A Machine Learning Approach for the Identification of Photovoltaic and Electric Vehicle Profiles in a Smart Local Energy System

1st Vihanga Peiris
Department of Mathematics, Physics & Electrical Engineering
Northumbria University
Newcastle,UK
jaysuriya.peiris@northumbria.ac.uk

2nd Goyal Awagan
Department of Mathematics, Physics & Electrical Engineering
Northumbria University
Newcastle,UK
goyal.awagan@northumbria.ac.uk

3rd Jing Jiang
Department of Mathematics, Physics & Electrical Engineering
Northumbria University
Newcastle,UK
jing.jiang@northumbria.ac.uk

***Abstract*—Smart local energy systems (SLESs) focus on integrating more renewable energy sources to the electrical distribution network. Digitalization of SLESs can be achieved through digital twins (DTs). The DT is a virtual replica of the physical energy system. The accurate recognition of the availability, capacity, and quantity of the distributed energy resources (DERs) connected to the electrical distribution network enables the development of a comprehensive DT of the local energy system. This research study proposes to use ML algorithms to identify the availability of DERs in the local area and it contributes to the field by providing a comparative analysis of the results of different classification based machine learning (ML) algorithms when recognising availability of solar energy generation and electric vehicles (EVs) in aggregated grid data using a comprehensive labelled dataset. The results presented in this paper show that with the application of sliding window method for photovoltaic (PV) and EV identification from aggregated data, the accuracy, precision, recall and F1 score metrics can reach an approximate value of 98%.**

***Index Terms*—Smart Local Energy System, Digital Twin, Distributed Energy Resource, Photovoltaic, Electric Vehicle, Non Intrusive Load Monitoring, Machine Learning**

I. INTRODUCTION

Smart local energy systems (SLESs) optimise distributed energy resource (DER) integration, generation and storage in a local area by exploiting the internet of things (IoT) and digital technology. SLESs achieve the dynamic balance of electricity supply and demand through intelligent digital coordination of data obtained from the local energy systems. Furthermore, SLESs provide flexibility in balancing energy among multiple energy vectors in energy supply, storage and generation [1]. The accurate recognition of the availability, capacity and the quantity of the DERs and the development of comprehensive digital twins (DTs) enable management of the local electrical distributed network during a breakdown of a DER or a set of DERs. The energy management in such a situation can be achieved by isolating the respective faulty DER or the set of DERs from distributed network.

The research area of SLES is expanding vigorously as a result of the emerging global goals of achieving net-zero carbon emissions and integration of more DERs in the electric distribution network [1]. Taking the optimal use of the DERs connected to the electric distribution network through accurate identification is essential to minimize the use of fossil fuels. Therefore, development of the DT has been identified as playing an important role in the functioning of SLES. The DT can be further used in predicting and forecasting overall energy supply and demand.

The data collection for the development of DTs is challenging. The collection of appliance-level data from households compromises consumers' privacy by revealing detailed information about their daily activities. The load monitoring methods can be broadly categorised into two namely, itrusive load monitoring (ILM) and non-intrusive load monitoring (NILM). ILM methods require several meters, hence multiple points of measurement to monitor the electric loads connected to the distribution network in a specific location whereas the NILM method requires only one measuring device per location [2]. Therefore, the NILM method has a simpler deployment of infrastructure compared to the ILM method and has been a popular research topic for energy management given its simplicity [3], [4]. Considering consumer privacy and the low cost of implementation, NILM techniques are preferable over ILM methods [5]. In [6], Monteiro et al. compared the classification accuracy of various machine learning (ML) algorithms for NILM using the reference energy disaggregation dataset (REDD). The REDD consists of low frequency labelled data relevant to six houses. The apparent power at the main circuit level of the residences, voltage and the current values are recorded in the REDD. However, as mentioned in [6], the dataset exhibited numerous instances of missing data points. Therefore, the data referring to the current signals of electrical

This work was supported by the Engineering and Physical Sciences Research Council [grant number EP/Y005376/1] – VPP-WARD Project (https://www.vppward.com).

devices respective to two houses has been used in [6].

In [7], a ML algorithm has been studied for NILM by using the data of REDD. The authors of [7] have used data from two residences of REDD for training and one residence for testing. According to the results presented in [7], the data of two houses show better results in identifying the refrigerators' electric signature, while it's comparatively low for Microwave and dishwashers. However, the results for house 1 are better compared to house 2 depending on the dataset. Therefore, the load identification results are dependent on the data set used to train the ML models.

The NILM methods utilise smart meter data or transformer data and analyse them with ML algorithms and statistical methods to recognize the availability of DERs. Classification-based supervised ML algorithms such as K-nearest neighbour (KNN), random forest (RF), and artificial neural network (ANN) have been employed to identify the presence of DERs among houses in a particular region using NILM methods in past research studies [8].

However, a comparison of the performance of supervised classification-based ML using a wider range of algorithms is unavailable in the up-to-date literature. Therefore, a comparison of various conventional ML algorithms in the identification of photovoltaics (PVs) and electric vehicles (EVs) is presented in this paper. This research work describes the methods of processing collected smart meter data and analyses the results obtained by training and testing the ML algorithms. The prime objective of conducting a comparison between wider range of ML algorithms in this conference paper is to assess their suitability for PV and EV identification and their feasibility in implementing DTs. Consequently, the contributions of the conference paper are:

1) A comparative analysis of different ML algorithms aimed at recognising the presence of PVs and EVs in aggregated grid data.
2) Use of a practical and comprehensive set of data for non intrusive DER identification to facilitate rigorous performance evaluations and precise assessments.
3) Propose to use DER identification for the construction of a DT, where meticulous data collection and precise DER identification are crucial for establishing a robust data-driven DT.

The rest of the paper is organised as follows: Section II provides a detailed description of the proposed system. Section III presents the methods used in PV and EV identification. Section IV presents identification results from different ML algorithms and discusses their performance. Finally, section V presents the conclusions.

II. Proposed System Model

The Fig. 1 illustrates the system model proposed in the paper. The rest of the section explains the proposed system model.

The DER data and grid data in the local area are used as input data for DER identification. In this research, historic DER and grid data is proposed to be used for offline training and testing of the DER identification tool. Subsequently, the DT, the virtual replica of the physical multi vector energy system is implemented and the real time data is fed to the DER identification tool to update the DT.

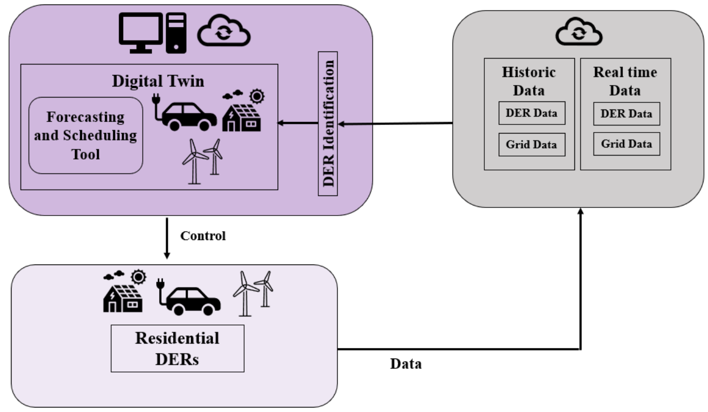

Fig. 1. System Model

The DT is used to monitor the DERs and calibrate and predict the energy demand of a particular region. Furthermore, it is proposed to obtain a dynamic balance of energy and demand by using outputs of energy forecasting and scheduling tools and improve the resilience of the electric grid.

The DER identification tool presented in this research work is composed of PV and EV identification tools. The PV and EV energy profile identification tools identify the availability of PV panels and EVs at homes by recognising the solar energy availability of houses from aggregated electric grid data. Therefore, the proposed system uses the NILM method described in Section I for DER identification.

The remainder of the section includes a comprehensive explanation of data collection, pre-processing and the MATLAB tools utilised in the development of the PV/EV identification and the blocks in the Fig. 1.

A. Data Acquisition for DER Identification

Historic power consumption data is used to train and test the PV/EV identification tool. The 1 minute data of residences of New York city available in Pecan Street Data port relevant to an hourly time slot is used to train and test the PV and EV profile identification tool studied in this paper [9]. The New York dataset consists of appliance level power consumption, solar power generation data, EV load power consumption and the aggregated electric grid data read from the smart meters. The aggregated electric grid data is the power drawn from or fed to the electrical grid. The solar energy generation data and EV data are used to label the dataset with PV and EV availability of the houses. 0.1 kW is used as the threshold to label PV and EV availability in households. The labelled data is used to train the EV and PV identification model [8]. The distribution of the number of houses with PV and EV data used in this research is presented in Table I. The total number of houses used to train and test PV model is 13, where 7 houses

TABLE I
INFORMATION OF NUMBER OF HOUSES WITH AND WITHOUT PV AND EV AVAILABLE IN THE DATASET.

Availability	PV	EV
Available	7 houses	5 houses
Unavailable	6 houses	5 houses

was with solar power generation and 6 was without PV. For the EV model training and testing, data of 10 houses were used, where 5 houses were with EVs and 5 houses without EVs.

B. Physical System

The physical system related to this research study comprises DERs (PVs and EVs) connected to the electric distributed network (the DERs connected to the households). The data from the smart meters of the residences is collected and stored in the cloud.

C. Forecasting and Scheduling Tool

The historic and real time data can be used to train and test the energy forecasting tools. The energy forecasting is known as the prediction of future power consumption (power demand) and power generation [10]. Energy scheduling refers to energy management, planning, decision-making, distribution of energy resources over a specific period [11].

III. PV/EV PROFILE IDENTIFICATION

This section presents different methods used to develop and improve the PV and EV identification tool.

A. Sliding Window (SW)

SW is a technique used in scenarios where the time series values of adjacent data points are essential for recognizing patterns. The method is known to improve the accuracy of the ML models. The SW overlaps the windows and improves recognition by repeating patterns and by reducing the effect of noise on recognition [12].

Based on previous research papers, a window size of 10 data points and an overlap of 5 is used to improve the PV and EV identification accuracy. Hence the window size and overlap consist of data points respective to 10 and 5 minutes respectively [8].

B. Extraction of time domain (temporal) features

Time domain features have been extracted from the windows during data processing and were used as inputs of the ML models. However, in this paper minimum, maximum, range, mean, variation and standard deviation are used as the the time domain or the temporal features [8], [13].

C. Dataset and ML models

MATLAB classifier application under the ML and deep learning toolbox of MATLAB 2023b version is used as the main tool for implementation of ML based PV/EV identification tool. The ML models available in the classifier application are trained and tested with 5 fold cross validation. The test dataset consists of 20% of the dataset. The labelled electric grid data and time domain features are used as the response and the predictors of ML models respectively. The data processing in this research work is carried in two methods;

1) Method I - direct use of the labelled dataset.
2) Method II - applying SW along with the time domain feature inputs.

IV. RESULT ANALYSIS

This section consists of the results obtained when training and testing the above ML models for the dataset explained in Section II-A. The metrics used in evaluating the performance of ML models are accuracy, precision, recall, and F1 score.

The definitions of accuracy, F1 score, recall, and precision can be expressed as in the equations (1), (2), (3) and (4) below;

$$Accuracy = \frac{TP\ +\ TN}{TP\ +\ TN +\ FP\ + FN} \tag{1}$$

$$Precision = \frac{TP}{TP\ + FP} \tag{2}$$

$$Recall = \frac{TP}{TP\ + FN} \tag{3}$$

$$F1\ Score = 2 \times \left(\frac{Precision\ \times\ Recall}{Precision + Recall} \right) \tag{4}$$

where, *TP* = True Positives, *FP* = False Positives, *TN* = True Negatives and *FN* = False Negatives.

The accuracy explains the prediction of the models which were correctly classified while the precision is the fraction of the positives which were correctly identified by the models. Recall explains the proportion of actual positives identified correctly. F1 score is a metric that is valuable when there is an uneven distribution between positives and negatives.

The ML models were trained and tested under two categories. The categorisation was made based on Method I and Method II as previously explained in Section III-C.

A. PV Identification Results

The results for PV identification with Method I and Method II are described below.

1) Method I: The results for PV identification without SW and time domain feature inputs are graphically illustrated in Fig. 2 below. According to the results, the highest accuracy and F1 score are shown in the results of the coarse tree algorithm and Neural Networks (Wide) models. Most of the models resulted in an accuracy above 90% and an F1 Score around 90%.

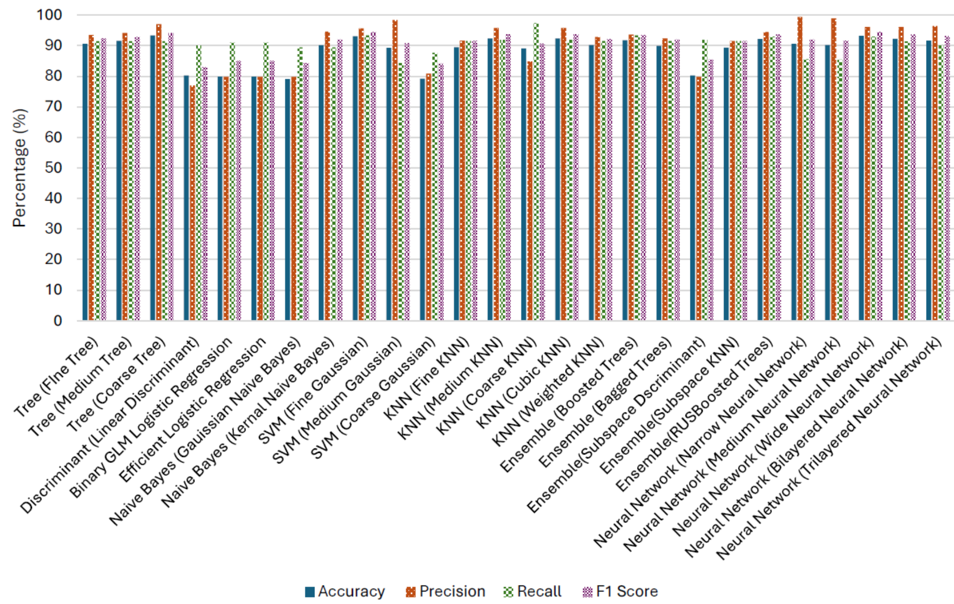

Fig. 2. PV identification results - Method I.

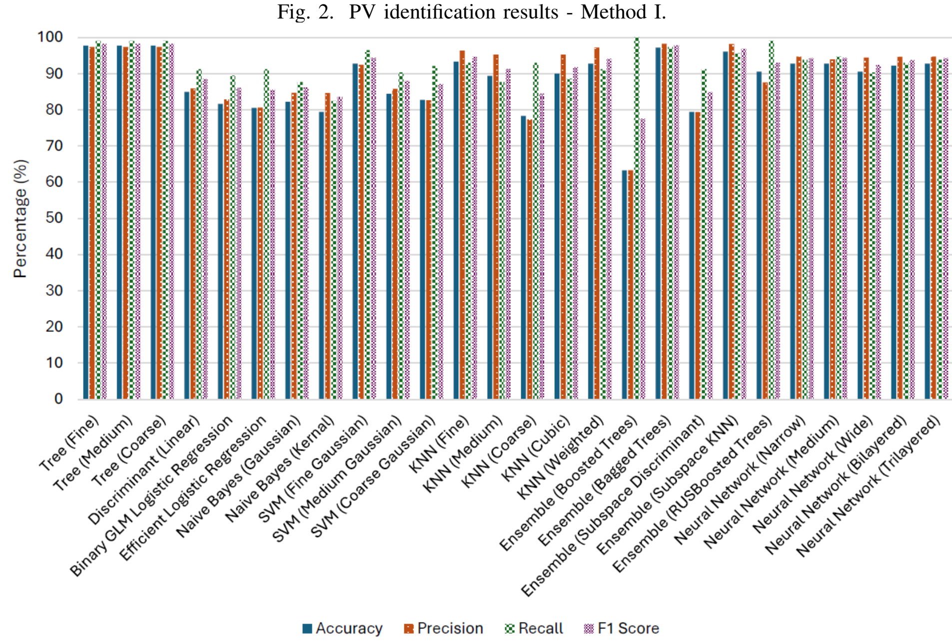

Fig. 3. PV identification results - Method II

2) Method II: The results for PV identification without SW and time domain feature inputs are graphically illustrated in Fig. 3 below. For PV identification, the application of the SW window method improved the performance metrics of accuracy, precision, recall and F1 score for most of the ML algorithms and worsened the performance of a few of the ML models. Accordingly, the ML model that showed inferior results in method II is Ensemble Boosted Trees. However, the Logistic Regression models, Naive Bayes models and KNN models showed similar performance in both Method I and Method II for PV identification. The highest performance for PV identification in Method II was recorded in Trees algorithms with accuracy, precision, recall and F1 score of 97.73%, 97.37%, 99.11% and 98.23% respectively.

B. EV Identification Results

The results for EV identification with Method I and Method II are described below.

1) Method I: The results for EV identification without SW and time domain feature inputs are graphically illustrated in Fig. 4 below. According to the results, the highest accuracy and F1 score resulted from Ensembled (Boosted Trees) and KNN (Cubic) models. The rest of the models resulted in an accuracy above 90%. However, the F1 score of all the models was below

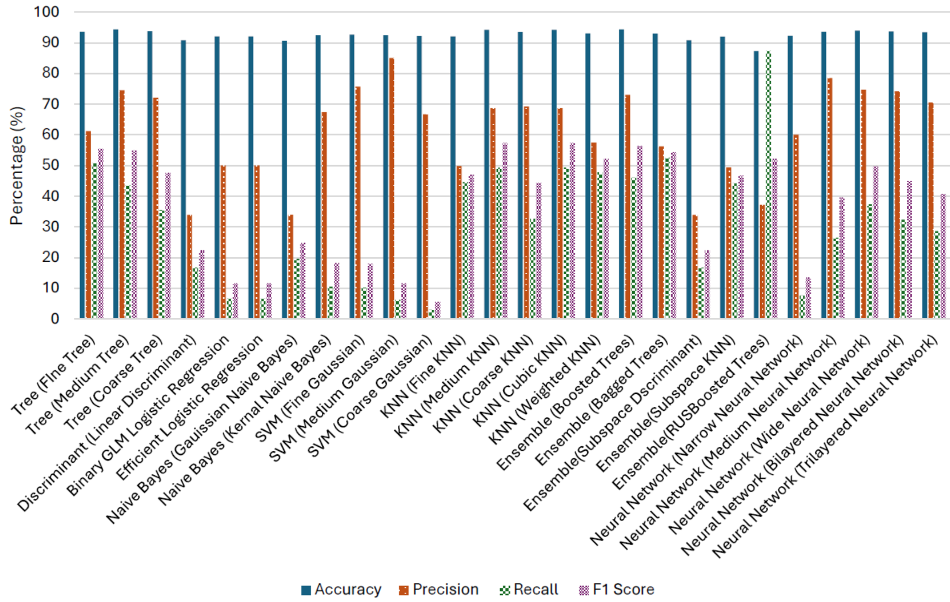

Fig. 4. EV identification results - Method I

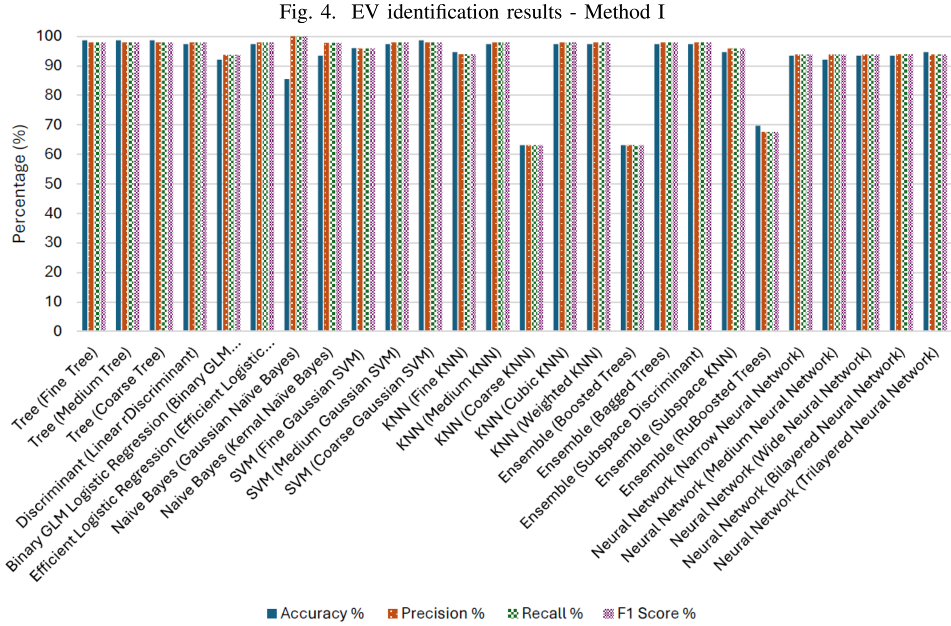

Fig. 5. EV identification results - Method II

50%. This is due to the inability of the algorithm to identify the minority class. After refining the dataset with a threshold of 0.1 kW, the resulting dataset with actual EV loading is imbalanced. Consequently, the model fails to correctly identify instances of the minority class, as evidenced by its low recall value for EV availability. This results in low F1 scores despite the high accuracy.

2) Method II: The results for EV identification with SW and time domain features are graphically illustrated in Fig. 5 below. The results show an increment in most of the EV identification model results in Method II. However, KNN (coarse), Ensemble (boosted trees), and Ensemble (RuBoosted) algorithms show a decrement in accuracy results compared to Method I. In conclusion, Tree algorithms and SVM (Coarse Gaussian SVM) algorithms show the highest performance with accuracy, precision, recall and F1 score of 97.92%, 97.92%, 97.92% and 97.92% respectively.

V. Conclusions

Results show that ML models can identify PVs easily with above 90% on the four performance metrics, i.e., accuracy, precision, recall and F1 score in both Method I and Method II. The application of the SW method improves the results of the ML algorithms used for PV and EV identification. The EV identification using Method I resulted in very low values for the four performance metrics whereas PV identification showed promising results with Method I. However, with the use of SW, the EV identification performance was significantly improved.The SW method improves the identification of the availability of EV loads (minority class) in the dataset, resulting in improvements in all four performance metrics. The ML models that resulted in the highest performance in PV and EV identification are listed in Table II. In conclusion, both PV and EV identification can reach an approximate value of 98% in all four metrics, accuracy, precision, recall and F1 score with the application of SW method.

TABLE II
PV AND EV RESULTS OF THE BEST PERFORMED ML MODEL.

DER Type	Method Type	Model Type	Accuracy	Precision	Recall	F1 Score
PV	Method I	Coarse tree algorithm	93.36%	96.99%	91.49%	94.16%
		Neural Network	93.25 %	96.17%	92.95%	94.53%
	Method II	Tree	97.73%	97.37%	99.11%	98.23%
EV	Method I	Ensemble (Boosted trees)	93.69%	72.99%	46.01%	56.44%
		KNN (Cubic)	93.46%	68.88%	49.28 %	57.38%
	Method II	Tree algorithm	97.92%	97.92%	97.92%	97.92%
		SVM (Coarse Gaussian)	97.92%	97.92%	97.92%	97.92%

REFERENCES

[1] J. Wu, Y. Zhou, and W. Gan, "Smart local energy systems towards net zero: Practice and implications from the uk," *CSEE Journal of Power and Energy Systems*, vol. 9, no. 2, pp. 411–419, 2023.

[2] R. Ramadan, Q. Huang, and O. Bamisile, "Non-intrusive load monitoring for household energy disaggregation: A state-of-the-art," in *2022 IEEE 6th Conference on Energy Internet and Energy System Integration (EI2)*, pp. 1397–1403, 2022.

[3] A. Ridi, C. Gisler, and J. Hennebert, "A survey on intrusive load monitoring for appliance recognition," in *2014 22nd International Conference on Pattern Recognition*, pp. 3702–3707, 2014.

[4] A. Zoha, A. Gluhak, M. A. Imran, and S. Rajasegarar, "Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey," *Sensors*, vol. 12, no. 12, pp. 16838–16866, 2012.

[5] S. Dai, F. Meng, Q. Wang, and X. Chen, "Federatednilm: A distributed and privacy-preserving framework for non-intrusive load monitoring based on federated deep learning," in *2023 International Joint Conference on Neural Networks (IJCNN)*, pp. 01–08, 2023.

[6] D. Zufferey, C. Gisler, O. A. Khaled, and J. Hennebert, "Machine learning approaches for electric appliance classification," in *2012 11th International Conference on Information Science, Signal Processing and their Applications (ISSPA)*, pp. 740–745, 2012.

[7] L. Mauch and B. Yang, "A new approach for supervised power disaggregation by using a deep recurrent lstm network," in *2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 63–67, 2015.

[8] A. F. Moreno Jaramillo, J. Lopez Lorente, D. Laverty, P. Brogan, S. Velasquez, J. Martinez-del Rincon, and A. Foley, "Distributed energy resources electric profile identification in low voltage networks using supervised machine learning techniques," pp. 1–19, Feb. 2023.

[9] "Pecan street dataport, residential data (2023), https://dataport.pecanstreet.org/academic."

[10] T. Hong, P. Pinson, Y. Wang, R. Weron, D. Yang, and H. Zareipour, "Energy forecasting: A review and outlook," *IEEE Open Access Journal of Power and Energy*, vol. 7, pp. 376–388, 2020.

[11] A. Moradmand, M. Dorostian, and B. Shafai, "Energy scheduling for residential distributed energy resources with uncertainties using model-based predictive control," *International Journal of Electrical Power Energy Systems*, vol. 132, p. 107074, 2021.

[12] M. Jaén-Vargas, K. Reyes Leiva, F. Fernandes, S. Gonçalves, M. Silva, D. Lopes, and J. Serrano Olmedo, "Effects of sliding window variation in the performance of acceleration-based human activity recognition using deep learning models," *PeerJ Computer Science*, vol. 8: e1052, 08 2022.

[13] A. Ridi, C. Gisler, and J. Hennebert, "Unseen appliances identification," in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications* (J. Ruiz-Shulcloper and G. Sanniti di Baja, eds.), (Berlin, Heidelberg), pp. 75–82, Springer Berlin Heidelberg, 2013.