Letters

Detection and Diagnosis of Data Integrity Attacks in Solar Farms Based on Multilayer Long Short-Term Memory Network

Fangyu Li, Qi Li, Jinan Zhang, Jiabao Kou, Jin Ye, WenZhan Song, and Homer Alan Mantooth

***Abstract*—Photovoltaic (PV) systems are becoming more vulnerable to cyber threats. In response to this emerging concern, developing cyber-secure power electronics converters has received increased attention from the IEEE Power Electronics Society that recently launched a cyber-physical-security initiative. This letter proposes a deep sequence learning based diagnosis solution for data integrity attacks on PV systems in smart grids, including dc–dc and dc–ac converters. The multilayer long short-term memory networks are used to leverage time-series electric waveform data from current and voltage sensors in PV systems. The proposed method has been evaluated in a PV smart grid benchmark model with extensive quantitative analysis. For comparison, we have evaluated classic data-driven methods, including K-nearest neighbor, decision tree, support vector machine, artificial neural network, and convolutional neural network. Comparison results verify performances of the proposed method for detection and diagnosis of various data integrity attacks on PV systems.**

***Index Terms*—Data integrity attack (DIA), deep learning, machine learning, smart grids, solar inverter.**

I. Introduction

POWER grids have become more vulnerable to cyber threats than before [1]. In response to this emerging concern, developing cyber-secure power electronics converters has received increased attention from the IEEE Power Electronics Society that recently launched a cyber-physical-security initiative. There are two main reasons: First, to improve the operation efficiency and eliminate human intervention, the power grid has been more and more connected, resulting in increasing challenges in reliability, security, and stability. Second, a significantly increasing amount of distributed energy resources, such as solar photovoltaic (PV) [2], that are typically power electronics converters that are being incorporated into smart grids.

Data integrity attacks (DIAs) attempt to insert or alter data to mislead the victim systems to make wrong decisions [3]. A considerable amount of literature works have been conducted in providing analysis of DIA on legacy power systems, such as dc microgrids [4], smart grids [5], etc.

To mitigate the vulnerability, model-based and data-driven methods have been proposed [6]. However, model-based methods that rely on the accurate mathematical models of the healthy systems are hard to be used in real applications because of an unavoidable model-reality mismatch for the complexity of power electronics based smart grids. Data-driven methods, on the other hand, employing measured data without an explicit mathematical model, are currently receiving attentions [7], [8]. To date, the grid security heavily focuses on the system level, and almost neglects the device level, particularly power electronics converters, which has not been well addressed [9]. In our previous work [10], we detected and diagnosed a variety of cyber-physical threats for distribution systems with PV farms, including cyber attacks on the solar inverter controller, cyber attacks on relays/switches, and other faults (e.g., short circuit faults).

Here, we propose a data-driven deep sequence learning method for automatic DIA diagnosis of smart grids with PVs. Unlike our previous approach, we propose to use only one voltage sensor and one current sensor at the point of common coupling (PCC) for PV systems to detect and diagnose more than 3000 cases of cyber attacks on dc–dc and dc–ac converters. Here, we assume that the waveform sensor at the PCC is secure and trustworthy. In real applications, its communication channel can be encrypted to ensure the security of waveform data. We propose to use multilayer long short-term memory (MLSTM) networks [11] to handle intrinsic sequential characteristics of streaming sensor data. Five data-driven methods are engaged as comparison methods, which are K-nearest neighbor (KNN), decision tree (DT), support vector machine (SVM), artificial neural network (ANN), and convolutional neural network (CNN). Our contributions can be summarized as follows.

1) This is one of the first attempts to analyze the DIA impacts on solar farms using electrical waveform data at PCC. We propose a well-suitable deep learning strategy to solve the issue.

Manuscript received May 15, 2020; revised June 22, 2020 and July 28, 2020; accepted August 15, 2020. Date of publication August 19, 2020; date of current version October 30, 2020. This work was supported in part by the U.S. Department of Energy's Solar Energy Technology Office under Award No. DE-EE0009026 and in part by the U.S. National Science Foundation under Grant ECCS-1946057. *(Corresponding author: Jin Ye.)*

Fangyu Li is with the Department of Electrical and Computer Engineering, Kennesaw State University, Marietta, GA 30060 USA (e-mail: fli6@kennesaw.edu).

Qi Li, Jinan Zhang, Jin Ye, and WenZhan Song are with the Center for Cyber-Physical Systems, University of Georgia, Athens, GA 30602 USA (e-mail: qi.li2@uga.edu; jinan.zhang@uga.edu; jin.ye@uga.edu; wsong@uga.edu).

Jiabao Kou is with the School of Electrical Engineering and Automation, Harbin Institute of Technology, Harbin 150001, China (e-mail: koujiabao_hit@163.com).

Homer Alan Mantooth is with the Department of Electrical Engineering, University of Arkansas, Fayetteville, AR 72701 USA (e-mail: mantooth@uark.edu).

Color versions of one or more of the figures in this letter are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TPEL.2020.3017935

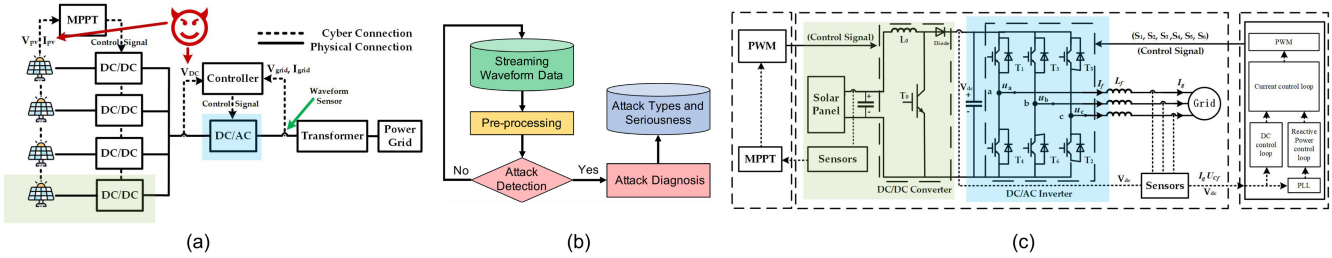

Fig. 1. (a) Cyber attacks to a solar farm in a power grid. (b) Proposed DIA detection and diagnosis workflow. The streaming data are acquired from the waveform sensor annotated in (a). (c) Cyber-physical models of PV systems.

2) We propose a novel deep learning framework to tackle a variety of DIA on dc–dc and dc–ac converters in PV systems (over 3000 attack scenarios).
3) Both attack detection and diagnosis have been developed, leading to sensitive attack awareness as well as accurate attack type and seriousness analysis.

II. DIA Detection and Diagnosis

DIA disrupt the system by manipulating data or introducing corruption. Attacks are assumed to happen between the end devices (or sensors) and the control center, e.g., smart grid measurement data can be attacked in conjunction with the solar panel measurement data, as shown in Fig. 1(a). DIA are usually defined as mixing the original data/measurements vector with a malicious vector [4], [5]

$$\mathbf{Z} = \alpha * \mathbf{W} + \mathbf{Z}_0 \tag{1}$$

where $\mathbf{Z}$ is the compromised data vector that is eventually used by the system, $\mathbf{Z}_0$ is the true measurement, $\mathbf{W}$ is a general compromised data vector, which can be independent or determined by $\mathbf{Z}_0$, and α is a multiplicative factor that defines the weight of the attack vector. The proposed attack detection and diagnosis workflow aims to achieve a real-time and effective attack detection as well as identify the attack types and seriousness when an attack occurs based on monitoring the electric waveforms of the smart grid, as shown in Fig. 1(b).

A. Problem Formulation

The smart grid measurement at a certain time point is influenced by the states of its previous time points. The electric waveform is recorded for every time interval, which can be represented using a time series model

$$x(t) = \mathcal{G}(x(t-\delta_t), x(t-\delta_t), \ldots) + \epsilon(t) \tag{2}$$

where $x(t)$ is the sensor reading at time t, $\mathcal{G}$ denotes the function that correlates the previous data samples to the present $x(t)$, δ_t is the system time interval, which is 1 ms in this letter, since 1-kHz sampling rate is adopted. In addition, $\epsilon(t)$ is the residual error, defined as $\epsilon(t) = \mathcal{F}(\lambda_1, \lambda_2, \ldots) + \varepsilon(t)$, where, without losing generality, $\mathcal{F}$ is a function describing controlling factors λ_k, which in our study indicate the critical variables in the dc–dc and dc–ac controllers, and ε is the random noise with a zero mean.

B. Attack Detection Model

There are various states of PV systems, including the normal state and underattack states with various attack types. Because it is difficult to accurately detect and identify various types of attacks simultaneously, we propose to first focus on detecting whether the PV system is under attack or not. We apply the one-class detection as the attack detection model, which has been widely applied for outlier detection to accurately classify the normal and underattack states [12]. Training one-class detection model only requires normal data, which is an advantage for a potentially large number of attacks.

Our proposed detection model is expressed as $g(\mathbf{x}(t)) = \text{sgn}(\mathcal{G}^*(\mathbf{x}(t)) - \rho)$, where $\mathbf{x}(t)$ denotes a vector of time series of smart grid sensor data from $t - L$ to t. $\mathcal{G}^*$ is the trained one-class model. ρ is the detection error threshold (DET), so if the prediction error is larger than DET, it may indicate an anomaly. As a sign function, $\text{sgn}(\alpha) := \begin{cases} 1 & \text{if } \alpha \geqslant 0 \\ -1 & \text{if } \alpha < 0 \end{cases}$.

C. Attack Diagnosis Model

The attack identification is actually a classification model based on a multiclassification model to identify attack types. Nevertheless, the seriousness of the same type of attack is also important, but has not been well explored. In addition, the cross-entropy loss function often in practice means a cross-entropy loss function for classification problems and a mean squared error loss function for regression problems [13]. Therefore, to analyze not only the attack types but also the seriousness, we propose a cross-entropy loss between the empirical distribution defined by the training set and the probability distribution defined by the model as follows:

$$J(\theta) = -\mathbb{E}_{x,y\sim\hat{p}_{\text{data}}} \log p_{\text{model}}(\mathbf{y}|\mathbf{x}). \tag{3}$$

D. Multilayer LSTM Based Deep Sequence Learning

Since we try to model electric waveform data, which have complicated nonlinear temporal characteristics, we leverage the LSTM model. The structure of recurrent neural network (RNN) utilizes the information memory at the previous time to apply to the current sequence data prediction. However, RNN training long sequences in a multilayer network will generate gradient disappearance and explosion [14]. While LSTM uses the guided gates for selectivity, remembering both short and long-term behaviors across many time series, which effectively solves the problem of gradient diffusion and explosion. Fig. 2 shows the proposed MLSTM architecture, which not only remembers sequential information but also carries out more rigorous screening of time information. So, we can generalize the behavior complexity of the PV system without a huge dataset. Specifically, hyberparameters for MLSTM models are batch size = 128,

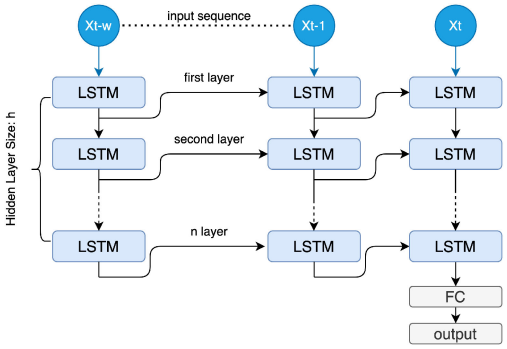

Fig. 2. Proposed multilayer LSTM architecture.

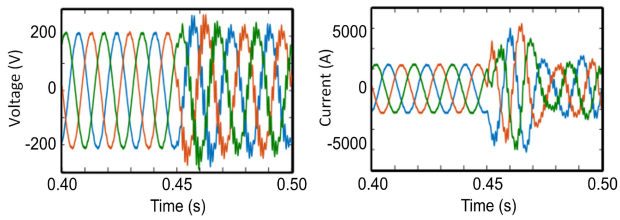

Fig. 3. Electric waveforms (voltage and current) simulations of a dc–ac controller attack.

learning rate = 0.001, hidden size = 32, optimizer = *Adam*, number of layers = 2 (detection) / 5 (diagnosis), which are obtained through experiments and trials. Note that CNN shares most of the hyperparameters of MLSTM in our study.

III. Cyber Attack Modeling for PV Systems

To evaluate the proposed method, we simulate comprehensive cyber attacks using a benchmark PV solar farm model, which has been used in our previous work [10]. The main power grid is modeled as an ideal voltage source, and the load is linear. One rate voltage of 260 V/25 kV, 400 kVA, transformer connects the PV farm, which includes four dc–dc converters and one dc–ac inverter, to the power grid. The topology of one converter circuit is shown in Fig. 1(c).

Here, cyber attacks on the dc–dc controller sensor only change the current and voltage of the PV panel. Following the DIA model in (1), α_V and α_I represent fake measurement coefficient of voltage and current in the PV panel. $(\alpha_V, \alpha_I) \in [(0,0),(2,3),(2,0.3),(0.5,3),(0.5,0.3)]$. For the dc–ac controller, the cyber attacks inject a time delay into sensor feedback, $t_{\text{delay}} \in [0,4,6,8,10,12,14\,\text{ms}]$. Considering the uncertainty of cyber attacks, the attacks happened at different time are simulated in our model, such as phase angles $0°, 30°, 60°, 90°, 120°, 150°,$ and $180°$. Besides, to test the robustness of the proposed method toward different conditions, we also consider the irradiation impact on the power generation. The irradiation on the PV panel varies in range of 900, 941, 967, 988, and 1000 w/m^2. Thus, more than 3900 training samples are simulated. The waveform at the PCC is obtained to verify our proposed method. The sampling frequency is 1 kHz, and 0.7-s data are captured for each scenario, which have 701 samples.

Fig. 3 demonstrates a dc–ac controller attack simulation, where both voltage and current show obvious distortions. Because of various attack seriousness parameters, DIA result in not only strong anomalies but also minor waveform distortions, which makes the attack detection and diagnosis challenging.

IV. Evaluation

A. Comparison Models

To validate the performances of the proposed MLSTM method, classic machine learning and deep learning models, such as KNN, SVM, DT, ANN, and CNN, are compared, which are powerful data-driven methods with a wide range of applications [13]. For the machine learning models, data features, such as frequency, amplitude, phase angle (because of ac waveform), spectrum properties, are extracted. For deep learning models, data streams are managed to be fed into models. We implemented them through Pytorch (1.3.1) [15] and Sklearn (0.22.1) [16] on a Ubuntu 16.04 server (CPU: i7-6850 K, 3.60 GHz, RAM 64 GB) armed with GPU (GeForce GTX 1080 Ti). For the validation purpose, we utilize a ten-fold randomized cross-validation with 80% training data and 20% testing data for the model training. To quantitatively evaluate method performances, we employ accuracy, precision, recall, and F_1 score, which are obtained from the confusion matrix for detection and classification evaluation [17]. We adopt an offline training and online testing strategy.

B. Attack Detection Performance Evaluation

In the attack detection stage, all data-driven models are trained under the one-class model structure, which is simple with efficient computations. So, the attack detection model has ensured its applicability in practice and, thus, achieves a real-time manner. Table I lists the evaluation metrics: accuracy, recall, precision, and F_1 score. In addition, in order to further characterize the model sensitivity, we also test the analysis window with different window lengths. It is clear that the proposed MLSTM achieves the best performances in terms of all metrics, with only two layers. SVM cannot achieve good performance, maybe because the data structure is too complicated. KNN and DT show acceptable performances, but not as good as CNN and MLSTM. Due to the shallow model depth, ANN does not show ideal performances, whereas CNN achieves very good performances also only with two layers. Compared with CNN, MLSTM achieves high detection accuracy even when the window size is 50 (0.05 s), and with longer analysis window length, MLSTM can even do better.

C. Attack Diagnosis Performance Evaluation

Different from attack detection where only normal and abnormal data are labeled, attack diagnosis requires more detailed data analysis. Because of the data unbalance that normal condition has a large amount of available data while each attach scenario only has limited available data, accuracies of all data-driven models are high, but some have really bad recall, precision, and F_1 scores, as listed in Table II. However, MLSTM and CNN still show the advantages of deep learning models even with five layers. Besides the slightly better performances in terms of metrics compared with CNN, MLSTM actually has another advantage. Fig. 4 displays the training and testing performances of CNN and MLSTM with the same analysis window length. MLSTM shows a smoother loss curve, which means it potentially has

TABLE I
DETECTION PERFORMANCE EVALUATION USING METRICS (ACCURACY, F_1, RECALL, AND PRECISION)

Window Size	50	80	100	140	160	200
SVM	0.79/0.47/0.31/0.96	0.77/0.43/0.28/0.97	0.75/0.42/0.27/0.96	0.71/0.36/0.22/0.96	0.69/0.36/0.22/0.97	0.67/0.34/0.21/0.98
KNN	0.90/0.83/0.83/0.84	0.91/0.85/0.86/0.85	0.91/0.87/0.87/0.87	0.90/0.87/0.87/0.87	0.89/0.86/0.87/0.86	0.88/0.86/0.85/0.87
DT	0.92/0.86/0.81/0.92	0.92/0.86/0.82/0.92	0.91/0.87/0.86/0.88	0.91/0.89/0.91/0.87	0.93/0.91/0.92/0.91	0.93/0.92/0.94/0.89
ANN	0.85/0.85/0.81/0.85	0.91/0.91/0.90/0.91	0.91/0.91/0.90/0.91	0.85/0.85/0.85/0.86	0.82/0.82/0.80/0.82	0.75/0.73/0.70/0.78
CNN	0.93/0.93/0.91/0.93	0.97/0.97/0.97/0.97	0.97/0.97/0.97/0.97	0.94/0.94/0.93/0.94	0.95/0.95/0.95/0.95	0.97/0.97/0.97/0.97
MLSTM	**0.97/0.97/0.96/0.97**	**0.98/0.98/0.97/0.98**	**0.98/0.98/0.97/0.98**	**0.97/0.97/0.97/0.97**	**0.97/0.97/0.96/0.97**	**0.98/0.98/0.98/0.98**

The bold numbers are the largest values in the same column, which is a typical representation in the machine learning related papers.

TABLE II
DIAGNOSIS PERFORMANCE EVALUATION USING METRICS (ACCURACY, F_1, RECALL, AND PRECISION)

Window Size	50	80	100	140	160	200
SVM	0.95/0.12/0.11/0.12	0.94/0.03/0.02/0.09	0.95/0.11/0.11/0.12	0.93/0.01/0.01/0.11	0.93/0.01/0.01/0.14	0.93/0.08/0.08/0.08
KNN	0.95/0.02/0.02/0.02	0.94/0.01/0.01/0.01	0.95/0.02/0.01/0.02	0.93/0.01/0.01/0.02	0.93/0.01/0.01/0.01	0.92/0.01/0.01/0.01
DT	0.95/0.12/0.12/0.12	0.95/0.06/0.05/0.06	0.95/0.12/0.12/0.12	0.93/0.04/0.03/0.04	0.93/0.04/0.03/0.05	0.93/0.06/0.06/0.06
ANN	0.95/0.10/0.09/0.10	0.95/0.09/0.08/0.10	0.96/0.11/0.11/0.11	0.94/0.06/0.03/0.13	0.94/0.06/0.05/0.08	0.93/0.12/0.12/0.12
CNN	0.91/0.83/0.83/0.84	0.95/0.90/0.87/0.93	0.95/0.94/0.91/0.97	0.95/0.92/0.90/**0.95**	0.96/0.93/0.90/0.96	0.97/0.96/0.96/0.96
MLSTM	**0.97/0.93/0.90/0.96**	**0.97/0.94/0.93/0.96**	**0.98/0.95/0.92/0.97**	**0.96/0.92/0.91**/0.94	**0.96/0.93/0.90/0.96**	**0.98/0.97/0.96/0.97**

The bold numbers are the largest values in the same column, which is a typical representation in the machine learning related papers.

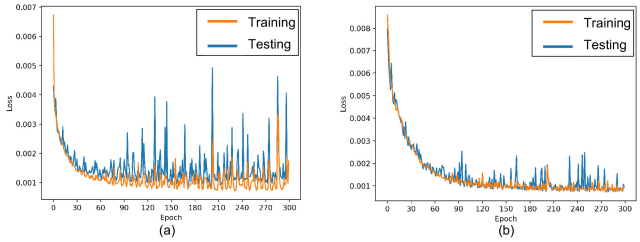

Fig. 4. (a) CNN and (b) MLSTM loss curves in the attack diagnosis with window length 100 (0.1 s).

better model robustness and stable performances. Notice that MLSTM demonstrates the best performances when the analysis window size is 80 or 100. Although the metrics achieved another peaks with window size 200, that would be clearly overfitting on interferences.

V. CONCLUSION

We propose a cyber security mechanism by combining a one-class detection model and an attack diagnosis model, which are tailored for electric waveform profiles of a solar PV smart grid for real-time attack detection and identification. First, an analysis was conducted on DIA on the smart grid with solar PV farm embedded. Then, an MLSTM-based comprehensive approach was developed. We apply the one-class detection model to detect whether a PV farm is under attack or not. When it is detected to be under attack, we identify the attack type by leveraging the attack diagnosis model. The proposed mechanism has been evaluated using a MATLAB Simulink solar farm model, and achieves much improved attack detection and diagnosis performances.

REFERENCES

[1] S. Sarangan, V. K. Singh, and M. Govindarasu, "Cyber attack-defense analysis for automatic generation control with renewable energy sources," in *Proc. North Amer. Power Symp.*, 2018, pp. 1–6.

[2] X. Liu, M. Shahidehpour, Y. Cao, L. Wu, W. Wei, and X. Liu, "Microgrid risk analysis considering the impact of cyber attacks on solar PV and ESS control systems," *IEEE Trans. Smart Grid*, vol. 8, no. 3, pp. 1330–1339, May 2016.

[3] B. Yang, L. Guo, F. Li, J. Ye, and W.-Z. Song, "Vulnerability assessments of electric drive systems due to sensor data integrity attacks," *IEEE Trans. Ind. Informat.*, vol. 16, no. 5, pp. 3301–3310, May 2020.

[4] O. A. Beg, T. T. Johnson, and A. Davoudi, "Detection of false-data injection attacks in cyber-physical dc microgrids," *IEEE Trans. Ind. Informat.*, vol. 13, no. 5, pp. 2693–2703, Oct. 2017.

[5] A. Giani, E. Bitar, M. Garcia, M. McQueen, P. Khargonekar, and K. Poolla, "Smart grid data integrity attacks: Characterizations and countermeasures π," in *Proc. IEEE Int. Conf. Smart Grid Commun.*, 2011, pp. 232–237.

[6] M. Esmalifalak, L. Liu, N. Nguyen, R. Zheng, and Z. Han, "Detecting stealthy false data injection using machine learning in smart grid," *IEEE Syst. J.*, vol. 11, no. 3, pp. 1644–1652, Sep. 2014.

[7] F. Li, Y. Shi, A. Shinde, J. Ye, and W.-Z. Song, "Enhanced cyber-physical security in internet of things through energy auditing," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 5224–5231, Jun. 2019.

[8] F. Li, A. Shinde, Y. Shi, J. Ye, X.-Y. Li, and W.-Z. Song, "System statistics learning-based IoT security: Feasibility and suitability," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 6396–6403, Aug. 2019.

[9] J. C. Balda, A. Mantooth, R. Blum, and P. Tenti, "Cybersecurity and power electronics: Addressing the security vulnerabilities of the internet of things," *IEEE Power Electron. Mag.*, vol. 4, no. 4, pp. 37–43, Dec. 2017.

[10] F. Li *et al.* "Detection and identification of cyber and physical attacks on distribution power grids with PVs: An online high-dimensional data-driven approach," *IEEE J. Emerg. Sel. Topics Power Electron.*, pp. 1–10, 2019.

[11] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," in *Proc. 9th Int. Conf. Artif. Neural Netw.*, 1999, pp. 850–855.

[12] L. A. Maglaras and J. Jiang, "A real time OCSVM intrusion detection module with low overhead for SCADA systems," *Int. J. Adv. Res. Artif. Intell.*, vol. 3, no. 10, pp. 45–53, 2014.

[13] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[14] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Trans. Neural Netw.*, vol. 5, no. 2, pp. 157–166, Mar. 1994.

[15] A. Paszke *et al.*, "Automatic differentiation in PyTorch," in *Proc. 31st Conf. Neural Inf. Process. Syst.*, 2017, pp. 1–4.

[16] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[17] F. Li, J. Clemente, M. Valero, Z. Tse, S. Li, and W. Song, "Smart home monitoring system via footstep-induced vibrations," *IEEE Syst. J.*, pp. 1–7, 2019.