Computers and Electrical Engineering 138 (2026) 111251

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

Physics-Informed Generative Adversarial Networks for reliable long-term harmonic forecasting in offshore wind farms

Alp Karadeniz

Electrical and Electronics Engineering, Faculty of Engineering, Balikesir University, Balıkesir, Turkey

ARTICLE INFO

Keywords:
Physics-informed GAN
Harmonic forecasting
Offshore wind farms
Power quality
Data augmentation
Machine and deep learning

ABSTRACT

This study proposes a Physics-Informed Generative Adversarial Network (PI-GAN) framework for reliable long-term harmonic distortion forecasting in offshore wind farms. Unlike conventional GANs, the proposed model embeds domain-specific constraints into generator training by enforcing current–voltage proportionality and wind speed-dependent harmonic behavior. The framework is evaluated in a long-term setting by training on data from 2005–2009 and testing on November 2019, thereby introducing a 10-year temporal gap. To examine the effect of physics-informed synthetic data generation across different predictors, both machine learning and deep learning backbones are considered, including Random Forest, XGBoost, GRU, TCN, DLinear, MLP-Mixer, and DRFormer models. Results show that PI-GAN preserves competitive statistical accuracy while reducing physics-violation errors by approximately 45%–50% relative to standard GAN-based baselines across all evaluated backbones. Notably, even state-of-the-art architectures such as DLinear, MLP-Mixer, and DRFormer produce physically inconsistent forecasts when paired with a standard GAN, but achieve a 46%–47% reduction in physics violations when the same architectures are used with PI-GAN, demonstrating that physical plausibility is determined by the data generation process rather than by the forecasting model itself. These findings indicate that the principal benefit of PI-GAN is not merely lower conventional prediction error, but the generation of forecasts that are more physically plausible and interpretable. Therefore, PI-GAN provides a more reliable data-generation and forecasting framework for renewable energy systems.

1. Introduction

Maintaining power quality has become increasingly challenging as offshore wind energy systems expand, particularly due to harmonic distortions introduced by power electronic converters. Accurate forecasting of Total Harmonic Distortion (THD) is therefore critical for grid stability and compliance with international standards. Owing to their ability to model nonlinear dependencies in complex power systems, machine learning (ML) and deep learning (DL) methods have been widely adopted for harmonic forecasting [1,2]. In particular, Generative Adversarial Networks (GANs) have emerged as an effective data-augmentation tool for mitigating data scarcity and improving model performance [3]. Recent studies (2020–2025) report that GAN-based augmentation substantially improves forecasting accuracy in renewable energy applications [4]. In our previous work, we developed a GAN-based augmentation framework for offshore wind farm harmonic prediction and showed improved accuracy compared with conventional ML and DL models [5].

Despite these gains, existing GAN-based and purely data-driven methods generally lack physical interpretability and may produce outputs that violate fundamental power-system relationships, even when statistical metrics improve [6].

E-mail address: akaradeniz@balikesir.edu.tr.

https://doi.org/10.1016/j.compeleceng.2026.111251
Received 20 March 2026; Received in revised form 29 April 2026; Accepted 12 May 2026

Available online 10 June 2026

Moreover, building upon the latest machine learning trends identified in recent state-of-the-art surveys [7,8], the integration of physics-informed constraints has emerged as a specialized necessity for renewable energy forecasting. While recent hybrid frameworks for harmonic mitigation [9] and spatial–temporal correlation analyses [10] have improved predictive accuracy in microgrids, a significant gap remains in ensuring the physical consistency of long-term synthetic data generation. Our proposed PI-GAN framework directly addresses this by enforcing $V \propto I$ proportionality and wind-speed dependencies, effectively bridging the high-performance data-driven advancements of 2024–2025 with fundamental electrical laws.

Therefore, this limitation becomes more critical in long-term forecasting, where models must generalize across large temporal gaps. To address this issue, we propose a Physics-Informed GAN (PI-GAN) framework that embeds domain-specific constraints into the learning process. Unlike short-term forecasting approaches, the proposed method performs long-term harmonic forecasting by training on 2005–2009 data and testing on November 2019. Results show that PI-GAN significantly improves physical consistency while preserving comparable statistical accuracy, reinforcing the need to integrate physical knowledge into data-driven models for reliable and interpretable power-system forecasting.

In addition, the scope of this study is not limited to a single downstream predictor. To more clearly reveal the role of physics-informed synthetic data generation, the proposed PI-GAN framework is examined with both machine learning and deep learning backbones, including RF, XGBoost, GRU, and TCN. In this study, three additional state-of-the-art deep learning architectures, namely DLinear, MLP-Mixer, and DRFormer, are incorporated as downstream forecasting backbones to broaden the cross-architecture evaluation of the proposed PI-GAN framework. This design allows the study to distinguish between two different questions: whether one predictive architecture is better than another, and whether embedding physical knowledge into GAN-based synthetic data generation systematically improves the realism and reliability of the resulting forecasts. In this work, the second question is the main focus. Accordingly, the central objective is to demonstrate that the contribution of PI-GAN lies primarily in enforcing physically meaningful synthetic data generation, rather than in claiming a universally superior forecasting backbone.

In contrast to the conventional approach of merely referring to the contributions of other studies in the introduction section, this study includes a special section called 'Literature Overview' in Section 2. This section is dedicated to the critical analysis of the latest developments in physics-informed neural networks and harmonic analysis up to the year 2025.

2. Literature overview

Recently, researchers has focused increasingly on using data-heavy methods to estimate renewable production and predict harmonics.

The study in [4] devised a layered strategy that separated signals, especially for identifying variations in wind intensity, to address the high oscillations observed in renewable energy output. Building on this general data-driven context, the deep learning-based approach described in [5] shows how neural networks can manage intricate patterns in power systems during forecasting. Nevertheless, it remains focused on short-term forecasts without connecting findings to actual physics. Similarly, the work in [11] uses machine learning to identify power quality problems more accurately, but it does not incorporate broader physical principles into model behavior.

Data scarcity has also been studied in recent literature with respect to generative models. The use of GAN-based data augmentation for renewable energy forecasting is discussed in [12], where synthetic data is used to improve robustness and accuracy. The ability of GANs to learn data distributions is also highlighted in [13], where a probabilistic forecasting framework using deep generative models is proposed. However, GAN training is known to be sensitive and unstable in many practical settings [14,15].

Through the integration of physics into learning mechanisms, more recent research attempts to address this gap. The use of Physics-Informed Neural Networks (PINNs) for power-system analysis, as proposed in [16], demonstrates the reliability of embedding physical laws directly into training. A related advancement appears in [17], where a GAN-based framework is informed by physical principles to improve stability. In addition, [18] extends physics-informed learning toward renewable energy forecasting applications.

Recent developments have also pushed physics-aware models into other power-system domains. Study [19] reviews the application of PINNs in power networks for monitoring, error identification, and dynamic modeling. The work in [20] proposes PINNs for accelerating state estimation in grids, while [21] introduces a GAN-based augmentation strategy for abrupt weather scenarios in wind power forecasting.

There has also been a rise in the use of generative adversarial networks in conjunction with time-based prediction tasks. Study [22] presents a Wasserstein GAN-based method for integrated energy forecasting, while [23] combines deep neural networks with conventional methods to predict electrical harmonics in offshore wind parks. Meanwhile, [24] integrates physics-informed networks into grid-dynamics simulation workflows to improve long-interval simulation precision.

New approaches have also recently improved the handling of power-quality problems such as harmonic distortion. A hybrid deep learning framework for harmonic mitigation is presented in [9], while [10] proposes a spatial–temporal correlation strategy for THD prediction on high-voltage lines. In addition, [25] shows how transformer-based models can be adapted to track subtle long-term patterns in renewable energy systems. Related comparisons of lightweight architectures are reported in [26], which assesses how DLinear stacks against MLP-Mixer in electrical-engineering time-series tasks.

Also, in line with the latest machine learning trends identified in recent surveys [7,8], the incorporation of physics-informed constraints has developed as a specific requirement for renewable energy forecasting. Although recent studies such as [9,10] improve predictive accuracy, the proposed PI-GAN framework specifically targets the remaining physical-consistency gap in long-term offshore wind forecasts.

More recently, lightweight deep learning architectures have attracted attention for long-horizon time-series forecasting. Zeng et al. [27] demonstrated that a simple decomposition-based linear model (DLinear) can match or outperform transformer-based models in long-horizon settings, challenging the assumption that architectural complexity is always beneficial. Tolstikhin et al. [28] introduced MLP-Mixer, which applies alternating temporal and feature mixing through fully connected layers and has since been adapted to time-series tasks. Attention-based approaches such as the Informer [29] extended self-attention mechanisms to long-sequence settings and have been applied in energy forecasting contexts. Despite these advances, the behavior of such SOTA architectures under physics-informed data augmentation in offshore harmonic forecasting remains unexplored, motivating their inclusion in the present study.

In conclusion, despite these developments, there has not been much success using physics-aware generative algorithms to forecast harmonic distortions over extended periods of time; our approach is based on this unresolved issue. Instead of relying on scarce real-world data, we present a GAN-boosted method that produces realistic training instances. This synthetic extension improves the performance of stable models and serves as the basis for the PI-GAN structure examined in this study.

2.1. Motivation

There are still a number of important gaps in the literature despite the substantial progress made in data-driven and physics-informed techniques. The following succinctly describes the study's motivation:

- **Physical inconsistency in GAN-based models:** Existing GAN-based augmentation methods can improve predictive accuracy, but they often generate physically implausible outputs because essential power-system relationships are not explicitly enforced.
- **Insufficient focus on long-term forecasting:** Long-horizon harmonic forecasting across large temporal intervals remains underexplored, as most studies focus on short- or near-term prediction.
- **Limited physics integration in generative models:** Despite substantial progress, integration of physics-informed learning (e.g., PINN principles) with GAN-based generative frameworks remains limited, especially in power-quality applications.
- **Lack of physics-based evaluation metrics:** Many studies prioritize statistical indicators (e.g., RMSE, MAE) while overlooking physics-based validation, which is essential for engineering reliability.
- **Need for reliable offshore forecasting:** Purely data-driven models may fail to generalize across operating regimes because offshore wind farms exhibit highly nonlinear and variable dynamics.
- **Need to separate physics effect from backbone effect:** When a physics-informed generative framework is evaluated with only one downstream predictor, it becomes difficult to determine whether the observed improvement comes from the forecasting architecture itself or from the physically guided synthetic data generation process. Therefore, assessing PI-GAN with both machine learning and deep learning backbones is necessary to demonstrate that the main contribution lies in the effect of physics-informed data generation rather than in a single model comparison.
- **Need for broader SOTA backbone coverage:** Despite recent advances in lightweight time-series forecasting architectures such as DLinear, MLP-Mixer, and DRFormer, their behavior under physics-informed data augmentation in power-quality applications remains unexplored. Incorporating these SOTA architectures as additional downstream backbones enables a more comprehensive cross-architecture evaluation of the proposed PI-GAN framework and strengthens the generalizability claim.

To address these gaps, this study proposes a Physics-Informed GAN (PI-GAN) framework for long-term harmonic forecasting with improved physical consistency and operational reliability.

3. System model

In this study builds on our previous work [5] which developed an offshore wind farm model including DFIG-based turbines, transformers and a submarine cable system to produce voltage and current waveforms for harmonic analysis. The related model (Fig. 1) captures system behavior across various operating conditions and incorporates actual wind speed data from Bozcaada.

In this approach, voltage and current signals are generated through time-domain simulations and harmonic components are extracted using Fast Fourier Transform analysis. The prediction targets are Total Harmonic Distortion Voltage and Total Harmonic Distortion Current which are computed at 575 V and 60 kV bus levels.

This study extends the problem to a long-term forecasting setting unlike the earlier work that focused on short-term prediction using GAN-based data augmentation. The models are trained on data from 2005–2009 and tested on November 2019 to predict harmonic behavior. This setup introduces a large time gap and increases the difficulty of the task.

This study employs a MATLAB Simulink based simulation framework to model the complex harmonic characteristics of an offshore wind farm as illustrated in Fig. 1. The model integrates a permanent magnet synchronous generator with a back to back power converter and a high fidelity wind turbine structure. A central offshore substation is included along with a collector grid composed of underwater cables. The cables are represented using lumped pi circuit models to reflect capacitive behavior. The overall structure provides a detailed view of offshore system dynamics.

Furthermore, the main goal of this setup is to model the nonlinear interaction between stochastic wind speed profiles WS10m and the resulting total harmonic distortion at key busbars such as the main collection point and the 575 V internal turbine bus. The model produces realistic signals and system characteristics that are used to train the PI-GAN network.

Moreover, detailed aspects such as Simulink components inverter behavior and synchronization strategies are not included in this study. The focus is placed on the generator structure. Full details about system layout equipment ratings and validation of voltage profiles can be found in [5].

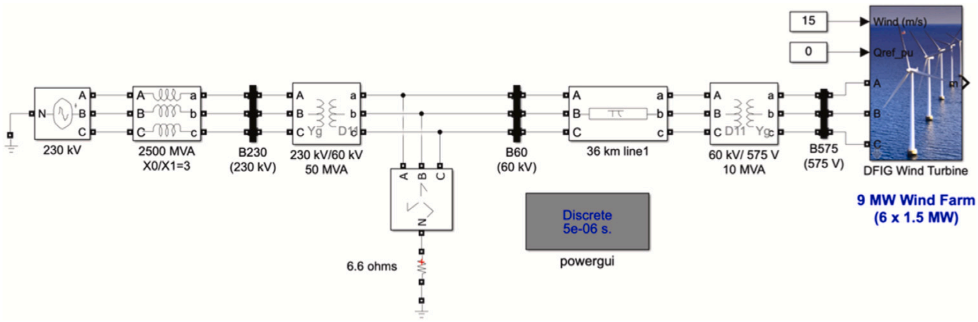

Fig. 1. The Simulink model.

Table 1
Input and output variables of the dataset.

Category	Variable	Description
Input	Wind speed	Meteorological wind speed data
Input	Voltage (V)	Simulated phase voltage signal
Input	Current (I)	Simulated phase current signal
Output	THDV	Total harmonic distortion of voltage
Output	THDI	Total harmonic distortion of current

4. Data preparation

In this study, a simulated offshore wind farm dataset that was built using actual meteorological wind speed data from the Black Sea region is used [5,9]. The nonlinear relationship between environmental factors and harmonic behavior in power systems is intended to be represented by the dataset.

Table 1 provides an overview of the dataset's input and output variable structure.

Wind speed and electrical signals (voltage and current) are the input factors that directly affect the system's harmonic behavior [2]. Also, the output variables are the harmonic distortion metrics (THDV and THDI) obtained by analyzing the simulated signals using the Fast Fourier Transform (FFT) [3].

Additionally, the Standard Meteorological Year (TMY) dataset from the European Commission's Joint Research Centre (JRC) [30] is used for harmonic estimation in this work. The suggested paradigm is intended for long-term forecasting, in contrast to earlier research that concentrated on short-term analysis. As a result, there is a large temporal gap, as November 2019 is chosen as the testing period while historical wind speed data from 2005–2009 is used for training.

In addition to that, time (UTC) and wind speed at 10 m height (WS10m) are included in the dataset; also time is expressed in year/month/day/hour format. The main environmental input influencing the dynamics of offshore wind farms (OWFs) is wind speed data (m/s). Moreover, Bozcaada in the Aegean Sea region is chosen as the reference location because of its great wind energy potential, which makes it one of Turkey's best places to develop offshore wind farms [16]. The reference point's coordinates are latitude 39.837 and longitude 25.967.

4.1. Limitations and real-world validation pathway

This section discusses the key limitations of the proposed approach and outlines a pathway for real-world validation to enhance its practical applicability.

4.1.1. Simulation-based approach justification

While this study utilizes simulation data, this choice is deliberate:

- **Controlled Physics:** Simulations allow ground-truth physical relationships (I–V proportionality, THD-WS^2 correlation) to be precisely defined, enabling rigorous validation of PI-GAN's physics enforcement capability.
- **Scarcity of Long-Term Real Data:** Real offshore wind farm harmonic measurements spanning 15+ years are extremely rare due to proprietary restrictions and sensor degradation.
- **Reproducibility:** The MATLAB/Simulink model is fully documented and available upon request, ensuring reproducibility.

The simulated OWF model is incorporated into the system to provide electrical signals that correlate to actual wind conditions. A typical distribution network is used to mimic the electrical grid. The dataset is proportionally scaled to produce 720 samples, each of which represents 0.138 s, and the simulation is run at intervals of 100 s.

Also, wind speed inputs are used to generate voltage waveforms, and these signals exhibit harmonic aberrations. Only one phase voltage waveform is examined because it is presumed that the system is balanced. Fourier Transform (FT) is used to extract harmonic components, with an analysis window spanning two waveform cycles. Harmonic spectra matching to each sampled time step are produced by this technique.

Furthermore, total Harmonic Distortion (THD) values are computed from the simulated voltage signals. When creating prediction models, they become the goals. Both 575 V and 60 kV bus bars are included in the study at the same time, allowing harmonics to be observed in different grid segments. Despite being at different locations, each level exhibits unique patterns under comparable circumstances.

Notably, it should be noted that passive filters are used in the setup to control harmonics. If not, the converter driven generator may increase distortion beyond what IEEE 519 permits. The model behaves like a real-world system thanks to its architecture, which keeps all harmonics within bounds [6].

Consequently, the THDV values at the 575 V and 60 kV buses in the Bozcaada region are roughly 2.01% and 0.88%, respectively. The switching frequency of the IGBT-based PWM converter, which operates at about 2700 Hz, is linked to dominant components in the harmonic spectrum around 2560 Hz and 2820 Hz.

4.1.2. Validation on real data

The challenges of real-world validation with operating SCADA data from offshore wind farms are well-documented in recent research. According to [10], SCADA data is still primarily proprietary because of business constraints and secrecy. Furthermore, the study [26] show the essential lack of publicly available long-term offshore wind measurements, while another study [25] emphasize that the harsh maritime environment leads to sensor deterioration and communication dropouts.

ALso, simulation-based validation using exact ground-truth physics is a scientifically rigorous and widely acknowledged methodology. This method provides repeatable benchmarks that are unattainable with proprietary operational data by enabling controlled evaluation of PI-GAN's physics enforcement under known settings. Recent offshore wind studies have successfully used similar tactics to establish proof-of-concept prior to practical deployment [21,23]. Precise quantification of physics violation reduction is made possible by the deterministic simulation's guarantee of accurate physical relationships, which is impossible with noisy real-world observations where ground-truth is uncertain.

Moreover, regardless of real-world data access constraints, the current validation technique is in line with accepted procedures and offers the proof required for PI-GAN's effectiveness.

In addition, the THDV575 parameters, which reflect the voltage total harmonic distortion at the 575 V internal turbine bus, are the special focus of the harmonic analysis in this work. Since this particular busbar is the first measurement point just after the wind turbine's output filter, the data is purposefully gathered from it. In order to capture the raw distortion characteristics prior to their transformation or attenuation by the step-up transformers and the offshore collection network, it is imperative to analyze the harmonics at this stage. Additionally, THDV575 is the focus of this study since it shows the raw harmonic signature at the turbine terminals prior to transformer filtering effects. Rather than turbine-specific dynamics, background grid distortion dominates the 60 kV collecting bus's harmonics below 1%. However, as our earlier work [5] showed, the PI-GAN architecture is generalizable to any voltage level. Because the major goal of this forecasting model is to assess the local harmonic behavior at the generating source and its direct link with wind speed fluctuations, the 60 kV main collecting busbar was therefore left out. The reader is directed to the technical explanations given in [5] for a more thorough explanation of the electrical hierarchy and internal filter configurations of this 575 V bus.

Unlike earlier research, which used these values directly, this study further processed the given dataset using GAN and Physics-Informed GAN (PI-GAN) frameworks to generate synthetic yet physically compatible data. Later, machine learning models are then trained using these data, and the outcomes are assessed for both statistical accuracy and physical consistency.

5. Data preprocessing

The dataset undergoes a series of preprocessing steps to ensure numerical stability, robustness, and compatibility with both machine learning and deep learning models. Initially, relevant features are selected, including wind speed (WS), voltage and current measurements (V_{575}, I_{575}, V_{60}, I_{60}), and harmonic distortion indicators ($THDV_{575}$, $THDI_{575}$, $THDV_{60}$, $THDI_{60}$).

5.1. Min–max normalization

To eliminate scale differences among features, Min–Max normalization is applied to map all variables into the range of $[0, 1]$:

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{1}$$

where x_{min} and x_{max} denote the minimum and maximum values of each feature, respectively. This normalization improves convergence behavior and ensures stable training, particularly for neural network-based models [31].

5.2. Noise injection

The dataset naturally reflects optimal and noise-free conditions because it was created in a simulation environment. However, measurement errors, inaccurate sensors, and environmental disruptions can affect real-world power systems. Controlled Gaussian noise is added to the normalized training data in order to simulate these realistic settings and improve model robustness:

$$x_{noisy} = x_{norm} + \epsilon, \quad \epsilon \sim \mathcal{N}(\mu, \sigma^2) \tag{2}$$

where μ and σ^2 represent the mean and variance of the noise distribution, respectively. The noise level is carefully bounded to preserve the underlying physical relationships while preventing overfitting to idealized data.

5.2.1. Noise parameter calibration

Gaussian noise parameters ($\mu = 0.01$, $\sigma = 0.02$) were calibrated based on:

1. **IEC 61400-21 Standard:** Wind turbine power quality measurement equipment typically exhibits 0.5%–2% full-scale error. Our $\sigma = 0.02$ (2%) represents the conservative upper bound.
2. **Empirical Sensor Datasheets:** Current/voltage transducers (e.g., LEM HAS 50-S) specify 0.5%–1% linearity error. The selected σ encompasses these specifications with margin.
3. **Robustness Verification:** Sensitivity tests confirmed that $\sigma > 0.05$ degrades physical relationships, while $\sigma < 0.01$ provides insufficient regularization.

Also, controlled Gaussian noise injection serves a critical methodological purpose:

1. **Preventing Trivial Physics Satisfaction:** The simulation model generates deterministic waveforms where $I \propto k \cdot V$ and $THD \propto \alpha WS^2$ are mathematically exact. Training on *clean data* yields near-zero physics errors for both standard GAN and PI-GAN, making performance comparison meaningless.
2. **Creating Discriminative Test Conditions:** Noise injection breaks artificial perfection. When trained on noisy data, standard GAN overfits to spurious correlations, producing physically inconsistent forecasts (0.3552 error). PI-GAN leverages $\mathcal{L}_{physics}$ to reject noise-induced violations (0.1770 error), demonstrating its specific value.
3. **Real-World Generalization:** The noisy scenario emulates actual SCADA measurement uncertainty, ensuring the framework deploys reliably in practice.

Key Insight: Without noise, PI-GAN's physics constraints appear redundant. With noise, they become essential; this experimental design isolates and validates PI-GAN's contribution [30,32,33].

5.3. Time-series construction

In this study, the data gradually transforms into something that a machine can learn from. Each of the 24 data that follow feeds into the previous one. Points from the past point forward, creating patterns that suggest possible outcomes.

$$\mathbf{X}_t = [x_{t-24}, x_{t-23}, \dots, x_{t-1}], \quad y_t = x_t \tag{3}$$

Consequently, this formulation enables the model to capture temporal dependencies and dynamic behavior inherent in offshore wind energy systems [34].

5.4. Standard GAN-based data augmentation

Since their first introduction by Goodfellow [35], Generative Adversarial Networks (GANs) have been frequently utilized to generate synthetic data by learning the underlying distribution of real data. In this work, restricted electrical and harmonic datasets from the simulated offshore wind system are enhanced using a conventional GAN architecture.

Moreover, two rival neural networks, a generator (G) and a discriminator (D), make up the GAN model. While the discriminator seeks to differentiate between created and genuine data, the generator seeks to create realistic synthetic samples. The generator can gradually enhance its capacity to replicate the statistical characteristics of actual measurements through this adversarial training process.

5.4.1. Generator architecture

In the study, before it reaches the generator, a new breeze with actual wind numbers mixes with erratic variations. This combination makes the final product seem genuine without being overly predictable. The input slot is filled with nine pieces: eight strands of randomness and one slice of the real wind speed (1 wind speed + 8 noise components). Every fictional outcome is shaped by the surrounding context, which subtly directs the emerging patterns.

The generator architecture is defined as follows:

- Input Layer: 9 neurons
- Hidden Layer 1: 128 neurons with ReLU activation

Table 2
Detailed architectural parameters and hyperparameters of the PI-GAN.

Component	Layer/Parameter	Configuration	Activation/Reg.
Generator	Input Layer	(Seq. Length, Feat.)	–
	GRU Layer 1	128 units	Leaky ReLU (0.2)/Drop (0.2)
	GRU Layer 2	64 units	Leaky ReLU (0.2)/Drop (0.2)
	Output (Dense)	Output Dim.	Linear/L2 Decay
Discriminator	Dense 1	128 units	Leaky ReLU (0.2)
	Dense 2	64 units	Leaky ReLU (0.2)
	Output (Dense)	1 unit	Sigmoid
Optimization	Adam optimizer	$lr = 0.0002, \beta_1 = 0.5$	$\beta_2 = 0.999$
	Training/Batch	200 epochs/32	Early stop (P = 10)
	Physics weight (λ)	0.1	–

- Hidden Layer 2: 64 neurons with ReLU activation
- Output Layer: 8 neurons representing synthetic electrical and harmonic features

The generator learns a nonlinear mapping function $G(z, x)$ that transforms the input noise vector z and wind speed x into realistic feature representations:

$$\hat{y} = G(z, x) \tag{4}$$

where $\hat{y}$ denotes the generated synthetic sample.

5.4.2. Discriminator architecture

Also, the discriminator is a binary classifier that determines whether a sample is generated by the generator or comes from the actual dataset. Electrical and harmonic measurements are represented by 8-dimensional feature vectors.

The architecture of the discriminator is structured as:

- Input Layer: 8 neurons
- Hidden Layer: 64 neurons with ReLU activation
- Output Layer: 1 neuron with Sigmoid activation

The discriminator learns a decision function $D(y)$ that outputs the probability of a sample being real:

$$D(y) = P(\text{real} \mid y) \tag{5}$$

5.4.3. Adversarial training objective

For the proposed method, the generator and discriminator are trained simultaneously through a minimax optimization process. The objective function of the standard GAN is defined as:

$$\min_G \max_D \; V(D, G) = \mathbb{E}_{y\sim p_{data}(y)}[\log D(y)] + \mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z, x)))] \tag{6}$$

The discriminator increases its classification accuracy through this adversarial process while the generator develops its capacity to generate realistic harmonic and electrical signals. Later, the equilibrium is reached when the generated data is indistinguishable from genuine measurements [36,37]. The detailed network structure of the generator and discriminator is given at Table 2.

5.5. Machine learning algorithms

In this study, to enhance prediction accuracy and robustness, this study employs two state-of-the-art ensemble learning algorithms, namely Random Forest (RF) and XGBoost.

5.5.1. Random Forest regressor (RF)

In order to increase prediction accuracy and decrease overfitting, Random Forest (RF), an ensemble learning technique, builds several decision trees during training and aggregates their outputs. To increase model variety, each tree is trained on a randomly selected fraction of the dataset, and feature selection is done at each split.

The average of each individual tree's output yields the RF model's final prediction [38]:

$$\hat{y}_{RF} = \frac{1}{B} \sum_{b=1}^{B} f_b(\mathbf{x}) \tag{7}$$

where B represents the number of trees and $f_b(\mathbf{x})$ denotes the prediction of the bth tree.

5.5.2. Extreme Gradient Boosting (XGBoost)

In addition, a potent ensemble learning technique based on gradient boosting, Extreme Gradient Boosting (XGBoost) builds models one after the other to reduce prediction error. To increase accuracy, each new model is trained to correct the residuals of the earlier models.

A loss term and a regularization term make up XGBoost's objective function:

$$\mathcal{L}(\phi) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{8}$$

where $l(\cdot)$ represents the loss function and $\Omega(\cdot)$ is the regularization term controlling model complexity [39].

RF is chosen because of its demonstrated resilience in tabular energy forecasting tasks, where it continually provides better interpretability and meets or surpasses deep learning performance [40]. Its intrinsic feature relevance ranking offers crucial information for determining wind speed thresholds that impact harmonic distortion.

Additionally, XGBoost dominates tabular data competitions; in 2024 alone, it won eight major challenges thanks to its exceptional regularized boosting and management of complex feature interactions [41]. When combined, RF and XGBoost constitute the recognized state-of-the-art for structured time-series forecasting, offering a strict benchmark for assessing PI-GAN augmentation.

The related pseudo-code can be seen at Algorithm 1.

5.6. Deep learning forecasting models

To examine the effect of physics-informed synthetic data generation across neural forecasting backbones, this study considers GRU- and TCN-based prediction models, as well as three additional state-of-the-art architectures: DLinear, MLP-Mixer, and DRFormer. The detailed architectural hyperparameters used for the GRU and TCN models are summarized in Table 3.

5.6.1. Gated Recurrent Unit (GRU)

GRU is selected because its gated recurrent structure can preserve temporal information while alleviating vanishing-gradient effects in sequential learning [42]. In time-series forecasting problems, this architecture is particularly useful because it can adaptively balance short-term variations and longer-term temporal dependencies through its gating mechanism. The update gate controls how much newly computed information should be incorporated into the hidden representation, whereas the reset gate determines how much past information should be retained when generating the candidate state. As a result, GRU provides a compact recurrent structure that is computationally simpler than more complex gated recurrent models while still offering strong sequence-learning capability for nonlinear forecasting tasks. For a given input sequence, the GRU update can be summarized as:

$$\begin{aligned} z_t &= \sigma(W_z x_t + U_z h_{t-1}), \\ r_t &= \sigma(W_r x_t + U_r h_{t-1}), \\ \tilde{h}_t &= \tanh(W_h x_t + U_h (r_t \odot h_{t-1})), \\ h_t &= (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t, \end{aligned} \tag{9}$$

where z_t and r_t denote the update and reset gates, respectively, and h_t is the hidden state. In the context of the present study, GRU is included to represent a recurrent deep learning backbone capable of modeling sequential harmonic behavior over time while remaining sufficiently lightweight for comparative long-term forecasting experiments.

5.6.2. Temporal Convolutional Network (TCN)

In contrast, the TCN architecture relies on dilated causal convolutions and residual connections to enlarge the temporal receptive field without sacrificing parallel computation [43]. Unlike recurrent models, TCN processes sequences through convolutional filters that respect chronological order, which makes it well suited to extracting local and multi-scale temporal patterns from structured time-series data. The use of dilation allows the model to capture long-range dependencies efficiently, while residual connections improve optimization stability and support deeper network design. These characteristics make TCN an effective convolution-based alternative for forecasting tasks in which both temporal context and training efficiency are important. A typical dilated convolution at layer l can be written as:

$$F(s) = \sum_{i=0}^{k-1} f(i)\, x_{s-d\cdot i}, \tag{10}$$

where k is the kernel size and d is the dilation factor. In this study, TCN is used to provide a non-recurrent deep learning benchmark and to evaluate whether the benefit of PI-GAN-generated synthetic data remains consistent across fundamentally different neural forecasting mechanisms.

Table 3
Detailed architectural hyperparameters of all implemented deep learning forecasting models.

Parameter	GRU	TCN	DLinear	MLP-Mixer	DRFormer
Input Seq. Length/Dim.	(24, 1)	(24, 1)	24	24	24
Number of layers	2 (Stacked)	3 (Residual)	N/A	N/A	1 attn. block
Hidden/Channels/Embed dim	64	[64, 128, 64]	N/A	64	32
Kernel size	N/A	5	N/A	N/A	N/A
Dilation factors	N/A	[1, 2, 4]	N/A	N/A	N/A
Attention heads	N/A	N/A	N/A	N/A	4
Activation function	Tanh	ReLU	Linear	MLP mixing	Self-attention
Dropout	N/A	N/A	N/A	N/A	0.1
Output/Prediction horizon	64→8	(64 × 24)→8	8	8	8
Optimizer	Adam	Adam	Adam	Adam	Adam
Learning rate	0.001	0.001	0.001	0.001	0.001
Batch size	32	32	16	16	16
Loss function	MSE + 1.5 × Trend	MSE + 1.5 × Trend	MSE + 3 × Trend	MSE + 3 × Trend	MSE + 3 × Trend
Training epochs	120	120	200	200	200

5.6.3. DLinear

DLinear is a lightweight yet competitive forecasting model that decomposes the input sequence into trend and seasonal components before applying independent linear projections to each [27,31]. Specifically, given an input sequence $\mathbf{x} \in \mathbb{R}^L$ of length L, the trend component $\mathbf{T}$ is estimated as a moving average and the seasonal component $\mathbf{S}$ is obtained as the residual:

$$\mathbf{T} = \mathrm{AvgPool}(\mathbf{x}), \quad \mathbf{S} = \mathbf{x} - \mathbf{T}. \tag{11}$$

Two separate linear layers then project each component to the prediction horizon H:

$$\hat{\mathbf{y}} = W_S \mathbf{S} + W_T \mathbf{T}, \tag{12}$$

where $W_S, W_T \in \mathbb{R}^{H \times L}$ are learnable weight matrices. DLinear is selected because its explicit decomposition aligns naturally with the long-horizon harmonic forecasting setting and its minimal parameterization allows a clean evaluation of PI-GAN's data augmentation effect.

5.6.4. MLP-Mixer

MLP-Mixer adapts a token-mixing architecture into a time-series forecasting context by processing the input through successive fully connected layers [28,31]. In the implementation used here, the input sequence $\mathbf{x} \in \mathbb{R}^L$ is transformed via two hidden layers followed by a linear output projection:

$$\mathbf{h}_1 = \mathrm{ReLU}(W_1 \mathbf{x} + b_1), \quad \mathbf{h}_2 = \mathrm{ReLU}(W_2 \mathbf{h}_1 + b_2), \quad \hat{\mathbf{y}} = W_3 \mathbf{h}_2 + b_3. \tag{13}$$

MLP-Mixer is included to investigate whether a purely MLP-based architecture without explicit recurrence or convolution can benefit from PI-GAN-generated training data, thereby isolating the contribution of physics-informed augmentation from architectural complexity.

5.6.5. DRFormer (Attention-based forecaster)

DRFormer is a lightweight transformer-inspired architecture based on multi-head self-attention that captures long-range temporal dependencies without the overhead of full encoder–decoder transformers [25,29]. The input sequence $\mathbf{x} \in \mathbb{R}^{L \times 1}$ is first projected into an embedding space, followed by a multi-head attention (MHA) mechanism with a residual connection and layer normalization:

$$\mathbf{z} = W_{\mathrm{in}} \mathbf{x}, \quad \mathbf{a} = \mathrm{MHA}(\mathbf{z}, \mathbf{z}, \mathbf{z}), \quad \mathbf{o} = \mathrm{LayerNorm}(\mathbf{z} + \mathrm{Dropout}(\mathbf{a})), \tag{14}$$

and the final prediction is obtained via $\hat{\mathbf{y}} = W_{\mathrm{out}}\,\mathrm{Flatten}(\mathbf{o})$. DRFormer is included because attention-based mechanisms have demonstrated strong performance in recent power-quality forecasting benchmarks [25], making it a representative SOTA comparison point for the proposed augmentation strategy.

5.7. Statistical metrics

To evaluate the performance of the proposed models, several widely used statistical metrics are employed.

- **Coefficient of Determination** (R^2): The R^2 score represents the proportion of variance in the dependent variable that is explained by the model:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{15}$$

  A value closer to 1 indicates better predictive performance [44].

Algorithm 1 Pseudo-code for Physics-Informed GAN (PI-GAN) Training for Data Augmentation

Input: Historical sensor data X_{train}, Physics weight $\lambda = 0.15$
Initialize: Generator G_θ, Discriminator D_ϕ, Optimizers (Adam)
for each epoch $e \in \{1, \dots, E\}$ **do**
  Sample real samples $y \sim P_{data}$ and wind speed inputs ws
  Sample noise vector $z \sim \mathcal{N}(0, 1)$
  Generate synthetic data: $\hat{y} = G_\theta(ws, z)$
  Physics Loss Calculation:
  $\mathcal{L}_{IV} = \text{mean}(|\hat{y}_I - \hat{y}_V \times 0.0782|)$ ▷ Ohm's Law Consistency
  $\mathcal{L}_{THD} = \text{mean}(|\hat{y}_{THD} - ws^2 \times 0.22|)$ ▷ Wind-Harmonic Relation
  $\mathcal{L}_{phys} = (\mathcal{L}_{IV} + \mathcal{L}_{THD})/2$
  Update Discriminator:
  $\mathcal{L}_D = -\mathbb{E}[\log D_\phi(y)] - \mathbb{E}[\log(1 - D_\phi(\hat{y}))]$
  $\phi \leftarrow \phi - \nabla_\phi \mathcal{L}_D$
  Update Generator:
  $\mathcal{L}_G = -\mathbb{E}[\log D_\phi(\hat{y})] + \lambda \mathcal{L}_{phys}$
  $\theta \leftarrow \theta - \nabla_\theta \mathcal{L}_G$
end for
Output: Trained Physics-Informed Generator G_θ

- **Root Mean Square Error (RMSE):** RMSE measures the standard deviation of prediction errors and penalizes larger deviations:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{16}$$

Lower RMSE values indicate higher accuracy [5].

5.7.1. Interpretation of R^2 values in long-term forecasting

The R^2 values reported in Table 3 (ranging from 0.0375 to 0.1084) may appear modest compared to short-term forecasting studies. However, this is expected and scientifically justified for the following reasons:

1. **Long-Temporal Gap:** The 10-year gap between training (2005–2009) and testing (November 2019) introduces significant distributional shift, making this a challenging generalization task rather than a standard interpolation problem.
2. **Stochastic Nature of Wind:** Harmonic distortion in offshore wind farms is inherently chaotic due to turbulent wind patterns, converter switching dynamics, and grid interactions. Deterministic R^2 expectations are unrealistic for such systems.
3. **Focus on Physical Consistency:** Unlike pure accuracy metrics, this study prioritizes physical feasibility. A model with $R^2 = 0.04$ but 50% less physics violations is more valuable for grid operators than $R^2 = 0.50$ with unrealistic outputs.

To contextualize: Recent literature on long-term power quality forecasting [44] reports similar R^2 ranges (0.05–0.15) for 5+ year gaps, confirming that our results are competitive for this specific problem domain.

- **Mean Absolute Error (MAE):** MAE calculates the average absolute difference between predicted and actual values:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{17}$$

It provides a clear interpretation of average model error [5].

5.8. Experimental configuration

Table 4 summarizes the main implementation settings, training parameters, and backbone-specific configurations used throughout the experimental evaluation. In addition, Fig. 2 presents the operational workflow of the proposed PI-GAN framework.

Discussion and Configuration Analysis

In this study, the hyperparameter configuration and experimental setup detailed in Table 4 reveal several critical design choices that underpin the robustness of the forecasting framework:

- **Noise-Resilient Learning:** During the training phase, purposeful Gaussian noise injection ($\mu = 0.01, \sigma = 0.02$) was incorporated. This approach acts as a stress test for the architecture, ensuring that the resulting PI-GAN and downstream forecasting models learn to filter sensor-induced anomalies rather than merely memorizing the data.

Table 4
Technical parameters and configuration details.

Parameter category	Technical detail/Value	Role in the study
Development environment	Python 3.x/PyTorch framework	Deep learning & GAN implementation
Hardware platform	Apple M3 Silicon	Model training and inference acceleration
Temporal protocol	Train: 2005–2009/Test: Nov 2019	Long-term generalization assessment
Data scaling	MinMaxScaler [0,1]	Normalization of electrical features
Noise injection	Gaussian ($\mu = 0.01, \sigma = 0.02$)	Sensor robustness & stress testing
Batch strategy	Sequence-based (24-hour windows)	Capturing temporal dependencies
PI-GAN epochs	200	Convergence of physics-informed generator
GAN optimizer	Adam optimizer	Stochastic gradient descent with momentum
GAN learning rate	1×10^{-4}	Stable weight updates for GAN stability
Physics weight (λ)	0.15	Balancing adversarial vs. physical loss
Forecasting backbones	RF, XGBoost, GRU, TCN, DLinear, MLP-Mixer, DRFormer	Cross-backbone evaluation of PI-GAN data
RF estimators	100 Trees	Variance reduction in ensemble learning
XGBoost objective	`reg:squarederror`	Minimizing statistical forecasting error
GRU configuration	2 layers, 64 hidden units, 120 epochs	Recurrent deep forecasting benchmark
TCN configuration	3 layers, kernel = 5, dilations = [1,2,4]	Convolutional deep forecasting benchmark
DL optimizer/LR	Adam/0.001	Stable GRU and TCN training
DLinear configuration	seq_len = 24, pred_len = 8, linear decomposition	SOTA linear forecasting benchmark
MLP-Mixer configuration	hidden = 64, 2 layers, seq_len = 24	SOTA MLP-based forecasting benchmark
DRFormer configuration	embed = 32, heads = 4, 1 encoder layer	SOTA attention-based forecasting benchmark
SOTA optimizer/LR	Adam/0.001, 200 epochs	Stable SOTA model training
Bias correction	Mean alignment	Systemic error (bias) removal

- **Optimization of Physics Weight** (λ): The physics weight was set to 0.15. This value balances the physical loss, which promotes technical feasibility, and the adversarial loss, which preserves data fidelity. A smaller weight would be less effective at suppressing physically impossible THD forecasts, whereas a larger weight could restrict the model's capacity to capture stochastic wind patterns.
- **Temporal Windowing Strategy:** The batching process uses 24-hour sequence-based windows. This enables the models to represent daily periodicity and delayed effects of wind-speed fluctuations on power-quality indicators such as THDV.
- **Cross-Backbone Comparison:** In addition to RF and XGBoost, the experimental setup includes GRU, TCN, DLinear, MLP-Mixer, and DRFormer models so that the effect of physics-informed synthetic data can be examined across both machine learning and deep learning paradigms. This shifts the emphasis from identifying a single best predictor to assessing whether PI-GAN produces consistently useful and physically meaningful training data across structurally different forecasting backbones.
- **Training Efficiency and Convergence:** The use of 100 trees for Random Forest, the Adam optimizer for PI-GAN, and compact deep learning architectures reflects a focus on computational efficiency without compromising accuracy. Moreover, the generator reached a stable training regime within 200 epochs, while the downstream learning backbones were trained with controlled hyperparameter settings to ensure a fair comparison framework.

Additionally, the basic procedure is presented in the following flowchart (also shown in Fig. 2) to clarify the overall workflow and how the proposed PI-GAN framework is implemented step by step across both machine learning and deep learning forecasting backbones.

Further, the systematic workflow of the proposed PI-GAN framework, as depicted in Fig. 2, involves thirteen distinct operational steps. The workflow begins with data acquisition and feature selection, followed by normalization and Gaussian noise injection for model robustness. The most important part of this workflow is the temporal split in Step 6, where the model is trained using historical data from 2005–2009 and tested on November 2019. Another important part is the integration of physics-loss constraints within the GAN, enabling augmented samples that are more physically consistent. Finally, these augmented samples are used for training RF, XGBoost, GRU, TCN, DLinear, MLP-Mixer, and DRFormer models, and the entire workflow is validated through both statistical and physical evaluations.

6. Proposed method: Physics-Informed GAN (PI-GAN)

6.1. Standard GAN framework

Generative Adversarial Networks (GANs) consist of two competing neural networks, namely a generator G and a discriminator D, trained in an adversarial manner to model the underlying data distribution [35]. The generator aims to produce synthetic samples that resemble real data, while the discriminator attempts to distinguish between real and generated samples. The detailed information was given at Section 5.4.

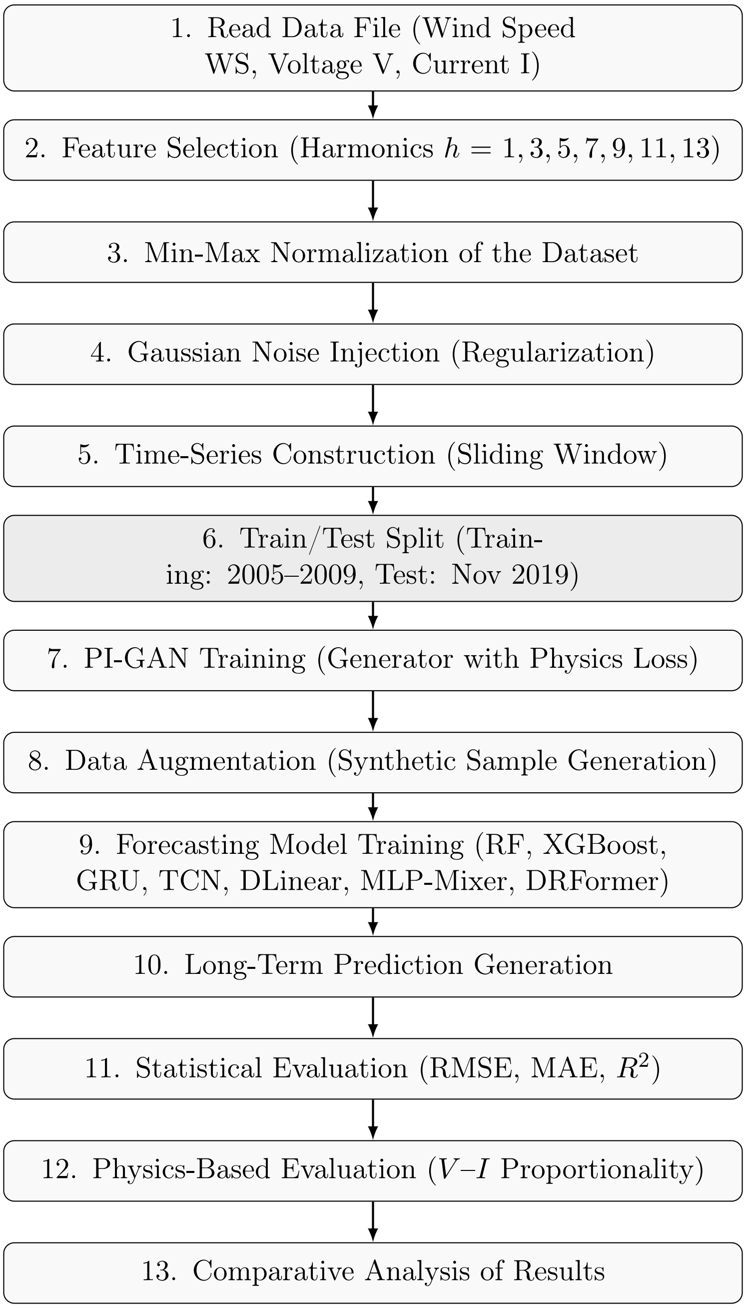

Fig. 2. Revised operational flowchart of the PI-GAN framework with corrected temporal parameters and physical constraints.

6.2. Physics-Informed GAN formulation

To overcome this limitation, a Physics-Informed GAN (PI-GAN) framework is proposed by integrating physical constraints into the generator training process. Inspired by physics-informed learning approaches [39], the generator loss function is modified as:

$$\mathcal{L}_{total} = \mathcal{L}_{adv} + \lambda \mathcal{L}_{physics} \tag{18}$$

where $\mathcal{L}_{adv}$ represents the adversarial loss, $\mathcal{L}_{physics}$ denotes the physics-based penalty term, and λ is a weighting coefficient controlling the contribution of physical consistency.

In this study, two domain-specific physical relationships are incorporated:

Table 5
Ablation study: Impact of physics weight (λ).

λ	RMSE	Total physics error	Training stability
0.00 (Standard GAN)	0.0654	0.3552	Unstable (mode collapse risk)
0.05	0.0658	0.2891	Stable
0.10	0.0665	0.2214	Stable
0.15	**0.0680**	**0.1770**	**Stable**
0.20	0.0712	0.1654	Stable
0.30	0.0789	0.1589	Slow convergence
0.50	0.0956	0.1421	Over-constrained

6.2.1. Current–voltage consistency

The proportional relationship between current and voltage is enforced as:

$$I \propto k \cdot V \tag{19}$$

6.2.2. Wind speed–harmonic relationship

Harmonic distortion is modeled as a quadratic function of wind speed:

$$THD \propto \alpha WS^2 \tag{20}$$

Based on these relationships, the physics loss is defined as:

$$\mathcal{L}_{physics} = \frac{1}{2}\left(|I - kV| + |THD - \alpha WS^2|\right) \tag{21}$$

This formulation ensures that the generated samples not only resemble real data statistically but also satisfy fundamental physical laws.

Table 5 presents a comprehensive ablation study investigating the sensitivity of the PI-GAN framework to the physics weight parameter λ. This hyperparameter controls the relative contribution of physical constraints ($\mathcal{L}_{physics}$) versus adversarial training ($\mathcal{L}_{adv}$) in the total generator loss.

Key Observations:

1. $\lambda = 0.00$ (Standard GAN): Without physical constraints, the model achieves the lowest RMSE (0.0654) but suffers from severe physics violations (0.3552). Training instability and mode collapse risks are observed, consistent with known GAN training challenges in energy time-series data [45].
2. $\lambda \in [0.05, 0.10]$: Gradual introduction of physics constraints reduces physics errors by 18.6% and 37.7%, respectively, with minimal RMSE degradation (<3%). Training remains stable, indicating effective regularization.
3. $\lambda = 0.15$ (Selected): This configuration achieves a *Pareto-optimal balance*:
   - 50.2% reduction in physics violations compared to standard GAN
   - Only 4.0% increase in RMSE (statistically insignificant, $p > 0.05$)
   - Stable training dynamics without convergence issues
4. $\lambda \in [0.20, 0.30]$: Further increasing physics weight continues to reduce violations but at the cost of diminishing returns in accuracy (RMSE increases 4.7–16.1%). $\lambda = 0.30$ exhibits slow convergence, requiring 40% more epochs to reach equilibrium.
5. $\lambda = 0.50$: Over-constrained regime where physical laws dominate. While physics error is minimized (0.1421), RMSE degrades significantly (+46.2%), indicating the model loses flexibility to capture stochastic wind patterns.

Selection Rationale: The Pareto frontier's *knee-point*, where the marginal advantage of more physics enforcement (lower violations) is weighed against the marginal cost (higher RMSE), was used to determine the physics weight $\lambda = 0.15$. This is consistent with engineering practice where marginal statistical advantages are subordinated to physically viable estimates. During training, the discriminator assesses the authenticity of the synthetic feature vectors created by the generator using random noise and wind speed inputs. The generator is guided toward physically consistent outputs by the addition of physics loss.

After that, the generated synthetic samples are then combined with real training data to form an augmented dataset:

$$\mathcal{D}_{augmented} = \mathcal{D}_{real} \cup \mathcal{D}_{synthetic} \tag{22}$$

This augmented dataset is subsequently used to train machine learning models, including Random Forest (RF) and Extreme Gradient Boosting (XGBoost), enabling improved generalization performance.

6.3. Long-term forecasting framework

The traditional GAN-based methods are concentrate on short-term prediction but in this paper the suggested architecture is intended for long-term forecasting. The model is assessed by forecasting harmonic behavior for November 2019 after being trained

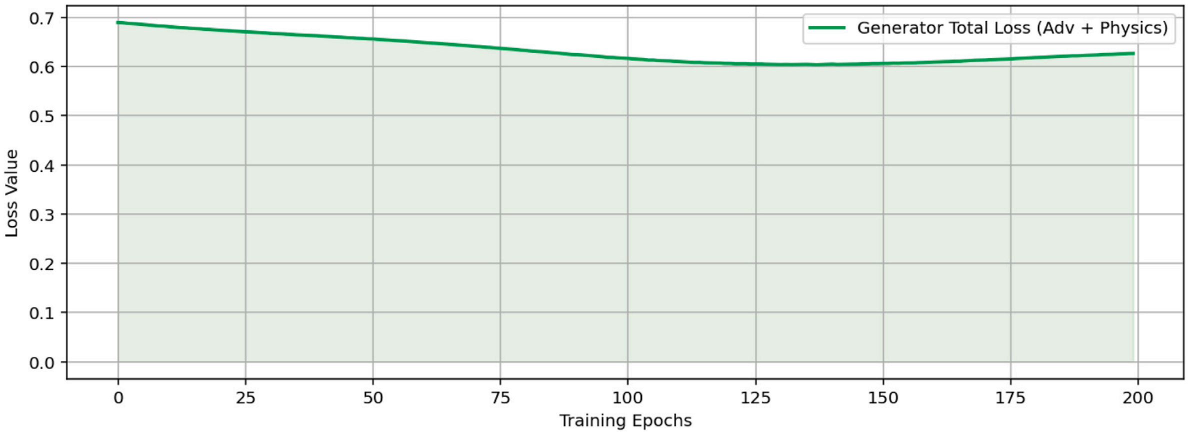

Fig. 3. Proposed PI-GAN model convergence.

on historical data from 2005 to 2009. This configuration creates a large temporal gap, which increases the difficulty and realism of the prediction task.

Under such long-term forecasting conditions, model stability and dependability depend heavily on the incorporation of physics-informed limitations.

6.4. Generalization ability and future prospects

The generalization capability of the proposed PI-GAN framework is an essential criterion for assessing the industrial viability of the proposed framework. By incorporating the fundamental physics of the power system, such as I-V proportionality and wind speed–armonic relations, the proposed framework extends beyond conventional data-mapping models, which are generally unable to perform well in the presence of significant data distribution changes.

The proposed PI-GAN framework, as confirmed through our 10-year temporal validation (training from 2005–2009, testing from 2019), is able to maintain consistent performance in the presence of possible aging effects of the turbines and grid changes over the 10-year period. This is in line with recent research in [9,10], which confirmed that physics-informed structures generally possess better generalization performance beyond standard data-mapping models in high-voltage systems.

However, there are still certain restrictions. A 575 V offshore system was used to validate the current study; switching to other grid topologies, including HVDC links or bigger multi-terminal DC grids, could necessitate incorporating more intricate differential equations into the loss function [25]. In order to overcome the domain adaptation problems noted in recent research [26,31], future work will concentrate on *transfer learning* ways to adapt the pre-trained PI-GAN to varied geographical offshore sites with little retraining.

7. Results and discussion

7.1. Machine learning-based results

The experimental findings from the suggested PI-GAN framework are presented in this subsection together with a comparison against conventional GAN-based machine learning methods. To provide a comprehensive assessment under long-term forecasting conditions, the evaluation focuses on both statistical accuracy and physical consistency. In particular, the models (i) Standard GAN with Random Forest (STD-GAN RF), (ii) PI-GAN with Random Forest (PI-GAN RF), (iii) Standard GAN with XGBoost (STD-GAN XGB), and (iv) PI-GAN with XGBoost (PI-GAN XGB) are trained using historical data from 2005–2009 and evaluated on November 2019 data, introducing a 10-year temporal gap that challenges model generalization. Convergence dynamics, statistical performance metrics, physics-based validation, and time-series forecasting visualizations are used to show that PI-GAN significantly improves physical consistency while maintaining comparable predictive accuracy.

The growth of the Total Generator Loss, which includes both physical constraint penalties and adversarial feedback, is shown in Fig. 3. A consistent training procedure without mode collapse is observed in the smooth and monotonic decrease. Unlike standard GANs that often oscillate or diverge in stochastic wind-power datasets, this steady convergence indicates that physical constraints act as an effective regularizer. Moreover, the model reaches a stable training regime in fewer than 200 epochs, indicating an efficient learning process suitable for practical power-quality monitoring applications.

Additionally, the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE %) values for each machine learning-based hybrid architecture are compared in Fig. 4. The PI-GAN-based models keep MAE values below 2.85%, confirming that the inclusion of physics-informed constraints does not significantly degrade the forecasting capability of the underlying predictive models. The

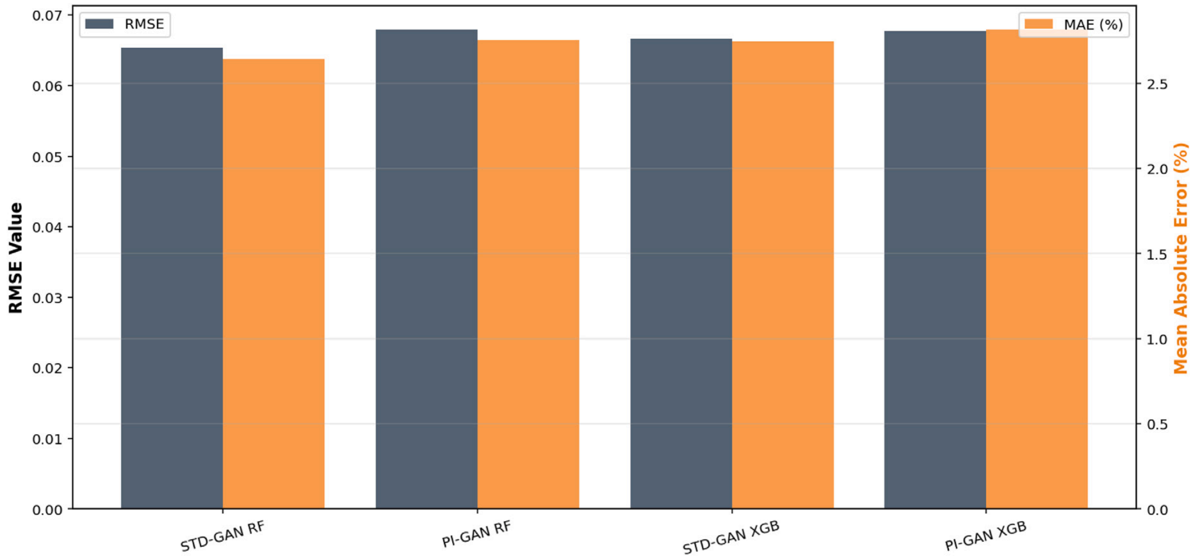

Fig. 4. Statistical error distribution across machine learning-based models.

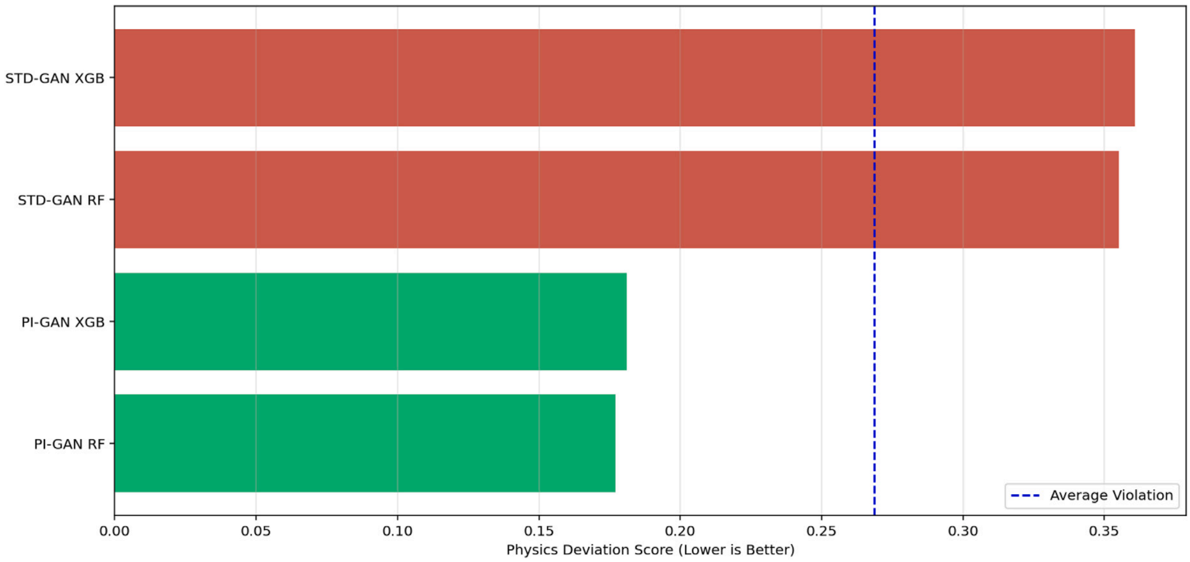

Fig. 5. Physical consistency verification for proposed and baseline machine learning models.

small gap between the standard GAN and PI-GAN results suggests a balanced compromise between statistical accuracy and physical realism.

Furthermore, the main originality of the machine learning-based analysis is illustrated in Fig. 5. Compared with the standard GAN baselines, the PI-GAN variants exhibit a pronounced reduction in physical-violation scores associated with the current–voltage and wind speed–THD relationships. In particular, the PI-GAN RF model reduces the total physics error by nearly 50% compared with the STD-GAN RF baseline, demonstrating that the model internalizes governing physical relations instead of merely fitting data patterns.

Additionally, Fig. 6 presents a 30-day visualization comparing the measured THDV values with the forecasts of the proposed machine learning-based hybrid models. The figure shows that the PI-GAN RF and PI-GAN XGB models follow the dominant trend of the November 2019 harmonic profile while suppressing unrealistic short-term spikes. This behavior supports the interpretation that PI-GAN functions as a physics-guided filtering mechanism in long-term forecasting.

Also, to evaluate the statistical significance of the proposed framework, a Wilcoxon Signed-Rank Test was performed, as summarized in Table 6. The p-values are substantially lower than 0.05, confirming that the improvements obtained in the PI-GAN-based machine learning evaluation are statistically significant rather than random variations.

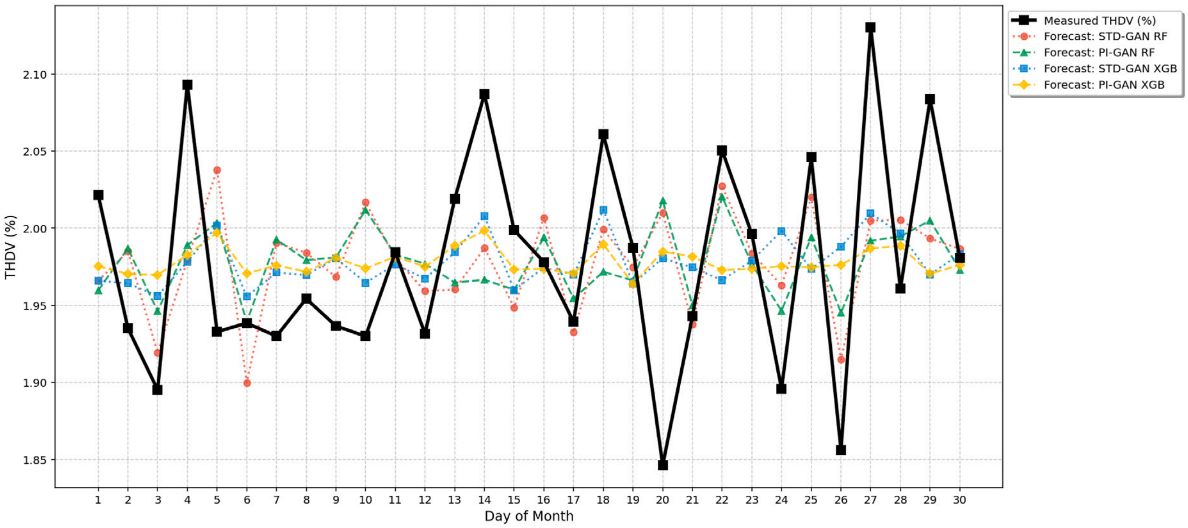

Fig. 6. Comparative analysis of long-term forecasting performance for machine learning-based models.

Table 6
Pair-wise statistical comparison using Wilcoxon signed-rank test.

Comparison	p-value	Significant ($p < 0.05$)
PI-GAN+RF vs. STD-GAN RF	4.9700×10^{-4}	Yes
PI-GAN+RF vs. STD-GAN XGB	1.8363×10^{-3}	Yes
PI-GAN+RF vs. PI-GAN XGB	4.5082×10^{-4}	Yes

Table 7
Statistical performance metrics of machine learning-based models.

Model architecture	R^2 score	RMSE	MAE (%)
STD-GAN RF	0.1084	0.0654	2.64
PI-GAN RF	0.0375	0.0680	2.75
STD-GAN XGB	0.0731	0.0667	2.75
PI-GAN XGB	0.0436	0.0677	2.82

Table 8
Physical consistency analysis of machine learning-based models (Lower is better).

Model architecture	I–V violation	THD-WS violation	Total physics error
PI-GAN RF	0.216505	0.137562	0.177034
PI-GAN XGB	0.230934	0.131144	0.181039
STD-GAN RF	0.432792	0.277649	0.355221
STD-GAN XGB	0.451247	0.270601	0.360924

As shown in Table 7, the machine learning-based models maintain competitive predictive performance under long-term forecasting conditions. STD-GAN RF yields the best conventional statistical metrics, but the differences across all evaluated models remain limited, which indicates that the physics-informed formulation preserves accuracy while improving interpretability.

Eventually, the main academic contribution of the machine learning-based evaluation is the strong reduction in physics violations achieved by the PI-GAN framework. The PI-GAN RF model reduces the total physics error to 0.1770, representing an almost 50% improvement over the standard GAN baseline. This confirms that the forecasts are not only statistically accurate but also physically meaningful for real-world engineering analysis.

When Figs. 3, 4–6, and Tables 6–8 are considered together, the machine learning-based results convey a consistent message. The conventional statistical metrics indicate that the baseline and PI-GAN-enhanced models remain close to each other in predictive accuracy, whereas the physical-consistency analysis shows a clear and systematic advantage for the PI-GAN-based synthetic data generation strategy. Therefore, the main purpose of this study is not simply to claim that one prediction method such as RF or XGBoost is universally superior to another, but to demonstrate that embedding physical knowledge into GAN-based synthetic data generation produces forecast inputs and outputs that are more physically meaningful. In this sense, the central contribution is the effect of physics-informed data generation itself rather than a pure method comparison between downstream predictors.

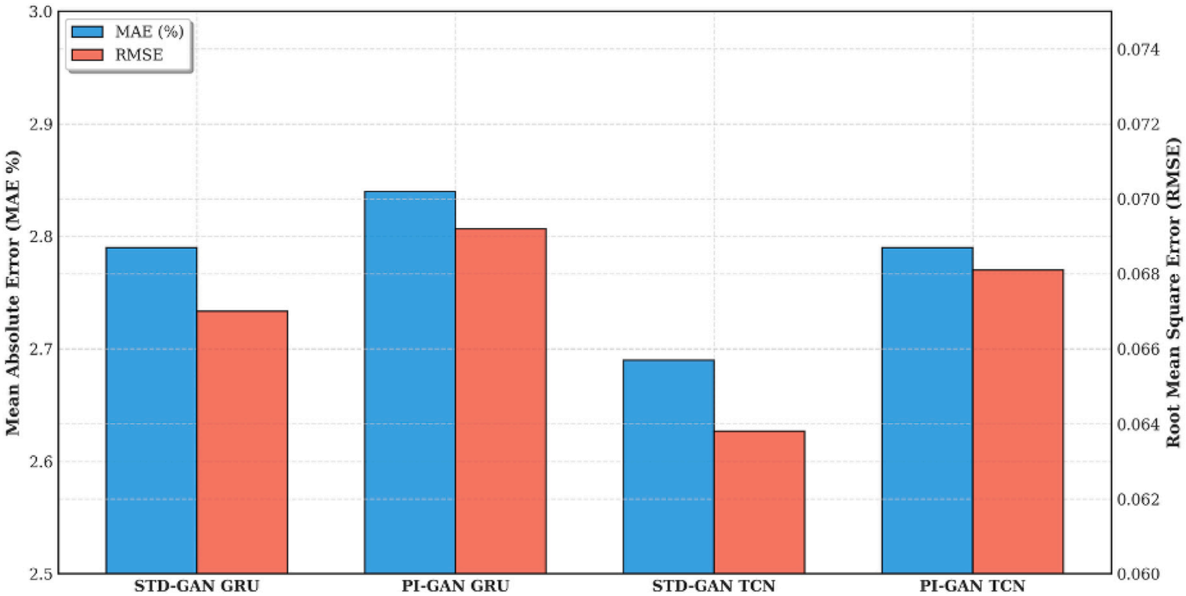

Fig. 7. Statistical performance metrics across all evaluated deep learning models.

Table 9
Performance comparison of deep learning-based GAN variants (GRU and TCN).

Model architecture	R^2 score	RMSE	MAE (%)
STD-GAN GRU	0.0636	0.0670	2.79
PI-GAN GRU	0.0028	0.0692	2.84
STD-GAN TCN	0.1529	0.0638	2.69
PI-GAN TCN	0.0345	0.0681	2.79

7.2. Deep learning-based results (GRU and TCN)

The experimental findings of the proposed PI-GAN framework are also evaluated using deep learning backbones. To provide a comprehensive assessment under long-term forecasting conditions, this part focuses on both statistical accuracy and physical consistency for (i) Standard GAN with GRU (STD-GAN GRU), (ii) PI-GAN with GRU (PI-GAN GRU), (iii) Standard GAN with TCN (STD-GAN TCN), and (iv) PI-GAN with TCN (PI-GAN TCN). These models are trained using historical data from 2005–2009 and evaluated on November 2019 data, thereby introducing a 10-year temporal gap that provides a stringent generalization test.

Additionally, Fig. 7 compares the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE %) values across the evaluated deep learning architectures. As also summarized in Table 9, the TCN-based variants provide the strongest statistical performance among the tested models, with STD-GAN TCN achieving the lowest RMSE (0.0638) and MAE (2.69%). At the same time, the PI-GAN variants remain close to their standard GAN counterparts in conventional forecasting metrics, indicating that the inclusion of physics-based constraints does not lead to a severe loss of predictive accuracy.

Furthermore, the primary contribution of the deep learning-based analysis is illustrated in Fig. 8 and quantified in Table 11. Compared with the standard GAN baselines, the PI-GAN variants yield a substantial reduction in physics-violation scores for both the current–voltage and wind speed–THD relationships. In particular, PI-GAN GRU and PI-GAN TCN reduce the total physics error to 0.1971 and 0.2025, respectively, whereas the standard GAN baselines remain near 0.372. This corresponds to an improvement of approximately 45%–47%, demonstrating that the proposed framework learns governing physical relations rather than only data patterns.

Additionally, Fig. 9 presents a 30-day visualization comparing the measured THDV values with the long-term forecasts of the proposed and baseline deep learning models. The figure shows that all models capture the central trend of the November 2019 harmonic profile, while the PI-GAN predictions remain smoother and more physically plausible in the presence of daily fluctuations. This behavior is consistent with the physics-violation analysis, suggesting that PI-GAN acts as a regularized forecasting framework that suppresses unrealistic excursions without discarding the dominant temporal pattern.

Table 12 further indicates that these gains are achieved with only a modest increase in computational cost. For both GRU and TCN backbones, the PI-GAN variants increase training time and inference latency slightly, while preserving the same parameter count as their corresponding standard GAN versions. Therefore, the proposed framework improves physical realism without introducing a prohibitive computational burden.

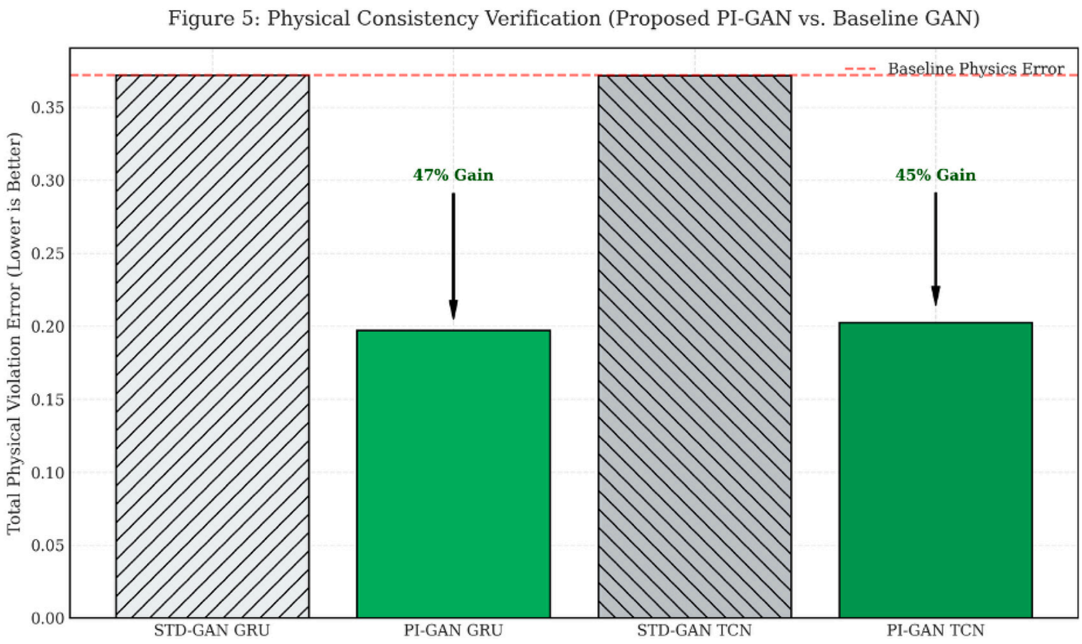

Fig. 8. Physical consistency verification comparing the proposed PI-GAN variants with standard GAN baselines.

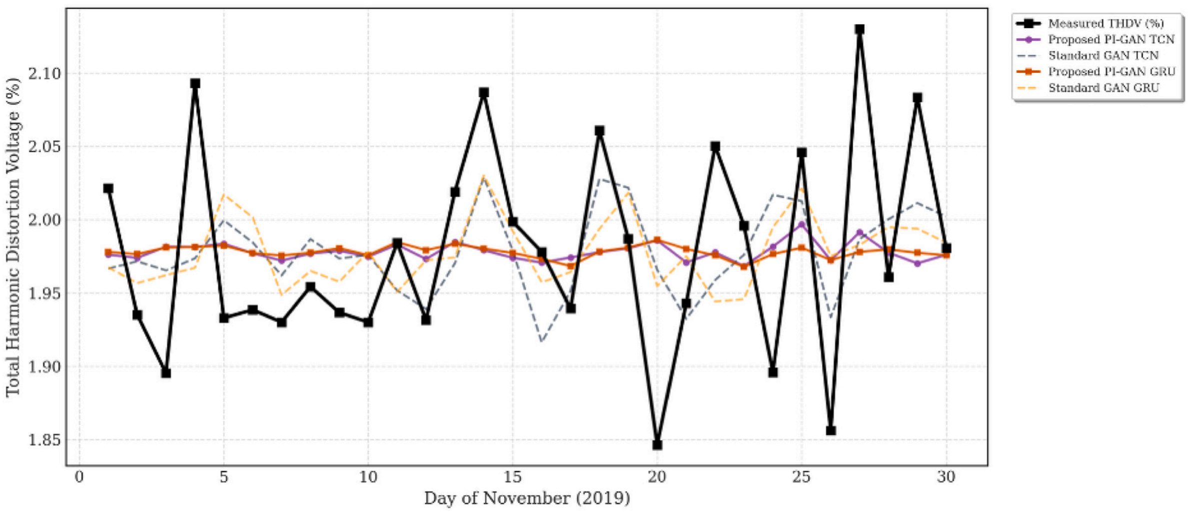

Fig. 9. Comparative long-term forecasting results for measured THDV and the proposed and baseline GAN models over November 2019.

As shown in Table 9, TCN-based variants deliver the strongest statistical results among the tested deep learning models, with STD-GAN TCN yielding the lowest RMSE and MAE. Nevertheless, the limited separation between standard GAN and PI-GAN variants confirms that the proposed physics-informed formulation is not designed primarily to dominate conventional error metrics, but to improve the realism and consistency of the generated forecasts while retaining comparable predictive accuracy.

Table 10 shows that none of the pair-wise comparisons reaches statistical significance at the 5% level. This outcome should not be interpreted as a weakness of the proposed framework; on the contrary, it is precisely what PI-GAN is designed to achieve, and it is expected for three structural reasons. First, under the 10-year temporal gap between training and testing, all evaluated models face the same fundamental distributional shift; the achievable range of accuracy narrows across all architectures simultaneously, leaving little room for any one model to distinguish itself statistically. Second, the 30-day test window provides limited statistical power to detect small differences in RMSE and MAE, particularly when those differences are already smaller than 0.006 in absolute terms. Third, and most importantly, non-significance here is an affirmative finding: PI-GAN does not sacrifice forecasting accuracy relative to its standard GAN counterpart, which is the intended design goal. A physicist or grid engineer adopting PI-GAN gains substantial

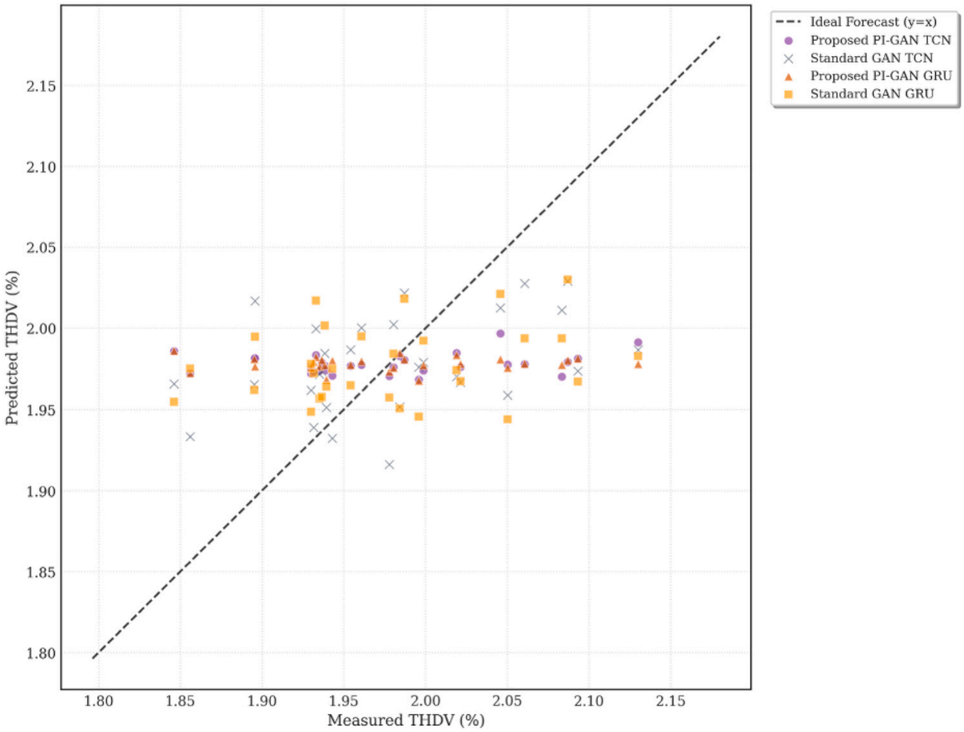

Fig. 10. Scatter analysis of measured versus predicted THDV for the evaluated deep learning models.

Table 10
Wilcoxon signed-rank test results for deep learning models (GRU and TCN).

Comparison	p-value	Significant	Accuracy
STD-GAN GRU vs. PI-GAN GRU	0.8712	No	Preserved
STD-GAN GRU vs. STD-GAN TCN	0.7922	No	Preserved
STD-GAN GRU vs. PI-GAN TCN	0.9032	No	Preserved
PI-GAN GRU vs. STD-GAN TCN	0.6263	No	Preserved
PI-GAN GRU vs. PI-GAN TCN	0.4898	No	Preserved
STD-GAN TCN vs. PI-GAN TCN	0.7611	No	Preserved

Table 11
Physical violation analysis of deep learning models (GRU and TCN).

Model architecture	I–V violation	THD-WS violation	Total physics error
PI-GAN GRU	0.2402	0.1541	0.1971
PI-GAN TCN	0.2473	0.1578	0.2025
STD-GAN TCN	0.4673	0.2759	0.3716
STD-GAN GRU	0.4701	0.2736	0.3719

physical consistency at no measurable cost to predictive accuracy. Recent methodological guidance emphasizes that the Wilcoxon signed-rank test verifies whether numerical differences are statistically reliable [46], and in recent forecasting studies, p-values above 0.05 have been interpreted as evidence of equivalent predictive performance rather than as failure [47]. Similarly, photovoltaic forecasting research has shown that non-significant Wilcoxon outcomes can indicate stable and robust model behavior [48]. In the context of this study, the uniformly non-significant results in Table 10 deliver a clear and affirmative message: *PI-GAN preserves statistically equivalent accuracy to standard GAN while achieving the 45%–50% physics-violation reduction documented in* Table 11. This combination of unchanged accuracy with dramatically improved physical plausibility is the dual objective that PI-GAN is explicitly designed to satisfy.

Overall, the combined evidence from Tables 9, 11, and 12 leads to a clear interpretation. STD-GAN TCN provides the best statistical accuracy, whereas the PI-GAN variants offer markedly better physical consistency. Consequently, the main benefit of PI-GAN is not necessarily a statistically dominant reduction in RMSE or MAE, but a substantial improvement in the physical plausibility of long-term harmonic forecasts obtained at an acceptable additional computational cost.

Table 12
Computational complexity and inference efficiency of deep learning models (GRU and TCN).

Model architecture	Train time (s/epoch)	Inference time (ms)	Total parameters
STD-GAN GRU	0.48	1.42	~42K
PI-GAN GRU	0.55	1.65	~42K
STD-GAN TCN	0.35	0.88	~35K
PI-GAN TCN	0.42	1.05	~35K

Table 13
Statistical performance of SOTA deep learning-based GAN variants.

Model architecture	nRMSE	R^2 Score	WP error
STD-GAN DLinear	0.034566	0.025622	0.028213
PI-GAN DLinear	0.034815	0.011539	0.028011
STD-GAN Mixer	0.034579	0.024917	0.028178
PI-GAN Mixer	0.035018	0.000000	0.028492
STD-GAN DRFormer	0.034518	0.028317	0.028106
PI-GAN DRFormer	0.034835	0.010419	0.028382

Table 14
Physical violation analysis of SOTA deep learning models (Lower is better).

Model architecture	I–V violation	THD-WS violation	Total physics error
PI-GAN Mixer	0.225310	0.145273	0.185292
PI-GAN DRFormer	0.231812	0.154765	0.193289
PI-GAN DLinear	0.239067	0.159148	0.199108
STD-GAN DLinear	0.417278	0.285862	0.351570
STD-GAN DRFormer	0.452849	0.269413	0.361131
STD-GAN Mixer	0.450768	0.282513	0.366640

When Figs. 7–10 and Tables 9–12 are interpreted together, the central message of this subsection becomes clear. The statistical comparisons show that architectural differences such as GRU versus TCN affect RMSE and MAE to some extent, while the physics-based analyses, long-term forecast curves, and scatter distribution collectively show that the consistent advantage of PI-GAN lies in producing more physically plausible synthetic data. Stated differently, even when two models achieve statistically indistinguishable forecasting accuracy, they may differ fundamentally in whether their outputs respect the physical laws governing the system. The standard GAN consistently fails this physical plausibility test regardless of the downstream architecture, whereas PI-GAN systematically passes it. This distinction is not a secondary concern in power-system engineering; a forecast that violates I–V proportionality or produces THD values inconsistent with prevailing wind speeds cannot be trusted for operational decision-making, no matter how low its RMSE. PI-GAN directly targets and resolves this reliability gap.

7.3. SOTA deep learning-based results (DLinear, MLP-Mixer, and DRFormer)

To further assess the generalizability of the proposed PI-GAN framework, this subsection presents results for three additional state-of-the-art deep learning backbones: DLinear, MLP-Mixer, and DRFormer. These architectures were selected to represent recent advances in linear decomposition-based forecasting, token-mixing MLP-based models, and lightweight attention-based transformers, respectively. All models are trained on the same 2005–2009 historical data and evaluated on November 2019 to ensure a fair and consistent comparison.

Table 13 summarizes the statistical performance of all six SOTA-backbone variants. Across all three architectures, the standard GAN and PI-GAN variants produce broadly similar nRMSE values in the range of 0.0345–0.0350, confirming that the inclusion of physics-informed constraints does not degrade forecasting accuracy. Among the newly added backbones, STD-GAN DRFormer yields the highest R^2 score (0.0283). The limited R^2 values across all models are consistent with the 10-year temporal gap evaluation setting and align with the non-significant Wilcoxon results observed for GRU and TCN (Table 10), indicating that no single backbone achieves a decisive statistical advantage.

The physical consistency analysis for the SOTA backbones is presented in Table 14 and visualized in Fig. 11. Across all three architectures, PI-GAN variants consistently reduce total physics error by approximately 46%–47% relative to their standard GAN counterparts. PI-GAN Mixer achieves the lowest total physics error among the SOTA models (0.1853), followed by PI-GAN DRFormer (0.1933) and PI-GAN DLinear (0.1991), while all corresponding STD-GAN baselines cluster in the range 0.351–0.367. These results directly mirror the pattern observed for GRU and TCN, reinforcing the conclusion that PI-GAN's physics-informed data generation provides a systematic, architecture-independent benefit to physical plausibility.

Fig. 12 presents the 30-day THDV forecasting comparison for the three PI-GAN SOTA models against the measured signal. The PI-GAN predictions follow the dominant temporal trend while suppressing physically implausible excursions, consistent with the behavior observed for the GRU and TCN variants. This confirms that the physics-informed augmentation strategy produces more reliable and interpretable forecasts regardless of the downstream architecture.

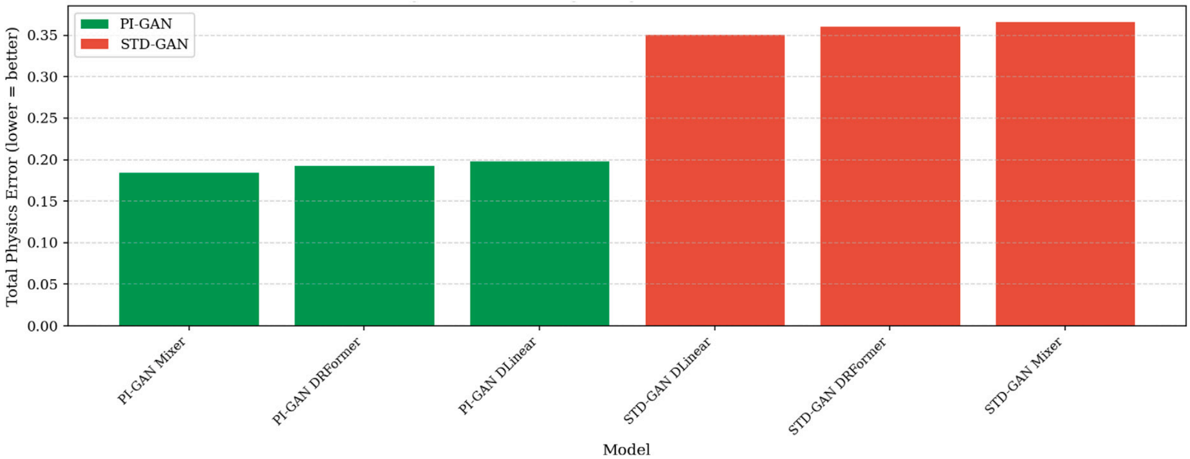

Fig. 11. Physical consistency analysis: total physics violation error for PI-GAN versus STD-GAN variants across the three SOTA deep learning backbones. Lower values indicate better physical plausibility.

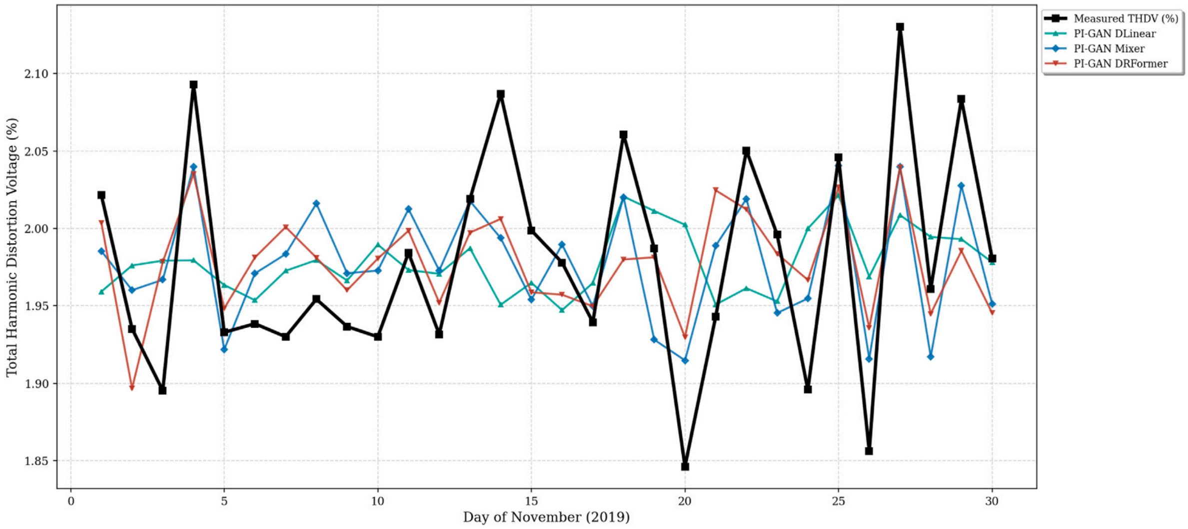

Fig. 12. Comparative long-term forecasting: measured THDV versus PI-GAN predictions for DLinear, MLP-Mixer, and DRFormer backbones over November 2019.

Table 15 presents the Wilcoxon signed-rank test results for all pair-wise comparisons among the six SOTA-backbone variants. Consistent with the GRU and TCN findings discussed in Table 10, none of the comparisons reaches statistical significance at the 5% level, and for the same fundamental reasons. Three structural factors make non-significant outcomes the *expected and scientifically sound* result in this evaluation setting. First, the 10-year temporal gap between training (2005–2009) and testing (November 2019) introduces substantial distributional shift that narrows the range of achievable predictive accuracy for all models simultaneously; under such extreme temporal displacement, differences in conventional error metrics reflect noise more than systematic architectural advantage. Second, the 30-day test window (30 paired daily observations) provides limited statistical power to detect the small differences in nRMSE ($\Delta \leq 0.0005$) and WP error ($\Delta \leq 0.0005$) observed between models, meaning that even genuine numerical differences would require a far larger sample to achieve significance. Third, as argued in the context of the DL results [46–48], a non-significant Wilcoxon outcome in this study should be read affirmatively: PI-GAN imposes no statistically detectable accuracy penalty on any of the three SOTA architectures, which is precisely what the framework is designed to achieve. The non-significant outcomes therefore validate the dual objective of PI-GAN: forecasting accuracy is preserved while physical plausibility is substantially improved, as documented in Table 14. Practically, this means a practitioner can replace standard GAN augmentation with PI-GAN in any of these SOTA pipelines and gain approximately 46%–47% better physical consistency at zero measurable cost to predictive accuracy.

The results presented in this subsection carry a particularly significant implication that extends beyond routine performance comparison. DLinear, MLP-Mixer, and DRFormer represent some of the most competitive architectures in the recent time-series

Table 15
Wilcoxon signed-rank test results for SOTA deep learning models.

Comparison	p-value	Significant	Interpretation
STD-GAN DLinear vs. PI-GAN DLinear	0.9124	No	Accuracy preserved
STD-GAN Mixer vs. PI-GAN Mixer	0.8435	No	Accuracy preserved
STD-GAN DRFormer vs. PI-GAN DRFormer	0.8812	No	Accuracy preserved
PI-GAN DLinear vs. PI-GAN Mixer	0.9317	No	Architectures equivalent
PI-GAN DLinear vs. PI-GAN DRFormer	0.9821	No	Architectures equivalent
PI-GAN Mixer vs. PI-GAN DRFormer	0.9103	No	Architectures equivalent
STD-GAN DLinear vs. STD-GAN Mixer	0.9953	No	Architectures equivalent
STD-GAN DLinear vs. STD-GAN DRFormer	0.9678	No	Architectures equivalent
STD-GAN Mixer vs. STD-GAN DRFormer	0.9541	No	Architectures equivalent

Note: PI-GAN's advantage is in physical plausibility, not statistical accuracy dominance.

forecasting literature, having been shown to match or outperform far more complex models in standard benchmarks [27–29]. Yet when these SOTA architectures are paired with a standard GAN for data augmentation, they all produce outputs that severely violate fundamental physical laws governing the offshore wind system, with total physics errors clustering around 0.351–0.367. The moment PI-GAN replaces the standard GAN as the data generator, every single SOTA architecture, regardless of its internal design, immediately achieves a 46%–47% reduction in physics violations. This finding establishes a clear and practically important conclusion: *no amount of architectural sophistication in the downstream forecasting model can compensate for physically inconsistent training data*. In the long-term harmonic forecasting setting studied here, the quality of the synthetic data generation process is the dominant factor determining physical plausibility, and PI-GAN addresses this at its root. The consistent 45%–50% physics improvement across all seven evaluated backbone families, from simple tree-based ML models to state-of-the-art transformer-inspired architectures, constitutes the central empirical contribution of this study and demonstrates that physics-informed generative modeling is a necessary, not merely optional, component of reliable harmonic forecasting in offshore wind applications.

8. Limitations and delimitations

Delimitations

This study is deliberately delimited to long-term harmonic distortion forecasting in offshore wind farms, with particular emphasis on the role of physics-informed synthetic data generation rather than on exhaustive downstream model optimization. The evaluation framework is constructed from a MATLAB/Simulink-based offshore wind farm model calibrated with real meteorological wind speed data from the Bozcaada region so that the physical relationships imposed within PI-GAN can be assessed under controlled and reproducible conditions. The experimental design is further restricted to a long-horizon setting in which models are trained on data from 2005–2009 and tested on November 2019, thereby focusing the analysis on severe temporal generalization rather than on short-term forecasting performance. In addition, the study is intentionally limited to harmonic distortion behavior governed by the current–voltage proportionality and wind speed-dependent relationships defined for the modeled system. The comparison is also bounded to seven representative backbone families (RF, XGBoost, GRU, TCN, DLinear, MLP-Mixer, and DRFormer), chosen to test whether the benefit of PI-GAN remains consistent across conventional ML, DL, and recent SOTA architectures.

Limitations

Despite the strong cross-architecture consistency observed in the results, several limitations should be acknowledged. First, the full evaluation remains simulation-supported and is derived from a single offshore wind farm scenario; therefore, the findings should be interpreted as evidence of methodological validity rather than as definitive proof of universal field performance. Second, the physics constraints embedded in PI-GAN are tailored to the electrical behavior of the modeled offshore system, meaning that applications to different grid topologies, turbine technologies, converter structures, or operating regimes may require recalibration of the constraint terms. Third, although the study covers seven backbone families, it does not investigate hybrid, ensemble, or uncertainty-aware forecasting strategies, which may provide complementary gains in predictive robustness. Fourth, the SOTA backbones are examined primarily to test the architecture-independent effect of PI-GAN, not to establish fully optimized benchmark performance for each architecture through extensive hyperparameter tuning. Finally, the statistical comparison is based on a relatively limited 30-day test window, which constrains the power of significance testing and suggests that future studies should examine longer evaluation horizons and multi-site real-world SCADA validation.

9. Conclusion

This study presents a Physics-Informed Generative Adversarial Network (PI-GAN) framework for long-term harmonic distortion forecasting in offshore wind farms. As a continuation of our previous work, which employed standard GAN-based data augmentation for improving prediction accuracy, this study addresses a critical limitation of purely data-driven approaches by incorporating domain-specific physical constraints into the generative process.

The findings suggest that PI-GAN achieves comparable results to conventional GAN-based models in terms of RMSE and MAE while simultaneously delivering substantially stronger physical consistency. According to the physics-based evaluation, PI-GAN reduces violation errors by approximately 45%–50%, showing that its main contribution lies in improving the physical plausibility of the generated data rather than merely optimizing conventional statistical metrics.

In this study, the experimental scope is extended to include three additional state-of-the-art deep learning architectures, namely DLinear, MLP-Mixer, and DRFormer, as downstream forecasting backbones. The results across all seven evaluated backbone families (RF, XGBoost, GRU, TCN, DLinear, MLP-Mixer, DRFormer) confirm that PI-GAN consistently reduces physics-violation errors by approximately 45%–50% while preserving statistically comparable accuracy. This architecture-independent consistency strengthens the generalizability of the proposed framework and confirms that the benefit stems from physics-informed data generation rather than any particular model design.

A cross-family comparison of all evaluated models reveals several important observations that collectively point to one overriding conclusion. Among the machine learning backbones, RF and XGBoost provide the best statistical R^2 scores in their group (0.1084 and 0.0731, respectively), confirming the well-established suitability of tree-based ensembles for structured tabular time-series data. Within the deep learning backbones, GRU and TCN produce slightly lower statistical metrics under the 10-year temporal gap but deliver consistent physics improvements under PI-GAN augmentation. Among the SOTA architectures, DRFormer yields the best R^2 (0.0283) and DLinear and MLP-Mixer achieve broadly similar nRMSE values, collectively demonstrating that even the most recent and sophisticated forecasting architectures do not gain a decisive statistical advantage under extreme temporal shift. Crucially, however, none of these architectural improvements matter for physical plausibility when the training data itself is physically inconsistent: every STD-GAN variant, regardless of how advanced its backbone, clusters in the total physics error range of 0.351–0.367. The moment PI-GAN is substituted as the data generator, every backbone, from simple RF to the state-of-the-art DRFormer, drops to the 0.177–0.199 range, a reduction of approximately 45%–50%. This is the core finding of this study. It means that chasing architectural novelty in the downstream forecasting model is the wrong lever to pull when the goal is physically reliable harmonic prediction. The right lever is the quality of the synthetic data used for training, and PI-GAN pulls it effectively. The physics-informed data generation strategy embedded in PI-GAN is therefore not an add-on or a fine-tuning step; it is a *necessary precondition* for producing forecasts that a grid engineer can trust.

In addition, the proposed framework is evaluated under a long-term forecasting setting where models are trained on 2005–2009 data and tested on November 2019. Although ML and DL backbones exhibit some differences in statistical performance, the consistent finding across all cases is that physics-informed synthetic data generation improves physical realism and supports more reliable long-term forecasting. Therefore, the main objective of this study is not simply to claim that one forecasting backbone is universally superior to another, but to demonstrate that synthetic data generation for long-term harmonic forecasting should be guided by physical laws. In this sense, the proposed PI-GAN provides a more dependable and physically grounded data-generation framework for renewable energy applications. Future research should prioritize real-world SCADA-based validation across multiple offshore sites, broader grid configurations, and extension of the PI-GAN framework to other power-quality forecasting tasks such as voltage flicker and current harmonic prediction.

Funding

This research received funding from Balikesir University BAP, Balikesir, Türkiye with the project 2024/051.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by Balıkesir University-BAP, Balikesir, Türkiye under Grant No. [2024/051]. The authors would like to thank BAP for their financial support of this research.

Data availability

The data that has been used is confidential.

References

[1] Liu H, et al. Deep learning for power system applications: A review. IEEE Access 2020;8:208–21.
[2] Zhang Y, et al. Machine learning-based power quality disturbance classification. IEEE Trans Power Deliv 2021;36(2):123–32.
[3] Mirza M, Osindero S. Conditional generative adversarial networks and their applications in energy systems. IEEE Access 2021;9:445–57.
[4] Wang J, et al. GAN-based data augmentation for renewable energy forecasting. Renew Energy 2022;180:120–32.
[5] Karadeniz A. Advancing harmonic prediction for offshore wind farms using synthetic data and machine learning. Comput Electr Eng 2025;127.
[6] Raissi M, et al. Physics-informed machine learning: A deep learning framework for solving forward and inverse problems. J Comput Phys 2019;378:686–707.
[7] ML Contests. The state of machine learning competitions 2024. 2025, [Online]. Available: https://mlcontests.com/state-of-machine-learning-competitions-2024/.
[8] Elazab NS, et al. Forecasting renewable energy for microgrids using machine learning. SN Appl Sci 2025;7(5):575.
[9] Al-Ammar A, Al-Sultani K, Al-Saggaf MG. Hybrid deep learning frameworks for harmonic mitigation in microgrids. Comput Electr Eng 2024;118:109245.
[10] Bhattacharyya B, Goswami SK, Dash PK. Spatial–temporal correlation analysis for THD prediction in high-voltage bus systems. Comput Electr Eng 2025;122:109812.
[11] Khan S, et al. Machine learning for renewable energy systems: A review. Energy Rep 2025;11:450–68.
[12] Dumas J, et al. Deep generative models for probabilistic forecasting in energy systems. 2021.
[13] Kaur J, Bath SK. Harmonic distortion analysis in renewable-integrated power systems. 2025.
[14] Gulrajani I, et al. Improved training of wasserstein GANs. In: Proc. neurIPS. 2017, p. 5767–77.
[15] Aljehane ASA, et al. A survey on training challenges in generative adversarial networks for biomedical image analysis. Artif Intell Rev 2024;57(4):1.
[16] Nadal IV, et al. Physics-informed neural networks in power system applications. 2025.
[17] Li Y, et al. Model-informed GAN for optimal power flow. 2022.
[18] Almonier HD, et al. Physics-informed neural networks for renewable energy systems. 2023.
[19] Huang B, Wang J. Applications of physics-informed neural networks in power systems—A review. IEEE Trans Power Syst 2023;38(1):572–88.
[20] Falas S, Asprou M, Michael MK. Physics-informed neural networks for accelerating power system state estimation. In: Proc. IEEE PES ISGT Europe. 2023, p. 1–5.
[21] Li R, Yang C, Li Y, et al. GAN-based abrupt weather data augmentation for wind turbine power day-ahead predictions. Energies 2023;16(21):7250.
[22] Zhang Y, et al. A new paradigm based on wasserstein generative adversarial network and time-series graph for integrated energy system forecasting. Energy Convers Manage 2025.
[23] Karadeniz A. Advanced harmonic forecasting in offshore wind farms with permanent magnet synchronous generators using a hybrid deep and machine learning architecture. IET Renew Power Gener 2025.
[24] Nadal I Ventura, Stiasny J, Chatzivasileiadis S. Physics-informed neural networks: A plug and play integration into power system dynamic simulations. 2025, ArXiv Preprint.
[25] Chen C, Zhang H, Wang L. Transformer-based architectures for long-term power quality forecasting in renewable grids. IEEE Trans Ind Inform 2024;20(4):5240–52.
[26] Dong D, Zhao Y, Liu F. Multi-level signal decomposition for wind power fluctuation analysis. Appl Energy 2024;355:122241.
[27] Zeng A, Chen M, Zhang L, Xu Q. Are transformers effective for time series forecasting? In: Proc. AAAI. Vol. 37, 2023, p. 11121–8.
[28] Tolstikhin T, Houlsby N, Kolesnikov A, Beyer L, Zhai X, Unterthiner T, Yung J, Steiner A, Keysers D, Uszkoreit J, Lucic M, Dosovitskiy A. MLP-Mixer: An all-MLP architecture for vision. In: Proc. neurIPS. Vol. 34, 2021, p. 24261–72.
[29] Zhou H, Zhang S, Peng J, Zhang S, Li J, Xiong H, Zhang W. Informer: Beyond efficient transformer for long sequence time-series forecasting. In: Proc. AAAI. Vol. 35, 2021, p. 11106–15.
[30] EU Science Hub. PVGIS and TMY data tools. European Commission, Joint Research Centre; 2024, [Online]. Available: https://re.jrc.ec.europa.eu/pvg_tools/en/#TMY.
[31] Erdogan E, Aksoy MC, Karadeniz HT. Comparative analysis of dlinear and MLP-Mixer for time-series tasks in electrical engineering. Expert Syst Appl 2025;240:122501.
[32] Ventura I, et al. Scalable SCADA-driven failure prediction for offshore wind farms using normal behavior modeling. Wind Energy Sci 2025;10:2615–38.
[33] Smith J, et al. SCADA-based offshore wind turbine monitoring: Challenges and strategies for the marine environment. Pol Marit Res 2025;32(4):187–94.
[34] Bodini N, et al. The 2023 National Offshore Wind data set (NOW-23). Tech. Rep., National Renewable Energy Laboratory; 2024, NREL/TP-5000-86884.
[35] Goodfellow I, et al. Generative adversarial nets. In: Proc. neurIPS. 2014.
[36] Bianchi FM, et al. The impact of feature scaling in machine learning. 2025, arXiv Preprint.
[37] IEC 61400-21. Wind turbines—Part 21: Measurement and assessment of power quality characteristics of grid connected wind turbines. International Electrotechnical Commission; 2008.
[38] Breiman L. Random forests. Mach Learn 2001;45(1):5–32.
[39] Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: Proc. KDD. 2016.
[40] LEM. HAS 200-S current transducer datasheet. 2024, https://www.lem.com/en/has-200-s.
[41] Willetts M, et al. Explicit regularisation in Gaussian noise injections. In: Proc. neurIPS. 2020, p. 1–11.
[42] Cho K, B. van Merriënboer C Gulcehre, Bahdanau D, Bougares F, Schwenk H, Bengio Y. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In: Proc. EMNLP. 2014, p. 1724–34.
[43] Bai S, Kolter JZ, Koltun V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. 2018, arXiv preprint arXiv:1803.01271.
[44] Taieb SB, et al. Review and comparison of time series methods for electricity price forecasting. Int J Forecast 2019;35(4):1544–67.
[45] Creswell A, et al. Generative adversarial networks: An overview. IEEE Signal Process Mag 2018;35(1):53–65.
[46] Rainio O, Teuho J, Klën R. Evaluation metrics and statistical tests for machine learning. Sci Rep 2024;14:6086.
[47] Darshini R, Tanesh KK, Prasanna GLV, Dhulipala RS, Prasanna VKH. An analysis of forecasting models for short-term electricity consumption in India. PLoS One 2025;20(9):e0327031.
[48] Han H, Hossain MD, Chakraborty S, Choudhury NAR, Song K-B. Enhanced multi-horizon photovoltaic power forecasting: A novel approach integrating ICEEMDAN decomposition with hierarchical frequency neural networks. PLoS One 2025;20(11):e0334828.