Intrusion Detection and Mitigation System for Smart Inverters Based on Machine Learning and Environmental Sensors

Zheyu Zhang, *Member*, *IEEE* and Saifur Rahman, *Fellow*, *IEEE*

***Abstract*—The rapid expansion of renewable energy deployment has accelerated the adoption of smart inverters in solar farms. However, their reliance on communication networks introduces cybersecurity vulnerabilities that make inverter measurements susceptible to manipulation. This paper presents an intrusion detection and mitigation system (IDMS) for smart inverters, which leverages machine learning algorithms and environmental sensor data to enhance data integrity and operational resilience. The proposed IDMS integrates the machine learning algorithms to combine a detection system with a mitigation system. The detection system integrates environmental sensor inputs with inverter measurements to accurately identify normal operating behavior and anomalous conditions in real time. To address the cumulative error challenges that arise during attacks, a prediction model based on long short-term memory (LSTM) and trained solely on environmental features is introduced in the mitigation system to reconstruct compromised voltage and current data, thereby preventing recursive cumulative error propagation and ensuring stable system operation. Case studies demonstrate that the proposed IDMS can promptly detect cyberattacks, replace compromised measurements with reliable estimates, and restore normal operation once attacks cease. Additional evaluations across varying data availability conditions in an open dataset confirm the robustness and scalability of the proposed IDMS, making it practical for diverse solar farm deployment environments.**

***Index Terms*—Artificial intelligence, solar farm, intrusion detection and mitigation system (IDMS), environmental sensor, smart inverter, cybersecurity, machine learning.**

I. Introduction

THE rapid growth of renewable energy sources (RESs), driven by increasing concerns over climate change, has led to the widespread deployment of distributed energy systems [1]. Among these advancements, smart inverters have emerged as a key innovation in solar energy systems [2]. With integrated monitoring and control functionalities, smart inverters facilitate the seamless integration and regulation of RESs within smart grids, thereby enhancing system efficiency and reliability [3]. However, their reliance on public communication channels for interaction with control centers and consumer servers introduces significant cybersecurity vulnerabilities [4].

Smart inverters are critical components of modern energy systems, making their cybersecurity a pressing concern. Two pivotal surveys published in 2023 provide a comprehensive overview of recent advances. The first study by [5] classified cyberattacks into three layers—hardware, controller, and network—highlighting threats such as firmware tampering, malicious command injection, and man-in-the-middle attacks. A particularly dangerous threat, the false data injection attack (FDIA), was examined in both surveys and further analyzed in [6], where it was shown to corrupt power system state estimation through strategic manipulation of meter measurements [7]. These vulnerabilities underscore the need for defense mechanisms capable of maintaining data integrity and detecting stealthy intrusions in real time.

Consequently, intrusion detection systems (IDSs) have become central to smart-grid cybersecurity frameworks [8], [9]. Machine learning (ML) significantly enhances IDS performance by learning complex attack patterns and adapting to evolving threats [10], [11]. A variety of ML algorithms have been proposed. Random forests have been used to detect disconnection and volt-var anomalies based on GridLAB-D simulation data [12], [13], illustrating the ML perspective in intrusion detection. However, the reliance on simulated data raises concerns about the applicability of the models to real-world operating conditions. Hybrid long short-term memory (LSTM) -isolation forest models establish a simple correlation between photovoltaic (PV) power output and environmental sensor data [14], but they overlook voltage and current, which are more detailed and informative electrical measurements. As a result, abnormal voltage-current combinations can still yield a seemingly normal power output, masking potential anomalies. Unsupervised techniques such as principal component analysis (PCA) combined with isolation forests effectively identify integrity breaches in high-dimensional communication systems [15], [16], but detection at the network layer, without a clear understanding of the normal behavior in local measurements,

Manuscript received: August 5, 2025; revised: October 21, 2025; accepted: December 15, 2025. Date of CrossCheck: December 15, 2025. Date of online publication: January 15, 2026.

This work was supported in part by the Advanced Research Institute at Virginia Tech.

Z. Zhang (corresponding author) and S. Rahman are with the Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University, Arlington, VA 22203, USA (e-mail: zheyuzhang21@vt. edu; srahman@vt.edu).

DOI: 10.35833/MPCE.2025.000713

makes it difficult to effectively mitigate the impact of attacks.

As attack complexity grows, hybrid IDS architectures have been developed to improve robustness. Density-based spatial clustering of applications with noise (DBSCAN) excels at identifying anomalies in multi-dimensional datasets [17], while convolutional neural network (CNN)-LSTM-transformer frameworks exploit both spatial and temporal correlations for high-accuracy detection [18], [19]. Similarly, [20] employed a CNN-LSTM model on a PMU-based monitoring system to detect abnormal voltage and frequency variations [21]. Hierarchical LSTM models also demonstrated reliable real-time detection of time-delay attacks [22]. Nevertheless, without a clear understanding of the normal operating behavior, it remains difficult to design effective mitigation strategies, which explains why most existing studies focus primarily on detection rather than recovery.

Despite these advances in IDSs, mitigation remains underdeveloped [23]. Recovering post-attack states is difficult due to inverter dynamics and environmental variations. Recent studies explored control-based mitigation and deep reinforcement learning (DRL) for adaptive volt-var and volt-watt regulation under FDIA [24]. Forecasting and behavior-based strategies have also been introduced: Gaussian conditional random fields for spatial-temporal prediction [25], cross-site validation across neighboring PV farms [26], and *I-V* curve analysis for early fault detection [27]. Beyond ML, trust-based mechanisms provide an alternative approach. For instance, [28] proposed a trust-based control scheme that maintains data integrity without ML dependence.

However, most existing studies focus on large-scale systems without considering distributed energy systems [29]. The study in [30] developed a secondary control scheme with compensation modules for DC bus voltage recovery, but it is limited to the controller level. In practice, network-layer attacks should also be considered, since during such attacks, local devices may continue operating normally while data transmission becomes compromised. With an appropriate detection and mitigation framework, it is possible to preserve stable and reliable operation even under network disturbances.

Collectively, these studies emphasized the importance of integrating advanced detection and robust mitigation to enhance inverter resilience [31]. With ongoing progress in ML [32] and real-time monitoring technologies [33], the cybersecurity landscape continues to evolve toward intelligent, data-driven defenses. Yet most existing studies lack a well-defined baseline of normal inverter behavior, making it difficult to distinguish genuine attacks from natural variations such as solar irradiance and temperature changes. In addition, network-level cybersecurity for small-scale distributed energy systems remains underexplored, leaving a critical research gap.

The intrusion detection and mitigation system (IDMS) proposed in this study tackles these challenges by combining an IDS based on Transformer algorithm with a mitigation system, ensuring accurate intrusion identification and dependable recovery after cyberattacks on smart inverters of solar farms at the network level. The system relies on environmental sensors operating independently of the inverters, each using a dedicated communication link to transmit data directly to a local server instead of the central control unit. These sensors monitor key parameters: voltage and current, which strongly correlate with inverter measurements. Because they also capture fluctuations in solar irradiance, temperature, and other ambient factors, the collected data allow the system to infer how the inverters should behave under normal conditions. This expected behavioral baseline underpins both the anomaly detection and mitigation processes.

To enhance IDS performance, we incorporate the previous time-step inverter measurements, assuming historical data are clean due to 24 months of normal operation. This improves detection accuracy and reduces estimation error. However, in the mitigation stage, when attacks are ongoing, inverter measurements become unreliable. Using estimated values recursively would lead to error accumulation and degraded performance. To address this, the interface layer discards inverter measurements during mitigation to prevent propagation of compromised data. Additionally, the proposed IDMS continuously compares real-time measurements with model estimates. Once the difference falls below a defined threshold, indicating the attack has ceased, the proposed IDMS halts mitigation and switches back to detection mode.

For the experiments, 24 months of real-world operational data from solar panels and inverters are used to develop and evaluate the proposed IDMS. To simulate more practical conditions where long-term data may not be available, additional experiments are conducted using reduced dataset sizes to assess the performance of the proposed IDMS under limited data conditions.

In summary, the contributions of this paper are listed below.

1) A detection system with IDS model based on the Transformer algorithm is developed, which leverages environmental sensor data and last-moment inverter data to accurately capture the normal behavior of smart inverters. This system effectively detects cyberattacks on real-world solar farm devices, ensuring the security and reliability of operations.

2) A mitigation system is proposed, in which a prediction model based on LSTM and trained solely on environmental features is introduced to prevent recursive cumulative error propagation and ensure stable system operation.

3) An open dataset is proposed [34] to demonstrate the performance of the proposed IDMS under varying data availability conditions. This provides users with flexible options for strategies based on their specific data conditions.

II. Methodology

A. FDIAs on Smart Inverters of Solar Farms

FDIAs on smart inverter measurements have been well studied and categorized in existing literature. They are generally classified into four types: over-inflation attacks, random attacks [35], replay attacks [36], and delay attacks. These attacks exploit vulnerabilities in the communication channel between the inverter and the control center server, which is

responsible for receiving operational measurements. By intercepting and altering the transmitted data, attackers can artificially inflate or randomly modify the smart inverter measurements, potentially leading to system instability and disruptions in normal operation.

1) Over-inflation attacks: attackers increase the output voltage and current of smart inverters.

$$\hat{x}_t = \begin{cases} x_t & t \notin \tau_t \\ (1+\lambda_f)x_t & t \in \tau_t \end{cases} \tag{1}$$

where τ_t is the attack period; x_t denotes the smart inverter measurement; $\hat{x}_t$ denotes the information received in the server through the communication; and λ_f denotes the over-inflation factor designed by the attackers.

2) Random attacks: attackers randomly alter the voltage and current measurements of smart inverters.

$$\hat{x}_t = \begin{cases} x_t & t \notin \tau_t \\ \lambda_r x_t & t \in \tau_t \end{cases} \tag{2}$$

where λ_r denotes the random factor designed by the attackers.

3) Replay attacks: attackers gain access to read historical data and replay the same period of data by using the stored information.

$$\hat{x}_t = \begin{cases} x_t & t \notin \tau_t \\ x_{t-d} & t \in \tau_t \end{cases} \tag{3}$$

where x_{t-d} denotes the smart inverter measurement after the replay attack.

4) Delay attacks: attackers create a channel to intentionally delay information transmission either to the inverter or from the inverter to the customer server, to disrupt the normal operation of the system.

$$\hat{x}_t = \begin{cases} x_t & t \notin \tau_t \\ x_{t-h} & t \in \tau_t \end{cases} \tag{4}$$

where x_{t-h} denotes the smart inverter measurement after the delay attack.

Since the attacks focus on the real-time communication layer, we assume the historical databases of both smart inverters and environmental sensors remain uncompromised. The environmental sensors, integrated into the solar panels, operate independently from smart inverters and use a separate communication channel. As non-intelligent devices without control capabilities, they are less attractive targets for attackers. Thus, we assume the environmental sensors function normally and provide trustworthy measurements.

B. ML Algorithms

1) Transformer Algorithm

The Transformer algorithm is a deep learning algorithm based on the attention mechanism proposed by [37]. Unlike traditional recurrent neural networks (RNNs) or LSTMs, the Transformer algorithm completely discards the recurrent structure. Instead, it uses self-attention and multi-head attention mechanisms to efficiently model global dependencies in sequential data. Its strong parallelization capabilities and sensitivity to long-range dependencies have made it a key algorithm in fields such as natural language processing (NLP) and time-series analysis.

The architecture of Transformer algorithm consists of two main components: the encoder and the decoder. Both components rely on the self-attention mechanism, which is the cornerstone of the algorithm. This mechanism enables the algorithm to compute contextual relationships between all elements in a sequence simultaneously.

The input sequence $\boldsymbol{X}$ is transformed into three separate representations: the query $\boldsymbol{Q}$ represents what the algorithm is searching for, the key $\boldsymbol{K}$ represents features used for matching, and the value $\boldsymbol{V}$ represents the information to be aggregated. These representations are computed through linear transformations using trainable weight matrices, denoted as $\boldsymbol{W}^{\boldsymbol{Q}}$, $\boldsymbol{W}^{\boldsymbol{K}}$, and $\boldsymbol{W}^{\boldsymbol{V}}$, respectively.

$$Attention(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = f_{\text{softmax}}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}}\right)\boldsymbol{V} \tag{5}$$

where $Attention(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V})$ computes a weighted sum of the value $\boldsymbol{V}$, where the weights are determined by the similarity between the query $\boldsymbol{Q}$ and the key $\boldsymbol{K}$; $\sqrt{d_k}$ is a scaling factor that prevents the inner product from becoming excessively large; and $f_{\text{softmax}}(\cdot)$ is a function that normalizes the attention scores.

Each encoder layer also includes a position-wise feed-forward network (FFN) in (6), which is applied independently to each position in the sequence. The FFN consists of two linear transformations with a ReLU activation:

$$FFN(\boldsymbol{x}) = ReLU(\boldsymbol{x}\boldsymbol{W}_1 + \boldsymbol{b}_1)\boldsymbol{W}_2 + \boldsymbol{b}_2 \tag{6}$$

where $\boldsymbol{W}_1$, $\boldsymbol{W}_2$, $\boldsymbol{b}_1$, and $\boldsymbol{b}_2$ are the trainable matrices; and $\boldsymbol{x}$ denotes the input features.

To incorporate the order of the sequence, the Transformer algorithm uses positional encodings (PEs), which are added to the input embeddings. These encodings allow the model to capture the relative and absolute positions of tokens within a sequence. The PEs are defined as:

$$PE(i, 2j) = \sin\left(\frac{i}{10000^{\frac{2j}{d_{\text{model}}}}}\right) \tag{7}$$

$$PE(i, 2j+1) = \cos\left(\frac{i}{10000^{\frac{2j}{d_{\text{model}}}}}\right) \tag{8}$$

where i is the position; j is the dimension; and d_{model} is the dimension of the input embeddings.

The Transformer algorithm leverages the self-attention mechanism and FFN to model complex dependencies in sequential data. It is adopted for anomaly detection due to its superior ability to capture long-range temporal relationships and global correlations among sensor variables. In smart inverters of solar farms, where abnormal patterns may evolve gradually or intermittently, modeling long-term dependencies is crucial. This is a capability that recurrent networks such as LSTM, gated recurrent unit (GRU), and RNN often fail to sustain due to vanishing gradients.

2) LSTM

For the sequential model, LSTM is designed with a unique architecture that significantly outperforms convention-

al RNNs in handling sequential data [38]. The governing equations of LSTM are as follows:

$$\boldsymbol{g}_r = \sigma(\boldsymbol{W}_r[\boldsymbol{s}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_r) \tag{9}$$

$$\boldsymbol{i}_p = \sigma(\boldsymbol{W}_p[\boldsymbol{s}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_p) \tag{10}$$

$$\hat{\boldsymbol{C}}_t = \tanh(\boldsymbol{W}_h[\boldsymbol{s}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_h) \tag{11}$$

$$\boldsymbol{C}_t = \boldsymbol{i}_p \hat{\boldsymbol{C}}_t + \boldsymbol{g}_r \boldsymbol{C}_{t-1} \tag{12}$$

$$\boldsymbol{o}_u = \sigma(\boldsymbol{W}_u[\boldsymbol{s}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_u) \tag{13}$$

$$\boldsymbol{h}_t = \boldsymbol{o}_u \tanh(\boldsymbol{C}_t) \tag{14}$$

where $\boldsymbol{g}_r$ is the forget gate; $\boldsymbol{i}_p$ is the input gate; $\sigma(\cdot)$ is the sigmoid function; $\boldsymbol{o}_u$ is the output gate; $\boldsymbol{x}_t$ is the input state; $\boldsymbol{s}_t$ is the hidden state; $\boldsymbol{C}_t$ is the cell state; $\hat{\boldsymbol{C}}_t$ is the candidate cell state; $\boldsymbol{W}_r$, $\boldsymbol{W}_p$, $\boldsymbol{W}_h$, $\boldsymbol{W}_u$, $\boldsymbol{b}_r$, $\boldsymbol{b}_p$, $\boldsymbol{b}_h$, and $\boldsymbol{b}_u$ are the trainable weight matrices; and $\boldsymbol{h}_t$ is the hidden state.

An LSTM consists of multiple units, each with three gates: $\boldsymbol{g}_r$, $\boldsymbol{i}_p$, and $\boldsymbol{o}_u$. The forget gate in (9) controls how much past information to retain or discard. It takes the previous hidden state $\boldsymbol{s}_{t-1}$ and the current input state $\boldsymbol{x}_t$, along with the learned parameters $\boldsymbol{W}_r$ and $\boldsymbol{b}_r$. The sigmoid function outputs values between 0 and 1, regulating the influence of the previous cell state $\boldsymbol{C}_{t-1}$. The input gate in (10) updates the cell state with new information. The sigmoid function determines input relevance, while the hyperbolic tangent in (11) generates the candidate cell state $\hat{\boldsymbol{C}}_t$. The cell state $\boldsymbol{C}_t$ is determined by the combined effects of the input gate and forget gate in (12). The input gate determines how much new information is written into the memory cell, while the forget gate controls how much past information is retained. The cell state $\boldsymbol{C}_t$ serves as memory, preserving information across time and mitigating vanishing or exploding gradients. The output gate in (13) decides the information passed to the current hidden state. The output gate uses $\boldsymbol{s}_{t-1}$, $\boldsymbol{x}_t$, and $\boldsymbol{C}_t$, where the sigmoid function controls gate activation and the hyperbolic tangent produces the hidden state $\boldsymbol{h}_t$ in (14) as the output representation at the current time step.

For the mitigation system, LSTM is chosen for its ability to capture short-term temporal patterns and perform smooth interpolation, which is vital for reconstructing compromised inverter measurements. Unlike the Transformer algorithm, it avoids overfitting on limited abnormal segments and ensures stable regression under noisy conditions.

C. System Design

Environmental factors such as ambient temperature, solar irradiance, and wind speed significantly influence the power output of PV panels, while remaining independent of the internal measurements of inverters. Owing to the strong correlation between inverter-side voltage/current and these environmental factors, sensor data can be effectively utilized to characterize the normal operating behavior of inverters.

To exploit this relationship, an IDS model based on the Transformer algorithm is developed to estimate the DC-side voltage and current of the inverter from environmental inputs, including ambient temperature, module temperature, solar irradiance, and wind speed, which exhibit strong temporal correlations. Capturing these nonlinear dynamics is inherently complex. Therefore, several deep learning algorithms including Transformer, LSTM, GRU, and RNN are evaluated.

Incorporating previous inverter voltage and current readings as additional inputs further improves the predictive accuracy. Based on this model, a dynamic anomaly detection threshold is defined using historical prediction errors and a tunable anomaly factor. During real-time operation, differences between actual and predicted values exceeding this threshold are flagged as anomalies.

Under cyberattack conditions, the PV panels may continue operating normally while the reported data of inverters become compromised. When an anomaly is detected, compromised measurements are replaced with predicted values by the mitigation system. However, the recursive use of the predictions of the IDS model leads to error accumulation and reduced long-term reliability.

To address this issue, a prediction model based on LSTM and trained solely on environmental features is introduced, making it immune to compromised inverter data to work in the mitigation system. Both models share identical architectures and training procedures, differing only in their input features.

The proposed IDMS functions in two stages: training and inference. In the training stage, historical inverter and environmental data are used to develop the proposed IDMS. In the inference stage, real-time data are monitored by the proposed IDMS—if no anomaly is detected, inverter data are transmitted directly. When an intrusion occurs, the IDS triggers an alert, and the mitigation system substitutes the compromised data with predicted values. Once normal conditions are restored, the proposed IDMS resumes using real inverter measurements. Figure 1 illustrates the workflow of the proposed IDMS. Smart inverter continuously measures voltage and current at the solar farm, while environmental sensors—installed separately and typically from different manufacturers—record ambient conditions. As these sensors are independent of the inverter, they are less likely to be compromised, making them a reliable foundation for the proposed IDMS.

The IDS is designed to monitor inverter measurements by comparing them with prediction data generated by an ML algorithm using sensor inputs. To enhance performance, the system incorporates previous inverter measurements as additional inputs. A threshold T is calculated based on validation dataset, which is expressed as:

$$T = a \cdot percentile(\{|y_i - \hat{y}_i| i = 1, 2, \ldots, n\}, p) \tag{15}$$

where a is the anomaly factor; p is the percentage factor of the residual dataset; $percentile(\cdot)$ is the percentage of errors occurring in the validation dataset; n is the total number of data points; y_i is the actual value; and $\hat{y}_i$ is the predicted value in the validation dataset. The residuals are calculated as the difference between y_i and $\hat{y}_i$, and p is used to determine the base threshold.

The anomaly factor a is introduced to balance the false positive and false negative. A higher anomaly factor reduces the number of false positives but may allow more attacks to remain undetected, while a lower anomaly factor increases the detection number of attack points but may also result in more false positives. In the experiment, the anomaly factor is set to be 1.5 to identify anomalies effectively.

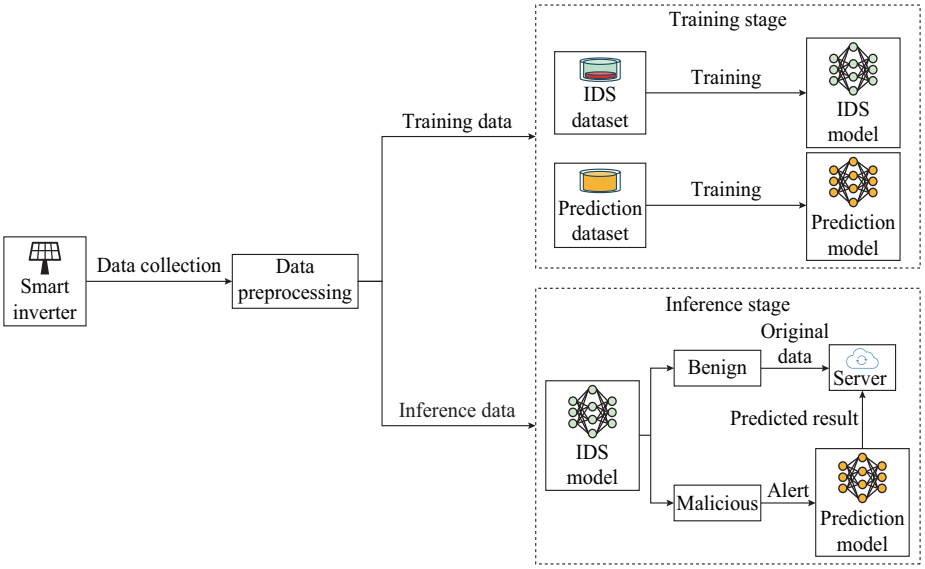

Fig. 1. Workflow of proposed IDMS.

If the difference between the actual and predicted values exceeds the threshold in the IDS model based on Transformer algorithm, the IDS flags an attack and triggers the mitigation system to replace compromised data with the results of the prediction model based on LSTM.

When the mitigation system is activated, it operates solely on environmental sensor inputs, excluding inverter measurements, which are presumed compromised. Incorporating prior inverter measurements in this phase would introduce error propagation and unnecessary computational overhead. By relying on independent sensor data, the prediction model based on LSTM ensures that predicted values remain unaffected by attacks or accumulated errors, thereby effectively mitigating the impact of compromised inverter measurements.

Simultaneously, the proposed IDMS continuously monitors the difference between the actual and predicted values to determine whether the attack has ceased. When the two values align closely enough to satisfy the predefined trigger conditions, the proposed IDMS resumes using actual inverter data. The trigger threshold plays a key role in reducing the false switching events and ensuring the stability of the proposed IDMS, allowing seamless transitions even under frequent switching conditions while maintaining reliable data transmission to the server.

The experiments were conducted using a comprehensive 24-month dataset. To account for conditions where such extensive data may not be available, additional evaluations were performed using 6-month and 12-month subsets. These tests revealed the influence of dataset size on detection and prediction accuracy. Although the IDS model based on Transformer algorithm and the prediction model based on LSTM trained on smaller datasets exhibited slightly reduced performance compared to the full 24-month dataset, the results remained satisfactory for real-time inference, demonstrating the adaptability of the proposed IDMS to varying data availability.

To assess the effectiveness of the IDS model based on Transformer algorithm under cyberattack conditions, various attack types were injected into 20% of the test data. This evaluation assesses the capability of the IDS model to differentiate between normal and abnormal operating conditions, with a focus on false positive occurrences and the accuracy of abnormal event detection. Random seeds were fixed to ensure reproducibility.

In the inference stage, a representative day was selected to visualize the behavior of the proposed IDMS under different attack conditions. The results illustrate how the proposed IDMS responds when an attack occurs and how the proposed IDMS dynamically transitions between normal and abnormal states.

D. Recursive Cumulative Error

The IDS model based on Transformer algorithm takes the feature vector $\boldsymbol{d}_t$ as the input, which includes the feature $f_{i,t}$ of data point i at time t, as defined in (16). The features include the ambient temperature, module temperature, solar irradiance, wind speed, and voltage and current from the previous time step. The voltage and current from the previous time step are derived from the inverter measurements, which are the target of the attacks.

$$\boldsymbol{d}_t=[f_{1,t},f_{2,t},\ldots,f_{n,t}] \tag{16}$$

If this IDS model is employed as a prediction model for the mitigation system during an attack, the inverter output becomes unreliable. In such cases, the input at the next time step $\boldsymbol{d}_{t+1}$ should be replaced with the predicted data $\hat{y}_t$ from the model instead of relying on the measurement, as expressed in (17).

$$\boldsymbol{d}_{t+1}=[\hat{y}_t,f_{1,t+1},f_{2,t+1},\ldots,f_{m,t+1}] \tag{17}$$

where m is the number of features that are not under attack.

This model then predicts $\hat{y}_{t+1}$ using the input $\boldsymbol{d}_{t+1}$, which includes the predicted value $\hat{y}_t$ as an input, along with the same system settings θ, as described in (18).

$$\hat{y}_{t+1}=f(\boldsymbol{d}_{t+1};\theta) \tag{18}$$

where $f(\cdot)$ is the prediction function from the model.

The prediction error e_{t+1} is defined in (19), and it can be further expressed as (20).

$$e_{t+1}=\hat{y}_{t+1}-y_{t+1} \tag{19}$$

$$e_{t+1}=f([\hat{y}_t,\boldsymbol{x}_{t+1}];\theta)-y_{t+1} \tag{20}$$

where y_{t+1} is the actual value.

Over k recursive steps, the error propagation e_{t+k} can be described as:

$$e_{t+k}=\sum_{i=1}^{k}\left(\prod_{j=i}^{k}\frac{\partial f}{\partial \hat{y}_{t+j-1}}\right)e_{t+i-1}+\sum_{i=1}^{k}\epsilon_i \tag{21}$$

where $\partial f/\partial \hat{y}_{t+j-1}$ denotes the error amplification; and ϵ_i denotes the random noise within the model.

Due to the cumulative error propagation, the recursive predictions of the IDS model, which are used to substitute compromised measurements, may progressively degrade in accuracy. As a result, this model cannot be reliably employed as the prediction model of the mitigation system.

III. Experimental Setup

A. Parameter Setting

The rooftop solar array used in this study has a capacity of 6.44 kW, and the data period spans 24 months from January 1, 2022 to December 31, 2023. The resolution of the inverter and sensor measurements is 1 min.

The first step focused on the normal operating period of the device through data cleaning. After the cleaning process, the dataset was subsequently split into training (80%), validation (10%), and test (10%) subsets. All algorithms were implemented in Python using PyTorch 2.4.0. For the algorithmic configuration, the architecture of Transformer algorithm consists of 6 encoder layers and 6 decoder layers, each equipped with multi-head attention mechanisms to effectively capture long-range temporal dependencies. The embedding dimension of the algorithm is set to be 32, with 8 attention heads, a feedforward dimension of 32, and a dropout rate of 0.2.

Three-layer LSTM, GRU, and RNN algorithms were also developed, incorporating layer normalization and dropout in each layer to enhance training stability and prevent overfitting. The hidden dimension is set to be 256, and the dropout rate is 0.2. For the CNN-LSTM, two convolutional structures were used, each with 16 output channels and 2 kernel filters.

B. Metrics

The mean absolute error (MAE) evaluates the average magnitude of the differences between actual and predicted values, disregarding their direction. It is calculated as:

$$MAE=\frac{1}{n}\sum_{i=1}^{n}|y_i-\hat{y}_i| \tag{22}$$

MAE is intuitive to interpret since it is expressed in the same units as the data, providing a clear sense of the average prediction error.

The mean squared error (MSE) computes the average of the squared differences between actual and predicted values. By squaring the errors, it gives more weight to larger discrepancies shown in (23).

$$MSE=\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2 \tag{23}$$

MSE is sensitive to outliers and is often used as a loss function in optimization due to its mathematical properties.

The root mean squared error (RMSE) is the square root of the MSE, representing the average magnitude of prediction errors in the same units as the original data, shown in (24).

$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2} \tag{24}$$

RMSE penalizes large errors more than MAE and is useful when outliers are significant.

The coefficient of determination R^2 measures the proportion of the variance in the dependent variable that is predictable from the independent variables. It is defined as:

$$R^2=1-\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y}_i)^2} \tag{25}$$

where $\bar{y}_i$ is the mean of the actual value.

An R^2 value of one indicates that the model explains all the variability of the dependent variable, signifying a perfect fit. Conversely, an R^2 value of 0 indicates that the model explains none of the variability, implying that it is no better than a simple average.

These metrics were used to evaluate model performance. For the IDS model, better performance indicates a higher ability to detect abnormal behavior, as the predicted values are closer to the actual values. For the prediction model based on LSTM in the mitigation system, better performance reflects lower error when compromised data are replaced with predicted values.

IV. Performance Evaluation

A. Performance in Training Stage

1) Performance of IDS Model in Detection System

The IDS model was designed with six inputs: ambient temperature, module temperature, wind speed, solar irradiance, last-moment interval voltage, and last-moment interval current. To determine the most suitable ML algorithm, five different algorithms were tested, and the one that achieved the best performance was used as the core algorithm of the IDS model.

Figure 2(a) illustrates the training loss of the five algorithms. The Transformer algorithm required more than 2000 epochs to achieve convergence, whereas the LSTM and RNN converged after approximately 600 epochs. In contrast, the GRU reached convergence in over 1500 epochs. Among all algorithms, the CNN-LSTM exhibited the fastest convergence, stabilizing within 50 epochs. The training loss for all the five algorithms decreased progressively with each epoch and reached its minimum point when early stopping was triggered.

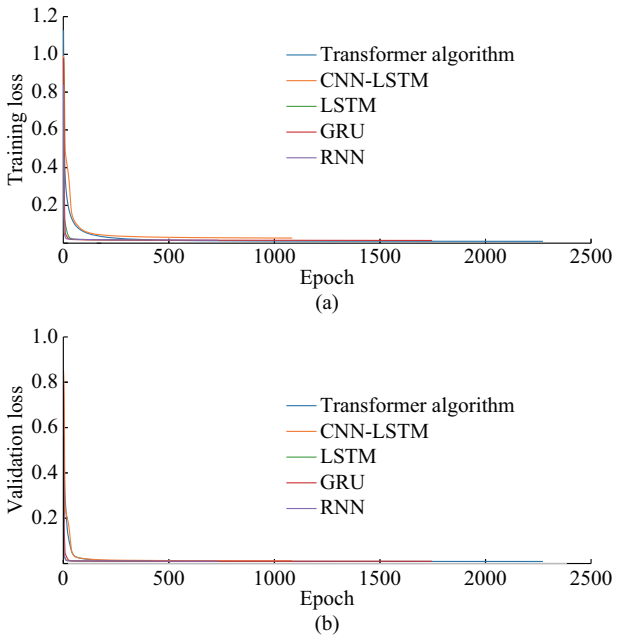

Fig. 2. Training and validation losses of different algorithms. (a) Training loss. (b) Validation loss.

Figure 2(b) illustrates the validation loss of the five algorithms, which also decreased steadily over epochs, aligning with the trend in the training loss. This indicates that all the five algorithms achieved good convergence performance without exhibiting overfitting. Consequently, the test data performance can be reliably used to evaluate the algorithms of the IDS model.

Table I presents the test data performance of the five algorithms.

TABLE I
TEST DATA PERFORMANCE OF DIFFERENT ALGORITHMS

Algorithm	Metric	Voltage	Current
Transformer	MAE	1.5205 V	0.2049 A
	MSE	7.4456 V^2	0.0904 A^2
	RMSE	2.7287 V	0.3007 A
	R^2	0.9777	0.9899
LSTM	MAE	1.6209 V	0.3447 A
	MSE	8.1129 V^2	0.2007 A^2
	RMSE	2.8483 V	0.4479 A
	R^2	0.9757	0.9776
GRU	MAE	1.5934 V	0.2516 A
	MSE	7.8023 V^2	0.1330 A^2
	RMSE	2.7933 V	0.3647 A
	R^2	0.9766	0.9852
RNN	MAE	1.6959 V	0.2583 A
	MSE	8.3570 V^2	0.1421 A^2
	RMSE	2.8908 V	0.3769 A
	R^2	0.9749	0.9842
CNN-LSTM	MAE	3.5007 V	0.5058 A
	MSE	21.9555 V^2	0.3648 A^2
	RMSE	4.6857 V	0.6040 A
	R^2	0.9342	0.9594

Among them, the Transformer algorithm achieved the best performance since it considered the overall parameters, with an MAE of 1.5205 V for voltage and 0.2049 A for current. The voltage range in the dataset spanned from 360.01 V to 506.77 V, while the maximum current value is 17.4 A. The performance of LSTM was slightly lower than that of the Transformer algorithm but better than GRU. GRU, a simplified version of LSTM, typically converged faster but may underperform LSTM on highly complex datasets. Both LSTM and GRU were designed to address the gradient explosion problem commonly encountered in traditional RNN. The CNN-LSTM exhibited the highest error among all algorithms, indicating that incorporating convolutional layers may not provide additional benefits. For the IDS model, a lower error indicated better performance in identifying abnormal conditions.

As shown in Table II, the Transformer algorithm achieved the highest F1 score of 0.9503, along with the lowest false-positive rate (FPR) of 3.75%. The LSTM, GRU, and RNN followed closely, while the CNN-LSTM exhibited the worst performance among all algorithms.

TABLE II
PERFORMANCE OF IDS MODEL BASED ON DIFFERENT ALGORITHMS IN DETECTION SYSTEM

Algorithm	F1 score	Precision	Recall	FPR
Transformer	0.9503	0.9616	0.9393	0.0375
LSTM	0.9447	0.9553	0.9343	0.0438
GRU	0.9471	0.9596	0.9349	0.0394
RNN	0.9456	0.9569	0.9346	0.0421
CNN-LSTM	0.9055	0.8819	0.9305	0.1247

In conclusion, the IDS model employing the Transformer algorithm achieved the lowest prediction error among the evaluated algorithms, demonstrating its effectiveness for accurate anomaly detection in the IDS.

2) Performance of Prediction Model in Mitigation System

While the IDS model showed promise as a prediction model for the mitigation system, its implementation during an attack where the output was replaced by the predicted value and subsequently used as input for the next prediction period will introduce cumulative errors, as demonstrated in Table III for the recursive algorithm.

Compared with the LSTM, the recursive algorithm performed worse on both the voltage and current predictions. Additionally, the recursive algorithm increased computation time because it required feeding the output back as input for the next prediction step. This iterative process significantly reduces computational efficiency compared with algorithms that rely solely on sensor data as input. For a test dataset of 37894 samples, the recursive algorithm took 330.5 s to complete predictions, while the LSTM required only 0.097 s for the same condition. Considering that the computing device used for deployment is likely to have lower performance than the experimental device, the 3407-fold computational differences should be accounted for to ensure the proposed IDMS can respond on time.

TABLE III
PERFORMANCE OF PREDICTION MODEL BASED ON DIFFERENT ALGORITHMS IN MITIGATION SYSTEM

Algorithm	Metric	Voltage	Current
LSTM	MAE	5.0087 V	0.2766 A
	MSE	58.3479 V^2	0.1637 A^2
	RMSE	7.6386 V	0.4046 A
	R^2	0.8251	0.9818
Transformer	MAE	5.3075 V	0.2807 A
	MSE	61.7027 V^2	0.1715 A^2
	RMSE	7.8551 V	0.4141 A
	R^2	0.8150	0.9809
CNN-LSTM	MAE	6.0395 V	0.6576 A
	MSE	76.7315 V^2	0.7069 A^2
	RMSE	8.7597 V	0.8408 A
	R^2	0.7700	0.9212
GRU	MAE	6.7555 V	0.3707 A
	MSE	79.2581 V^2	0.2276 A^2
	RMSE	8.9027 V	0.4771 A
	R^2	0.7624	0.9746
RNN	MAE	8.3908 V	0.2745 A
	MSE	109.5292 V^2	0.1775 A^2
	RMSE	10.4656 V	0.4213 A
	R^2	0.6716	0.9802
Recursive	MAE	5.1863 V	0.3423 A
	MSE	57.5333 V^2	0.2176 A^2
	RMSE	7.5851 V	0.4664 A
	R^2	0.8275	0.9758

In the prediction model, inputs from the inverter were discarded because, during an attack, the inverter is compromised by the attacker, and cumulative errors render it unnecessary to adapt. Instead, sensor data were used as input. In this case, LSTM demonstrated superior performance across all metrics compared with the Transformer algorithm, CNN-LSTM, GRU, and RNN.

The performance of the prediction model based on LSTM in the mitigation system was worse than that of the IDS model based on Transformer algorithm because the IDS model incorporated more features, leading to higher accuracy. However, the output of the prediction model based on LSTM was not influenced by the status of the inverter, which made it independent of the inverter and free from cumulative error issues.

In conclusion, for the mitigation system, the prediction model employing the LSTM achieved the lowest prediction error, demonstrating its effectiveness for accurate anomaly detection in smart inverters of solar farms.

3) Performance Under Different Dataset Sizes

In the proposed IDMS, improved model performance directly translates to stronger detection capabilities and reduced prediction errors. However, model performance is highly dependent on dataset size, with larger datasets generally yielding better performance. To ensure robustness, the experiments leveraged a comprehensive two-year dataset as a reliable evaluation baseline.

To reflect real-world conditions where such extensive data may be unavailable, the dataset was segmented into 6-month and 12-month subsets. Both models were evaluated in consistent settings, providing practical information on the performance of the proposed IDMS under varying dataset sizes.

Table IV shows the performance of the IDS model based on Transformer algorithm under different dataset sizes. As expected, larger dataset sizes yield better detection accuracy, emphasizing the value of long-term data for model effectiveness. Notably, even with only six months of data, the model maintains reasonable performance—MAE increases from 1.5205 to 2.2265 V—which remains acceptable given the system voltage and current ranges.

TABLE IV
PERFORMANCE OF IDS MODEL BASED ON TRANSFORMER ALGORITHM UNDER DIFFERENT DATASET SIZES

Time (month)	Metric	Voltage	Current
6	MAE	2.2265 V	0.3261 A
	MSE	11.8896 V^2	0.2479 A^2
	RMSE	3.4481 V	0.4979 A
	R^2	0.9414	0.9864
12	MAE	1.9396 V	0.3134 A
	MSE	8.7321 V^2	0.2267 A^2
	RMSE	2.9550 V	0.4355 A
	R^2	0.9633	0.9871
24	MAE	1.5205 V	0.2049 A
	MSE	7.4456 V^2	0.0904 A^2
	RMSE	2.7287 V	0.3007 A
	R^2	0.9777	0.9899

Similarly, Table V shows the performance of prediction model based on LSTM under different dataset sizes.

TABLE V
PERFORMANCE OF PREDICTION MODEL BASED ON LSTM UNDER DIFFERENT DATASET SIZES

Time (month)	Metric	Voltage	Current
6	MAE	6.4331 V	0.3788 A
	MSE	68.9778 V^2	0.3030 A^2
	RMSE	8.3053 V	0.5505 A
	R^2	0.7873	0.9834
12	MAE	5.1933 V	0.3766 A
	MSE	54.6614 V^2	0.2714 A^2
	RMSE	7.3933 V	0.5210 A
	R^2	0.7936	0.9666
24	MAE	5.0087 V	0.2766 A
	MSE	58.3479 V^2	0.1637 A^2
	RMSE	7.6386 V	0.4046 A
	R^2	0.8251	0.9818

The MAE increases from 5.0087 to 6.4331 V as the dataset size shrinks, reflecting the model sensitivity to dataset size. Nonetheless, this level of error is tolerable for fault da-

ta replacement, as it remains significantly lower than the impact of compromised measurements and is unaffected by potentially manipulated inverter data.

The experimental results show that larger dataset sizes enhance the performance of the prediction model. However, under practical conditions where only six months of data are available instead of the full 24 months of data, the system still maintains acceptable performance suitable for reliable operation.

B. Performance in Inference Stage

This subsection demonstrates the performance of the proposed IDMS during an attack on the inverter and its subsequent response. The proposed IDMS was designed to detect attacks and replace compromised data caused by the attack with prediction data. In order to demonstrate the prediction accuracy of the system, two similar days (November 13, 2023 and November 14, 2023) were selected to perform a comparative validation. An examination of the available dataset revealed that the performance (i.e., DC voltage and DC current) of the solar array on these two days was identical. November 14 was designated as the attack day, while the November 13 served as the reference day for performance comparison. The contaminated data of November 14 were passed through the filter algorithm and were replaced with data generated from sensor-based information.

A test window was established to evaluate the performance of the proposed IDMS to detect and mitigate four types of attacks on inverter measurements of voltage and current, as illustrated in Fig. 3.

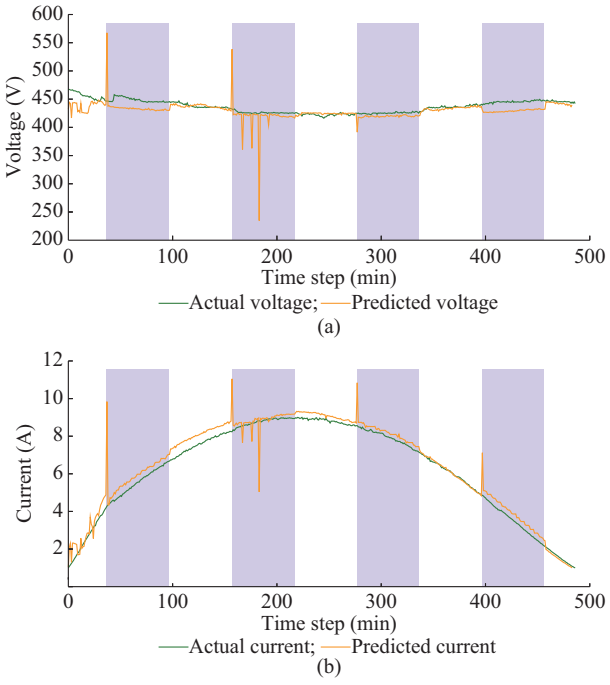

Fig. 3. Performance of proposed IDMS to detect and mitigate attacks on inverter measurements of voltage and current. (a) Voltage. (b) Current.

The attack period was highlighted with the purple bar, beginning with the over-inflation attack in the first attack period. When the attack occurred, all voltage and current values increased significantly. Without the proposed IDMS, this increase would persist throughout the entire attack period. However, with the proposed IDMS, the voltage and current values dropped once the IDS identified abnormal behavior. The mitigation system then replaced the compromised data with predicted data generated by the mitigation system. Although the predicted data still contained some errors, they were acceptable relative to the impact of the attack.

The attack in the second attack period was a random attack, meaning that both voltage and current values were randomly changed up and down. Once the IDS detected the compromised data, the mitigation system generated synthetic data using meteorological data and characteristics of the solar array and passed the information on to the designated server for use by the power system operators. It is important to note that the proposed IDMS was seamlessly interconnected. Even when the mitigation system was stopped, the IDS remained operational, continuously monitoring for abnormal behavior while inverter data flowed uninterrupted. When a new attack was detected, the mitigation system reactivated and repeated the process. This explains the multiple trigger points observed in the second attack period.

The attack in the third attack period was the replay attack, where attackers replaced the actual data with historical data from the same day and time in the previous month. Since power generation behaviors vary from day to day, the proposed IDMS can easily detect and mitigate the impact of this attack.

The attack in the final attack period was the delay attack, where attackers introduced a one-hour delay to compromise the operations of the inverter. Voltage and current were time-series measurements that evolved over time. The IDS, trained on time-series data, can effectively track temporal changes to identify normal behavior or detect anomalies caused by the delay attack. The mitigation system is also capable of adapting to temporal variations, generating acceptable values that can function as normal data.

The case study demonstrated that the proposed IDMS responds promptly and effectively, significantly reducing the impact of various types of cyberattacks through timely detection and mitigation.

V. Conclusion

This study presents an IDMS that identifies and counteracts cyberattacks on smart inverters in solar farms through ML and environmental sensors. Experimental results show that integrating inverter and sensor features yields accurate predictions and strong intrusion detection performance. However, when inverter data are included in the prediction model, cumulative error propagation occurs. To overcome this issue, a prediction model based on LSTM in the mitigation system is developed using only environmental inputs, and independent of inverter measurements.

The detection system and mitigation system in the proposed IDMS operate collaboratively. During an attack, the IDS activates the IDS model to generate substitute values. Once the threat subsides, the IDS seamlessly resumes using actual inverter data to maintain accuracy and prevent long-

term drift. A case study conducted on a representative day validates this mechanism, confirming that the proposed IDMS can detect anomalies, replace compromised data, and restore normal operation effectively. In addition, the open dataset is proposed to test the effectiveness of the proposed IDMS under different data sizes.

The proposed IDMS depends on environmental sensors and historical operational data. In real-world settings, sensor readings may be noisy, incomplete, or compromised. The proposed IDMS exhibits resilience to such noise by exploiting temporal correlations, maintaining stable estimates despite occasional measurement errors. Nevertheless, if environmental data are entirely missing or severely distorted, accurate prediction cannot be guaranteed. In such cases, spatial correlations among nearby sensors can provide useful redundancy—neighboring devices within the same area can serve as reference sources. Future work will focus on improving sensor reliability and strengthening system robustness against data uncertainty.

REFERENCES

[1] A. S. L. V. Tummala, R. Inapakurthi, and P. V. Ramanarao, "Observer based sliding mode frequency control for multi-machine power systems with high renewable energy," *Journal of Modern Power Systems and Clean Energy*, vol. 6, no. 3, pp. 473-481, May 2018.

[2] R. Zafar and H. R. Pota, "Multi-timescale coordinated control with optimal network reconfiguration using battery storage system in smart distribution grids," *IEEE Transactions on Sustainable Energy*, vol. 14, no. 4, pp. 2338-2350, Oct. 2023.

[3] K. Zeb, S. U. Islam, W. Uddin *et al*., "High-performance and multifunctional control of transformerless single-phase smart inverter for grid-connected PV system," *Journal of Modern Power Systems and Clean Energy*, vol. 9, no. 6, pp. 1386-1394, Nov. 2021.

[4] M. Ni, M. Li, J. Li *et al*., "Concept and research framework for coordinated situation awareness and active defense of cyber-physical power systems against cyber-attacks," *Journal of Modern Power Systems and Clean Energy*, vol. 9, no. 3, pp. 477-484, May 2021.

[5] B. Ahn, T. Kim, S. Ahmad *et al*., "An overview of cyber-resilient smart inverters based on practical attack models," *IEEE Transactions on Power Electronics*, vol. 39, no. 4, pp. 4657-4673, Apr. 2024.

[6] J. Chen, G. Liang, Z. Cai *et al*., "Impact analysis of false data injection attacks on power system static security assessment," *Journal of Modern Power Systems and Clean Energy*, vol. 4, no. 3, pp. 496-505, May 2016.

[7] H. Zhang, B. Liu, and H. Wu, "Smart inverter enabled meter encoding for detecting false data injection attacks in distribution system state estimation," *Journal of Modern Power Systems and Clean Energy*, vol. 13, no. 5, pp. 1776-1786, Sept. 2025.

[8] J. F. Gaviria, G. Narváez, C. Guillen *et al*., "Machine learning in photovoltaic systems: a review," *Renewable Energy*, vol. 196, pp. 298-318, Aug. 2022.

[9] H. Gao, D. Yang, G. Cai *et al*., "Machine learning-based reliability improvement of ambient mode extraction for smart grid utilizing isolation forest," *IEEE Transactions on Power Systems*, vol. 38, no. 5, pp. 4752-4760, Sept. 2023.

[10] Y. Li and J. Yan, "Cybersecurity of smart inverters in the smart grid: a survey," *IEEE Transactions on Power Electronics*, vol. 38, no. 2, pp. 2364-2383, Feb. 2023.

[11] B. Yang, F. Li, J. Ye *et al*., "Condition monitoring and fault diagnosis of generators in power networks," in *Proceedings of 2019 IEEE PES General Meeting*, Atlanta, USA, Aug. 2020, pp. 1-5.

[12] D. M. Shilay, K. G. Lorey, T. Weiz *et al*. (2017, Sept.). Catching anomalous distributed photovoltaics: an edge-based multi-modal anomaly detection. [Online]. Available: https://arxiv.org/abs/1709.08830

[13] D. P. Chassin, K. Schneider, and C. Gerkensmeyer, "GridLAB-D: an open-source power systems modeling and simulation environment," in *Proceedings of 2008 IEEE/PES Transmission and Distribution Conference and Exposition*, Chicago, USA, Apr. 2008, pp. 1-5.

[14] M. Ibrahim, A. Alsheikh, F. M. Awaysheh *et al*., "Machine learning schemes for anomaly detection in solar power plants," *Energies*, vol. 15, no. 3, pp. 1082-1091, Feb. 2022.

[15] S. Ahmed, Y. Lee, S. H. Hyun *et al*., "Unsupervised machine learning-based detection of covert data integrity assault in smart grid networks utilizing isolation forest," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 10, pp. 2765-2777, Oct. 2019.

[16] A. M. Saber, A. Youssef, D. Svetinovic *et al*., "Anomaly-based detection of cyberattacks on line current differential relays," *IEEE Transactions on Smart Grid*, vol. 13, no. 6, pp. 4787-4800, Nov. 2022.

[17] A. Parizad and C. J. Hatziadoniu, "Cyber-attack detection using principal component analysis and noisy clustering algorithms: a collaborative machine learning-based framework," *IEEE Transactions on Smart Grid*, vol. 13, no. 6, pp. 4848-4861, Nov. 2022.

[18] Z. Wang, Q. Cui, Z. Gong *et al*., "Anomaly identification for photovoltaic power stations using a dual classification system and gramian angular field visualization," *Processes*, vol. 12, no. 4, pp. 690-708, Mar. 2024.

[19] E. M. Al-Ali, Y. Hajji, Y. Said *et al*., "Solar energy production forecasting based on a hybrid CNN-LSTM-transformer model," *Mathematics*, vol. 11, no. 3, p. 676, Feb. 2023.

[20] M. Zia and H. Nazaripouya, "Detection of fire-ignition electrical faults for preventing electrical wildfires," in *Proceedings of 2023 North American Power Symposium*, Asheville, USA, Oct. 2023, pp. 1-5.

[21] E. Khaledian, S. Pandey, P. Kundu *et al*., "Real-time synchrophasor data anomaly detection and classification using isolation forest, KMeans, and LoOP," *IEEE Transactions on Smart Grid*, vol. 12, no. 3, pp. 2378-2388, May 2021.

[22] P. Ganesh, X. Lou, Y. Chen *et al*., "Learning-based simultaneous detection and characterization of time delay attack in cyber-physical systems," *IEEE Transactions on Smart Grid*, vol. 12, no. 4, pp. 3581-3593, Jul. 2021.

[23] F. Li, R. Xie, B. Yang *et al*., "Detection and identification of cyber and physical attacks on distribution power grids with PVs: an online high-dimensional data-driven approach," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 10, no. 1, pp. 1282-1291, Feb. 2022.

[24] C. Roberts, S. T. Ngo, A. Milesi *et al*., "Deep reinforcement learning for mitigating cyber-physical DER voltage unbalance attacks," in *Proceedings of 2021 American Control Conference*, New Orleans, USA, May 2021, pp. 2861-2867.

[25] B. Zhang, P. Dehghanian, and M. Kezunovic, "Spatial-temporal solar power forecast through use of Gaussian conditional random fields," in *Proceedings of 2016 IEEE PES General Meeting*, Boston, USA, Jul. 2016, pp. 1-5.

[26] K. G. Lore, D. M. Shila, and L. Ren, "Detecting data integrity attacks on correlated solar farms using multi-layer data driven algorithm," in *Proceedings of 2018 IEEE Conference on Communications and Network Security*, Beijing, China, May-Jun. 2018, pp. 1-9.

[27] A. F. Amiri, H. Oudira, A. Chouder *et al*., "Faults detection and diagnosis of PV systems based on machine learning approach using random forest classifier," *Energy Conversion and Management*, vol. 301, p. 118076, Feb. 2024.

[28] L. Ma, G. Xu, L. Ma *et al*., "Distributed resilient voltage and reactive power control for islanded microgrids under false data injection attacks," *Energies*, vol. 13, no. 15, pp. 3828-3856, Jul. 2020.

[29] L. Zhao, J. Li, Q. Li *et al*., "A federated learning framework for detecting false data injection attacks in solar farms," *IEEE Transactions on Power Electronics*, vol. 37, no. 3, pp. 2496-2501, Mar. 2022.

[30] D. Zhou, Q. Zhang, F. Guo *et al*., "Distributed resilient secondary control for islanded DC microgrids considering unbounded FDI attacks," *IEEE Transactions on Smart Grid*, vol. 15, no. 1, pp. 160-170, Jan. 2024.

[31] M. Beikbabaei, M. Montano, A. Mehrizi-Sani *et al*., "Mitigating false data injection attacks on inverter set points in a 100% inverter-based microgrid," in *Proceedings of 2024 IEEE PES Innovative Smart Grid Technologies Conference*, Washington, USA, Feb. 2024, pp. 1-5.

[32] R. Divya, S. Umamaheswari, and A. A. Stonier, "Machine learning based smart intrusion and fault identification (SIFI) in inverter based cyber-physical microgrids," *Expert Systems with Applications*, vol. 238, p. 122291, Mar. 2024.

[33] Y. He, G. J. Mendis, and J. Wei, "Real-time detection of false data injection attacks in smart grid: a deep learning-based intelligent mechanism," *IEEE Transactions on Smart Grid*, vol. 8, no. 5, pp. 2505-2516, Sept. 2017.

[34] Z. Zhang. (2025, Mar.). Smart inverter detection and response dataset. [Online]. Available: https://github.com/zzy986/IDMS-dataset

[35] M. Sun, L. He, and J. Zhang, "Deep learning-based probabilistic

anomaly detection for solar forecasting under cyberattacks," *International Journal of Electrical Power & Energy Systems*, vol. 137, p. 107752, May 2022.
[36] D. Hou, Y. Sun, V. Dinavahi *et al*., "Adaptive two-stage unscented Kalman filter for dynamic state estimation of synchronous generator under cyber attacks against measurements," *Journal of Modern Power Systems and Clean Energy*, vol. 12, no. 5, pp. 1408-1418, Sept. 2024.
[37] A. Vaswani, N. Shazeer, N. Parmar *et al*. (2017, Jun.). Attention is all you need. [Online]. Available: https://arxiv.org/abs/1706.03762
[38] J. He, J. Salazar, K. Yao *et al*. (2024, Feb.). Zero-shot end-to-end spoken language understanding via cross-modal selective self-training. [Online]. Available: https://arxiv.org/html/2305.12793v2

Zheyu Zhang received the M.S. degree from Virginia Polytechnic Institute and State University, Arlington, USA, in 2023, where he is pursuing the Ph.D. degree. His research interests include cybersecurity, smart inverter, and machine learning.

Saifur Rahman is the founding director of the Advanced Research Institute at Virginia Polytechnic Institute and State University, Arlington, USA, where he is the Joseph R. Loring Professor of electrical and computer engineering. He also directs the Center for Energy and the Global Environment, Arlington, USA. He is a Life Fellow of the IEEE and an IEEE Millennium Medal winner. He was the 2023 IEEE President and CEO and was the President of the IEEE Power and Energy Society (PES) for 2018 and 2019. He is the Founding Editor-in-Chief of the IEEE Electrification Magazine and the IEEE Transactions on Sustainable Energy. His research interests include alternate energy, smart grid, and environmental impact.