranging from 37.4 to 49.8 epochs) compared to the S2P-CNN (which stabilizes between 13.0 and 20.4 epochs). This indicates that the structural feature mapping of the CNN-based architecture is likely to facilitate a more robust convergence rate and learning efficiency under temporal sampling constraints. These experimental findings align with the structural challenges of low-resolution smart meter data noted by Vavouris et al. (2022) and suggest that the S2P-CNN serves as an optimized and robust framework for practical energy disaggregation.

5. Discussion

This study investigates the feasibility of a multi-output S2P-CNN framework for simultaneous EV and PV disaggregation under temporal sampling constraints. The experimental findings generally validate the initial hypotheses and highlight practical engineering applications. Three primary discussion points are presented below. The results support the hypothesis regarding the inherent trade-off between shape reconstruction and energy estimation. For transient loads like EV charging, temporal down-sampling linearly degrades point-wise accuracy (MAE and RMSE) due to the attenuation of high-frequency signal edges. However, the SAE remains relatively stable, successfully achieving the targeted sub-10% threshold even at the 15-minute resolution. This suggests that while coarse sampling rates are insufficient for precise behavioral profiling, the network adapts by expanding the activation base to conserve the overall energy mass. Conversely, for continuous PV generation, the 15-minute sampling window acts as a natural low-pass filter, mitigating

environmental noise and allowing the model to track the macroscopic generation trend efficiently.

Furthermore, the multi-seed variance analysis indicates that optimization stability is sensitive to the relationship between the fixed input window size and the sampling rate. Intermediate resolutions (5 and 10 minutes) appear to introduce optimization ambiguity, leading to higher variance in SAE. The baseline comparison also validates the architectural advantages of the proposed framework. As the sampling rate decreases, sequence-based models like the LSTM exhibit a progressive increase in error and require significantly more training epochs to converge. This implies that the structural feature extraction inherent in the CNN framework provides greater resilience under coarse sampling conditions and offers higher computational efficiency compared to conventional sequence-modeling approaches.

Achieving an SAE below 10% at the 15-minute resolution carries direct implications for real-world utility deployment. The statistical equivalence demonstrated by the Cohen's d effect size confirms that existing AMI can support reliable aggregate energy estimation without necessitating high-frequency hardware upgrades. Additionally, the simultaneous multi-output design effectively mitigates computational overhead. However, these implications must be considered within the study's specific context. Limitations include the use of a single household (UK-DALE House 1) as a controlled baseline, and the superposition of synthetic load profiles. Future work should prioritize validating these findings across diverse real-world datasets, such as the Pecan Street Dataport, and exploring advanced architectures like Attention mechanisms or Transformer models to further enhance low-frequency disaggregation.

6. Conclusion

This research evaluated a multi-output S2P-CNN framework designed for simultaneous EV and PV disaggregation. By analyzing performance across multiple random seeds, the study revealed that prediction accuracy is determined not only by the data sampling rate but also significantly by the initial network weights (seed selection). The findings suggest a clear engineering trade-off. While lowering the sampling rate generally reduces signal shape accuracy due to the loss of high-frequency transients, the model appears capable of preserving the total energy mass. Additionally, baseline comparisons indicate that the CNN architecture can offer greater robustness and computational efficiency under low-sampling

conditions compared to conventional sequence models like the LSTM. Furthermore, the experimental results highlighted specific challenges in optimization stability. At the intermediate 10-minute resolution, the PV energy estimation exhibited high variance, yielding a Signal Aggregate Error (SAE) of $9.63\% \pm 9.61\%$. These intermediate resolutions tend to complicate the feature extraction process, thereby increasing the sensitivity of the optimization process to seed initialization. However, at the standard 15-minute resolution, the model demonstrated more consistent behavior, keeping the average SAE safely below the 10% target. This provides preliminary evidence that existing AMI might support aggregate-level NILM applications, such as energy billing and grid load monitoring, without requiring costly hardware upgrades. Consequently, utility companies could potentially prioritize software-based algorithms over physical meter replacements, offering a cost-effective approach for smart grid energy management.

The primary limitation of this foundational study is the use of synthetic EV and PV profiles superimposed on a single household's background load (UK-DALE House 1). While this creates a strictly controlled testing environment, it lacks the dynamic variations of real-world multi-household deployments. Future work will focus on addressing these limitations by validating the framework on comprehensive real-world datasets, such as the Pecan Street Dataport. Additionally, we plan to explore advanced architectures, including Attention mechanisms and Transformer models, and compare them against this S2P-CNN baseline to further enhance disaggregation accuracy at lower sampling rates.

7. Abbreviations

Abbreviation	Full Term
AMI	Advanced Metering Infrastructure
CNN	Convolutional Neural Network
DERs	Distributed Energy Resources
EV	Electric Vehicle
HEMS	Home Energy Management System
HMM	Hidden Markov Model
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MSE	Mean Squared Error
NILM	Non-Intrusive Load Monitoring
PV	Photovoltaic
ReLU	Rectified Linear Unit
RMSE	Root Mean Square Error

Abbreviation	Full Term
RNN	Recurrent Neural Network
S2P-CNN	Sequence-to-Point Convolutional Neural Network
SAE	Signal Aggregate Error
SD	Standard Deviation
SNR	Signal-to-Noise Ratio

8. CRediT Statement

Peerapon Chanhom: Conceptualization, Methodology, Software, Formal Analysis, Investigation, Data Curation, Visualization, Writing – Original Draft.

Jiranat Tangchittichariya: Validation, Writing – Review & Editing.

9. References

- Al-Shetwi, A. Q. (2022). Sustainable development of renewable energy integrated power sector: Trends, environmental impacts, and recent challenges. *Science of The Total Environment*, 822, Article 153645. <https://doi.org/10.1016/j.scitotenv.2022.153645>
- Dash, S., & Sahoo, N. C. (2023). Attention-based deep learning approach for nonintrusive and simultaneous state detection of multiple appliances in smart buildings. *IEEE Journal of Emerging and Selected Topics in Industrial Electronics*, 5(3), 1248-1258. <https://doi.org/10.1109/JESTIE.2023.3333308>
- Dobos, A. P. (2014). *PVWatts version 5 manual* (NREL/TP-6A20-62641). National Renewable Energy Laboratory. <https://doi.org/10.2172/1158421>
- Fang, X., Misra, S., Xue, G., & Yang, D. (2012). Smart grid—The new and improved power grid: A survey. *IEEE Communications Surveys & Tutorials*, 14(4), 944-980. <https://doi.org/10.1109/SURV.2011.101911.00087>
- Hart, G. W. (1992). Nonintrusive appliance load monitoring. *Proceedings of the IEEE*, 80(12), 1870-1891. <https://doi.org/10.1109/5.192069>
- Himeur, Y., Ghanem, K., Alsalemi, A., Bensaali, F., & Amira, A. (2021). Artificial intelligence based anomaly detection of energy consumption in buildings: A review, current trends and new perspectives. *Applied Energy*, 287, Article 116601. <https://doi.org/10.1016/j.apenergy.2021.116601>
- Huber, P., Calatroni, A., Rumsch, A., & Paice, A. (2021). Review on deep neural networks

- applied to low-frequency nilm. *Energies*, 14(9), Article 2390.
<https://doi.org/10.3390/en14092390>
- IEEE Standards Association. (2018). *IEEE Standard for Smart Energy Profile Application Protocol*. New York, US: The Institute of Electrical and Electronics Engineers, Inc.
- International Electrotechnical Commission (IEC). (2020). *Electricity metering data exchange—The DLMS/COSEM suite—Part 1-0: Smart metering standardisation framework* (IEC 62056-1-0:2014+AMD1:2020). IEC.
- Jaime, H. C., de Souza, A. D., Santos Machado, R. C., & Gomes, O. d. S. M. (2026). Machine learning models for non-intrusive load monitoring: A systematic review and meta-analysis. *Inventions*, 11(2), Article 29.
<https://doi.org/10.3390/inventions11020029>
- Kelly, J., & Knottenbelt, W. (2015). The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes. *Scientific Data*, 2(1), Article 150007. <https://doi.org/10.1038/sdata.2015.7>
- Nizami, M. S. H., Hossain, M. J., Amin, B. R., & Fernandez, E. (2020). A residential energy management system with bi-level optimization-based bidding strategy for day-ahead bi-directional electricity trading. *Applied Energy*, 261, Article 114322.
<https://doi.org/10.1016/j.apenergy.2019.114322>
- Sykiotis, S., Kaselimi, M., Doulamis, A., & Doulamis, N. (2022). Electricity: An efficient transformer for non-intrusive load monitoring. *Sensors*, 22(8), Article 2926.
<https://doi.org/10.3390/s22082926>
- Todic, T., Stankovic, V., & Stankovic, L. (2023). An active learning framework for the low-frequency non-intrusive load monitoring problem. *Applied Energy*, 341, Article 121078.
<https://doi.org/10.1016/j.apenergy.2023.121078>
- Vavouris, A., Garside, B., Stankovic, L., & Stankovic, V. (2022). Low-frequency non-intrusive load monitoring of electric vehicles in houses with solar generation: Generalisability and transferability. *Energies*, 15(6), Article 2200.
<https://doi.org/10.3390/en15062200>
- Völker, B., Reinhardt, A., Faustine, A., & Pereira, L. (2021). Watt's up at home? Smart meter data analytics from a consumer-centric perspective. *Energies*, 14(3), Article 719.
<https://doi.org/10.3390/en14030719>
- Yin, Y., Xu, B., Zhong, Y., Bao, T., & Wang, P. (2023). A non-intrusive load monitoring model for electric vehicles based on multi-kernel conventional neural network. *World Electric Vehicle Journal*, 14(2), Article 51.
<https://doi.org/10.3390/wevj14020051>
- Zeifman, M., & Roth, K. (2011). Nonintrusive appliance load monitoring: Review and outlook. *IEEE Transactions on Consumer Electronics*, 57(1), 76-84.
<https://doi.org/10.1109/TCE.2011.5735484>
- Zhang, C., Zhong, M., Wang, Z., Goddard, N., & Sutton, C. (2018). Sequence-to-point learning with neural networks for non-intrusive load monitoring. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1), 2604-2611.
<https://doi.org/10.1609/aaai.v32i1.11873>
- Zoha, A., Gluhak, A., Imran, M. A., & Rajasegarar, S. (2012). Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey. *Sensors*, 12(12), 16838-16866.
<https://doi.org/10.3390/s121216838>