

Research Article

FEMUS-Nowcast: A Robust Deep Learning Model for Sky Image–Based Short-Term Solar Forecasting Under Adversarial Attacks

Animesh Sarkar Tusher ¹, M. A. Rahman ², Md. Rashidul Islam ¹, Sushanto Bosak ¹,
 and M. J. Hossain ³

¹Department of Electrical & Electronic Engineering, Rajshahi University of Engineering & Technology, Kazla, Rajshahi 6204, Bangladesh

²School of Information Technology, Deakin University, Waurn Ponds, Victoria 3216, Australia

³School of Electrical and Data Engineering, University of Technology Sydney, Ultimo, New South Wales 2007, Australia

Correspondence should be addressed to Md. Rashidul Islam; rashidul@eee.ruet.ac.bd

Received 27 March 2025; Revised 27 July 2025; Accepted 30 August 2025

Academic Editor: D. P. Rai

Copyright © 2025 Animesh Sarkar Tusher et al. International Journal of Energy Research published by John Wiley & Sons Ltd. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Accurate short-term solar power forecasting (nowcasting) facilitated by smart devices and cyberinfrastructure, which uses sky images and artificial intelligence (AI)–based models, is susceptible to cyberattacks. This study investigates the vulnerabilities of deep learning (DL) and artificial neural network (ANN)–based sky image–based nowcasting models to adversarial attacks such as fast gradient sign method (FGSM), projected gradient descent (PGD), and a mixed attack template, along with proposing a feature extraction–based multi-unit solar (FEMUS)-Nowcast model. Results reveal that adversarial attacks significantly degrade all models' accuracy and lead them to an unusable state. Moreover, FGSM is found to be the most severe attack, with root mean square error (RMSE) increasing by 5–16 times and mean absolute error (MAE) increasing by 4–12 times compared to the normal scenario under maximum perturbation. As the proposed FEMUS-Nowcast outperforms models of existing literature, reducing RMSE by 48% and 25% under normal conditions, adversarial training is adapted to enhance its robustness in the presence of cyberattacks. Furthermore, adversarially trained (AT) FEMUS-Nowcast shows no RMSE or MAE trade-offs under all scenarios. Additionally, the AT FEMUS-Nowcast model demonstrates high resilience against advanced attacks, including iterative FGSM (I-FGSM) and momentum I-FGSM (MI-FGSM), confirming its reliability and robustness across diverse attack scenarios.

Keywords: adversarial training; cyberattacks; nowcasting; sky images; solar energy

1. Introduction

The increasing global demand for energy, coupled with a growing emphasis on clean energy solutions driven by environmental concerns and ambitious climate targets, has led to a significant rise in renewable energy generation [1]. Among available renewable energy sources, solar power is one of the most prominent ones, as indicated in [2, 3]. Solar photovoltaic (PV) energy has taken center stage, accounting for 60% of the global annual renewable capacity additions in 2021 [4]. However, solar energy's intermittent and unpredictable nature

necessitates accurate solar radiation prediction for various applications such as grid connections, electricity markets, power system operations, and solar power plants [4–6]. Consequently, the demand for short-term solar irradiance forecasting or nowcasting has witnessed exponential growth since 2011 [7], driven by the significant costs associated with errors in solar irradiance prediction [6]. At the same time, smart devices, Internet of Things (IoTs), machine learning (ML) and deep learning (DL)–based models and communication technologies are being used to connect smart grids with renewable energy sources [8].

The fast adoption of cutting-edge technologies in smart grids has enabled data-driven techniques while raising serious concerns regarding cybersecurity issues [9]. Despite offering many benefits, the bidirectional information flow and data reliance in the generation, transmission, and distribution areas of digitalized and automated power systems have raised the risk of cyberattacks [10]. Renewable energy generation stations are already within the attack surface due to the integration of cyberinfrastructure with physical systems. For instance, in 2022, three wind energy companies in Germany experienced cyberattacks that resulted in the collapse of several digitally operated wind turbines [11], and a cyberattack on a hydro and nuclear power plant in Korea happened in 2014 [12]. Due to these alarming events, investigations regarding cybersecurity are emphasized for renewable energy sectors, especially while forecasting their generations [13]. As accurately predicting solar power generation is very important for smart grid reliability and efficiency, this study investigates the associated potential risk along with mitigation techniques.

Accurate solar power nowcasting is crucial for managing short-term variability in solar generation and maintaining grid stability. In recent years, there has been a noticeable shift from purely data-driven methods to image-based approaches, particularly those that utilize sky images from all-sky imagers (ASIs). This shift is motivated by several key challenges and opportunities as follows:

1. Data-driven methods rely on on-site measurements, which often lack spatial awareness of incoming cloud formations. This makes it difficult to predict sudden drops or rises in solar output caused by fast-moving clouds [14].
2. Numerical weather prediction (NWP) offers broader spatial forecasting, but its temporal resolution (typically hourly or more) is too coarse for short-term solar forecasting needs [15].
3. Moreover, NWP models update slowly, which fails to meet the rapid response requirements of grid-integrated solar systems [16].
4. In contrast, sky images provide rich spatial and temporal information about cloud coverage, movement, and formation in near real-time, making them better suited for rapid forecasting [17].
5. Clouds are the primary source of uncertainty in solar power generation. Being able to observe and interpret cloud dynamics visually allows for more informed and localized forecasting [18].
6. ASI-based systems capture continuous hemispheric views of the sky, enabling high-resolution temporal updates that are difficult to achieve using conventional sensors or weather stations [19].
7. DL models applied to sky images can automatically extract meaningful cloud features and learn temporal patterns to improve forecasting accuracy [20, 21].

As a result, ASI-based image-driven approaches have emerged as a promising and rapidly growing area of research

in solar forecasting [22]. These methods offer the potential to bridge the gap between spatial awareness and real-time responsiveness—two essential aspects for future smart grid and microgrid applications. Moreover, the ASI-based solar power forecasting has earned attention in recent years due to its ability to identify solar positions, estimate cloud heights, predict cloud movements with velocity, and its cost-effectiveness compared to sensor-based forecasting [4, 21]. DL models for short-term solar forecasting (nowcasting) have gained popularity due to their ability to extract complex underlying structures, their implementation without requiring feature engineering, and their reduced sensitivity to missing data [23]. Recent work by Hategan et al. [24] introduced a dynamic weighted ensemble model combining ML algorithms and sky image-based inputs, showing notable improvements over traditional methods, especially in high-variability sky conditions and longer forecast horizons (>15 min). Their integration of both direct and indirect forecasting strategies, including sky image processing with ExtraTrees and physical models, addresses limitations such as persistence and the lack of environmental awareness in conventional models. Similarly [25] proposed a PatchDLinear-based framework that fuses cloud image features (extracted using Cloud Y-Net) with meteorological and historical power data, while emphasizing the importance of combining deep image representations with context-aware time series models. Furthermore, a recent work [26] presented a residual vision reformer framework that integrates a multi-stream video vision transformer with an attentive residual fully connected layer and a fused reformer module. This model captures both localized and long-range spatiotemporal patterns from sky image sequences.

Besides, Hou et al. [27] proposed a unique approach by explicitly incorporating sun visibility and cloud type classifications into the nowcasting model. Using multivariate long short-term memory (LSTM)-based architectures (LSTNet and TPA-LSTM), this work forecasted short-term cloud modification factors and sun visibility up to 1 h ahead. Also, Wei et al. [28] introduced a novel multimodal data fusion and fuzzification framework for ultrashort-term global horizontal irradiance forecasting. Their method fuses normal and underexposed ASIs with a time-information layer converted into spatial image format using astronomical models. A Swin-Transformer backbone then extracts predictive features, and a fuzzy decoder is used to model the global horizontal irradiance in a probabilistic form, mitigating high nonlinearity due to cloud dynamics.

However, existing DL-based models struggle to nowcast accurately on cloudy days, exhibit limited precision in cloud cover dynamics, and demonstrate reduced persistence on overcast days. Also, deep CNN models, such as [5], have the risk of being underfitted due to their shallow nature. Hence, a research study is required to improve forecasting accuracy for ASI-based forecasting schemes. Furthermore, despite the economic and grid reliability importance of accurate solar power forecasting and the advancements in ASI-based artificial intelligence (AI) methods, the impact of cyberattacks on these forecasts remains largely unexplored, even though the risk and vulnerabilities to cyberattacks have already been proven. For

TABLE 1: Overview of relevant research works.

Ref.	Method	Category	Cybersecurity	Attack type
[24]	Weighted ensemble learning	●	×	—
[25]	Multimodal fusion framework	●	×	—
[26]	Deep hybrid transformer framework	●	×	—
[27]	Multivariate DL models (LSTNet and TPA-LSTM)	●	×	—
[28]	Swin-transformer backbone with fuzzy decoder	●	×	—
[5]	CNN	●	×	—
[21]	DNN	●	×	—
[29]	DNN	○	✓	FGSM and PGD
[30]	ANN	○	×	FGSM and R-FGSM
[31]	LSTM and Bi-LSTM	○	✓	FGSM

Note: ●, ASI-based; ○, data-based.

instance, the impacts of cyberattacks on forecasting methods can be severe, as highlighted in a recent review study [13], which reviews the vulnerabilities of forecasting methods along with security measures for different areas of smart grids. In particular, the vulnerabilities of DL and artificial neural network (ANN) models in sensor-based solar PV power generation forecasting by adversarial attacks have been addressed in previous studies [29–31] due to the feature of bidirectional information flow and data dependency inherent in these technologies. However, notable limitations remain across these studies. For instance, Kuzlu et al. [30] primarily focused on fast gradient sign method (FGSM) and regression-based FGSM (R-FGSM) attacks. However, it did not explore more advanced methods such as projected gradient descent (PGD) or iterative FGSM (I-FGSM), thereby limiting the generalizability of its findings across diverse threat scenarios. Moreover, while the study highlighted the susceptibility of ANNs to adversarial perturbations, it did not propose or evaluate any defense mechanisms, leaving a critical gap in practical applicability. Another study [29] demonstrated that adversarial training with PGD enhances model robustness, but compromises performance on clean data. Despite identifying this trade-off, the work did not explore alternative training strategies that might better balance accuracy and resilience. Additionally, its exploration of black-box attacks was restricted to data-level transferability and did not incorporate more sophisticated query-based methods, such as surrogate model construction, which could provide deeper insights into real-world attack feasibility. Similarly, Kuzlu et al. [31] evaluated LSTM and BiLSTM models under adversarial conditions, and also relied solely on I-FGSM for both attack and defense analysis. Although this approach offers simplicity, it restricts the broader understanding of model behavior under varied and more complex adversarial threats. Collectively, these limitations underscore the need for more comprehensive evaluations using diverse attack strategies and robust defense mechanisms to better secure solar forecasting models against adversarial perturbations. Moreover, an overview of existing approaches in Table 1 shows that while ASI-based models have gained popularity compared to data-driven methods,

cybersecurity considerations remain largely unaddressed in ASI-based solar forecasting research.

Therefore, to the best of the authors' knowledge, no studies have focused on analyzing the impacts of cyberattacks on ASI-based AI methods for solar irradiance nowcasting or forecasting. As well as considering the limitations of existing research that rely solely on data-driven methods and those that address cybersecurity in isolation, the following key research problems (RPs) remain largely unresolved and demand further investigation.

- (RP1) How can the performance of solar nowcasting models be improved under cloudy sky conditions, which introduce high variability and uncertainty in solar irradiance?
- (RP2) How do ASI-based AI models respond to cybersecurity threats, particularly adversarial attacks, and what is their level of resilience in operational environments?
- (RP3) Are adversarially trained (AT) ASI-based AI models robust against unseen or adaptive cyberattacks and can they generalize to evolving threat scenarios?

Hence, to address the abovementioned RPs, this work investigates the vulnerability of DL and ANN-based solar nowcasting models to cyberattacks through a comparative analysis while proposing a countermeasure. The contributions of this work can be summarized as follows:

1. This study proposes a novel DL architecture, feature extraction-based multi-unit solar (FEMUS)-Nowcast, for solar power nowcasting. The model extracts and fuses spatial cloud features across multiple pathways, resulting in improved forecasting accuracy and robustness under normal conditions compared to existing CNN-based models.
2. This work presents one of the first systematic evaluations of adversarial attacks on ASI-based solar nowcasting models. Also, the study analyzes the impact of both single and mixed attack templates on forecasting

reliability across varying weather conditions and perturbation strengths.

3. An adversarial training framework tailored to the sky image forecasting domain is introduced. This framework significantly enhances the robustness of forecasting models against deceptive data attacks and mitigates performance degradation effectively.
4. This work demonstrates that the AT FEMUS-Nowcast model maintains high forecasting performance under strong adversarial scenarios. This highlights its potential as a resilient and cyber-secure solution for real-world solar energy systems.

The structure of this paper is organized as follows: An overview of methodologies employed in this work is discussed briefly in Section 2, which includes modeling of cyberattacks and algorithms for countermeasures along with a brief discussion on image quality assessment (IQA) metrics. The proposed model and framework are introduced in Section 3. Then, a detailed discussion of the model training process, cyberattack implementation procedures, and performance analysis of all models under various scenarios are presented in Section 4. A summarized discussion based on the experimental results is also presented in Section 4. Finally, Section 5 concludes the paper.

2. Background

Sky image-based PV power forecasting methods have gained significant attention in recent years, as mentioned earlier. These methods include ANN- or DL-based models to extract features such as cloud coverage, motion, and density from sky images. However, the application of ANN- or DL-based models in smart grids is susceptible to adversarial attacks, as addressed in [13]. Hence, to investigate the vulnerability of solar power nowcasting models, this work employs two different adversarial attack models. On the other hand, considering the wide application of adversarial training as a countermeasure against such attacks, this method is employed in this work. Furthermore, as images are utilized for solar power nowcasting, IQA metrics are employed to ensure the quality of service. Two different IQA metrics are utilized to assess mathematical and perceptual errors. All these methodologies are summarized as follows.

2.1. Cyberattack Models. To assess the impact of cyberattacks on the existing established models and proposed models, this study adopts established adversarial attack methods due to their significant impact on DL in recent years. Adversarial attacks involve generating malicious examples to deceive trained ML models. Given a supervised dataset (x, y) , a deep neural network (DNN) with parameters θ is trained to predict label y as $f_\theta(x)$. However, adversarial examples are crafted by slightly perturbing x to maximize the change in the inference result.

Adversarial attacks can be categorized as targeted or non-targeted based on their specific goals. Targeted attacks aim to mislead the model into producing a specific incorrect result, while non-targeted attacks simply seek to cause any incorrect

inference. Additionally, adversarial attacks can be classified as single-step or iterative methods based on their design. Single-step attacks, like the FGSM, calculate the loss gradient once to generate the perturbation for each example. In contrast, iterative attacks, like PGD, iteratively refine the perturbation to achieve an effect. FGSM and PGD were also employed in the literature [29] to investigate the effects of cyberattacks on data-driven solar forecasting. Hence, this study utilizes the same cyberattack frameworks to assess their impact on ASI-based forecasting.

On the other hand, to validate the reliability of the proposed AT method, advanced attack methods such as I-FGSM [32] and momentum I-FGSM (MI-FGSM) [33] were employed additionally. These iterative attacks provide a more comprehensive assessment of the proposed model's robustness against adversarial scenarios.

2.1.1. FGSM Attacks. Goodfellow et al. [34] demonstrated that, unlike nonlinear models, high-dimensional linear models are more accurate and capable of generating adversarial examples. The FGSM was proposed to quickly produce adversarial examples [34]. Let x denote the original or "clean" samples, ζ represent the perturbation applied to each x , and y_1 denote the supervised learning label. The perturbation ζ should satisfy the constraint $\|\zeta\|_\infty < \delta$, where δ is the magnitude constraint of the perturbation. The adversarial examples are generated as $x' = x + \zeta$. The perturbation ζ can be computed by Equation (1).

$$\zeta = \delta \cdot \text{Sign}(\nabla_x J_\theta(x, y_1)), \tag{1}$$

where J_θ represents the Jacobian matrix as a function of x and y_1 , θ represents the model parameters, and $\text{Sign}(\cdot)$ ensures the maximized change caused by the perturbation.

Once the perturbation ζ is obtained, the weighted augmented perturbation is given by $\omega^T x' = \omega^T x + \omega^T \zeta$. If the weight ω has dimension d and mean m , it can be observed that the activation could be increased by δdm . In high-dimensional problems, small perturbations to each dimension could add up to make a large change in the output.

2.1.2. PGD Attacks. In [35], the iterative PGD technique was presented as a multi-step FGSM variation. PGD is capable of generating adversarial examples and provides a potential method to defend against first-order adversarial attacks. When used as a defense mechanism, PGD generates adversarial examples and incorporates them into the training process to enhance the robustness of the DNN model [29]. Equation (2) represents the saddle point problem that PGD was originally designed to tackle.

$$\min_{\theta} R(\theta), \tag{2}$$

where $R(\theta) = \mathbb{E}_{(x,y_1) \sim D} [\max_{s \in S} \mathcal{L}(\theta, x + \xi, y_1)]$ represents the population risk, which is also the objective function. Here, D denotes the distribution of samples that defines the distribution of x and y_1 , and S is a non-empty compact topological space. The inner optimization aims to maximize the loss

Input: Initial input x , target label y , step size α , number of steps K , perturbation limit ϵ , model parameters θ ;
Output: Adversarial example x_a ;
1 Initialize $x_a \leftarrow x$;
2 for $k = 1$ to K **do**
3 Calculate gradient: $g_k \leftarrow \nabla_x J(\theta, x_a, y)$;
4 Update adversarial example: $x_a \leftarrow x_a + \alpha \cdot \text{Sign}(g_k)$;
5 Clip to permissible set: $x_a \leftarrow \text{Clip}_{x,\epsilon}(x_a)$;
6 Final output is the adversarial example: x_a ;

ALGORITHM 1: K -step PGD attack.

Input: Original input x , true label y , model f , perturbation bound ϵ , step size α , number of iterations T ;
Output: Adversarial example x^* ;
1 Initialize $x^{(0)} \leftarrow x$;
2 for $t = 1$ to T **do**
3 Compute gradient: $g \leftarrow \nabla_x J(x^{(t-1)}, y)$;
4 Update adversarial example: $x^{(t)} \leftarrow \text{Clip}_{x,\epsilon}(x^{(t-1)} + \alpha \cdot \text{sign}(g))$;
5 end
6 Return $x^* \leftarrow x^{(T)}$;

ALGORITHM 2: I-FGSM.

function $\mathcal{L}(\cdot)$ over S . In Equation (2), PGD seeks to find the model parameters that minimize the loss of the adversarial attack, thereby creating the most robust DNN network possible. Similar to FGSM, the inner optimization aims to maximize the loss function $\mathcal{L}(\cdot)$. Samples are more likely to become adversarial examples if they satisfy the maximization condition. With these two optimization components, the saddle point problem integrates both the generation of adversarial examples and the enhancement of DNN robustness against adversarial attacks. In practical implementation, a K -step PGD attack of L_∞ norm is executed in Algorithm 1.

2.1.3. I-FGSM. The I-FGSM is an extension of the FGSM attack, where adversarial examples are generated iteratively to refine the perturbations. At each iteration, the adversarial example is updated by adding the gradient sign of the loss with respect to the previous step. The process ensures that perturbations are small and constrained within an ϵ -ball around the original input. The update rule is expressed by Equation (3), and the entire process is expressed in Algorithm 2.

$$x^{(t+1)} = \text{Clip}_{x,\epsilon}(x^{(t)} + \alpha \cdot \text{sign}(\nabla_x J(x^{(t)}, y))), \quad (3)$$

where $x^{(t)}$ is the adversarial example at iteration t , α is the step size, $\nabla_x J(x^{(t)}, y)$ is the gradient of the loss function J with respect to the input x , and $\text{Clip}_{x,\epsilon}$ ensures the perturbations remain within the allowable range ϵ .

2.1.4. MI-FGSM. The MI-FGSM attack builds upon the I-FGSM approach by incorporating a momentum term into the gradient computation. This momentum term accumulates

gradients over iterations, which helps the adversarial example escape local minima and improves the attack's effectiveness. At each iteration t , the accumulated velocity vector in the gradient direction is updated by Equation (4).

$$g_{t+1} = \mu \cdot g_t + \frac{\nabla_x J(x_t, y)}{\|\nabla_x J(x_t, y)\|_1}, \quad (4)$$

where g_t is the momentum vector at iteration t , μ is the decay factor, $\nabla_x J(x_t, y)$ is the gradient of the loss function J with respect to the input x_t , and $\|\cdot\|_1$ represents the L_1 norm of the gradient. Then, the adversarial example x_{t+1} is then updated by Equation (5).

$$x_{t+1} = x_t + \alpha \cdot \text{sign}(g_{t+1}), \quad (5)$$

where α is the step size. The entire process is expressed in Algorithm 3.

2.2. Countermeasure Technique. A DL model trained on clean data is often vulnerable to adversarial attacks. The robustness of DL-based solar forecasting can be enhanced through two approaches: preprocessing and improving the model's robustness. Preprocessing techniques involve detecting corrupted data and processing it before feeding it to the DL model. However, these approaches may fail against well-designed attacks. Alternatively, improving the robustness of the model itself makes it resilient against cyberattacks. Therefore, this work adopts an adversarial training approach to enhance robustness against attacks.

Adversarial training has emerged as a leading defense strategy against perturbations, where a model is trained on both

Input: A classifier f with loss function J , a real example x and ground-truth label y . The size of perturbation ε , number of iterations T , decay factor μ ;

Output: An adversarial example x^* with $\|x^* - x\|_\infty \leq \varepsilon$;

- 1 Set $\alpha = \varepsilon/T$;
- 2 Initialize $g_0 = 0$ and $x_0 = x$;
- 3 **for** $t = 0$ to $T - 1$ **do**
- 4 Input x_t to f and obtain the gradient $\nabla_x J(x_t, y)$;
- 5 Update g_{t+1} by accumulating the velocity vector:

$$g_{t+1} = \mu \cdot g_t + \frac{\nabla_x J(x_t, y)}{\|\nabla_x J(x_t, y)\|_1};$$
- 6 Update x_{t+1} by applying the sign gradient:

$$x_{t+1} = x_t + \alpha \cdot \text{sign}(g_{t+1});$$
- 7 **end**
- 8 **return** $x^* = x_T$;

ALGORITHM 3: MI-FGSM.

clean and adversarial examples [10]. By exposing the model to perturbed inputs during training, it learns to generalize better and becomes less sensitive to adversarial attacks [36, 37]. This approach intrinsically strengthens the model's resistance to attacks, making it a highly effective defense method in the solar nowcasting domain [29].

The foundation of adversarial training was discussed in [36], where neural networks were trained on a combination of clean and adversarial examples. During attack generation, an adversarial example \mathbf{X}_{adv} is created to maintain minimal distance from the original input \mathbf{X} while affecting the model's performance. Adversarial training utilizes both corrupted and normal data to train the model, and the predictions can be expressed as in Equations (6) and (7). This dual training approach aims to improve the model's resilience against attacks, although it may slightly reduce accuracy under normal conditions.

$$f_{N_{\text{adv}}}^{\text{aux}} = \text{Train}(\mathbf{X}_{\text{adv}} + \mathbf{X}). \quad (6)$$

$$f = \mathbf{X} \mapsto f_{N_{\text{adv}}}^{\text{aux}}(\mathbf{X}). \quad (7)$$

A crucial consideration in adversarial training is the selection of the attack model. A model trained against a specific attack may not generalize well to different attack scenarios. In this work, FGSM and PGD attacks were selected for adversarial training due to their well-established roles as foundational techniques in adversarial training. FGSM is widely recognized for its simplicity and efficiency in generating adversarial examples, making it a popular choice in initial adversarial training frameworks [38, 39]. Conversely, PGD introduces more refined perturbations, significantly improving model robustness when used in training [40, 41]. Together, these

methods provide a representative and effective subset for evaluating model resilience against adversarial attacks [42, 43].

The effectiveness of FGSM and PGD in adversarial training is well-accepted in the literature, with numerous studies demonstrating their ability to enhance model robustness against adversarial perturbations while maintaining generalization performance on clean data [44, 45]. Additionally, their computational efficiency ensures their practical applicability in real-world scenarios, where training time and resource constraints are critical considerations [45, 46].

Moreover, to validate the generalization capability of the AT FEMUS-Nowcast model, its performance was further evaluated against more advanced attacks, namely, I-FGSM and MI-FGSM. The results as demonstrated in Section 4 indicate that the AT FEMUS-Nowcast model demonstrates strong resilience to these attacks, further supporting the effectiveness of FGSM and PGD-based adversarial training in enhancing robustness against adversarial attacks in sky image-based solar power nowcasting.

2.3. IQA Metrics. For any application of visual media, quality assessment is one of the required criteria to ensure the performance of any technological developments. Among different IQA metrics, the metrics belonging to the objective category are more widely employed compared to subjective metrics [47], as objective metrics offer convenience in application, time efficiency, and cost-effectiveness. As the requirements of computer vision applications differ from the requirements of human vision-based applications, appropriate metric selection is critical for any technological advancement, as indicated in [48]. Hence, this work employs two metrics to account for both physical and perceptual distortions in sample images. From a physical perspective, peak signal-to-noise ratio (PSNR) is used, which is a widely employed mathematical approach for objective metrics due to its low computational complexities,

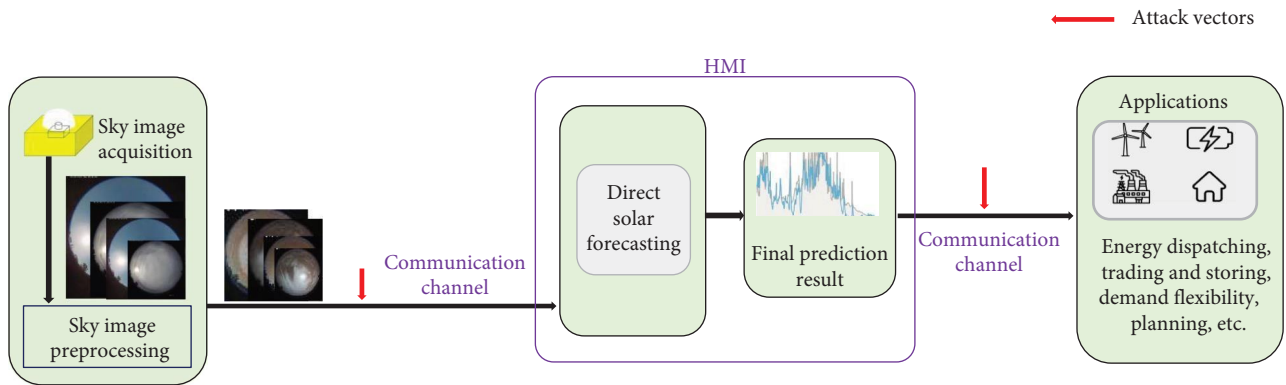


FIGURE 1: Framework for sky image-based PV power forecasting with the potential cyberattacks.

simplified analytical features, and easy interpretation [49]. A value of 30 dB can be considered a benchmark for PSNR to ensure reasonable image quality, as indicated in [50]. On the other hand, the multiscale structural similarity index measure (MS-SSIM) [51] is used as an example of the perceptual approach of objective metrics, which is widely accepted for its superior performance in resembling human perception. This metric accounts for the luminance, contrast, and structural sensitivity of the human visual system (HVS) in the spatial domain while considering the frequency sensitivity of HVS through downsampling and low filtering of sample images. As indicated in [52], a score around 0.95 for MS-SSIM is sufficient to ensure reasonable perceptual image quality while maintaining notable performance for computer vision applications. As these two metrics assess image quality from two different aspects, they are used to control the implementations of cyberattacks for making the distortions in corrupted images barely noticeable.

3. Proposed Framework

Accurate and timely solar irradiance nowcasts are essential for both power system operators and energy market participants [5]. The cost of forecast errors can be significant, ranging from \$0.02/MW h to \$105.28/MW h [6]. As a result, sky image-based solar nowcasting has garnered substantial attention in recent years [53–55], focusing on developing more accurate models. However, the potential impact of cyberattacks on these models' prediction errors has often been overlooked. This issue underscores the need for a comprehensive approach to nowcasting, considering not only accuracy but also cybersecurity.

The framework illustrated in Figure 1 depicts a sky image-based PV power forecasting system. Images are captured and preprocessed, and solar irradiance forecasts are generated. This information is critical for decision-making in applications such as energy dispatching, trading, and storage. Figure 1 highlights two potential vectors for cyberattacks: the communication channel between the camera panel and the human-machine interface (HMI), and the channel between the HMI and the end-user (application unit). While it is theoretically possible to implement cyberattacks on the communication channel between the HMI and the application, this work

specifically focuses on analyzing and mitigating the impact of cyberattacks targeting the communication channel between the camera and the HMI. The objective is to ensure the accuracy and reliability of solar power nowcasting predictions by safeguarding the integrity of the image-based data, which is critical for downstream processes. The risk of cyberattacks on the communication channel between the HMI and the application is a valid concern and is identified as an important area for future research.

3.1. PV Power Generation Nowcast. The task of nowcasting PV power generation can be defined as establishing a function (f_N) that maps sky images to their corresponding real-time PV power output [21], as expressed in Equation (8).

$$f_N : I_i \rightarrow P_i, \text{ where } \{I_i, P_i\} \in D. \quad (8)$$

Here, the benchmark dataset, denoted as D , is divided into two parts: a development set (D_d) and a test set (D_t). Each sample within D consists of paired data, aligning sky images (I) with corresponding PV power generation values (P).

The nowcast model presented in this study was developed using dataset D_d and evaluated on dataset D_t . Tenfold cross-validation was employed during the model development phase to enhance its generalizability. In the testing phase, the final prediction is generated by averaging the predictions of 10 sub-models, an ensemble approach aimed at improving overall accuracy. Figure 2 illustrates the model architecture of FEMUS-Nowcast, proposed by this work for PV power generation nowcasting. The reason behind the success of deep CNNs is the multilayer stacking of convolution processes, which improves the abstraction and understanding of complicated data (such as graphic imagery) [56].

Given a sky image sequence $\mathbf{X} \in \mathbb{R}^{N \times W \times H \times D}$, where N , W , H , and D represent the sample number, image width, image height, and image channel size, respectively, and the model output $\mathbf{y} \in \mathbb{R}^{N \times 1}$, the FEMUS-Nowcast model utilizes multiple units for feature extraction from these sky images. Each unit comprises a series of convolutional layers, batch normalization layers, and max pooling layers. With a step length of one and same-value padding, the convolutional layer uses a 3×3 filter size. Rectified linear unit (ReLU) [57] is the applied activation function. 64 filters are used in the first convolution-pooling

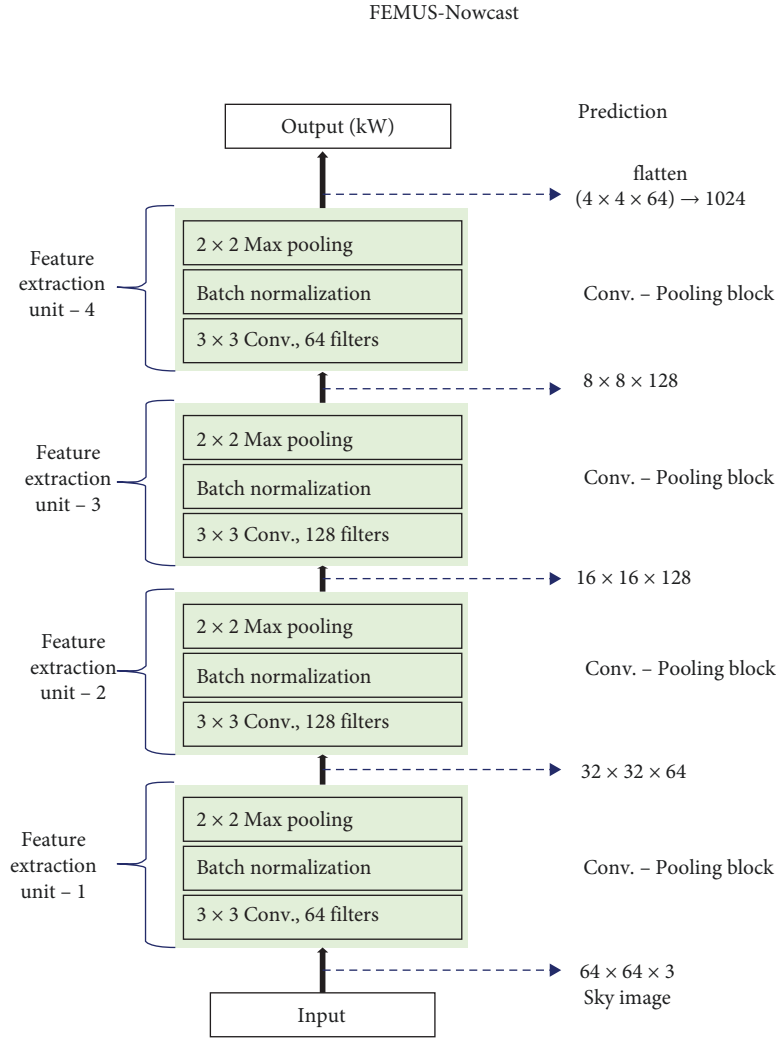


FIGURE 2: The architecture of proposed model for PV power generation nowcast.

block, 128 filters are utilized in the second and third blocks, and 64 filters are used in the final block. The core operation in a convolutional layer is a 2D convolution, which can be represented by Equation (9).

$$\mathbf{y}[i, j] = f \left(\sum_{u=-k/2}^{k/2} \sum_{v=-k/2}^{k/2} \omega[u, v] * \mathbf{x}[i + u, j + v] + b \right). \quad (9)$$

In essence, the convolutional layer produces an output feature map \mathbf{y} by applying a filter ω to the input image \mathbf{X} . The filter slides across the image, performing element-wise multiplications and summations at each position. A bias term b is added to the result, and an activation function f is then applied to introduce nonlinearity. After convolution, batch normalization is applied to speed up and improve the robustness of the model's training [58]. Batch normalization adjusts the distribution of activations to improve training stability, and the processes can be expressed by Equations (10) and (11).

$$\mathbf{X}_{\text{norm}} = \frac{\mathbf{X} - \mu}{\sqrt{\sigma^2 + \psi}}. \quad (10)$$

$$\mathbf{y} = \gamma \times \mathbf{X}_{\text{norm}} + \beta. \quad (11)$$

In addition to these operations, batch normalization includes a small constant, denoted as ψ , for numerical stability. The process also involves learnable scaling γ and shifting β parameters to further adjust the normalized activations. 2×2 max pooling is then applied as expressed in Equation (12), selecting the maximum value from a window within the input to downsample the feature maps. In addition to aggregating spatially neighboring features, this layer is important to generate translation-invariant features [56].

$$\mathbf{y}[i, j] = \max_{u, v \in \text{window}} \mathbf{X}[i \times s + u, j \times s + v], \quad (12)$$

where $\mathbf{y}[i, j]$ represents the output element at index i, j in the matrix or array \mathbf{y} . The operation $\max_{u, v \in \text{window}}$ involves finding the maximum value among elements u and v within a

specified window. The term $\mathbf{X}[i \times s + u, j \times s + v]$ denotes an element within the array or matrix \mathbf{X} , where s is the stride factor, and u and v are offsets.

After the convolutional blocks, a flattened layer transforms the output into a one-dimensional vector, facilitating subsequent processing steps as expressed in Equation (13).

$$\mathbf{y} = f_N(\omega \times \mathbf{X} + b). \quad (13)$$

The Adam optimizer [59], a common stochastic gradient descent optimizer in ML, is used in this study to train the model. Equation (14) expresses the mean square error of the PV output prediction, which is the loss function that has to be minimized.

$$\text{Loss} = \frac{1}{N} \sum_{i \in S} (P_i - Q_i)^2, \quad (14)$$

where $N = |S|$ is the number of samples, P_i is the predicted output, and Q_i is the true output for sample i .

Section 4 presents the performance of the FEMUS-Nowcast model on the test set, evaluated using root mean square error (RMSE) and mean absolute error (MAE) metrics. As Lago et al. [60] highlight, MAE and RMSE are reliable metrics in forecasting tasks, while MAPE is less popular due to giving more importance to data points close to zero. Hence, this work chose RMSE and MAE as measuring metrics. The results demonstrate the model's effectiveness in extracting relevant information from sky images and correlating it with local PV panel generation. Notably, FEMUS-Nowcast accurately approximates sun angle equations in clear sky conditions and provides reasonable estimates of PV power generation under various cloudy conditions, outperforming the adopted established models. The FEMUS-Nowcast model has the potential to serve as an alternative to traditional sensor measurements of solar irradiance or PV panel power output, which are generally expensive [21].

4. Experimental Setup and Analysis

The experiments were conducted using Google Colaboratory on a computer equipped with an Intel i5 processor and 8 GB of RAM. This section provides a detailed description of the model training process, cyberattack implementation, and an analysis of experimental observations. It includes information about data collection and the attack implementation process to support the investigations. The experimental analysis focuses on assessing the impacts of adversarial attacks on PV power generation nowcast models and evaluating the effectiveness of the proposed AT FEMUS-Nowcast model against cyberattacks. As performance evaluation metrics, MS-SSIM and PSNR are used to assess image quality, while RMSE and MAE are used to evaluate forecasting accuracy.

4.1. Model Training and Cyberattack Implementation. The open-source dataset provided by Nie et al. [21] served as the foundation for developing and testing the SUNSET Nowcast [21], ANN [19], and proposed FEMUS-Nowcast models. To evaluate the robustness of these forecasting models, FGSM,

PGD, and a mixed template attack of them were simulated on PV power generation nowcasts. Specifically, processed sky images were subjected to adversarial attacks during transmission over the communication channel. Model performance was then assessed using RMSE and MAE metrics.

Even though hybrid architectures such as CNN-LSTM [61] and transformer-based models [62] have gained attention in recent literature for temporal forecasting tasks, this study focuses on a CNN-based architecture and compares it against other CNN-based models. This decision is grounded in the nature of sky image-based solar irradiance prediction, which primarily involves (1) identifying cloud structures, (2) estimating their movement, and (3) assessing their potential to obstruct direct sunlight [55]. These tasks rely on spatial visual features such as cloud pixels, edges, corners, and cloud types, all of which can be automatically extracted by CNN models without the need for manual feature engineering [63, 64]. Moreover, CNNs offer architectural simplicity compared to LSTM or recurrent neural network (RNN)-based networks [63], and in certain scenarios, they have even demonstrated better performance than models incorporating explicit temporal dependencies [65]. Additionally, transformer-based models, despite their promise, generally involve greater training complexity [66]. Since this study also investigates the cyber vulnerability of forecasting models, an emerging and underexplored challenge in ASI-based solar nowcasting, the CNN-based approach of this study provides a robust and interpretable baseline.

The test dataset D_t consists of sky images captured at 1-min intervals. This study simulated a 5-min continuous attack for each of the considered templates and their combined template, and overall a total contamination of 30% of the entire test dataset. The impact of varying attack strengths (represented by the parameter (ϵ) with values 0.01, 0.015, and 0.02) was analyzed and visualized in Figures 3–8. These results reveal a clear trend, as ϵ increases, the severity of the cyberattack intensifies. However, an upper limit for ϵ was established based on the MS-SSIM and PSNR values presented in Tables 2–4. This upper limit was chosen to ensure that the adversarial attacks did not significantly degrade the image quality, as image quality is maintained above the benchmark values mentioned in Section 2.3 for these two metrics.

This work finds that the proposed FEMUS-Nowcast model outperforms the existing SUNSET Nowcast and ANN models in terms of RMSE and MAE, but is still susceptible to cyberattacks. This is also elaborated in the subsequent section. Hence, adversarial training procedures were adopted to enhance the model's robustness against cyberattacks. However, to mitigate the potential loss of accuracy when an AT model encounters normal (unperturbed) data, FEMUS-Nowcast was trained on a mixture of FGSM and PGD attacks, with 20% of the training data intentionally contaminated and the remainder left as clean data. Algorithm 4 illustrates the entire process, including cyberattack implementation, training of both normally trained (NT) and AT models, and the subsequent evaluation of their performance.

4.2. Performance Analysis. This subsection evaluates the performance of each model under single attack templates (FGSM

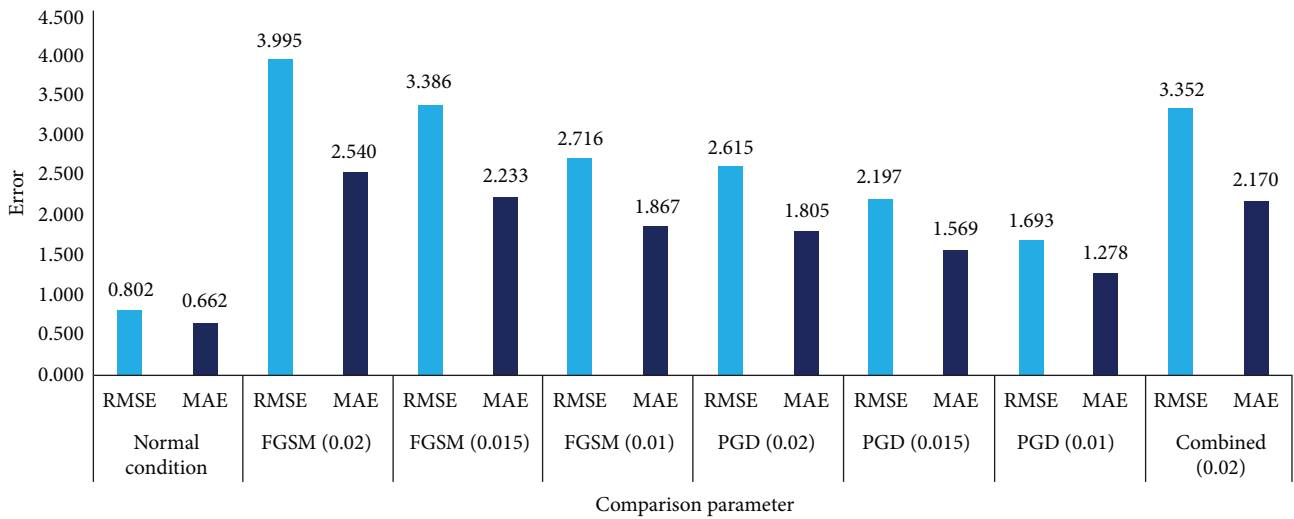


FIGURE 3: Comparison of SUNSET Nowcast under different attack scenarios for sunny days.

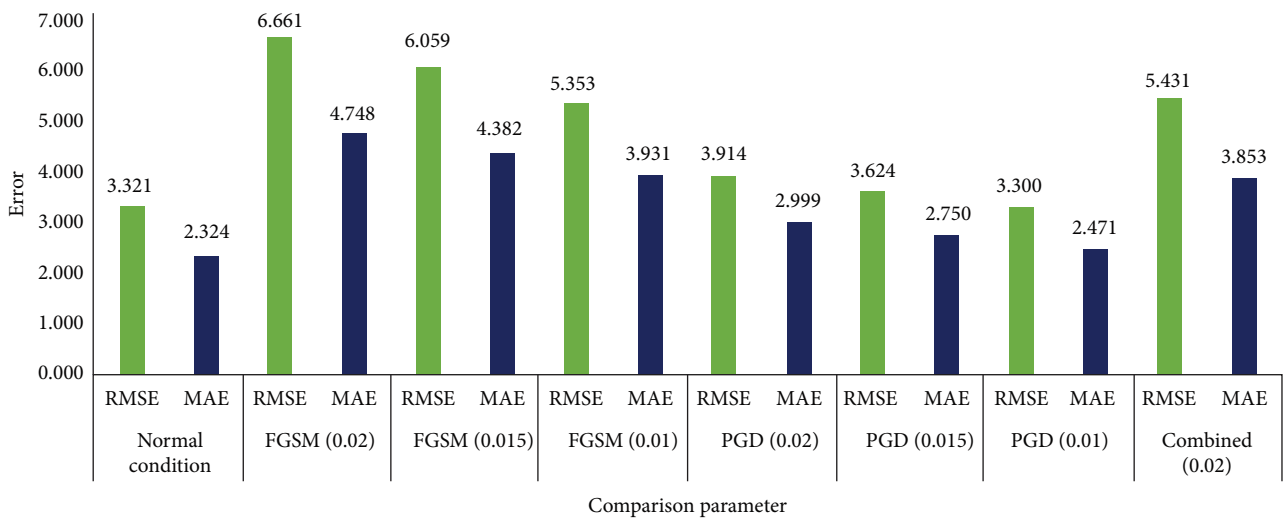


FIGURE 4: Comparison of SUNSET Nowcast under different attack scenarios for cloudy days.

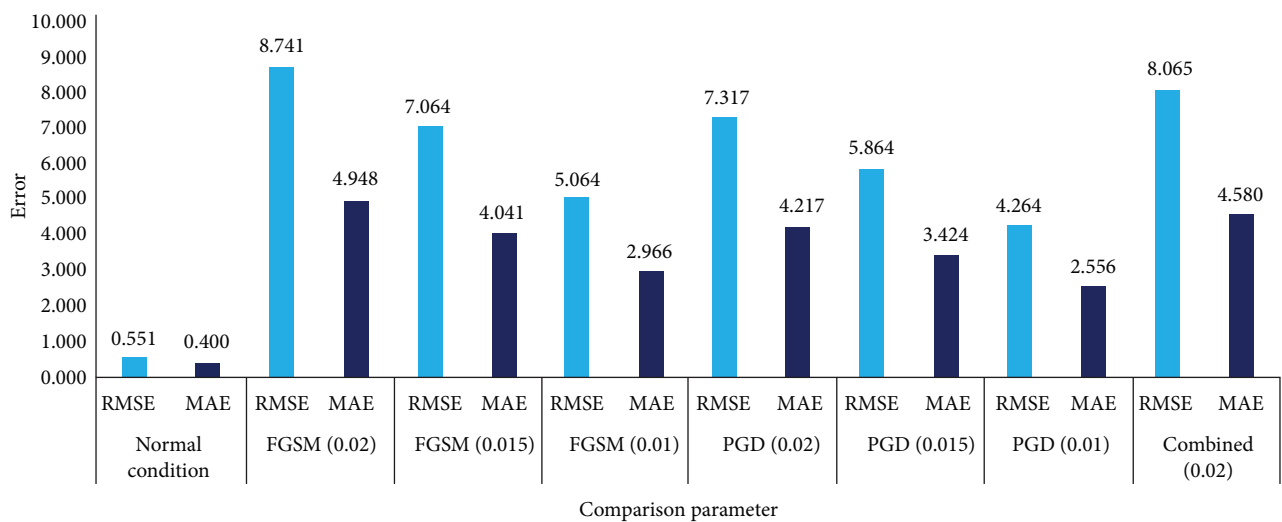


FIGURE 5: Comparison of ANN prediction under different attack scenarios for sunny days.

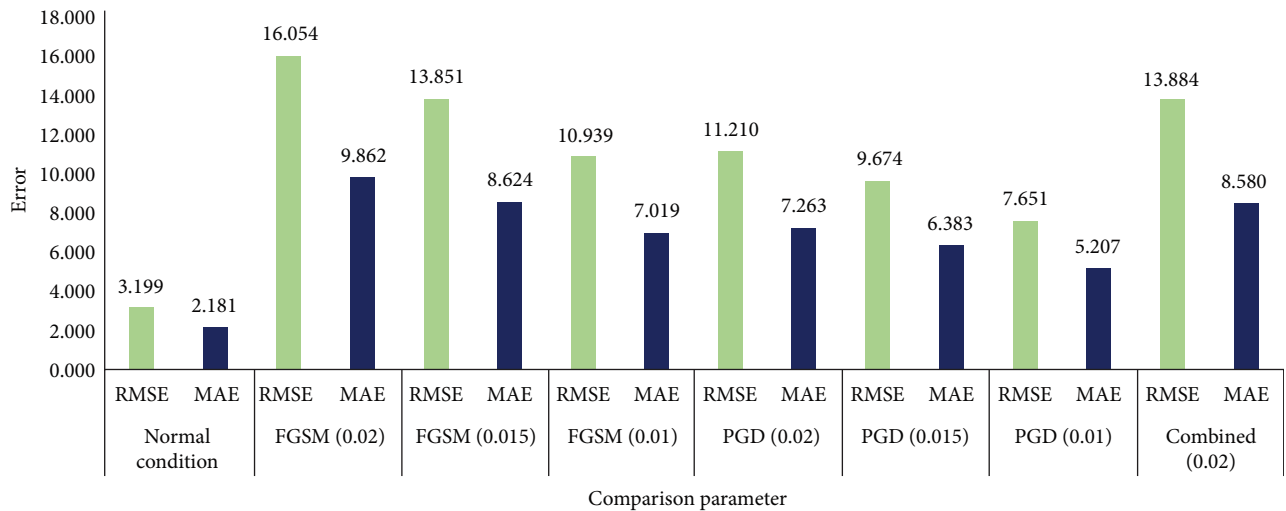


FIGURE 6: Comparison of ANN prediction under different attack scenarios for cloudy days.

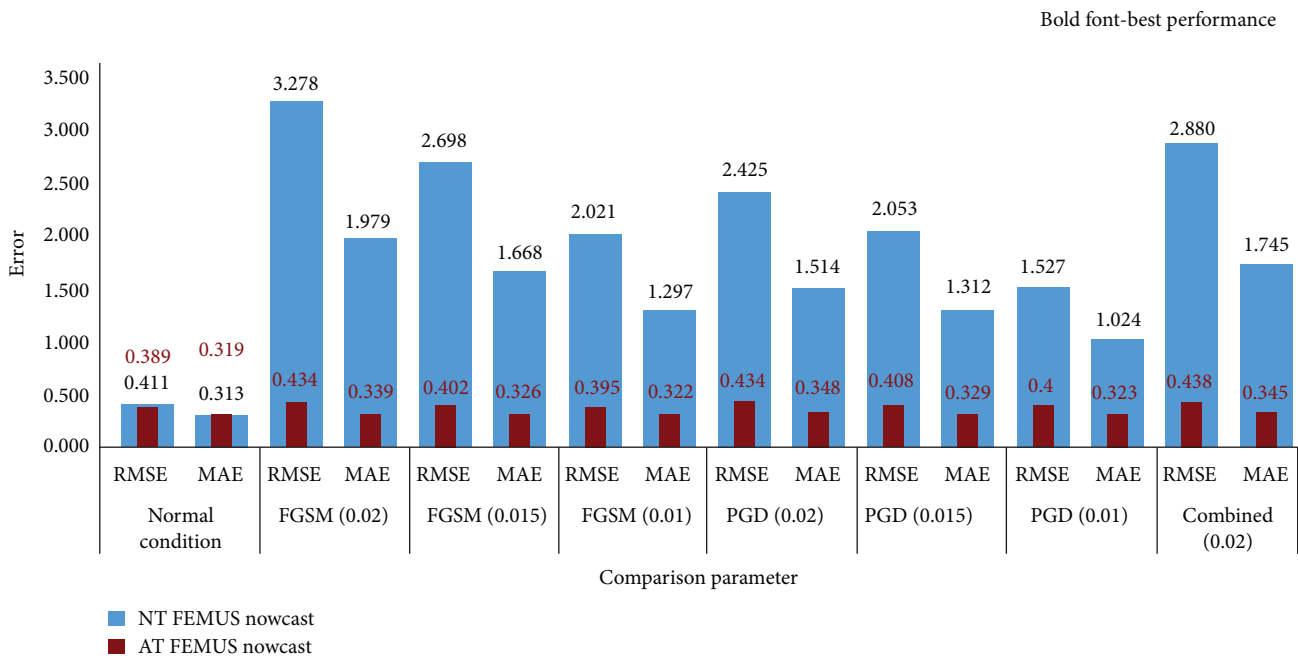


FIGURE 7: Comparison of FEMUS-Nowcast under different attack scenarios for sunny days.

and PGD) as well as their combined template to assess the severity of these attacks. The analysis considers multiple values of the perturbation parameter (ϵ). The performance of each benchmark model and the proposed model are analyzed individually at first, while their differences for the considered cases are summarized at the end of this section.

4.2.1. SUNSET Nowcast. First, the SUNSET Nowcast [21], a highly cited model, is taken to analyze under diverse scenarios. The performance of the SUNSET Nowcast model [21] under various adversarial attacks is quantitatively analyzed using RMSE and MAE, as shown in Figures 3 and 4. These figures summarize the impact of FGSM, PGD, and a combined attack template across different perturbation levels ($\epsilon = 0.01, 0.015,$ and 0.02) for both sunny and cloudy days.

Under cyberattack conditions, FGSM attacks exert the most pronounced impact on sunny days (Figure 3). RMSE becomes almost five times and MAE is almost four times higher than normal conditions at maximum perturbation. PGD attack leads to three times compared to normal conditions in both error metrics. While the combined attack template is less severe than FGSM, but more severe than PGD alone, it still causes much greater errors than those observed under normal operation. Importantly, error levels decline as the perturbation parameter ϵ decreases.

On cloudy days (Figure 4), FGSM and the combined attack template approximately double both RMSE and MAE, while PGD attack results in a small error increase. Predictions by SUNSET Nowcast are naturally prone to fluctuations in PV power generation (Figure A2), which is amplified by

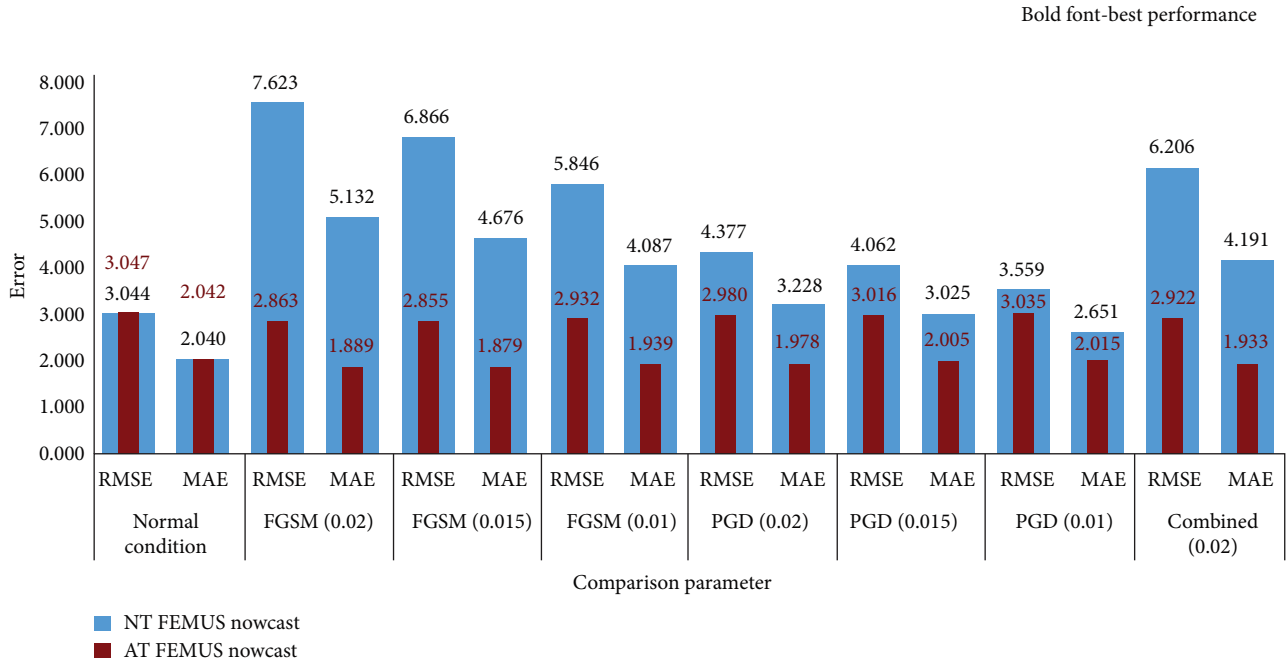


FIGURE 8: Comparison of FEMUS-Nowcast prediction under different attack scenarios for cloudy days.

TABLE 2: MS-SSIM and PSNR value for different cyberattack scenarios.

Attack (ϵ)	Sunny days		Cloudy days	
	MS-SSIM	PSNR	MS-SSIM	PSNR
FGSM (0.02)	0.94	34.415	0.94	34.433
FGSM (0.015)	0.962	36.911	0.963	36.929
FGSM (0.01)	0.983	40.429	0.983	40.446
PGD (0.02)	0.955	35.987	0.953	35.958
PGD (0.015)	0.973	38.446	0.973	38.452
PGD (0.01)	0.988	41.941	0.988	41.889
Combined (0.02)	0.946	35.199	0.945	35.201

TABLE 3: MS-SSIM and PSNR value for different cyberattack scenarios for ANN.

Attack (ϵ)	Sunny days		Cloudy days	
	MS-SSIM	PSNR	MS-SSIM	PSNR
FGSM (0.02)	0.941	34.616	0.934	34.373
FGSM (0.015)	0.964	37.109	0.961	36.867
FGSM (0.01)	0.985	40.623	0.982	40.383
PGD (0.02)	0.944	34.859	0.947	34.954
PGD (0.015)	0.967	37.197	0.968	37.357
PGD (0.01)	0.986	40.633	0.985	40.811
Combined (0.02)	0.942	34.736	0.939	34.662

cyberattacks. Paradoxically, at the lowest PGD perturbation level, this can lead to slightly reduced RMSE as fluctuations move predictions closer to ground truth. However, increased MAE across all attack scenarios confirms their overall significant negative impact. These results highlight the SUNSET

TABLE 4: MS-SSIM and PSNR value for different cyberattack scenarios for FEMUS-Nowcast.

Attack (ϵ)	Sunny days		Cloudy days	
	MS-SSIM	PSNR	MS-SSIM	PSNR
FGSM (0.02)	0.940	34.408	0.939	34.418
FGSM (0.015)	0.963	36.904	0.964	36.915
FGSM (0.01)	0.983	40.421	0.983	40.432
PGD (0.02)	0.953	35.735	0.953	35.717
PGD (0.015)	0.973	38.242	0.973	38.209
PGD (0.01)	0.987	41.642	0.988	41.573
Combined (0.02)	0.947	35.058	0.944	35.059

Nowcast model's vulnerability to adversarial perturbations, particularly under FGSM attacks, on both sunny and cloudy days. For visual inspection of prediction trends on each test day under different attacks, detailed figures have been presented in Appendix A1.

In terms of image quality, FGSM degrades the image quality more than PGD attacks, which is indicated by the lower MS-SSIM and PSNR values for FGSM attacks compared to PGD attacks, as illustrated in Table 2. Also, for the same value of ϵ , the impacts of the combined attack on image quality are found to be less severe than FGSM attacks and more severe than PGD attacks, as presented in Table 2. Hence, it is evident from Figures 3 and 4 and Table 2 that FGSM is the most severe attack, followed by the combined template, while PGD has the least impact among the three.

4.2.2. ANN. This work analyzes the performance of an ANN model as a second benchmark model. The model's response to normal conditions, FGSM, PGD, and combined attacks is

Input: D_d, D_t , train, test, FGSM, PGD
Output: $f_N, f_{ANN}, f_{FN}, f_{FN}^{adv}$, RMSE, MAE

- 1 $f_N, f_{ANN}, f_{FN} = \text{Train}(D_d)$
- 2 $\varepsilon = \{0.01, 0.015, 0.02\}$
- 3 Set step size $\alpha = 5$ minutes.
- 4 **for** $i \in \varepsilon$ **do**
- 5 $D_{t_{FGSM}}^i = \text{FGSM}(\frac{3}{10}D_t, i, \alpha)$
- 6 $\mathbf{D}_t^{\text{FGSM}} = \{D_{t_{FGSM}}^i | i \in \varepsilon\}$
- 7 **for** $i \in \varepsilon$ **do**
- 8 $D_{t_{PGD}}^i = \text{PGD}(\frac{3}{10}D_t, i, \alpha)$
- 9 $\mathbf{D}_t^{\text{PGD}} = \{D_{t_{PGD}}^i | i \in \varepsilon\}$
- 10 $D_t^{\text{combined}} = \text{FGSM}(\frac{3}{20}D_t, 0.02, \alpha) + \text{PGD}(\frac{3}{20}D_t, 0.02, \alpha)$
- 11 **for** $i \in \varepsilon$ **do**
- 12 $D_t^{\text{adv}, i} = \text{FGSM}(\frac{3.3}{100}D_t, i, \alpha) + \text{PGD}(\frac{3.3}{100}D_t, i, \alpha)$
- 13 $\mathbf{D}_t^{\text{adv}} = \{D_t^{\text{adv}, i} | i \in \varepsilon\}$
- 14 $f_{FN}^{\text{adv}} = \text{Train}(\mathbf{D}_t^{\text{adv}})$
- 15 RMSE, MAE = $\{f_N, f_{ANN}, f_{FN}, f_{FN}^{\text{adv}}\} \rightarrow (D_t, \mathbf{D}_t^{\text{FGSM}}, \mathbf{D}_t^{\text{PGD}}, D_t^{\text{combined}})$

ALGORITHM 4: Model development and evaluation.

illustrated using performance metrics in Figures 5 and 6, which show RMSE and MAE for sunny and cloudy days, respectively.

The ANN model's performance varies notably between sunny and cloudy conditions. Under normal sunny conditions, it achieves a reasonable level of accuracy. However, cloudy days introduce substantial errors, suggesting the model struggles with complex weather patterns or shifts in lighting. On the other hand, under sunny conditions (Figure 5), FGSM attacks drastically impact the model. RMSE becomes almost 16 times and MAE becomes more than 12 times higher at maximum perturbation compared to normal operation. PGD attacks, while less severe than FGSM, still lead to a more than thirteen-fold increase in RMSE and over tenfold in MAE. The combined attack template also results in significant error inflation. Importantly, errors decrease proportionally as the perturbation level (ε) is reduced, but every attack scenario is enough to lead the model to an unusable state.

Even on cloudy days (Figure 6), where the model's baseline accuracy is weaker, cyberattacks substantially degrade its performance. FGSM and the combined attack approximately quintuple both RMSE and MAE. PGD attacks also inflict damage, with errors raised more than three times at maximum perturbation.

These results indicate that the ANN model is particularly susceptible to FGSM attacks. While PGD is slightly less severe, it still causes critical degradation. For a day-wise visualization of the ANN model's predictions under various attack scenarios, Figures A3 and A4 have been presented in Appendix A2. At the same time, FGSM significantly degrades image quality compared to PGD, and the combined attack template also proves highly detrimental. Figures 5 and 6 and Table 3 provide representations of each of these, where the image quality metrics presented in Table 3 also support these findings. FGSM attacks consistently result in lower MS-SSIM and PSNR values

compared to PGD, indicating more noticeable visual distortion in the sky images. The combined attack yields intermediate quality degradation, aligning with the observed forecast performance.

4.2.3. FEMUS-Nowcast. This work proposes a nowcasting model named FEMUS-Nowcast, which leverages the advantages of feature extraction from sky images by multiple units. The architecture of the FEMUS-Nowcast model is depicted in Figure 2. While the model generally tracks the ground truth closely under normal conditions, adversarial attacks still cause notable performance degradation. And the impact of varying perturbation levels (ε) is demonstrated in Figures 7 and 8.

Under sunny conditions, FGSM attacks inflict the most severe damage on the NT FEMUS-Nowcast model. At maximum perturbation, RMSE increases by almost 3 and MAE by a little more than 1.6. PGD and combined attacks also lead to substantial error increases, with RMSE rising by over 2 and MAE by 1.2–1.5. Cloudy days exacerbate the problem. Baseline errors are already higher for the NT FEMUS-Nowcast model and attacks amplify them further. FGSM and combined attacks with maximum possible intensity lead to a dramatic increase in both RMSE (over 4) and MAE (over 3). In contrast, PGD appears less severe than them. Importantly, the impact of error decreases as the perturbation level decreases.

It is evident from Figures 7 and 8 and Table 4 that the NT FEMUS-Nowcast model appears most vulnerable to FGSM attacks, while PGD attacks, although less severe, are still capable of rendering the NT FEMUS-Nowcast model unusable. FGSM degrades the quality of images more significantly than PGD attacks, as indicated by reduced MS-SSIM and PSNR values in Table 4. For example, at the highest intensity in sunny conditions, the MS-SSIM for FGSM is 0.940, whereas it is 0.953 for PGD. The mixed attack is more severe than PGD alone, as

TABLE 5: Performance of AT FEMUS-Nowcast model against I-FGSM and MI-FGSM attacks.

Attack (ϵ)	Metrics	Sunny days	Cloudy days
I-FGSM (0.02)	RMSE	0.543	3.248
	MAE	0.429	2.256
MI-FGSM (0.02)	RMSE	0.472	2.974
	MAE	0.375	1.998

presented in Table 4. And for a day-wise visualization of the model’s predictions under different attacks, Figures A5 and A6 are presented in Appendix A3.

While the NT FEMUS-Nowcast model outperforms benchmark models in normal conditions, however, it remains vulnerable to cyberattacks. To mitigate this, this work employs adversarial training. A mixed attack template, combining FGSM and PGD, was used during training (Algorithm 4). The AT model demonstrates remarkable adherence to ground truth under normal conditions and significantly reduced deviations under attacks. Robustness across varying perturbation levels (ϵ) is demonstrated in Figures A7 and A8 and a day-wise visualization is presented in Appendix A4.

Figures 7 and 8 clearly show that adversarial training dramatically enhances model performance, enabling near-normal operation even during cyberattacks. Interestingly, on sunny days, the AT FEMUS-Nowcast outperforms the NT model in terms of RMSE, incurring only a minor trade-off of 0.06 increase in MAE under normal conditions. The maximum trade-off for the AT model under any attack scenario is approximately 0.05 for RMSE and only 0.03 for MAE. On cloudy days, the trade-offs of AT FEMUS-Nowcast are even smaller (0.03 for RMSE and 0.02 for MAE under normal conditions). Remarkably, the AT FEMUS-Nowcast shows no RMSE or MAE trade-offs under any attack type during cloudy days while still reducing errors compared to the NT version. Given the inherent challenges of PV power nowcasting on cloudy days, this enhanced robustness is particularly significant; adversarial training counterintuitively moves predictions closer to ground truth in the presence of attacks.

As the AT FEMUS-Nowcast model was trained using a combination of FGSM and PGD attacks, it was also evaluated under more advanced cyberattack templates, specifically I-FGSM and MI-FGSM attacks, as shown in Table 5. The results demonstrate that the AT model exhibits significant resilience against these two additional attack methods. For I-FGSM with perturbation $\epsilon = 0.02$, the RMSE trade-offs are only 0.154 and 0.201 on sunny and cloudy days, respectively. Similarly, the MAE trade-offs for sunny and cloudy conditions are 0.11 and 0.214, respectively. For the MI-FGSM attack under the same perturbation level, the RMSE trade-off is only 0.083 on sunny days, and the RMSE improves by 0.073 on cloudy days. On the other hand, the MAE trade-off is only 0.056, and the MAE improves by 0.044 for sunny and cloudy conditions, respectively. These results highlight the effectiveness of the AT FEMUS-Nowcast model in defending against a wide range of attack methods, even beyond FGSM and PGD.

4.2.4. *Summary.* Performance analysis of this work reveals a clear hierarchy among the evaluated nowcasting models. The

SUNSET Nowcast model serves as the initial benchmark and exhibits moderate accuracy, but remains highly susceptible to cyberattacks. FGSM attacks prove particularly devastating, significantly magnifying both RMSE and MAE across all weather conditions. For instance, under the FGSM attack with the maximum perturbation considered in this work, its RMSE increases by approximately 5 times and MAE by 4 times in sunny conditions. The ANN model, while demonstrating some potential under normal sunny days, however, it fares even worse, with errors increasing by up to 16 times under the maximum perturbation case of FGSM attacks, rendering it unusable in practical scenarios. Although PGD attacks are conventionally considered more severe than FGSM due to their iterative nature, this study consistently finds FGSM to be more detrimental across all models and test conditions. This deviation from expectation is primarily attributed to the relatively small values of the perturbation parameter (ϵ) used in our experiments, which were carefully selected to preserve image quality based on MS-SSIM and PSNR constraints. Notably, this observation aligns with prior research [67], which experimentally demonstrated that under small perturbations, FGSM can produce greater model degradation than PGD. However, as ϵ increases, the severity of PGD surpasses that of FGSM, reaffirming the conventional understanding in high-perturbation regimes.

The proposed FEMUS-Nowcast model, in both its NT and AT forms, represents a significant advancement. The NT FEMUS-Nowcast model demonstrates superior resilience to cyberattacks compared to the existing SUNSET Nowcast and ANN models. This robustness can be attributed to the model’s advanced feature extraction capabilities, facilitated by multiple feature extraction units. These units enable FEMUS-Nowcast to extract more comprehensive and relevant features from sky images, enabling the model to follow the ground truth more closely while there is no cyberattack and less deviation in the presence of cyberattacks compared to the existing models.

However, under the maximum FGSM attack case of the NT FEMUS-Nowcast model, the RMSE increased by approximately three times. Then, adversarial training dramatically improves its resilience. The AT FEMUS-Nowcast maintains accuracy comparable to, or exceeding, its performance on uncorrupted data, even under attack. Notably, it exhibits minimal trade-offs, with differences in RMSE and MAE remaining within 0.06 under all attack scenarios. This robustness, particularly significant during cloudy conditions and resiliency against I-FGSM and MI-FGSM attacks, highlights the effectiveness of adversarial training in ensuring reliable nowcasting

across a wide range of scenarios. These findings underscore the FEMUS-Nowcast model's reliability and robustness, making it a practical and resilient choice for accurate solar power nowcasting, even in adversarial attack conditions.

Also, it is important to note that the forecasting performance of sky image-based models also depends on the quality and clarity of the input images. Real-world challenges such as raindrops, condensation, soiling, or sensor obstructions (e.g., from dirt or unfavorable angles) can significantly degrade image quality and impact model reliability. Prior studies have suggested both hardware solutions, such as weather-protected domes with ventilated heaters and cleaning mechanisms [7], as well as algorithmic techniques like rain streak removal through L_0 -gradient minimization [68], to mitigate such effects. Additionally, optimal camera placement strategies have been proposed to reduce environmental interference [53]. While these aspects are kept beyond the scope of the current work, as this study primarily investigates adversarial attacks during data transmission, they represent important directions for improving real-world robustness in future deployments.

5. Conclusions

Solar energy nowcasting, a critical component for integrating solar power into the energy grid, increasingly relies on computer vision and DL techniques to analyze sky images for accurate forecasts. However, the vulnerability of these image-based nowcasting models to cyberattacks poses a significant risk. This study bridges the gap between solar irradiance nowcasting and cybersecurity by introducing adversarial attacks on DL and ANN-based models. This study exposes the vulnerabilities of image-based solar nowcasting models (SUNSET Nowcast, ANN, and FEMUS-Nowcast) to adversarial attacks, such as FGSM, PGD, and a mixed attack template, rendering them unreliable in real-world scenarios. Among the evaluated models, the proposed FEMUS-Nowcast model outperformed existing established models (SUNSET Nowcast and ANN) in terms of RMSE and MAE under normal conditions. Specifically, FEMUS-Nowcast reduced MAE by 53% compared to SUNSET Nowcast, and by 22% compared to the ANN model during sunny days. Additionally, during cloudy days, FEMUS-Nowcast achieved a 12% reduction in MAE over SUNSET Nowcast and a 6% reduction over the ANN model. Hence, to enhance its robustness against cyberattacks, adversarial training was applied exclusively to the FEMUS-Nowcast model. The AT FEMUS-Nowcast model demonstrates remarkable resilience, maintaining accuracy comparable to or exceeding its performance on uncorrupted data, even under a variety of attacks, such as FGSM, PGD, and mixed attacks as well as more advanced attack templates such as I-FGSM and MI-FGSM. The superiority of the AT FEMUS-Nowcast model comes from its ability to learn from corrupted data during the adversarial training process, allowing it to mitigate the impact of adversarial attacks, and ensuring reliable nowcasts across a wide range of scenarios and threat levels. In the domain of critical solar energy nowcasting, this yields a more dependable and trustworthy nowcasting solution. Future research will focus on improving nowcasting accuracy by

integrating additional feature extractors and addressing a wider range of cyberattacks with corresponding defense strategies. It will also explore the robustness of hybrid models like CNN-LSTM and transformer-based architectures under adversarial conditions. Additionally, to support real-time deployment on resource-constrained edge devices, future work may involve optimizing model complexity through techniques such as compression, knowledge distillation, or lightweight architecture design.

Nomenclature

ANN:	Artificial neural network
ASI:	All-sky imager
AT:	Adversarial training
CNN:	Convolutional neural network
DNN:	Deep neural network
DL:	Deep learning
FEMUS:	Feature extraction-based multi-unit solar
FGSM:	Fast gradient sign method
HMI:	Human-machine interface
I-FGSM:	Iterative-fast gradient sign method
IQA:	Image quality assessment
LSTM:	Long short-term memory
MAE:	Mean absolute error
MI-FGSM:	Momentum iterative-fast gradient sign method
ML:	Machine learning
MS-SSIM:	Multiscale structural similarity index measure
NT:	Normally trained
NWP:	Numerical weather prediction
PGD:	Projected gradient descent
PSNR:	Peak signal-to-noise ratio
PV:	Photovoltaic
ReLU:	Rectified linear unit
RNN:	Recurrent neural network
RMSE:	Root mean square error.

Appendix A. Extended Prediction Results of All Nowcast Models Under Adversarial Attacks

Appendix A1. Detailed Visualization of SUNSET Nowcast Predictions Under Adversarial Attacks. Figure A1 presents the SUNSET Nowcast model's predicted PV output across each sunny test day under FGSM, PGD, and combined adversarial attack scenarios. The impact of each attack on prediction trends is visually evident, with FGSM showing the greatest deviation from the ground truth. And Figure A2 shows the model's predictions for each cloudy test day under the same attack settings.

Appendix A2. Detailed Visualization of ANN Predictions Under Adversarial Attacks. Figures A3 and A4 illustrate the ANN model's response to normal conditions, FGSM, PGD, and combined attacks at maximum perturbation levels ($\epsilon = 0.02$) across each sunny and cloudy day of the test set, respectively.

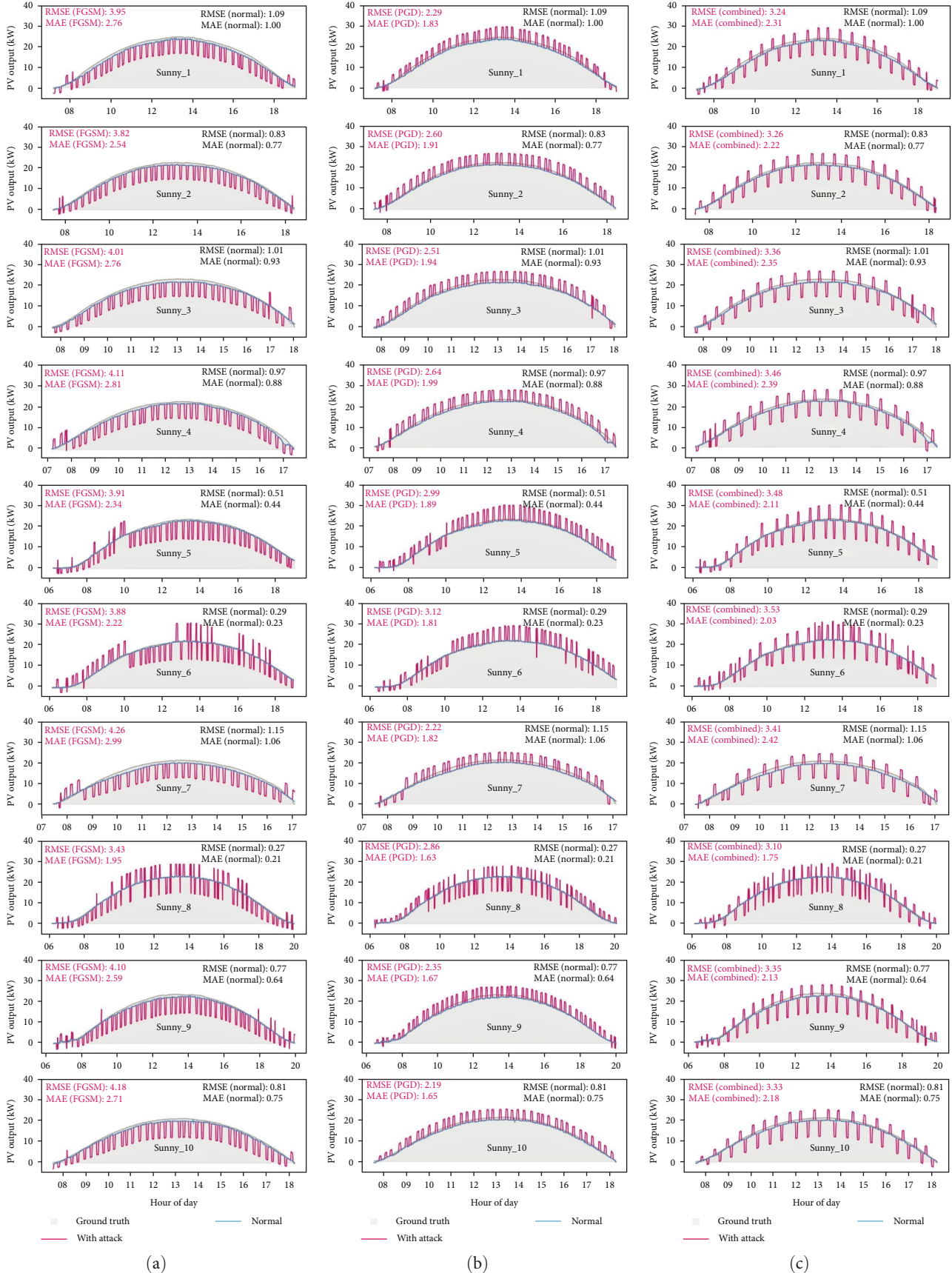


FIGURE A1: SUNSET Nowcast predictions on each of the sunny test days under various attack scenarios: FGSM (a), PGD (b), and combined (c).

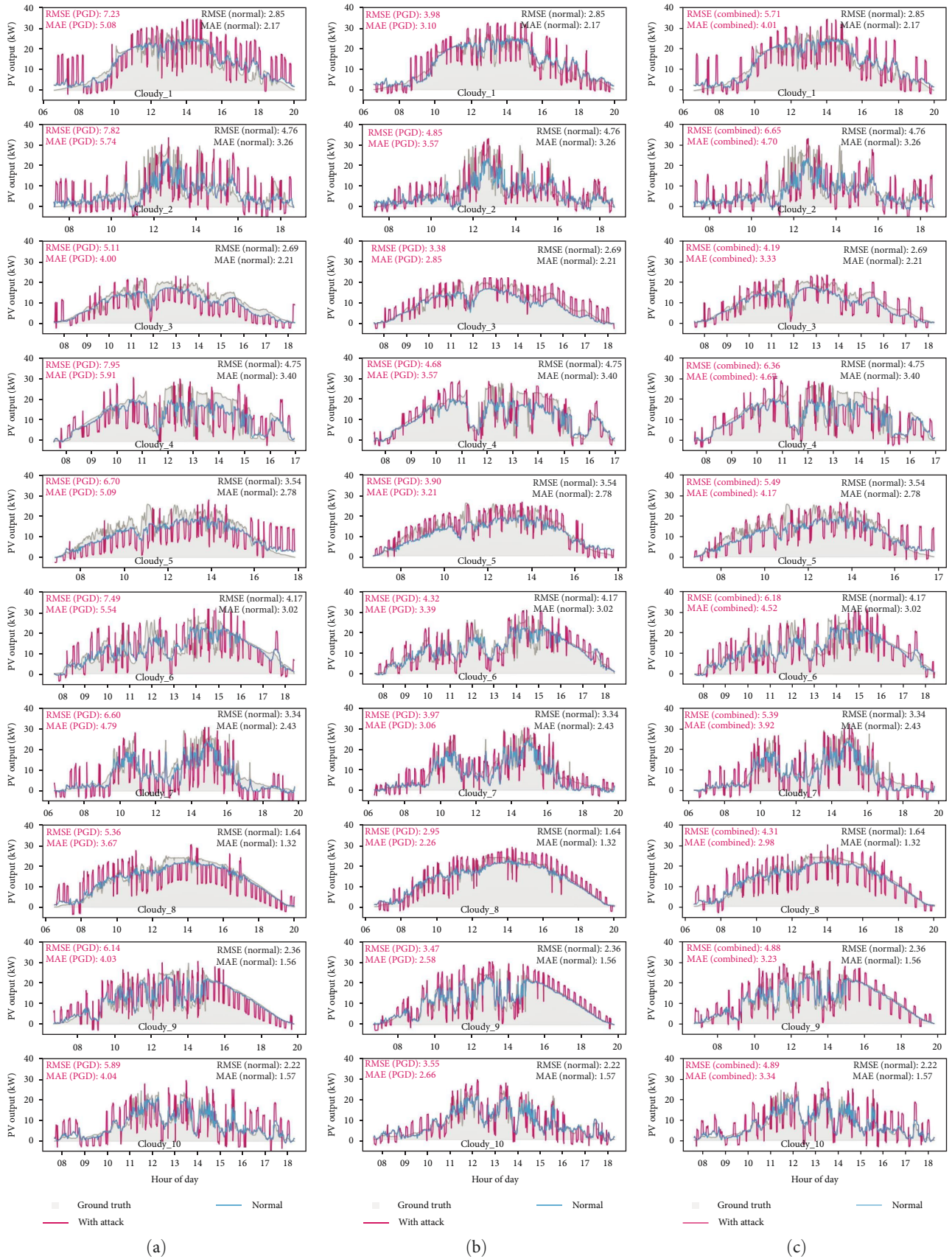


FIGURE A2: SUNSET Nowcast predictions on each of the cloudy test days under various attack scenarios: FGSM (a), PGD (b), and combined (c).

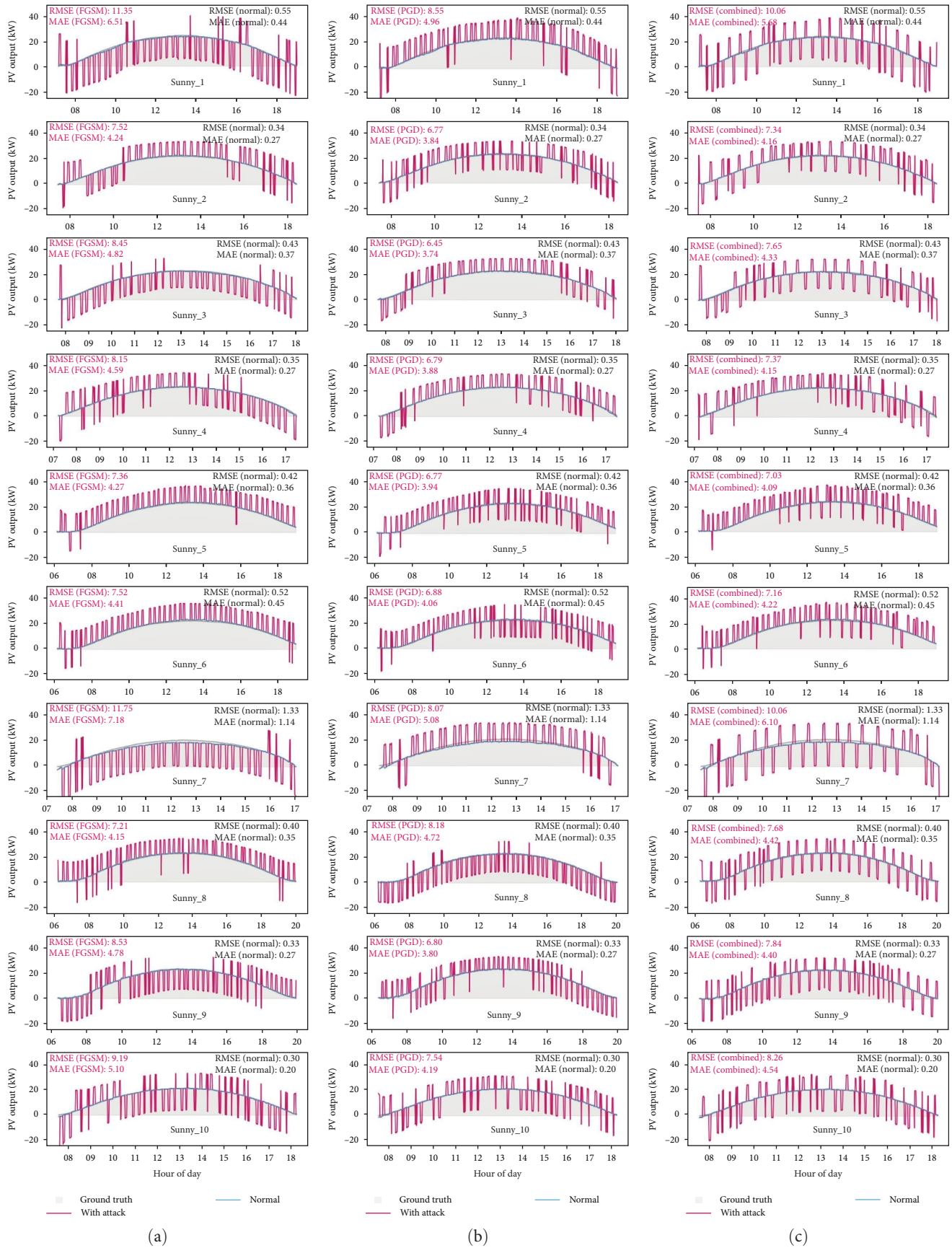


FIGURE A3: ANN predictions on each of the sunny test days under various attack scenarios: FGSM (a), PGD (b), and combined (c).

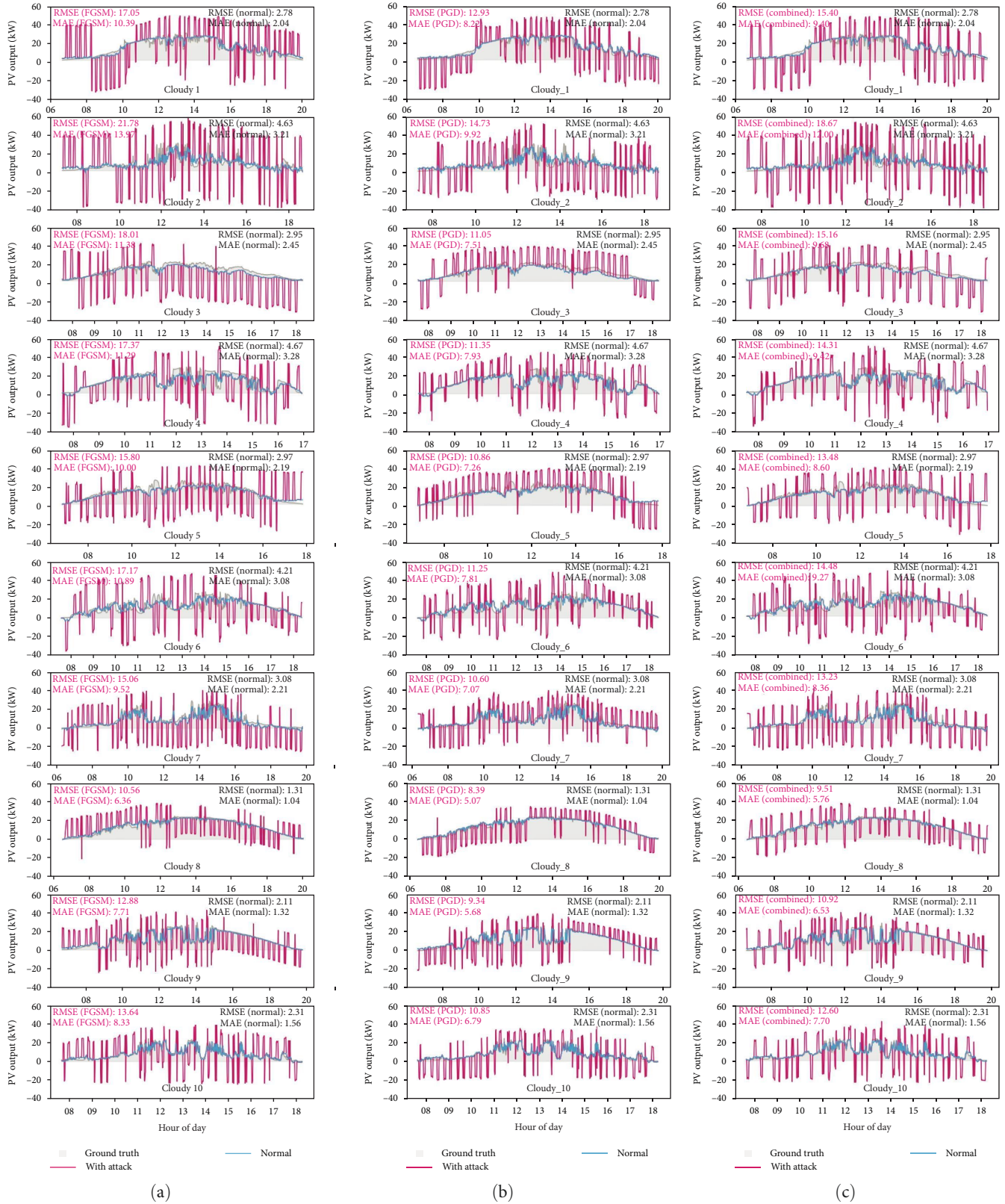


FIGURE A4: ANN predictions on each of the cloudy test days under various attack scenarios: FGSM (a), PGD (b), and combined (c).

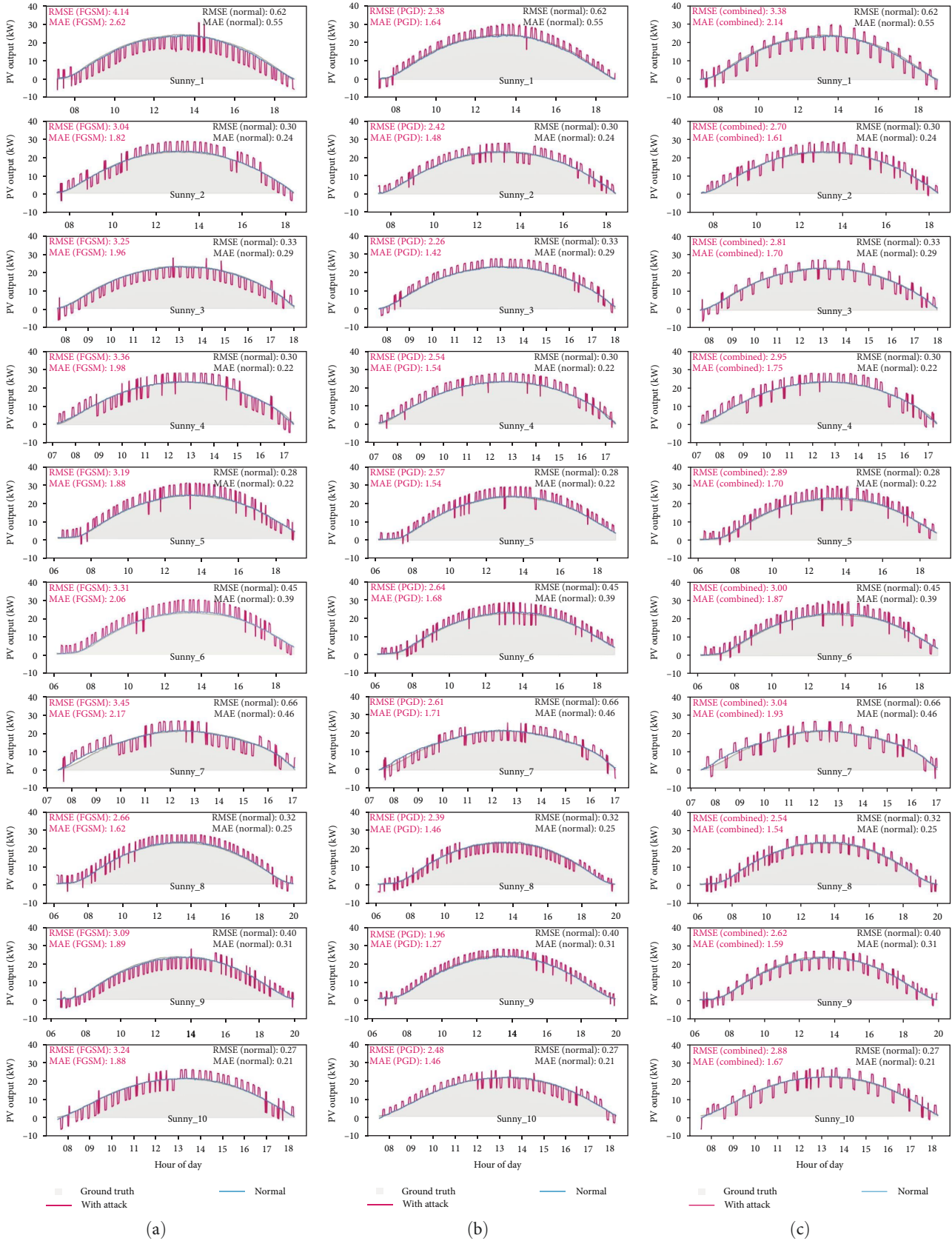


FIGURE A5: NT FEMUS-Nowcast predictions on each of the sunny test days under various attack scenarios: FGSM (a), PGD (b), and combined (c).

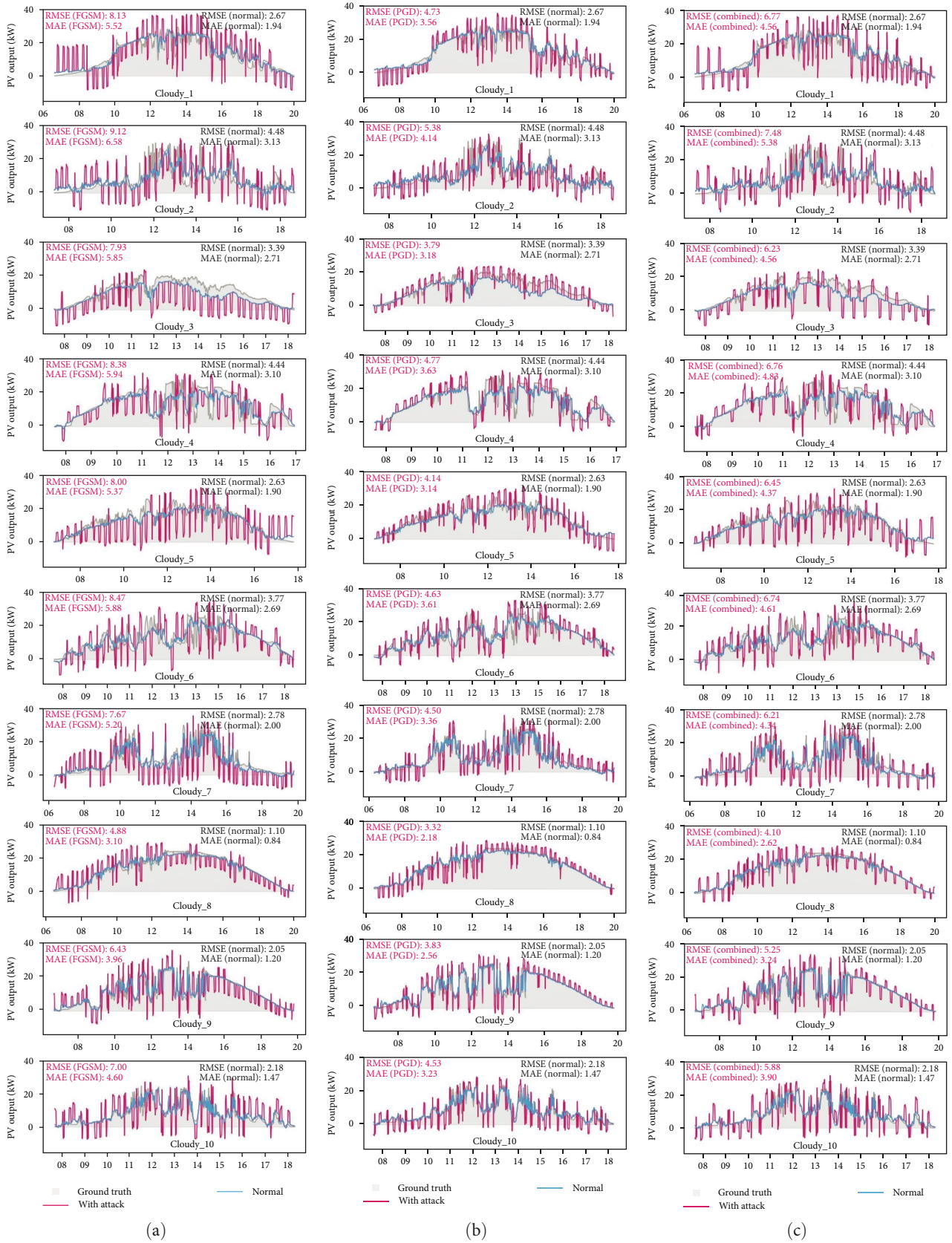


FIGURE A6: NT FEMUS-Nowcast on each of the cloudy test days under various attack scenarios: FGSM (a), PGD (b), and combined (c).

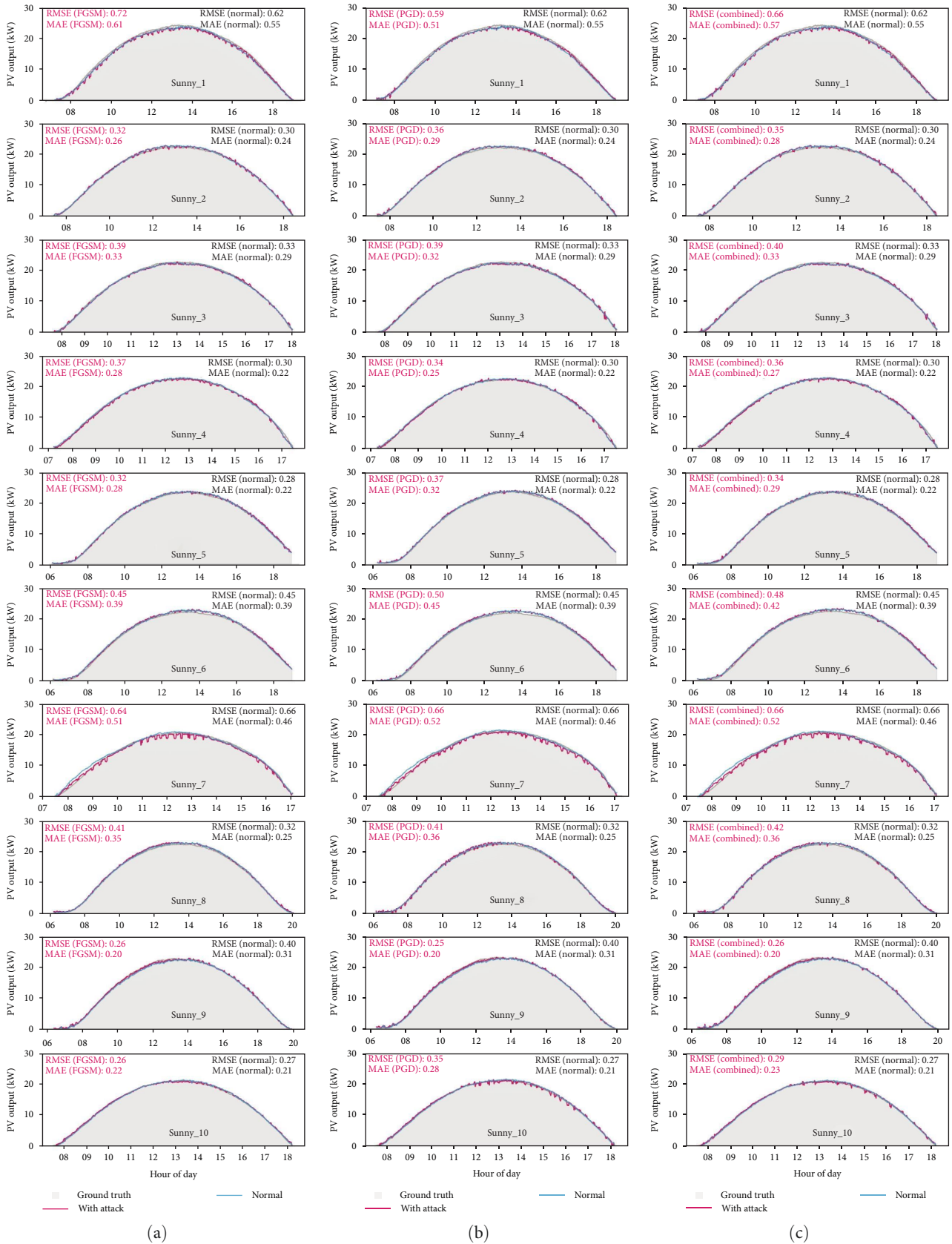


FIGURE A7: AT FEMUS-Nowcast predictions on each of the sunny test days under various attack scenarios: FGSM (a), PGD (b), and combined (c).

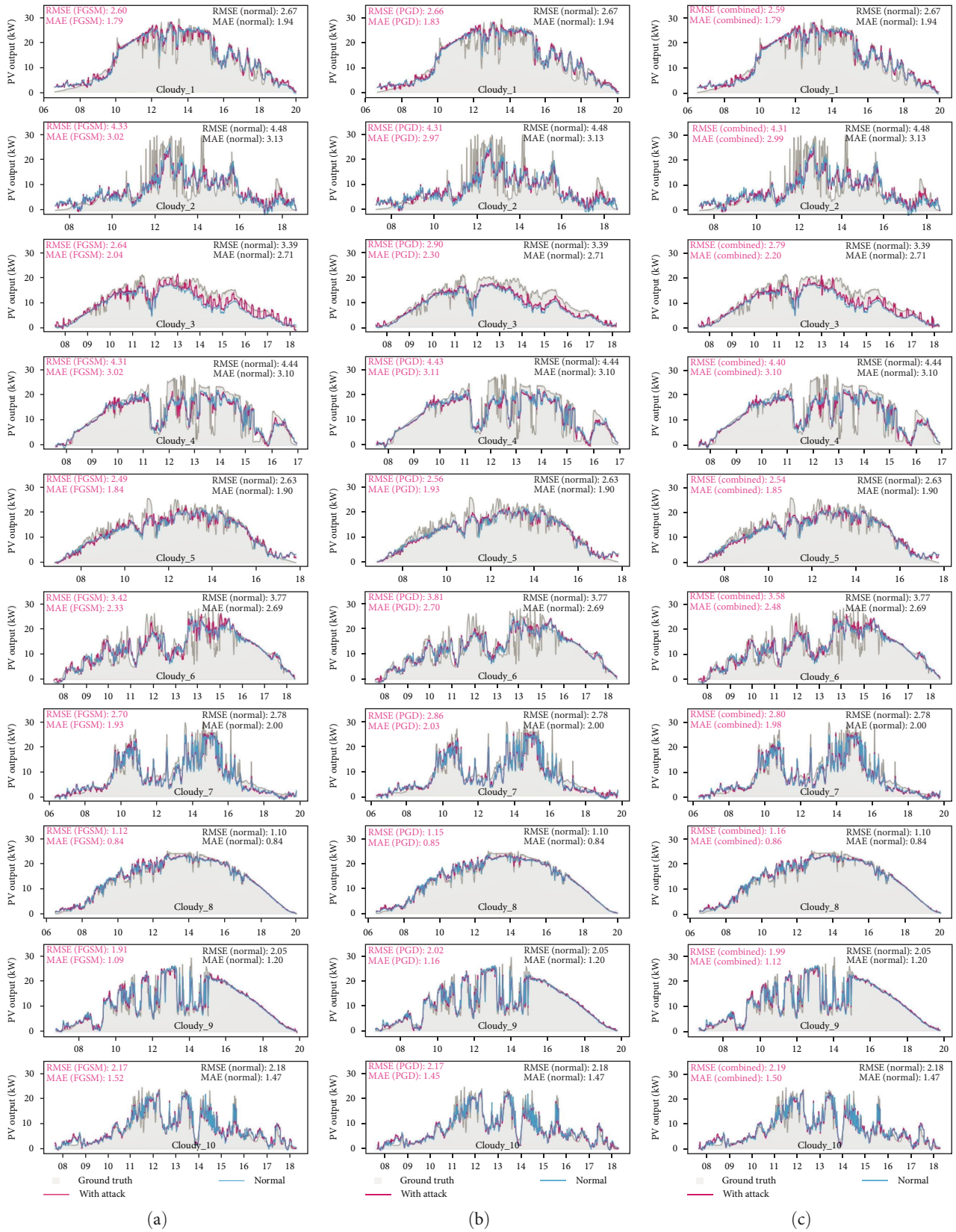


FIGURE A8: AT FEMUS-Nowcast on each of the cloudy test days under various attack scenarios: FGSM (a), PGD (b), and combined (c).

- Methods,” *International Journal of Forecasting* 39, no. 1 (2023): 244–265.
- [23] R. Belu, “Artificial Intelligence Techniques for Solar Energy and Photovoltaic Applications,” in *Handbook of Research on Solar Energy Systems and Technologies*, (IGI Global, 2013): 376–436.
- [24] S.-M. Hategan, N. Stefu, D. Petreus, E. Szilagyi, T. Patarau, and M. Paulescu, “Short-Term Forecasting of PV Power Based on Aggregated Machine Learning and Sky Imagery Approaches,” *Energy* 316 (2025): 134595.
- [25] Q. Zhang, B. Hou, W. Zhu, et al., “Distributed Photovoltaic Ultra-Short-Term Power Prediction Using Whole-Sky Images and Multi-Source Data,” *Electrical Engineering* (2025): 1–19.
- [26] R. Rastgoo, N. Amjadi, S. Lin, and S. M. Muyeen, “Ultra-Short-Term Solar Power Prediction Using Sky Image Sequences by a Residual Vision Reformer,” *IEEE Transactions on Sustainable Energy* (2025): 1–18.
- [27] X. Hou, I. Fountoulakis, P. Blanc, C. Aebi, and S. Kazadzis, “Intrahour Solar Radiation Forecasting Based on Sun Visibility for Different Cloud Types,” *Solar Energy* 294 (2025): 113477.
- [28] X. Wei, D. Yue, G. P. Hancke, C. Dou, H. Li, and Y. Qiu, “Ultra Short-Term Solar Irradiance Forecast Based on Multimodal Data Fusion and Fuzzification,” *IEEE Transactions on Industrial Informatics* 21, no. 4 (2025): 3256–3265.
- [29] N. Tang, S. Mao, and R. M. Nelms, “Adversarial Attacks to Solar Power Forecast,” in *2021 IEEE Global Communications Conference (GLOBECOM)*, (Madrid, Spain: IEEE, 2021), 1–6.
- [30] M. Kuzlu, S. Sarp, F. O. Catak, et al., “Analysis of Deceptive Data Attacks With Adversarial Machine Learning for Solar Photovoltaic Power Generation Forecasting,” *Electrical Engineering* 106, no. 2 (2024): 1815–1823.
- [31] M. Kuzlu, B. E. Tamayo, S. Sarp, F. O. Catak, U. Cali, and Y. Zhao, “Security Concerns of Adversarial Attack for LSTM/BiLSTM Based Solar Power Forecasting,” in *2023 IEEE Power & Energy Society General Meeting (PESGM)*, (Orlando, FL, USA: IEEE, 2023), 1–5.
- [32] A. Kurakin, I. J. Goodfellow, and S. Bengio, “Adversarial Examples in the Physical World,” in *Artificial Intelligence Safety and Security*, (Chapman and Hall/CRC, 2018): 99–112.
- [33] Y. Dong, F. Liao, T. Pang, et al., “Boosting Adversarial Attacks With Momentum,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (Salt Lake City, UT, USA: IEEE, 2018).
- [34] I. J. Goodfellow, J. Shlens, and C. Szegedy, “Explaining and Harnessing Adversarial Examples,” arXiv: 1412.6572 (2015).
- [35] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu, “Towards Deep Learning Models Resistant to Adversarial Attacks,” arXiv: 1706.06083 (2019).
- [36] C. Szegedy, W. Zaremba, I. Sutskever, et al., “Intriguing Properties of Neural Networks,” arXiv preprint arXiv: 1312.6199 (2013).
- [37] T. Bai, J. Luo, J. Zhao, B. Wen, and Q. Wang, “Recent Advances in Adversarial Training for Adversarial Robustness,” (2021).
- [38] B. Li, “Towards Understanding Fast Adversarial Training,” (2020).
- [39] A. Kurakin, “Adversarial Machine Learning at Scale,” (2016).
- [40] H. Kim, W. Lee, and J. Lee, “Understanding Catastrophic Overfitting in Single-Step Adversarial Training,” (2020).
- [41] N. Kaur, “Robustness and Security in Deep Learning: Adversarial Attacks and Countermeasures,” *Journal of Electrical Systems* 20, no. 3s (2024): 1250–1257.
- [42] C. Hsu, P. Chen, S. Liu, and C. Yu, “Adversarial Examples can be Effective Data Augmentation for Unsupervised Machine Learning,” (2021).
- [43] S. Li, “Learning More Robust Features With Adversarial Training,” (2018).
- [44] C. Xie, Y. Wu, L. Maaten, A. Yuille, and K. He, “Feature Denoising for Improving Adversarial Robustness,” (2019).
- [45] N. Papernot, P. McDaniel, X. Wu, S. Jha, and A. Swami, “Distillation as a Defense to Adversarial Perturbations Against Deep Neural Networks,” in *2016 IEEE Symposium on Security and Privacy (SP)*, (San Jose, CA, USA: IEEE, 2016), 582–597.
- [46] H. Wu, Z. Hu, and B. Gu, “Fast and Scalable Adversarial Training of Kernel SVM via Doubly Stochastic Gradients,” *Proceedings of the AAAI Conference on Artificial Intelligence* 35, no. 12 (2021): 10329–10337.
- [47] Z. Wang, “Applications of Objective Image Quality Assessment Methods [Applications Corner],” *IEEE Signal Processing Magazine* 28, no. 6 (2011): 137–142.
- [48] M. A. Rahman and H. Bhuiyan, “Is the Development of Objective Image Quality Assessment Methods Keeping Pace With Technological Developments?” *International Journal of Electrical and Computer Engineering Systems* 15, no. 6 (2024): 491–497.
- [49] U. Sara, M. Akter, and M. S. Uddin, “Image Quality Assessment Through FSIM, SSIM, MSE and PSNR—A Comparative Study,” *Journal of Computer and Communications* 7, no. 3 (2019): 8–18.
- [50] A. Rahimi, A. Ghofrani, K.-T. Cheng, L. Benini, and R. K. Gupta, “Approximate Associative Memristive Memory for Energy-Efficient GPUs,” in *2015 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, (Grenoble, France: IEEE, 2015), 1497–1502.
- [51] Z. Wang, E. P. Simoncelli, and A. C. Bovik, “Multiscale Structural Similarity for Image Quality Assessment,” in *The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers*, 2003, 2, (Pacific Grove, CA, USA: IEEE, 2003), 1398–1402.
- [52] M. A. Rahman, A. Huq, S. S. Tunny, M. H. Anik, M. T. Pervin, and M. R. Islam, “A Feasibility Analysis of Image Approximation With Image Quality Assessments,” *IET Image Processing* 18, no. 4 (2024): 897–913.
- [53] R. Samu, M. Calais, G. Shafullah, et al., “Applications for Solar Irradiance Nowcasting in the Control of Microgrids: A Review,” *Renewable and Sustainable Energy Reviews* 147 (2021): 111187.
- [54] B. Nouri, S. Wilbert, N. Blum, et al., “Probabilistic Solar Nowcasting Based on All-Sky Imagers,” *Solar Energy* 253 (2023): 285–307.
- [55] V. A. Martinez Lopez, G. van Urk, P. J. F. Doodkorte, M. Zeman, O. Isabella, and H. Ziar, “Using Sky-Classification to Improve the Short-Term Prediction of Irradiance With Sky Images and Convolutional Neural Networks,” *Solar Energy* 269 (2024): 112320.
- [56] Y. Sun, G. Szűcs, and A. R. Brandt, “Solar PV Output Prediction From Video Streams Using Convolutional Neural Networks,” *Energy & Environmental Science* 11, no. 7 (2018): 1811–1818.
- [57] V. Nair and G. E. Hinton, “Rectified Linear Units Improve Restricted Boltzmann Machines,” in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, (Omnipress, 2010): 807–814.
- [58] S. Ioffe and C. Szegedy, “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,”

- in *International Conference on Machine Learning, Pmlr*, (JMLR.org, 2015): 448–456.
- [59] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” arXiv preprint arXiv: 1412.6980 (2014).
- [60] J. Lago, G. Marcjasz, B. De Schutter, and R. Weron, “Forecasting Day-Ahead Electricity Prices: A Review of State-of-the-Art Algorithms, Best Practices and an Open-Access Benchmark,” *Applied Energy* 293 (2021): 116983.
- [61] M. Ansong, G. Huang, T. N. Nyang’onda, R. J. Musembi, and B. S. Richards, “Very Short-Term Solar Irradiance Forecasting Based on Open-Source Low-Cost Sky Imager and Hybrid Deep-Learning Techniques,” *Solar Energy* 294 (2025): 113516.
- [62] J. Liu, H. Zang, L. Cheng, T. Ding, Z. Wei, and G. Sun, “A Transformer-Based Multimodal-Learning Framework Using Sky Images for Ultra-Short-Term Solar Irradiance Forecasting,” *Applied Energy* 342 (2023): 121160.
- [63] C.-L. Fu and H.-Y. Cheng, “Cheng, Predicting Solar Irradiance With All-Sky Image Features via Regression,” *Solar Energy* 97 (2013): 537–550.
- [64] I. Vasilev, D. Slater, G. Spacagna, P. Roelants, and V. Zocca, *Python Deep Learning: Exploring Deep Learning Techniques and Neural Network Architectures With Pytorch, Keras, and TensorFlow* (Packt Publishing Ltd., 2019).
- [65] W. Kong, Y. Jia, Z. Y. Dong, K. Meng, and S. Chai, “Hybrid Approaches Based on Deep Whole-Sky-Image Learning to Photovoltaic Generation Forecasting,” *Applied Energy* 280 (2020): 115875.
- [66] Q. Paletta, G. Terrén-Serrano, Y. Nie, et al., “Advances in Solar Forecasting: Computer Vision With Deep Learning,” *Advances in Applied Energy* 11 (2023): 100150.
- [67] U. Mahapatra, M. A. Rahman, M. R. Islam, M. A. Hossain, M. R. I. Sheikh, and M. J. Hossain, “Adversarial Training-Based Robust Model for Transmission Line’s Insulator Defect Classification Against Cyber-Attacks,” *Electric Power Systems Research* 245 (2025): 111585.
- [68] B. N. Manu, “Rain Removal From Still Images Using L0 Gradient Minimization Technique,” in *2015 7th International Conference on Information Technology and Electrical Engineering (ICITEE)*, (Chiang Mai, Thailand: IEEE, 2015), 263–268.