Accurate Joint Detection of False Data Injection Attacks on Islanded PV Output Power and State of Health Estimation of Lithium-Ion Batteries

Mohamed Massaoudi[1,2], Katherine R. Davis[1], Haitham Abu-Rub[2], Ali Ghrayeb[2], and Tingwen Huang[3]
[1]Department of Electrical and Computer Engineering, Texas A&M University, College Station, USA
[2]Department of Electrical and Computer Engineering, Texas A&M University at Qatar, Doha, Qatar
[3]Science Department, Texas A&M University at Qatar, Doha, Qatar
Email: {mohamed.massaoudi, katedavis, haitham.abu-rub, ali.ghrayeb, tingwen.huang}@tamu.edu

***Abstract*—With the widespread adoption of islanded DC microgrids incorporating photovoltaic (PV) systems and battery energy storage systems (BESSs), safeguarding the cyber-physical security of PV-BESS configurations becomes crucial. While numerous false data injection attack (FDIA) detection techniques have been presented in existing literature, these methods are often tailor-made for specific attack types. This paper introduces an innovative approach for concurrently detecting FDIAs on islanded PV power plants and BESSs. A bespoke two-layer random forest model, meticulously crafted and trained, is deployed for identifying FDIAs instigated by malicious adversaries. Our method uniquely caters to three distinct cyber-attack scenarios: tampered attacks, poisoned attacks, and replay attacks. The efficacy of our FDIA detection approach is rigorously evaluated utilizing real-world datasets and synthetically generated attacks. The results underscore the method's impressive performance in identifying a spectrum of cyber-attacks on both PV system output and the state of health output from the battery data.**

***Index Terms*—Cyber-physical system, false data attacks, photovoltaic power, Lithium-ion battery, state of health.**

I. Introduction

The evolving prevalence of PV systems and large battery energy storage systems (BESS) within microgrid (MG) architectures has been significantly increasing [1]. Equipped with smart inverters and meters, the PV-BESS can function in two distinct modes: grid-connected and standalone. It operates under the management of and connectivity to the MG, facilitated by inter-communicating devices fitted with sensors. These sensors consistently communicate with control centers, enabling monitoring and control activities. However, the integration of a myriad of sensors and devices in renewable-heavy MGs, all engaging in continuous data transmission and reception, inadvertently creates a larger target for malicious cyber activities [2]. This continuous data exchange over the network is susceptible to cyber-physical attacks, making it vulnerable to exploitation by adversaries. Malicious manipulation of sensor data by these adversaries can disrupt system operations and decision-making processes. This risk is especially high when sensors are not equipped with tamper-resistant hardware. Thus, accurately detecting such attacks is crucial for maintaining the system's integrity and reliability [3].

False data injection attacks (FDIAs) represent a prominent threat to MG supervisory control and data acquisition (SCADA) systems. FDIAs introduce altered sensor readings to deceive the decision-making process within control centers [4]. These attacks meticulously erode the data integrity within the grid, insidiously manipulating the system's perception of its own operational status and, consequently, its strategic responses [5].

The realm of BESS in smart distribution networks (SDNs) has witnessed a surge in research, particularly focusing on the vulnerabilities and security challenges associated with FDIAs. Paper [6] significantly contributes to this discourse, unveiling the intricate mechanisms of static and sequential FDIAs and their consequential manipulations on the State-of-Charge (SoC) estimations—a critical component for optimal BESS operation and management within SDNs. Despite its analytical prowess, the paper stops short of presenting a holistic mitigation strategy, spotlighting an imperative research gap in fortifying the cybersecurity framework of BESS.

Concurrently, the study in [7] proposed a convolutional neural networks to adeptly detect and classify spurious battery data. This approach not only enhances the system's safety and reliability but also mitigates potential threats emanating from deceptive battery data. Delving into stealthier attack paradigms, [8] delineates a sophisticated assault using artificial neural networks and a man-in-the-middle (MitM) strategy, aiming at the communication nexus between the BESS's local supervisory controller and its battery control units. This revelation of undetectable attacks further amplifies the urgency for advanced and resilient cybersecurity mechanisms.

In light of the previous research gaps, this paper presents a novel method for FDIA detection (FDIAD) in PV-BESS. This study embarks on a meticulous exploration of FDIAs, signaling a transition from vulnerability exposure to the development of robust defense mechanisms. To the authors' best knowledge, this is the first attempt to secure an islanded PV system with BESS against cyber FDIAs. The main contributions of this paper are three-fold:

1) A novel two-layer random forest (TLRF) model is introduced and tailored for FDIAD.

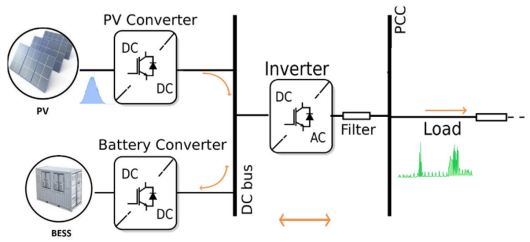

Fig. 1: Flowchart of the microgrid system.

2) A multimodal classifier of three malicious attacks, specifically, the tempered attacks, poisoning attacks, and replay attacks is proposed.
3) The comprehensive validation of the proposed model's effectiveness and superiority over existing methods is conducted using real-world datasets.

II. Cyberattacks types

This paper tackles three types of cyber-attacks: tempered attacks, poisoning attacks, and replay attacks. This section briefly describes these types. The PV-BESS system considered in this work for these attack scenarios is shown in Fig. 1.

A. Tempered Attacks

A tempered data attack refers to a malicious action where an adversary introduces false data or alters the legitimate data being used by the system, as shown in Fig. 2.(a). This can mislead the control system, causing inefficiencies, damage, or even system failures. An adversary can introduce a bias b to the solar irradiance data, leading to an altered G' given by $G' = G + b$. Similarly, they can introduce a bias in the power readings of the battery, leading to altered P'_{in} and P'_{out}. With the tempered solar irradiance data, the PV output becomes

$$P'_{PV} = G' \times A \times \eta. \tag{1}$$

This can cause the system to either overproduce or underproduce power, leading to wastage, inefficiencies, or unmet load demands. With altered power readings, the battery's state of health (SoH) calculation is calculated as

$$SoH'(t) = SoH(t-1) + \frac{\Delta t}{C} \times (P'_{in} - P'_{out}) \tag{2}$$

This can lead to overcharging or deep discharging of the battery, reducing its lifespan and potentially causing damage or hazards. Tempered data attacks in an islanded PV-BESS can have severe repercussions on the efficiency, safety, and longevity of the system.

B. Poisoning Attacks

Poisoning data attacks refer to the intentional injection of malicious or incorrect data into the system's monitoring or control datasets. These attacks can be targeted at any component of the PV system, but when directed at the battery's SoH, they can have particularly detrimental effects as shown in Fig. 2.(b). A compromised SoH reading can lead to improper

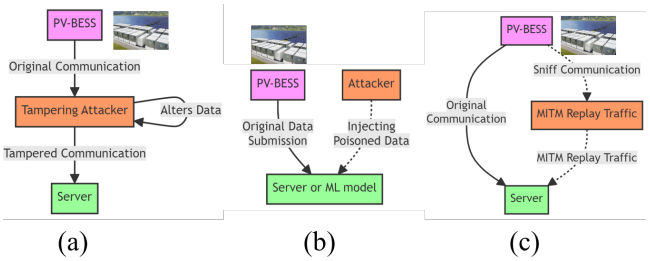

Fig. 2: Schematic of the depicted a) Tempering attacks b) Poisoned attacks c) Replay attacks.

charging or discharging, thus reducing the battery's lifespan or even causing immediate damage. In a poisoning data attack aimed at the SoH, the attacker manipulates the SoH value, leading to a miscalculation in the battery's charging or discharging power. The manipulated SoH is represented as

$$SoH' = SoH + \delta, \tag{3}$$

where δ is the poisoned data. This can lead to an incorrect calculation of P_B, represented as P'_B. The new relationship is computed as

$$P_{PV} + P'_B = P_L. \tag{4}$$

Given that P'_B is not the true required power for the battery, continued operation under this condition can lead to overcharging or over-discharging of the battery, affecting its lifespan and performance. Poisoning data attacks in islanded PV-BESS can have serious implications. By manipulating the battery SoH, attackers can induce improper operation, leading to reduced battery lifespan or immediate damage.

C. Replay Attacks

A replay data attack refers to a malicious activity where an attacker captures legitimate data from the system and then retransmits it at a later time. The purpose of the attack is to deceive the system controllers or monitors into believing that the retransmitted data is currently being generated by the sensors as illustrated in Fig. 2.(c). An attacker could potentially manipulate the readings related to the battery's SoH, PV generation, load consumption, or other essential parameters. Let's denote $P_{PV}(t)$ the Power generated by the PV panels and $P_{battery}(t)$ power charged or discharged by the battery. During a replay attack, an attacker captures data at time t_1 and replays it at a later time t_2. The system then receives

$$P_{PV}(t_2) = P_{PV}(t_1); SoH(t_2) = SoH(t_1). \tag{5}$$

The mismatch between the actual/replayed conditions can lead to incorrect decisions by the system controller. For instance, the battery might be charged or discharged unnecessarily based on the replayed SoH, leading to faster degradation. To detect such attacks, one can monitor the residuals as

$$r_P(t) = P_{PV}(t) + P_{battery}(t) - P_{load}(t), \tag{6}$$

$$r_{SoH}(t) = SoH_{measured}(t) - SoH_{expected}(t). \tag{7}$$

Under normal operation, both residuals $r_P(t)$ and $r_{SoH}(t)$ should be close to zero. Significant deviations can be an indication of a potential replay attack.

III. Proposed Architecture

The TLRF consists of two sequential layers that aim to enhance predictive performance by reducing both bias and variance. This method involves base Random Forest (RF) models in the first layer and a meta-model in the second layer, which leverages the strengths of the base models [9]. Given a dataset $\mathcal{D}$ consisting of feature vectors and corresponding labels, i.e., $\mathcal{D} = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$, where x_i is a feature vector and y_i is the corresponding label, we train M base RF models at the first layer. Each model h_m for $m = 1, \ldots, M$, is trained on a subset of the data $\mathcal{D}_m \subset \mathcal{D}$ and potentially a subset of the features $\mathcal{F}_m \subset \mathcal{F}$, where $\mathcal{F}$ is the set of all features. The prediction of the m-th base model for a new input vector x can be defined as

$$h_m(x) = f_m(\mathcal{D}_m, \mathcal{F}_m, x). \tag{8}$$

The predictions from the base models are used as input for a second layer which trains a meta-model. Defining the combined feature vector from the first layer for an input vector x as

$$z(x) = [h_1(x), h_2(x), \ldots, h_M(x)]. \tag{9}$$

where $z(x)$ is used as the input feature vector for the second layer model. The second layer model is then trained on a new dataset as

$$\mathcal{D}_z = \{(z(x_1), y_1), (z(x_2), y_2), \ldots, (z(x_n), y_n)\}. \tag{10}$$

The function learned by the second layer model is defined as

$$H(z(x)) = f_z(\mathcal{D}_z, z(x)). \tag{11}$$

The final prediction for a new input vector x is given by

$$\hat{y}(x) = H(z(x)). \tag{12}$$

The TLRF model integrates base learners to form a more robust predictive model. By utilizing the strengths and mitigating the weaknesses of various base models, the ensemble model $\hat{y}(x)$ is expected to produce more accurate and generalized predictions than any of the individual base models. Fig. 3 illustrates the structure of a TLRF method designed for FDIAD in PV-BSS systems. In the first layer, multiple bootstrap databases are used to train individual trees. The outputs of these trees undergo majority voting, producing an aggregate decision for each classifier. In the second layer, a similar process is followed, leveraging the outcomes from the first layer. The ensemble approach of the TLRF enhances its robustness and precision in detecting FDIAs. Upon completion of the training phase, the model is ready for implementation, either at the control center or directly on the sensors, to identify FDIAs. For any new measurement x_j, the model functions as follows.

$$g(x_j) = \begin{cases} 1, & \text{if an attack is detected on the system,} \\ 0, & \text{if the system is operating normally.} \end{cases} \tag{13}$$

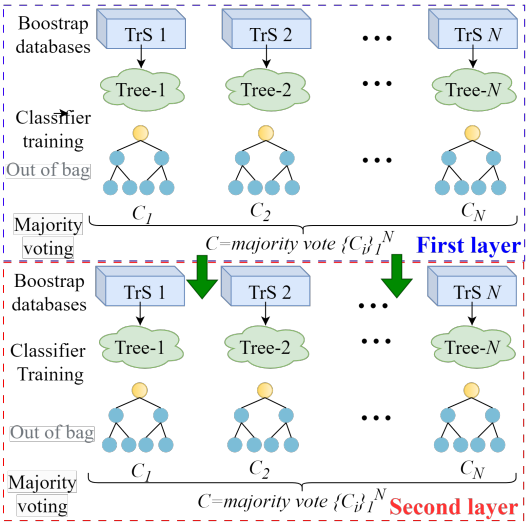

Fig. 3: Flowchart of the TLRF method of PV-BSS.

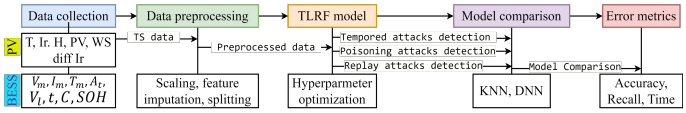

Fig. 4: Flowchart of the TLRF method for FDIA detection.

Fig. 4 presents a comprehensive flowchart detailing the methodology behind the TLRF approach for FDIAD, specifically tailored for an islanded PV system integrated with a BESS. Data from PV and BESS, including temperature, irradiance, and voltage, is collected and preprocessed. The TLRF model detects tempered, poisoning, and replay attacks. Its performance is compared with KNN and DNN models, using metrics such as accuracy and recall to measure efficiency. For FDIA, a common problem for ML techniques is the imbalanced datasets. Therefore, oversampling is a technique used to handle class imbalance by increasing the number of instances in the minority class to balance the class distribution [10]. The goal is to enhance the performance of a model by increasing the representation of the minority class, often by duplicating instances or generating synthetic samples. Let n_1 be the number of instances in the majority class and n_2 be the number of instances in the minority class, with $n_1 > n_2$. To balance the classes perfectly, we would need $n_1 - n_2$ new instances of the minority class. But, an oversampling ratio r can be defined to determine the number of new instances as

$$\text{New instances} = r \times (n_1 - n_2). \tag{14}$$

For $r = 1$, the classes will be perfectly balanced. But r can be greater or smaller than 1 depending on the desired balance level. Let S_2 be the set of instances in the minority class. The oversampled set $S_{2,\text{oversampled}}$ is created as

$$S_{2,\text{oversampled}} = S_2 \cup \{s_i : s_i \in S_2, i = 1, 2, \ldots, \text{New instances}\}, \tag{15}$$

where s_i are instances randomly chosen from S_2 with replacement. Instead of just duplicating instances, synthetic samples can be created using algorithms such as synthetic minority over-sampling technique [10]. Given two instances a and b from the minority class, a synthetic sample s is generated as

$$s = a + \lambda \times (b - a), \tag{16}$$

where λ is a random number between 0 and 1. This approach creates a sample that lies along the line segment joining a and b. Oversampling helps in equalizing the number of instances across classes, either by duplicating existing instances or by introducing synthetic ones, to improve the classifiers' performance on imbalanced datasets.

IV. Case study

The proposed model is implemented in the TensorFlow library. This work is implemented on a personal computer with Inter Core i7-10885H CPU @2.40 GHz and 32.0 GB RAM.

A. Data description

The simulation leverages publicly available data from the Desert Knowledge Alice Springs Center (DKASC), located in Central Australia at latitude 23.7618° S and longitude 133.8749° E [11]. This dataset originates from comprehensive monitoring of various types, models, and configurations of PV technologies. An in-depth analysis of two years (2018-2019) of DKASC's data is conducted. The input parameters for our system include ambient temperature in Celsius (T in $^\circ$C), wind direction (Wd in degrees), horizontal radiation (Ir in W/m^2), diffuse horizontal radiation (DIr in W/m^2), and relative humidity (Rh in %); while the system outputs the active power in kW.

Additionally, the study made use of a publicly available dataset. Generated by the National Aeronautics and Space Administration's (NASA) Prognostics Center of Excellence, the process of charging a battery adheres to the constant-current constant-voltage (CCCV) methodology. The feature vector in the database is formulated as $\Psi_k = [V_m(k), I_m(k), T_m(k), A_t, C_l, V_l, t, C]$, where $V_m(k)$, $I_m(k)$, and $T_m(k)$ signify the voltage measured, current measured, and temperature measurement of the battery at each time step k, respectively; A_t is the ambient temperature, C_l is the current load, V_l is the voltage load, t is the time, and C represents capacity, all of which contribute to a comprehensive evaluation and monitoring of the performance and health status of a battery system. The SoH is calculated as, SoH (%) $= \frac{C_k}{C_0} \times 100$, where C_0 and C_k denote the initial capacity and the measured capacity at cycle k, respectively. Four battery types (B0005, B0006, B0007, B0018) with different capacity degradation behavior are provided. In Both datasets, 10% of the samples are falsified and 70% of data is used for training the models.

B. Evaluation measures

In this study, the effectiveness of the proposed method is evaluated using the accuracy (Acc), precision (Prec), recall (R), and F1-score (F1). The mathematical formulas for these measures are provided as [12]

$$Prec = \frac{T_P}{T_P + F_P}, R = \frac{T_P}{T_P + F_N}, \tag{17}$$

$$Acc = \frac{T_P + T_N}{T_P + T_N + F_P + F_N}, F1 = 2 \times \frac{Prec \times R}{Prec + R}, \tag{18}$$

where T_P, T_N, F_N, and F_P denote true positive, true negative, false negative, and false positive, respectively.

C. Simulation Results

The simulation results consist of evaluating the classifier performance from the battery side and PV side separately. The proposed TLRF is compared to the K-nearest neighbors (KNN), deep neural network (DNN), and RF. The competitive models also witness the impact of the oversampling (OS) technique on the classification results.

1) FDIAD of battery SoH: In the proposed TLRF architecture, the first layer comprises a series of base models, the quantity of which is defined by `num_base_models`. Each of these models is an RF classifier, initialized with a balanced class weight and a random state of 42 to ensure reproducibility. To introduce diversity, 75% of features and instances are randomly selected for training each base model. The second layer, serving as the meta-model, is also an RF classifier, employing default settings and the same random state of 42.

Fig. 5 displays a series of six confusion matrices, which are used to evaluate the performance of the TLRF model. Each matrix represents the model's performance for a specific type of attack scenario or classification problem. Matrices (a) to (c) cover battery SoH attacks, with high accuracy and minimal misclassifications for Healthy, Tempered, Poisoned, and Replayed categories. For instance, matrix (a) pertains to Tempered battery SoH attacks, where there are 6178 correct predictions for Healthy and 20 incorrect, and for Tempered, there are 6346 correct predictions with 8 that were misclassified. Overall, the TLRF model shows exceptional resistance against tempered attacks, and good resistance against poisoning attacks, but reveals some vulnerabilities against replay attacks.

Table I presents the performance of the TLRF against cyber-attacks in the BESS. Across three types of attack scenarios, tempering, poisoning, and replied attacks. The model showcases impressive accuracy rates, with tempering attacks being detected with an accuracy of 99.78%, Poisoning at 98.25%, and Replied at 99.09%, respectively. Moreover, the execution time is fairly consistent, hovering a little over 2.4 seconds for each attack type. This suggests that the TLRF not only provides a comprehensive detection system against FDIA but does so efficiently.

Table II provides a comprehensive evaluation of FDIAD in a PV-BSS. The table indicates that the TLRF model consistently outperforms the other models across different attack types and scenarios, showcasing its robustness and reliability in detecting FDIA. For the tempered BESS attacks, the TLRF

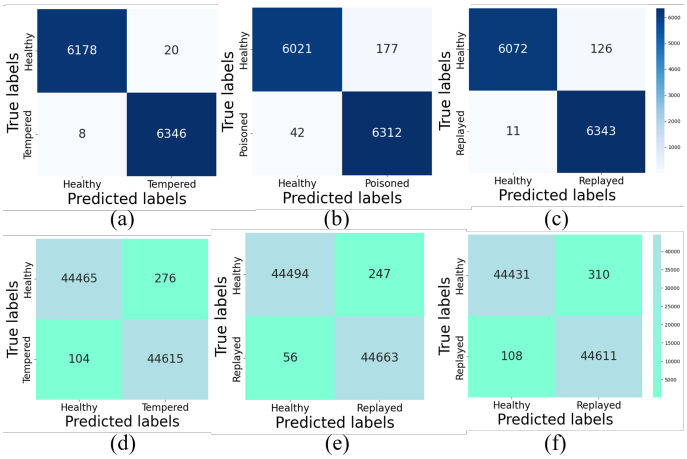

Fig. 5: Confusion matrix of the different classifications for (a) Tempered battery SoH attacks (b) Poisoning battery SoH attacks (c) Replay battery SoH attacks (d) Tempered PV attacks (e) Poisoning PV attacks (f) Replay PV attacks.

TABLE I: Simulation results for FDIA detection in the BESS.

Attack types	Acc	Prec	Recall	F-1	Time
Tempering attacks	99.78	99.70	99.87	99.78	2.37
Poisoning attacks	98.25	97.27	99.33	98.29	2.44
Replied attacks	99.09	98.65	99.55	99.10	2.53

exhibits exceptional performance with an accuracy of 99.78%, precision of 99.70%, recall of 99.87%, and an F1 score of 99.78%. The model takes only 2.37 seconds to execute.

TABLE II: Simulation results for FDIA detection in PV-BESS.

FDIA detection assessment on the battery side					
Model	**Acc (%)**	**Prec (%)**	**R (%)**	**F1 (%)**	**Time (s)**
Tempered attacks					
TLRF	99.78	99.70	99.87	99.78	2.37
KNN	61.01	29.73	20.82	24.49	1.77
DNN	37.43	30.88	85.82	45.42	0.02
Poisoning attacks					
TLRF	98.25	97.27	99.33	98.29	2.44
KNN	88.92	82.35	99.41	90.08	1.64
DNN	51.07	50.86	98.33	67.04	0.03
Replay attacks					
TLRF	99.09	98.65	99.55	99.10	2.53
KNN	88.59	81.87	99.49	89.82	2.36
DNN	51.15	50.91	97.79	66.96	0.03
FDIA detection assessment on the PV side					
Tempered attacks					
TLRF	99.71	99.51	99.91	99.71	18.73
KNN	92.87	87.75	99.65	93.32	12.60
DNN	72.57	65.65	94.63	77.52	1.357
Poisoning attacks					
TLRF	99.32	98.82	99.83	99.32	17.14
KNN	92.86	87.70	99.70	93.32	10.35
DNN	72.86	89.08	52.10	65.75	12.48
Replay attacks					
TLRF	99.64	99.46	99.82	99.64	22.75
KNN	72.72	22.91	20.76	21.78	7.41

2) FDIAD in the PV system side: For the PV data, matrices (d) to (f) in Fig. 5 evaluate PV attacks, showing substantial accuracy for healthy, tempered, and replayed labels, but

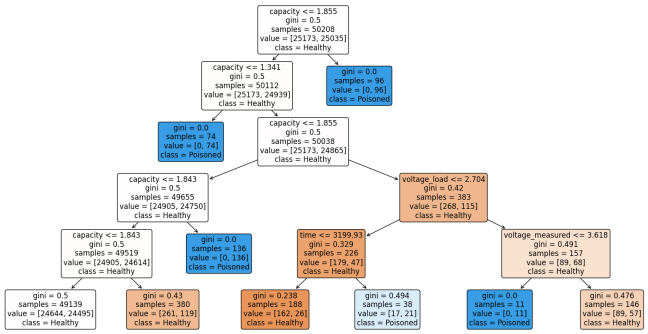

Fig. 6: Global surrogate model using DT for TLRF for FDIAD in PV data.

with a few errors. For instance, matrix (f) for replay PV attacks depicts 44431 correct healthy classifications with 310 errors and 44611 correct classifications for replayed attacks against 108 errors. Focusing on the explainability impact on cybersecurity, decision Trees, as visualized, inherently offer transparency and explainability, making them valuable tools in the cybersecurity domain. Fig. 6 illustrates a Decision Tree (DT) as a global surrogate model tailored for the TLRF approach, specifically addressing the detection of FDIA in PV systems. The tree's nodes are partitioned based on specific thresholds of various features, including capacity, voltage, and time, with Gini impurity scores (GIS) provided to measure the node's purity. The GIS is calculated as [13]

$$Gini = \frac{\sum_{a=1}^{\omega}(2a-\omega-1)\cdot h_a}{\kappa+\omega\cdot\sum_{a=1}^{\omega}h_a}. \tag{19}$$

h_a represents the value at position a within the TLRF output that contains ω elements, i denotes the ordinal position when arranged from smallest to largest, and κ, which equals $1e^{-8}$, is a minor constant introduced to preserve numerical stability. Terminal nodes, or leaves, categorize the samples into healthy or falsified. Table III illustrates the performance of the TLRF against cyber-attacks in PV systems. For Tempering attacks, the system boasts an accuracy of 99.71%, with precision at 99.51%, recall at 99.91%, an F-1 score mirroring the accuracy at 99.71%, and a processing time of 18.73 units. Poisoning attacks show slightly reduced accuracy at 99.32%, with precision dipping to 98.82%, though recall remains impressively high at 99.83%. The F-1 score for Poisoning attacks stands at 99.32% and takes 17.14 units of time. Lastly, replied attacks have an accuracy of 99.64%, a precision at 99.46%, a recall of 99.82%, an F-1 score of 99.64%, and the longest processing time of 22.75 units. In analyzing these results, the TLRF demonstrates exemplary robustness to correctly identify genuine attacks against FDIA in PV systems.

TABLE III: Simulation results for FDIAD in the PV system.

Attack types	Acc	Prec	Recall	F-1	Time
Tempering attacks	99.71	99.51	99.91	99.71	18.73
Poisoning attacks	99.32	98.82	99.83	99.32	17.14
Replied attacks	99.64	99.46	99.82	99.64	22.75

Table IV showcases the performance of various models in detecting FDIAs on PV systems under tempered attacks. The models compared include TLRF, KNN, KNN-OS, RF-OS, DNN, and DNN-OS. The TLRF and RF-OS models exhibit outstanding performance in FDIAD for PV systems, with TLRF achieving 44,680 true positives and 44,525 true negatives, and RF-OS achieving even higher true positives at 44,708 and true negatives at 44,255, both with minimal false positives and negatives. In contrast, models such as standard KNN, DNN, and their oversampled versions (KNN-OS and DNN-OS) show a trade-off between high true positive rates and increased false positives, with KNN-OS and DNN-OS significantly reducing false negatives to 156 and 2,398 respectively, at the cost of increasing false positives to 6,220 and 22,138 respectively. From Table IV, the OS technique significantly enhances the performance of ML models for FDIAD in PV systems. It is worth noting that the OS technique while improving the true positive rate, does tend to increase the false positive rate, which is an important consideration in practical applications.

TABLE IV: Confusion matrix of the proposed model and benchmarks with tempered PV attacks.

Model	Label	Healthy	Tempered
TLRF	Healthy	**44525**	**216**
	Tempered	**39**	**44680**
KNN	Healthy	43836	940
	Tempered	4408	516
KNN-OS	Healthy	38521	6220
	Tempered	156	44563
RF-OS	Healthy	44255	486
	Tempered	11	44708
DNN	Healthy	41919	2899
	Tempered	4025	857
DNN-OS	Healthy	22603	22138
	Tempered	2398	42321

From Table II, the TLRF model consistently excels with over 99% in accuracy, precision, and F1 score across all attack types, albeit with longer execution times (22.75 seconds for replay attacks). The KNN model shines in tempered and poisoning attacks with accuracy levels of 92.87% and 92.86% respectively, but its performance drops to 72.72% accuracy in replay attacks. The DNN model, while faster with execution times as low as 1.357 seconds in tempered attacks, shows varied performance with a notable 94.63% recall in tempered attacks but lower accuracy and precision across the board. The execution time for models is generally higher on the PV side, indicating a possible increase in computational demand or complexity in these scenarios. Overall, the TLRF model stands out as the most reliable choice for cyber-attack detection in MGs.

V. Conclusions

This paper proposed a TLRF model to proactively identify and counteract FDIAs that jeopardize the integrity of critical operational data. This paper has provided a comprehensive review of some cyber-attacks affecting MGs. Consequently, synthetic cyber-attack datasets based on the actual PV-BESS dataset were generated and implemented. On average, the TLRF model provides a high detection accuracy of 99% and a low false negative rate. In both scenarios, the oversampling technique is mandatory to tackle the FDIAD problem due to the severely imbalanced data and the scarcity of falsified samples. Our findings highlight the effectiveness and potential of the proposed model in establishing a resilient framework for detecting threats. The proposed method outperformed competitive models (DNN, RF, and KNN) with the adoption of the oversampling technique. In the future, we will extend our work by streamlining the proposed TLRF with an adaptive mode to keep pace with the increasing cyber threats.

Acknowledgments

This publication was made possible by NPRP12C-33905-SP-213 and NPRP12C-33905- SP-220 from the Qatar National Research Fund (a member of Qatar Foundation). The authors gratefully acknowledge the project "Research Impact Initiative SP6" by TAMUQ. The statements made herein are solely the responsibility of the authors.

References

[1] M. Massaoudi, I. Chihi, H. Abu-Rub, S. S. Refaat, and F. S. Oueslati, "Convergence of photovoltaic power forecasting and deep learning: State-of-art review," *IEEE Access*, vol. 9, pp. 136 593–136 615, 2021.

[2] T. Kim *et al.*, "An overview of cyber-physical security of battery management systems and adoption of blockchain technology," *IEEE Trans. Emerg. Sel. Topics Power Electron.*, early access, Jan. 21, 2020.

[3] H. Huang, Z. Mao, A. Layton, and K. R. Davis, "An ecological robustness oriented optimal power flow for power systems' survivability," *IEEE Trans. Power Systems*, vol. 38, no. 1, pp. 447–462, 2022.

[4] A. Adhikaree, T. Kim, J. Vagdoda, A. Ochoa, P. J. Hernandez, and Y. Lee, "Cloud-based battery condition monitoring platform for large-scale lithium-ion battery energy storage systems using internet-of-things (iot)," in *Proc. IEEE ECCE*, 2017, pp. 1004–1009.

[5] A. B. Lopez, K. Vatanparvar, A. P. Deb Nath, S. Yang, S. Bhnia, and M. A. Al Faruque, "A security perspective on battery systems of the internet of things," *J. Hardw. Syst. Security*, vol. 1, no. 2, pp. 188–199, 2017.

[6] P. Zhuang and H. Liang, "False data injection attacks against state-of-charge estimation of battery energy storage systems in smart distribution networks," *IEEE Trans. Smart Grid*, vol. 12, no. 3, pp. 2566–2577, 2020.

[7] H.-J. Lee, K.-T. Kim, J.-H. Park, G. Bere, J. J. Ochoa, and T. Kim, "Convolutional neural network-based false battery data detection and classification for battery energy storage systems," *IEEE Trans. Energy Conversion*, vol. 36, no. 4, pp. 3108–3117, 2021.

[8] M. Pasetti, P. Ferrari, P. Bellagente, E. Sisinni, A. O. de Sá, C. B. do Prado, R. P. David, and R. C. S. Machado, "Artificial neural network-based stealth attack on battery energy storage systems," *IEEE Trans. Smart Grid*, vol. 12, no. 6, pp. 5310–5321, 2021.

[9] M. Massaoudi, I. Chihi, L. Sidhom, M. Trabelsi, S. S. Refaat, and F. S. Oueslati, "Enhanced random forest model for robust short-term photovoltaic power forecasting using weather measurements," *Energies*, vol. 14, no. 13, p. 3992, 2021.

[10] M. Massaoudi, S. S. Refaat, and H. Abu-Rub, "Intrusion detection method based on smote transformation for smart grid cybersecurity," in *2022 3rd International Conference on Smart Grid and Renewable Energy (SGRE)*. IEEE, 2022, pp. 1–6.

[11] "Data Download — DKA Solar Centre." [Online]. Available: http://dkasolarcentre.com.au/locations/alice-springs,Accessedon23-09-2019

[12] M. Massaoudi, S. S. Refaat, A. Ghrayeb, and H. Abu-Rub, "Short-term dynamic voltage stability status estimation using multilayer neural networks," in *2023 IEEE Texas Power and Energy Conference (TPEC)*. IEEE, 2023, pp. 1–6.

[13] D. Batic, V. Stankovic, and L. Stankovic, "Towards transparent load disaggregation–a framework for quantitative evaluation of explainability using explainable ai," *IEEE Trans. Consumer Electronics*, 2023.