Identification of Photovoltaic and Electric Vehicle Profiles in Distribution Networks Using Long Short-Term Memory Network

Goyal Awagan, Jing Jiang, Vihanga Peiris
Dept. of Mathematics, Physics& Electrical Engineering,
Northumbria University
Newcastle, UK
{goyal.awagan, jing.jiang,jayasuriya.peiris}@northumbria.ac.uk

Hongjian Sun, Pratik Harsh
Dept. of Engineering,
Durham University
Durham, UK
{hongjian.sun, pratik.harsh}@durham.ac.uk

***Abstract*—The widespread implementation of global initiatives focused on achieving net-zero carbon emissions and the electrification of transportation has resulted in the extensive deployment of distributed energy resources (DERs) within the low-voltage distribution network. The rapid integration of DERs has introduced technical challenges, altering the electrical characteristics of conventional distribution networks. This challenge is exacerbated by the absence of monitoring infrastructure on the low-voltage side. Non-intrusive load monitoring (NILM) methods offers a chance to enhance the traditional electric measurements and boost the visibility of distribution network. The present work proposes a long short-term memory based NILM framework for the disaggregation of photovoltaic and electric vehicle profiles from the aggregated measurements in the distribution network. The comparative analysis has also been carried out with other machine learning classifiers Random Forest and k-Nearest Neighbors for the same dataset. The proposed approach has been rigorously validated for dataset with different input time frames to ensure robustness and reliability and found to achieve average F-scores in excess of 99.52% and 92.29% for identification of PV and EV profiles respectively.**

Keywords— Distributed energy resources, non-intrusive load monitoring, long short-term memory, photovoltaics, electric vehicles

I. INTRODUCTION

As per the COP21 Paris Agreement held on December 12, 2015, there is a requirement for greenhouse gas emissions to reach their peak by 2025, undergo a gradual reduction of 43% by 2030, and ultimately progress towards achieving net-zero emissions by 2050 [1]. At the COP28 UN climate summit convened in Dubai on December 12, 2023, an agreement has been ratified to shift away swiftly and ambitiously from reliance on fossil fuels. This initiative aims to move from agreement to action and achieve the goal of limiting global warming to 1.5°C, as articulated in the Paris Agreement. In this regard, the deal has been made to triple the integration of renewable energy resources and to submit stronger carbon cutting plans by 2025 [2].

A growing global focus on reaching net-zero emissions in the energy sector requires massive deployment of clean energy technologies such as solar photovoltaics (PVs), wind, electric vehicles (EVs). As per the latest findings of International Energy Agency, there has been huge surge in the deployment of solar PV and wind power driven by higher fossil fuel prices, global energy crisis and growing policy momentum which is set to continue till the global renewable electricity capacity reaches to 4500 GW [3,4]. Similarly, EVs play a crucial role in clean energy transitions by decarbonizing road transport, a sector responsible for approximately one-sixth of the global energy-related emissions [5]. The sale of EVs has experienced remarkable growth in recent years, accompanied by enhanced range, broader model availability, and improved performance. Specifically, electric car sales have achieved unprecedented records, and this momentum is anticipated to persist in the years to come [6].

The adoption of global initiatives related to net-zero carbon emissions and electrification of transportation has led to the massive deployment of distributed energy resources (DERs) particularly PVs and EVs in the distribution network. The rapid integration of these DERs has caused technical challenges to power grid operations modifying the electrical characteristics of the conventional distribution network [7,8]. Furthermore, the intermittent and stochastic behavior associated with renewable-based DERs and lack of visibility at the system level can lead to unanticipated disconnection of DER units, as evidenced by the UK power outage on 9th August 2019 [9]. As such, the proliferation of millions of DERs necessitates effective management strategies and improved visibility at the distribution system level. Also, incorporating these new technologies in the present distribution system demands for innovative methods to interact between the utility and the consumers via internet-connected devices. These smart appliances along with smart metering will enable the network operators to monitor the infrastructure on the low-voltage side of the distribution network. The real-time monitoring will thus enhance the observability of the network. Moreover, using these advanced monitoring techniques will increase the accuracy of the electrical maps, encompassing the locations of DER and network facilitates the implementation of effective energy management strategies [10].

Load monitoring methods can be classified into two categories: intrusive load monitoring (ILM) and non-intrusive load monitoring (NILM). ILM entails the installation of sensors on each individual device, enabling the acquisition of the electrical profile associated with each load.

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

NILM however, involves aggregated load measurements representing load profiles of several electrical devices connected to a common power supply. Thus, NILM methods focus on classification/disaggregation of individual loads from aggregated electrical measurements [11]. Early studies on NILM have mainly focused on the disaggregation of conventional household appliances. The majority of NILM algorithms documented in the literature have emphasized on distinguishing residential loads at the end-user distribution side. To date, only a limited number of research papers have addressed the identification of DERs on the low-voltage side of the distribution network. A few research studies that have already tackled the challenges of disaggregating EV charging patterns and identifying PV activity based on power consumption data from smart meters can be found in the literature. In this context, a predominant number of papers have used the public Pecan Street dataset [12]. In [13,14], NILM methods have been developed utilizing traditional machine learning algorithms like multilayer perceptron (MLP), random forest (RF), k-Nearest neighbors (kNN) and artificial neural network (ANN) for the identification of EV and PV profiles from smart meter net demand data. A similar approach for the detection of EV and PV from smart meter data using convolutional neural network (CNN) and MLP can be found in [15]. The primary shortcomings of the state-of-the-art techniques related to NILM for EV and PV revolve around inadequate transparency in disclosing critical information, including precise house locations i.e., *dataid* which is the unique identifier for each house, number of houses, exact span of data employed in training and testing. Hence, it becomes challenging to reproduce and compare results for various classification algorithms. Moreover, the inability of conventional machine learning classifiers (RF and kNN) to capture complex non-linear dynamics and increased sensitivity to noisy data during EV charging and PV generation leads to reduced accuracy in identification tasks. Similarly, the inability of MLP to capture spatial relationship in PV and EV data and requirement of feature engineering results in poor performance. Also, the constraint of CNNs in capturing long-term temporal dependencies within time-series data, coupled with the necessity for extensive labelled data, leads to suboptimal performance.

In this regard, the present work develops a long short-term memory (LSTM) based NILM framework for the disaggregation of EV charging and PV generation profiles in a distribution network with sufficient transparency in the dataset used. The LSTM network has the ability to capture temporal long-term dependencies and stochastic variations associated with the penetration of different EV and PV on the low-voltage side of the distribution network [16]. Also, LSTM exhibits the ability to adjust to varying lengths of input sequences, rendering it well-suited for a range of diverse scenarios. For distribution networks having scenarios with fluctuations profiles and intermittent behaviors, LSTMs demonstrates reduced sensitivity to noisy data. Moreover, unlike conventional machine learning algorithms, LSTM is free from challenges related to generalizability when applied to larger datasets [17]. Motivated by the significance of attaining improved accuracy in the identification of EV and PV profiles from the smart meter net demand data with robustness to noise, a LSTM-based NILM approach is adopted in the present work.

The contributions/highlights of the proposed work can be summarized as:

1) A LSTM-based NILM framework has been proposed for the effective identification of EV and PV profiles from the smart meter power consumption data.
2) Evaluation on larger dataset (Pecan Street Dataport) comprising of 6 months of data from 25 households in New York using LSTM classifier and comparison with existing RF and kNN based NILM techniques.
3) The proposed NILM approach is proved to be effective in identifying PV and EV profiles for dataset with different input time spans i.e., daily, weekly and monthly and can achieve F-scores in excess of 99.52% and 92.29% respectively.

The remaining paper is organized in four sections. Section II provides the brief description of dataset and describes the proposed LSTM-based NILM framework. Section III presents the performance of the LSTM classifier and its comparative analysis. Finally, Section IV presents the conclusions and directions for future work.

II. Methodology

The methodology for identification of EV and PV profiles in distribution network using LSTM basically involves two parts: (i) Data acquisition and pre-processing (ii) Implementation of proposed LSTM-based NILM framework.

A. Data Acquisition and Pre-processing

In this study, PV and EV profiles sourced from the public dataset, ‘Pecan Street Dataport’, recognized as the world's largest repository for residential energy usage data, are used to simulate the training and testing data. Time-series dataset with 15-minute energy for 15 individual homes located across New York region are used [12]. The detailed description of dataset for individual households with PV and EV integration is shown in Table I. The Pecan Street Dataset undergoes filtration to disaggregate DER electrical profiles, specifically those related to PV and EV along with aggregated household load measurement. This curated data is essential for both the

TABLE I. SUMMARY OF NEW YORK HOUSES WITH PV AND EV INTEGRATION

Dataset Description	**Dataset for PV Generation**	**Dataset for EV Demand**
Number of individual houses	With PV- 14 Without PV- 11	With EV- 5 Without EV- 20
Exact location of House IDs with PV and EV respectively	Ithaca- 387, 914, 950, 1222, 3000, 3488, 3517, 5058, 5587, 5679, 5997 Groton- 1240 Trumansburg- 142 Brooktondale- 27	Ithaca-1222, 3000, 5679 Brooktondale- 27 Lansing- 9053
Frequency	15-minute	15-minute
Data span	6 months of data of year 2019 (May, June, July, Aug, Sep, Oct)	6 months of data of year 2019 (May, June, July, Aug, Sep, Oct)

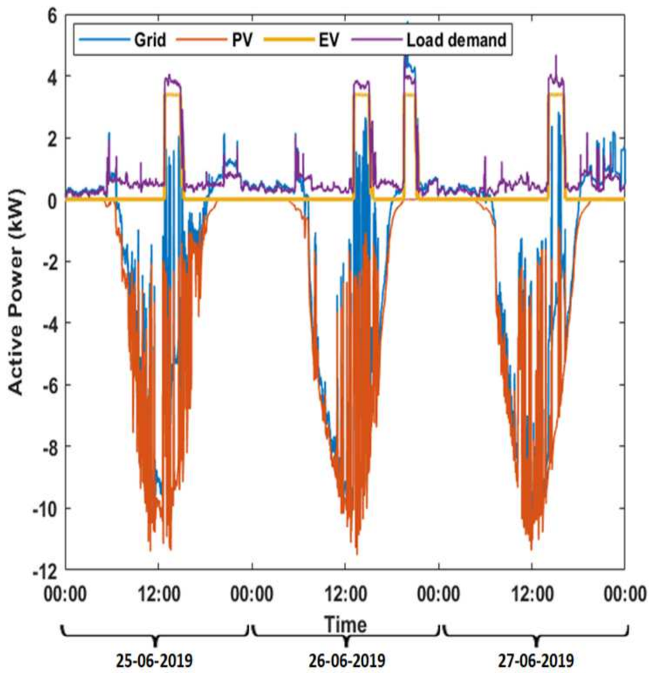

Fig. 1. Electrical profiles for House ID 27 from 25th to 27th June 2019.

training and testing phases. Consequently, different electrical profiles are derived, as depicted in Fig.1 for House ID 27 in New York from 25th to 27th June 2019. The EV charging process and the power generation of the PV system are depicted proportionally along with the energy provided by the grid and load demand. It can be observed that the identification of EV charging profile becomes more challenging in the presence of PV generation.

B. *Proposed LSTM-based NILM framework*

LSTM belongs to the family of recurrent neural networks (RNNs), designed to adeptly capture and learn long-term dependencies in data. However, vanishing gradient problem in traditional RNNs poses challenges in effectively learning and retaining information over long sequences. LSTMs have a more complex structure consisting of the memory cell state (C_t), input gate (I_t), forget gate (F_t), and output gate (O_t) as depicted in Fig.2 below which allow LSTM to selectively memorize or forget long term data. The memory cell state stores information over long periods. The input gate controls the flow of information into the memory cell whereas forget gate manages the removal of information from the memory cell. The output gate controls the information that is output from the memory cell and entering the next hidden state. At time t, X_t represents the input to an LSTM unit, while H_t denotes the output of the hidden layers of the LSTM unit. The weight values connecting LSTM input (X) and output (H) respectively to input gate, forget gate, output gate, memory cell state are denoted as W_{IX}, W_{FX}, W_{OX}, W_{CX} and W_{IH}, W_{FH}, W_{OH}, W_{CH}. The bias vectors for the input gate, forget gate, output gate and memory cell are denoted as b_I, b_F, b_O and b_C respectively [17].

$$I_t = sigmoid(W_{IX}X_t + W_{IH}H_{t-1} + b_I) \quad (1)$$

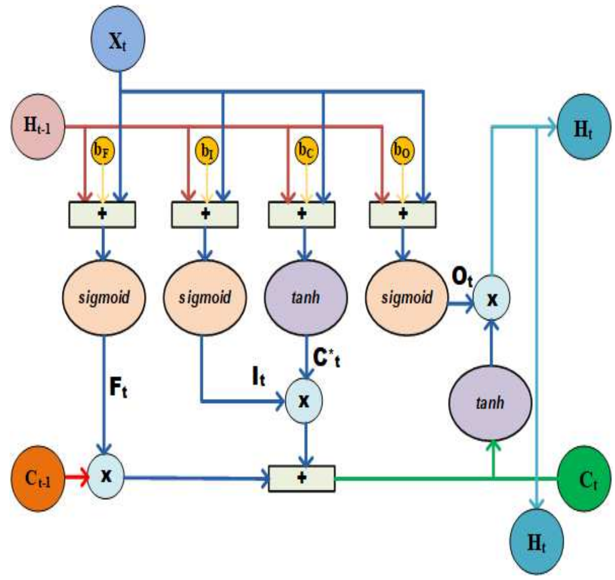

Fig. 2. LSTM network architecture.

$$F_t = sigmoid(W_{FX}X_t + W_{FH}H_{t-1} + b_F) \quad (2)$$

$$O_t = sigmoid(W_{OX}X_t + W_{OH}H_{t-1} + b_O) \quad (3)$$

$$C_t^* = tanh(W_{CX}X_t + W_{CH}H_{t-1} + b_C) \quad (4)$$

$$C_t = C_t^* \odot I_t + C_{t-1} \odot F_t \quad (5)$$

$$H_t = tanh(C_t) \odot O_t \quad (6)$$

where, *sigmoid*(.) denotes the sigmoid function, *tanh*(.) denotes the hyperbolic tangent function, the operator $\odot$ denotes the element-wise product.

Fig.3 depicts the flowchart of the proposed LSTM-based NILM framework. The NILM process initiates with the acquisition of data from Pecan Street Dataport for households in New York region followed by pre-processing. The data processed above is then randomly divided into training and testing scenarios. The dataset has been assessed across varying training/testing ratios, specifically (90%)/(10%) (80%)/(20%), (70%)/(30%) and (50%/50%). The accuracy of the LSTM classifiers i.e., LSTM_EV and LSTM_PV for the identification of EV and PV respectively has been checked for the above training-testing ratios. However, among the various combinations explored, the LSTM classifiers demonstrated notable accuracy, particularly with a split of 80% training and 20% testing scenarios. A summary of the hyperparameters selected during the training of each LSTM classifiers is provided in Table II. The hyperparameters were chosen through pilot runs, aiming to maximize the accuracy of LSTM classifiers.

Further, to assess the effectiveness of LSTM-based NILM in accurately detecting the integration of EV and PV profiles within the low-voltage distribution network, various metrics including precision, recall, accuracy, and F-Score are employed. These metrics typically rely on four parameters: true positive (TP), true negative (TN), false positive (FP), and false negative (FN) [13].

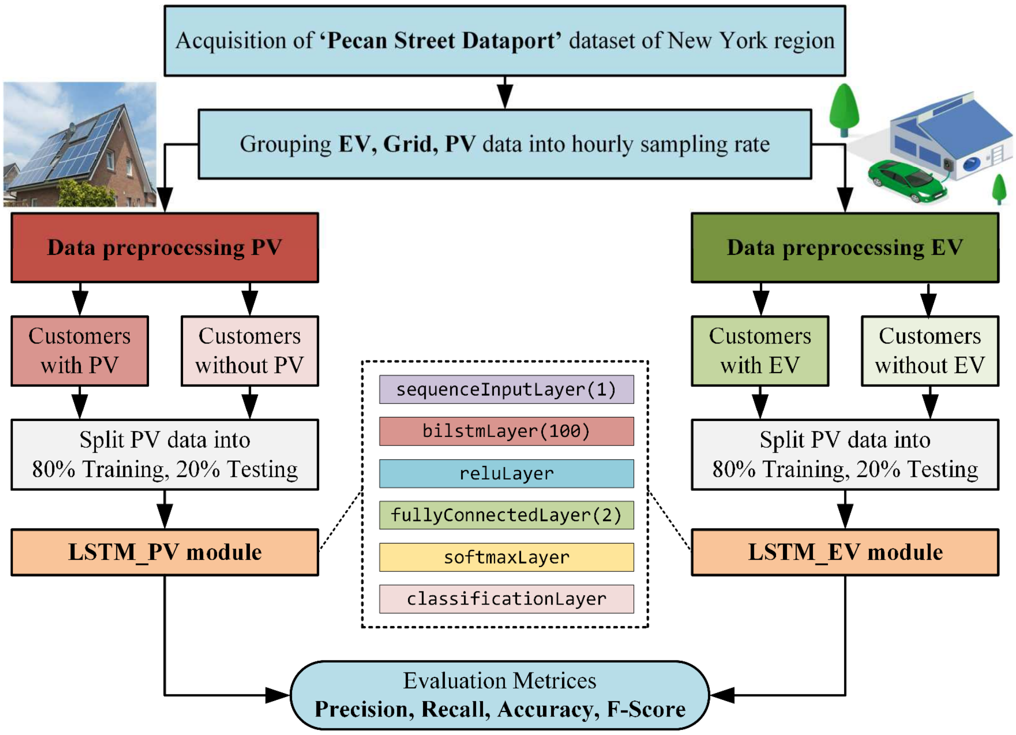

Fig. 3. Flowchart of proposed LSTM-based NILM framework.

TABLE II. SUMMARY OF HYPERPARAMETERS FOR TRAINING THE LSTM_PV AND LSTM_EV

Parameters	**LSTM_PV**	**LSTM_EV**
Input layer size	1	1
No. of Hidden Units	100	100
Input Activation Function	ReLU	ReLU
No. of Fully Connected Layers	2	2
Output Activation Function	Sigmoid	Sigmoid
Solver	Adam	Adam
Maximum Epochs	400	600
Mini Batch Size	100	100
Initial Learning Rate	0.001	0.001
Target Labels	PV, No PV	EV, No EV

$$Precision = TP/(FP + TP) \tag{7}$$

$$Recall = TP/(FN + TP) \tag{8}$$

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \tag{9}$$

$$F - Score = 2TP/(2TP + FP + FN) \tag{10}$$

III. PERFORMANCE ANALYSIS

The efficacy of the proposed LSTM-based NILM technique for identifying EV and PV profiles from aggregated measurements of smart meter data has been evaluated in this section. In this regard, all the time series simulations have been tested on an Intel Core i5-2400 @ 3.10 GHz PC MATLAB R2021a, version 9.10. The performance assessment of the LSTM_PV and LSTM_EV modules for different input time spans (daily, weekly and monthly) is discussed in the following subsections.

A. Identification of PV and EV Profiles

A comparative analysis of the evaluation metrices for LSTM-based NILM in classifying PV and EV has been carried out with conventional machine learning algorithms RF and kNN as depicted in Fig.4 and Fig.5 respectively for daily input time spans. Overall, the proposed LSTM-based NILM approach outperformed both RF and kNN classifiers. It can be observed that the F-Score takes into account both FP and FN, providing a more rigorous evaluation than accuracy, especially in the context of datasets with a substantial imbalance. As depicted in Fig.4, LSTM outperformed both RF and kNN, achieving F-Score of 99.22%, while RF and kNN obtained F-Scores of 94.62% and 90.92% respectively during the classification of PV generation profiles. Similarly, for the classification of EV load profile, LSTM provided F-Score of 93.99% when compared to 92.66% and 89.91% for RF and kNN as in Fig.5.

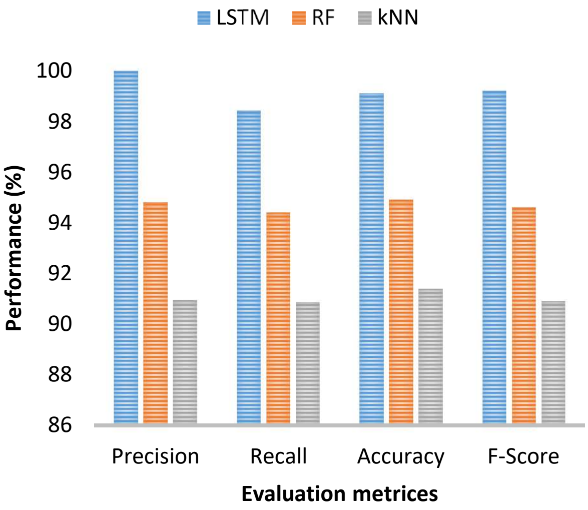

Fig. 4. Evaluation metrices for PV.

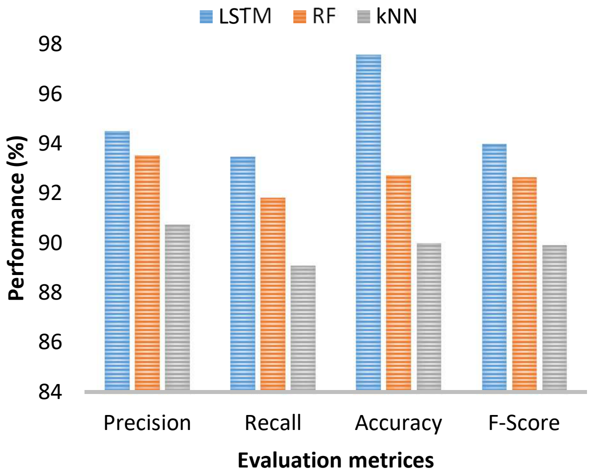

Fig. 5. Evaluation metrices for EV.

While the above conventional metrices illustrate the performance of each classifier models in NILM studies, misclassification may arise during the interpretation of imbalanced data as input to those classification models. It is observed that the imbalanced dataset can create an illusion of good classification accuracy when either TP or TN values dominate. Hence, it is advisable to employ a confusion matrix to obtain a more comprehensive understanding of the obtained predictions. This confusion matrix serves as a valuable tool for assessing the effectiveness of a classification model by offering a detailed breakdown of its performance. It illustrates specific instances in which the model encounters confusion while making predictions. Employing this metric to assess a classification algorithm offers the benefit of clearly pinpointing any deviations or biases the algorithm may exhibit towards certain class over the other.

Fig.6 shows the confusion matrix with and without PV generation. Considering the daily input time span, it can be observed that the LSTM_PV classifier is able to classify 98.45% of the households with PV generation correctly. Moreover, the LSTM_PV classifier classifies 100% of the houses without PV generation. Similarly, the LSTM_EV classifier detects 93.48% of the households with EV load demand and 98.64% of the houses with no EV power consumption as depicted in Fig.7.

B. *Performance Evaluation of Proposed Technique for Different Input Time Spans*

The LSTM classifiers' classification accuracies are evaluated by comparing them across various input time spans, including daily, weekly, and monthly intervals. Upon examination of the classification accuracies of LSTM_PV for detecting PV from power consumption data, as detailed in Table III, it is evident that the classification accuracy diminishes. This decline is noticeable particularly when considering a shorter input time span. The classification accuracies for the monthly input period are significantly higher as compared to weekly and daily inputs.

For the detection of PV systems, the LSTM_PV classifier achieves F-Score of 100% for the monthly input time span, while comparable F-Scores of 99.22% and 99.34% have been observed for the daily and weekly input time spans, respectively. Within the daily input span, 1.55% of households equipped with a PV system are misclassified. However, within the weekly input period, this figure decreases to 1.32%. Moreover, the LSTM_PV classifier demonstrates good performance for all input samples from households without PV, regardless of the input time span. In contrast to PV detection, LSTM_EV shows reduced performance for the identification of EV charging patterns for the different input time spans.

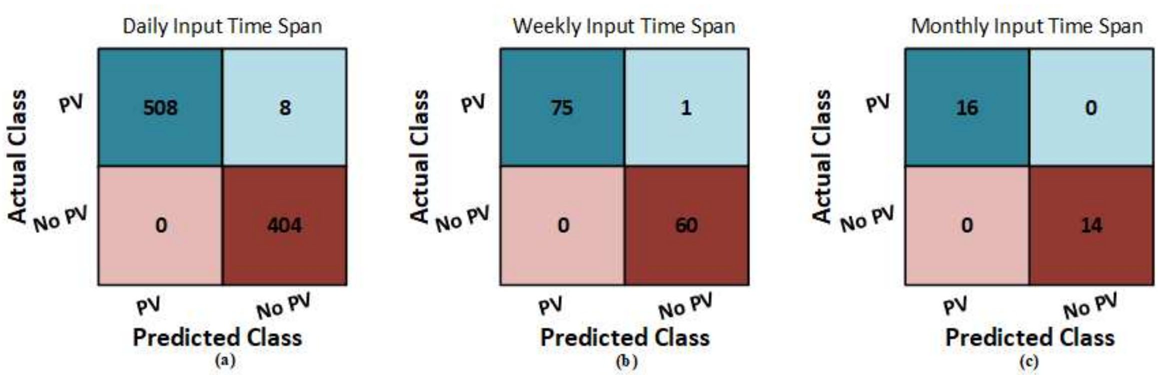

Fig. 6. Confusion matrix for LSTM_PV.

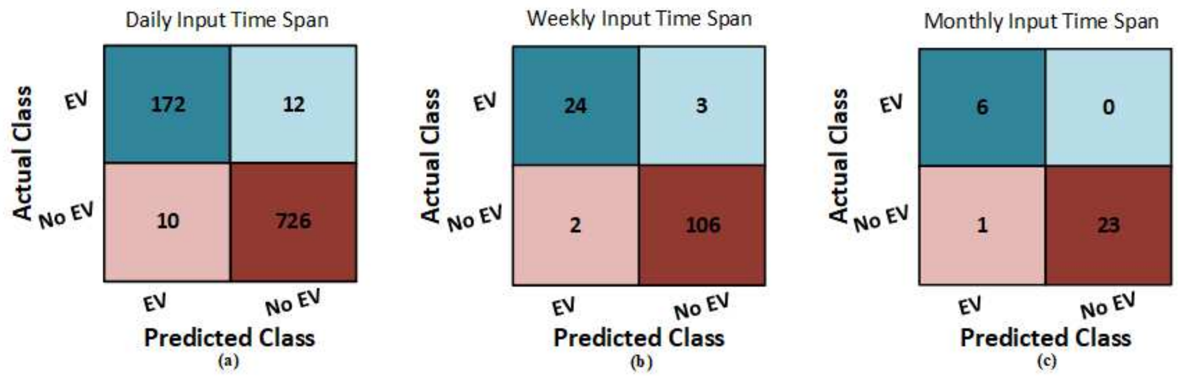

Fig. 7. Confusion matrix for LSTM_EV.

LSTM_EV module attains F-Score of 92.31% for monthly whereas F-Scores of 90.57% for weekly and 93.99% for daily input time spans.

TABLE III. ACCURACY OF LSTM_PV AND LSTM_EV MODULES FOR DIFFERENT INPUT TIME SPANS

LSTM classifier Modules	**Input time spans**		
	Daily	**Weekly**	**Monthly**
LSTM_PV	99.22%	99.34%	100%
LSTM_EV	93.99%	90.57%	92.31%

IV. CONCLUSION

The present work proposes a LSTM-based NILM framework for disaggregating EV load demand and PV generation profiles from the aggregated measurements within a distribution network. The dataset 'Pecan Street Dataport' employed in the present study ensures ample transparency in the analysis. Moreover, the ability of the proposed LSTM model to capture temporal dependencies and stochastic variations associated with the penetration of DERs in the distribution network has significantly improved the performance as compared to other conventional machine learning techniques. Significantly improved F-Scores have been obtained for identifying PV generation profiles (99.52%) and EV load profiles (92.29%) using LSTM approach, thus increasing the observability of distribution network operators. Future work is planned on imparting resilience to distribution network against extreme weather events by incorporating climatic and environmental data.

ACKNOWLEDGMENT

This work is supported by the Engineering and Physical Sciences Research Council [grant number EP/Y005376/1] – VPP-WARD Project (https://www.vppward.com). The authors would also like to thank Pecan Street Inc.(https://dataport.pecanstreet.org) for making freely available residential data with PV and EV integration.

REFERENCES

[1] T. Jayaraman, "The Paris Agreement on Climate Change: Background, Analysis, and Implications," *Rev. Agrar. Stud.*, vol. 5, no. 2, 2015.

[2] UN Climate Press Release, "COP28 Agreement Signals 'Beginning of the End' of the Fossil Fuel Era," 2023. [Online]. Available:https://unfccc.int/news/cop28-agreement-signals-beginning-of-the-end-of-the-fossil-fuel-era.

[3] IEA, "World Energy Outlook 2023," 2023. [Online]. Available: https://www.iea.org/reports/world-energy-outlook-2023.

[4] IEA, "Renewables 2023," 2024. [Online]. Available: https://www.iea.org/reports/renewables-2023.

[5] Y. Cao *et al.*, "Towards cyber security for low-carbon transportation: Overview, challenges and future directions," *Renew. Sustain. Energy Rev.*, vol. 183, pp. 0–3, 2023.

[6] IEA, "Global EV Outlook 2023," 2023. [Online]. Available: https://www.iea.org/reports/global-ev-outlook-2023.

[7] E. S. Obobisa, "Achieving 1.5 °C and net-zero emissions target: The role of renewable energy and financial development," *Renew. Energy*, vol. 188, pp. 967–985, 2022.

[8] A. G. Rameshrao, E. Koley, and S. Ghosh, "An optimal sensor location based protection scheme for DER-integrated hybrid AC/DC microgrid with reduced communication delay," *Sustain. Energy, Grids Networks*, vol. 30, p. 100680, 2022.

[9] Wilde S. Ofgem, "9 August 2019 power outage report," *Technical report*, 2020.

[10] S. Bahramara, A. Mazza, G. Chicco, M. Shafie-khah, and J. P. S. Catalão, "Comprehensive review on the decision-making frameworks referring to the distribution network operation problem in the presence of distributed energy resources and microgrids," *Int. J. Electr. Power Energy Syst.*, vol. 115, p. 105466, 2020.

[11] I. Abubakar, S. N. Khalid, M. W. Mustafa, H. Shareef, and M. Mustapha, "Application of load monitoring in appliances' energy management – A review," *Renew. Sustain. Energy Rev.*, vol. 67, pp. 235–245, 2017.

[12] Pecan Street Inc., "Pecan Street Dataport, Residential Data (2024)." https://dataport.pecanstreet.org/academic.

[13] A. F. Moreno Jaramillo, J. Lopez-Lorente, D. M. Laverty, J. Martinez-del-Rincon, D. J. Morrow, and A. M. Foley, "Effective identification of distributed energy resources using smart meter net-demand data," *IET Smart Grid*, vol. 5, no. 2, pp. 120–135, 2022.

[14] A. F. M. Jaramillo *et al.*, "Distributed Energy Resources Electric Profile Identification in Low Voltage Networks Using SupervisedMachine Learning Techniques," *IEEE Access*, vol. 11, pp. 19469–19486, 2023.

[15] M. Neubert, O. Gnepper, O. Mey, and A. Schneider, "Detection of Electric Vehicles and Photovoltaic Systems in Smart Meter Data," *Energies*, vol. 15, no. 13, 2022.

[16] A. G. Rameshrao, E. Koley, and S. Ghosh, "A LSTM-based approach for detection of high impedance faults in hybrid microgrid with immunity against weather intermittency and N-1 contingency," *Renew. Energy*, vol. 198, no. August, pp. 75–90, 2022.

[17] H. Hwang and S. Kang, "Nonintrusive Load Monitoring Using an LSTM With Feedback Structure," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–11, 2022.