Journal of Computer Science and Technology Studies
ISSN: 2709-104X
DOI: 10.32996/jcsts
Journal Homepage: www.al-kindipublisher.com/index.php/jcsts

JCSTS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

AI-Driven Predictive Modeling for Solar Power Generation Using Real-Time Photovoltaic Sensor Data

Rayhanul Islam Sony[1], Md Ariful Islam Bhuiyan[2] and Dipta Roy[3]
[1]*Trine University, 1 University Ave, Angola, IN 46703, USA, rsony23@my.trine.edu*
[2]*California State University Northridge, 18111 Nordhoff St, Northridge, CA 91330, USA,* md-ariful-islam.bhuiyan.390@my.csun.edu
[3]California State University Northridge, 18111 Nordhoff St, Northridge, CA 91330, USA, *dipta.roy.229@my.csun.edu*
Corresponding Author: Rayhanul Islam Sony, **E-mail**: rsony23@my.trine.edu

| ABSTRACT
Solar PV systems, as part of modern power grids, demand accurate forecasting in real time to address the intermittency issues and ensure a stable grid. Current forecasting models, however, underperform when put into practice because of incomplete sensor coverage and cloud-induced ramp events, lack of cross-site generalization, and lack of quantified uncertainty. In this article, we propose a new end-to-end hybrid network, namely Missingness-Aware Physics-Guided Graph Transformer (MAPGFormer), to tackle these practical challenges. The core pillars of MAPGFormer are: introducing a novel sensor tokenization layer to separate out power, weather, physics, temporal and missingness signals; developing a missingness-aware reconstruction encoder to effectively exploit the structured missing patterns for the robust imputation and forecasting; designing a static-dynamic graph learner to fuse historical similarity with topologies learned by real-time frame attention for reconstruction and inverter/site level spatial modeling; designing a multi-scale temporal Transformer to perform multi-resolutions temporal signal reconstruction and prediction; designing a novel weather-regime Mixture-of-Experts module with specific experts for clear-sky, cloudy/ramp, low-irradiance and missing-sensor conditions; finally designing a probabilistic forecasting head that is able to generate calibrated quantile outputs. The framework was carefully tested on two complementary open Kaggle datasets: 1) Solar Power Generation Data for primary inverter-level development and benchmarking against classical, recurrent and Transformer baselines, and 2) UNISOLAR dataset for external multi-site generalization and transfer learning with validation using Transformer baselines. On the primary data set, MAPGFormer produced remarkable performance in predicting 15-minute electricity demand, achieving an MAE of 12.44 kW, RMSE of 18.09 kW, nRMSE of 2.13%, sMAPE of 4.02%, and R^2 of 0.9827, whereas the Vanilla Transformer gave a MAE of 13.61 kW and reduced the MAE by 18.6%. The accuracy was consistently demonstrated in multi-horizon and cross-plant analyses, and in UNISOLAR leave one out and held out site analysis, the R2s were 0.9791-0.9803. The model was found to have good robustness when missingness was simulated up to 40% and good probabilistic intervals. The contribution of each architectural component was validated using ablation studies and make the model explainable by SHAP analysis. MAPGFormer sets a new standard in practical PV forecasting for uncertainty-aware real-time grid management systems.

| KEYWORDS

MAPGFormer, Photovoltaic power forecasting, Missing data robustness, Graph neural network, Mixture-of-Experts, Multi-scale Transformer, Probabilistic forecasting, Ramp events

| ARTICLE INFORMATION

ACCEPTED: 01 April 2026 **PUBLISHED:** 16 May 2026 **DOI:** 10.32996/jcsts.2026.8.7.2

1. Introduction

There is significant transformation of the energy sector, due to the explosive growth of solar photovoltaic (PV) plants across the globe. Recent reports have showed the growth in PV capacity at the rate of around 32% in 2024, resulting in a share of 6.7% of all electricity produced in the world [1]. While shifting towards renewable sources of energy, countries seek to implement

measures that would decarbonize electricity production and make their grids climate neutral [2]. At the same time, the main attributes of solar energy – its availability and the lack of emissions during production of electricity – lead to significant technical problems. Firstly, solar energy is highly intermittent, and PV power depends on a variety of factors such as variability of sunlight irradiance, dynamics of clouds cover, seasonality and temperature change [3]. Secondly, high levels of penetration result in significant stress on the stability of electricity distribution system, electricity market operations, optimization of energy dispatch and scheduling of reserve capacity [4]. It becomes critical for these problems to find accurate solutions in terms of real-time PV power forecasting.

PV forecasting poses several problems that are distinct from the usual controlled lab or offline experiments conducted at individual sites. At any deployed PV plant, perfect data availability is not a certainty since sensor faults, communication delay, logging errors, and maintenance activities could lead to missing information in a significant portion of the RT inputs. According to previous works reported in the literature, approximately 40% of the PV generated data may have missing data points in a realistic scenario [5]. Besides this, PV generation dynamics are characterized by high nonlinearities, nonstationarities, and multiscale properties, as well as being extremely sensitive to cloud passing, also referred to as ramp events that pose difficulties for modeling during sunny weather conditions [6]. Moreover, many PV plants consist of geographically distributed inverters whose outputs have intricate spatial interdependencies based on common weather and cloud dynamics effects [7]. Another practical requirement is that instead of point predictions, calibrated probabilistic forecasts are needed for operational purposes, like reserve and intraday trading [8].

The rise of the deep learning models such as LSTM, GRU, and TCN (Temporal Convolutional Network) is a major milestone in improving the forecasting performance by capturing temporal dependency patterns that exist already in historical data. Last but not this way least good overflow for time series modeling up to revolutionized Transformer-based approaches; Models such as PatchTST, adopting patch-based tokenization for larger locality of the temporal context [9] and iTransformer, implementing inverted attention to model the multivariate correlation across all time steps for each variable separately [10], have exhibited state-of-the-art performance in bettering prediction quality of long-horizon forecasting or single-variates aggregation tasks in multi-variate time-series. These rely on multi-scale and probabilistic variants of Transformer networks that with temporal, spatial attention, and conformalized quantile regression useful for PV generation modelling [11]. Nevertheless, existing models face several drawbacks with respect to their application on real PV forecasting problems.

First and foremost, nearly all existing PV forecasting studies are validated on a single dataset collected under fairly controlled conditions, resulting in models that may learn dataset-specific behaviors and generalize poorly to new plants or even different geographies or system configurations [12]. Second, the conventional input to any given machine learning algorithm is a concise assignment of heterogeneous weather sensor signals, power generation readings and temporal features to all reside in a flat input vector that treats each of these distinct types of information uniformly despite the fact that those ground truth power observations, meteorological measurements, even physics-derived quantities and missingness indicators have different semantic roles (e.g. [13]). Thirdly, and arguably most importantly with respect to deployment in real time, the vast majority of deep learning forecasters use the assumption that we have all input data at inference time. Naive preprocessing steps such as linear interpolation or forward-fill imputation for missing sensor values result in losing the informative structure of missingness patterns and degrade the reliability of forecasts [13] [14]. Fourth, models that predict every inverter or PV site independently cannot utilize the structural and spatial dependencies of inverters within a plant or geographically proximate sites within a region to improve prediction accuracy [15] [16]. Fifth, while standard Transformer architectures learn an optimization for a single temporal scale, solar irradiance dynamics exist across multiple resolutions at the same time. Clouds can take seconds to tens of minutes to pass overhead creating variations on sub-hourly scales that are continuous with hourly weather patterns and evolution over day cycles spanning solar noon resolution, amongst others [17]. Sixth, although the statistical relationships between inputs and PV output vary fundamentally in clear-sky, overcast, low-irradiance and missing-sensor conditions [18], current models use a single joint forecast for all weather regimes. 7th: The predominantly focus of deterministic point-forecast models provides system operators with no significant uncertainty intervals for risk-aware grid operation and reserve planning [19]. Finally, few studies offer detailed explanations of model decisions which undermines their trustworthiness and deployment in operational settings.

All of these restrictions in unison necessitate an architecture with deep connections to real world limitations when predicting PV, one that is inherently integrated into the modeling process. Competing submitted models referred to the aforementioned openly available, reproducible Kaggle-hosted datasets which supports principled (and transparent) comparison of results and independent validation or ideally community-driven replication — principles increasingly appreciated as being integral to credible applied machine learning research. Our work uses two such datasets playing complementary roles: (1) Solar Power Generation Data — inverter-level data from 2 PV plants in India collected over 34 days, and more generally (2) the UNISOLAR dataset which consists of high-granularity multi-site PV generation and weather records covering each site across 42 sites located at five locations. The former acts as the main development benchmark for inverter-level forecasting, while the latter is an external validation benchmark for cross-site generalization that enables a more rigorous evaluation of transferability than what is typically achieved in existing literature.

To fill the identified voids, we present MAPGFormer (Missingness-Aware Physics-Guided Graph Transformer), a modular hybrid forecasting framework that consists of seven complementary components: focusing on significant sensor types, a sensor tokenization layer has been designed which assigns distinct learned representations for power, weather, physics, temporal and missingness-mask tokens; to not treat incompleteness as mere noise, we design a missingness-aware reconstruction encoder jointly learns to impute the missing sensor values along with forecasting output using structured missingness signals; fusing historical-similarity graphs with real-time attention-derived graphs characterizing inverter-level or site-level spatial dependencies is made possible using static-dynamic graph learner; propagating spatial context across the learned graph topology is also achieved by utilizing a graph attention encoder with interconnected multi-scale Temporal Transformers operating at four distinct resolutions (15 min., 1 h., 3 h., 24 hour) fused via cross-scale attention; finally coupled with this component is an architecture where each sub-network in dedicated-expert-type Mixture-of-Experts (MoE) modules governed by learnable gating network conducts inference given particular weather-regime scenarios such as clear-sky conditions versus cloudy/ramp/moisture-limited pathways through expert branches for low inputs or tree blooming plus abstract space concerns resulting in steady-state transitions yet calibrated sampling generate point estimates alongside probabilistic forecasts of skillful P10 percentile quantiles P50 annotation validation calibrators. The entire framework is trained end-to-end by a composite objective with ramp-weighted losses to enhance sensitivity to abrupt generation discontinuities.

The main contributions of this paper are as follows:

- We propose a sensor-token representation scheme that fully decouples the power, weather, physics, time and missingness information into separate learnable token types so that the model learns to predict which type of information is most predictive under each operating condition.
- We propose a missingness-aware reconstruction encoder that treats a missing-sensor condition as a type of structured input signal rather than pre-processing artifact, which enables real-time robustness under high percentage missing-data rates.
- We propose a static-dynamic graph learning mechanism, which fuses historical similarity graphs with real-time attention derived graph by learnable fusion weight to support inverter-level and PV-site-level spatial dependency modeling.
- We combine a multi-scale temporal Transformer with a Mixture-of-Experts module for weather-regime inference to jointly resolve solar dynamics across multiple resolutions and behavioral variation governed by weather.
- Multi-horizon risk-aware grid operations with probabilistic multi-horizon forecasts e.g., with PICP, PINAW metrics and CRPS based quantile outputs
- We validate this framework on two open Kaggle PV datasets, the Solar Power Generation Data for inverter-level development and UNISOLAR for external multi-site generalization, which shows consistent performance across diverse PV configurations, weather regimes and missing-data conditions.

The rest of this paper is organized as follows. Next, Section 2 summarizes relevant literature in relation to ML/DL-based PV forecasting, Transformer architectures, missing-data methods of interest within PV frameworks, graph-based forecasting and probabilistic techniques driven by both uncertainty modeling and quantification for all types of data. In Section 3 we describe the datasets and preprocessing pipeline. The MAPGFormer architecture is described in detail in Section 4. Section 5 details the experimental setup. Results are presented and analyzed in Section 6. Closing this article, we provide limitations and future directions in Section 7.

2. Literature Review

The application of Artificial Intelligence (AI) has proved to be significant in terms of enhancing the efficiency, reliability, and effectiveness of solar PV power plants. With the aid of AI-based predictive modeling, analysis can be done using real-time PV sensor data for predicting solar power generation, fault detection, energy management, and smart grid interfacing purposes. Over the past few years, several studies have been conducted by integrating machine learning, IoT, deep learning, and predictive analytics approaches to address the issues related to intermittency, maintenance, and energy optimization in solar energy systems.

2.1 AI-Based Solar Power Prediction and Forecasting

AI's role in photovoltaic technology has revolutionized solar efficiency and the management of renewable resources. Energy generation optimization, anomaly detection, and integration with smart grids are some of the applications of AI models like machine learning and predictive analytics. These also support sustainable development of solar infrastructures [20] .A systematic review of predictive models for perovskite and tandem PV systems based on IoT has shown the benefits of using real-time sensor data and AI algorithms for more accurate predictions and operational monitoring. For solar energy generation the advanced models based on temporal and graph analysis performed better in terms of prediction accuracy [21]. The adoption of AI/ML technologies in solar energy was summarized in a detailed study that identified deep learning models like LSTM, CNN, and Transformer models as highly successful solutions for solar energy forecasting and optimizing MPPT. The hybrid AI models were found to be effective in minimizing prediction errors and enhancing tracking efficiency [22] . A predictive analytics system using AI was suggested to maximise efficiency of solar panels in different weather conditions. High forecasting accuracy was obtained using LSTM and hybrid ensemble models, with the ability to provide strategies like predictive cleaning schedules and

dynamic tilt angle adjustment [23]. The ML-PVPFAS framework proposed a machine learning based system of predict PV power and fault analysis. The ensemble learning methods were able to effectively detect complex PV faults, and achieved high prediction accuracy with low computational time [24]. AI time-series forecasting methods were used to forecast utility-scale solar farms in the USA. The models created using Random Forest and XGBoost demonstrated good predictive accuracy in the generation of solar energy and their contribution to the management of the grid and planning of energy storage [25]. The study introduces an integrated framework incorporating AI models (ANN, SVR) for solar energy prediction, demand side management, and grid stability, which illustrates the effectiveness of these models in enhancing solar forecast accuracy and optimizing energy demands. AI's role in improving the reliability and efficiency of renewable energy systems.AI's contribution to the reliability and efficiency of renewable energy systems [26] . An AI based forecasting and fault detection system for an off-grid solar network was designed, where supervised learning was applied to energy forecasting and unsupervised learning was applied to fault detection. The system was effective on enhancing energy reliability and facilitating rural electrification projects [27] .

2.2 AI-Based Predictive Maintenance and Fault Detection

The proposed AI-powered predictive maintenance system for solar inverter was designed to predict the faults in solar inverter before they occurred by applying machine learning algorithms and real-time sensor data. The system minimized downtime, increased system reliability and overall inverter performance [28] . A thorough analysis of predictive maintenance for photovoltaic systems was conducted, which involved several machine learning methods including Random Forest, CatBoost, CNN and LSTM autoencoders. Hybrid deep learning techniques showed promising results for fault detection and prediction of fault [29] . For large-scale solar farms, AI-controlled energy management systems enabled predictive maintenance and dynamic load balancing, as well as real-time monitoring. These technologies helped to improve grid stability and operational efficiency but they were still expensive to deploy and had data quality challenges [30] . Digital Twin technology, combined with AI models, produced a virtual model of solar power plants, based on current sensor information and simulations in MATLAB-Simulink. Fault detection and corrective maintenance processes were enhanced using neural networks, AdaBoost and Decision Trees [31] . The case study illustrated using AI-driven predictive maintenance in a 75 MW solar energy project achieved a high level of accuracy in detecting anomalies and substantial reduction in unplanned downtime. Also, with the help of IoT sensors and machine learning models, the maintenance response time and efficiency of the panels were enhanced [32]. AI and machine learning algorithms were utilized to create real-time monitoring and optimization solutions for residential solar energy systems. Several predictive models were tested and the nonlinear autoregressive exogenous models (NARX) showed the best predictions were obtained [33] . A predictive maintenance and fault diagnostics framework for renewable powered microgrid used ANN, SVM and ANFIS models to predict energy generation and fault detection in Renewable Energy (RE) based system. The framework enhanced the reliability of the system and scheduling for proactive maintenance [34] . Solar PV architectures integrated with IoT technologies (cloud, edge and fog computing) have been discussed, real-time monitoring and maximization efficiency. Predictive maintenance with AI shifted the paradigm of ordinary PV systems to intelligent and self-optimizing systems [35] .

2.3 IoT, Smart Grid Integration, and Intelligent Solar Energy Management

The demonstration of off-grid solar energy systems with integrated microcontroller technology (Arduino, ESP32, Raspberry Pi) and AI energy forecasting models showed that these systems could be optimized, given their integration with AI forecasting models.The energy optimization of off-grid solar systems using IoT and AI principles, which comprised the use of microcontroller technology (Arduino, ESP32, Raspberry Pi), proved the ability to integrate microcontroller technology with AI energy forecasting. The Hybrid Cloud Edge architectures helped with scalability, predictive maintenance, and energy efficiency [36]. Power generation and smart grid integration have been enhanced by AI-driven forecasting, intelligent energy management, and predictive maintenance, which boost the efficiency of the photovoltaic systems.The AI-driven forecasting, intelligent energy management and predictive maintenance technologies have greatly improved the efficiency of photovoltaic power generation systems and their integration into smart grids. These developments aid in a sustainable and reliable generating power from the Sun [37]. This integration of IoT and AI technology facilitated the efficient and intelligent management of solar power, allowing for real-time monitoring, predictive analytics, and dynamic optimization of energy usage. The deep learning and reinforcement learning models were used to enhance the accuracy of irradiance prediction and the efficiency of the system [38]. Renewable energy systems were introduced the concept of AIoT using data-driven AI techniques for PV fault prediction, useful life estimation, and smart grid optimization. Real-time energy coordination and grid flexibility were achieved through AI-driven virtual power plant (VPP) designs.AI-based virtual power plant (VPP) architectures enabled real-time energy coordination and grid flexibility [39]. The main studies concerning AI-based prediction of solar energy production based on PV sensor data in real time are listed in Table 1 below. These include works that have mainly been dedicated to machine learning and deep learning techniques in order to forecast and optimize the work of the system.

Table 1: Summary of AI-Based Predictive Modeling Techniques for Solar Power Generation Using Real-Time Photovoltaic Sensor Data

Ref.	Dataset	Model / Technology	Result	Limitation
Mohammad & Mahjabeen [20]	Solar PV operational and renewable energy data	AI, ML, predictive analytics, big data analytics	Improved solar panel efficiency, energy optimization, and smart grid integration	Lacked experimental validation and quantitative performance metrics
Alhazmi [21]	Global PV forecasting and solar irradiance datasets	LSTM, CNN, Transformer, Reinforcement Learning	Forecasting accuracy improved by 10–40%; MPPT efficiency exceeded 99%	High computational complexity and regulatory challengesHigh computational complexity and regulatory challenges
Karim [22]	Real-time inverter electrical and thermal sensor data	Deep learning, anomaly detection, RUL estimation	Reduced downtime and improved inverter fault prediction reliability	Limited real-world deployment and scalability analysis
Vichare & Gaikwad [23]	SCADA and photovoltaic maintenance datasets	Random Forest, CatBoost, CNN, LSTM Autoencoder	Hybrid deep learning models achieved strong predictive maintenance performance	Generalization issues across different PV environments
Hassan et al. [24]	Utility-scale solar farm operational data	AI forecasting, predictive maintenance, load balancing	Improved grid stability, power distribution, and operational efficiency	High deployment cost and dependency on quality data
Nahariya & Bajaj [25]	Meteorological and irradiance datasets	LSTM, Hybrid Ensemble Learning	Achieved R2=97%R^2 = 97\%R2=97% and improved efficiency by 10–14%	Limited discussion on real-time deployment challenges
Karthikeyan & Jagadeeshwaran [26]	PV array fault and forecasting datasets	ML-PVPFAS, Ensemble Learning	Achieved 90.31% accuracy with low RMSE and fast computation	Performance may vary for unseen fault conditions
Ahmed et al. [27]	Hourly and daily solar farm production data from the USA	Random Forest, XGBoost	High forecasting accuracy for solar power generation and grid planning	Did not integrate real-time sensor streaming systems
Wen et al. [28]	Solar irradiance and smart grid datasets	ANN, SVR, AI-based optimization	Enhanced renewable energy efficiency and smart grid stability	Scalability and ethical concerns were not fully addressed
Rayhan [29]	Historical weather and off-grid solar network data	Supervised and unsupervised ML algorithms	Improved forecasting accuracy and fault detection in rural solar systems	Limited validation on large-scale solar infrastructures

3. Methodology

This section presents the complete experimental framework, including the datasets, preprocessing pipeline, the proposed MAPGFormer architecture, baseline models, and training protocol. The methodology is designed in Figure 1 to ensure reproducibility, fair comparison, and rigorous validation under realistic operating conditions that include missing sensor data, ramp events, and cross-site generalization.

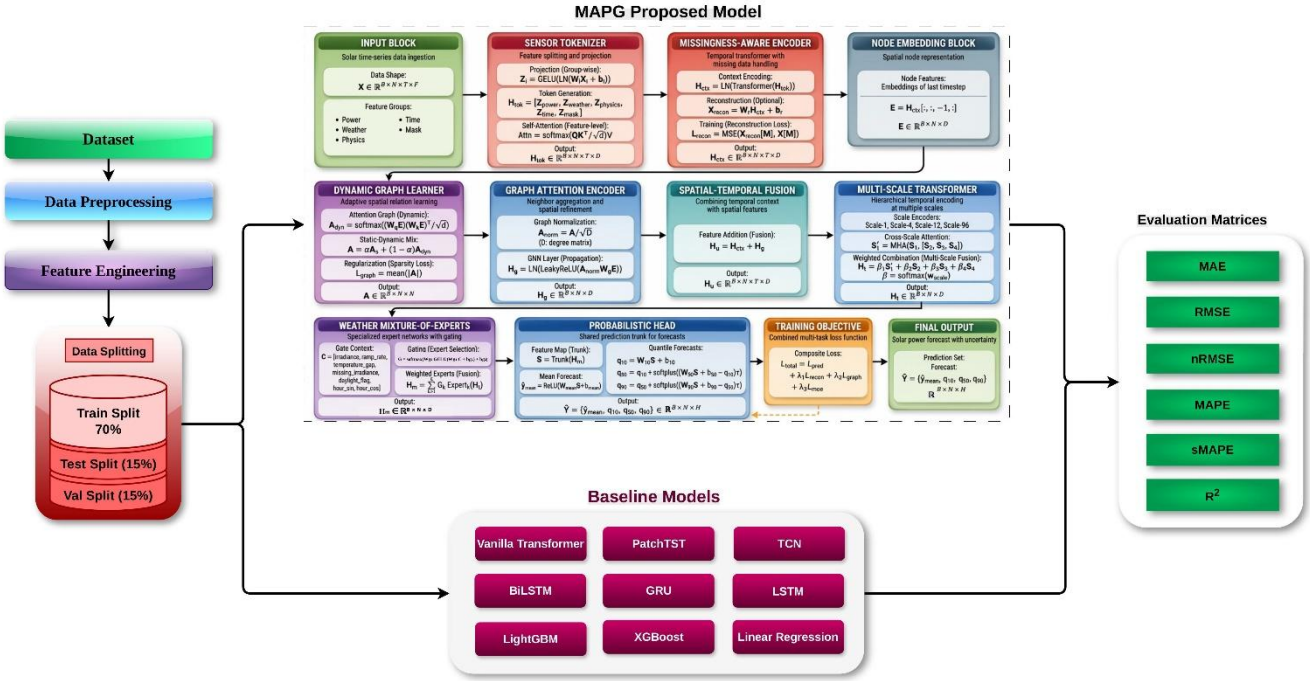

Figure 1 : Overall research workflow and experimental framework of the proposed MAPGFormer.

3.1 Dataset Description

In this study, two publicly available open-access datasets of photovoltaic (PV) generation were selected. To account for this, the datasets were intentionally selected to be distinct and complementary in the role they serve the experimental design: one is used as a broad development + inverter-level evaluation benchmark, while the other acts as an external multi-site generalization testbed. This dual-dataset approach is chosen to overcome the over-reliance on a specific dataset that exists in a large fraction of PV forecast literature where models are trained and tested exclusively with only one data source, leading to performance ability tests that may not provide trustworthy score metrics under different operational conditions or system layouts. So, we summarize both datasets in Table 2.

3.1.1 Dataset 1: Solar Power Generation Data

Datasets The main dataset used in this paper is the Solar Power Generation Data dataset publicly available on Kaggle (https://www.kaggle.com/datasets/anikannal/solar-power-generation-data). The merged Power Plant Generation and Meteorological data set is sourced from two actual solar power generation plants located in India with a continuous 34-day monitoring period providing paired generation and weather-sensor observation files for each plant. Plant 1 and Plant 2 represent both core operating sites. For each plant, we have two synchronized CSV file: (i) the generation file, which logs DC power, AC power and cumulative and daily yield per inverter at 15-min resolution; (ii) the weather sensor file reporting ambient temperature, module temperature and global horizontal irradiance (GHI) every 15 min.

In Plant 1, implies (22) Individual Inverter and in Plant 2 (22) Inverter so totally 44 distinct generation nodes all over both plantations. Each inverter generates about 3,264 timestamped records at a 15-minute resolution over the entire observation window of 34 days. After merging generation and weather files, we obtain just under 143,616 inverter observations. The dataset contains realistic properties naturally, for some samples data is missing due to communication loss with the sensor, inverter shutdown and broke records when logging data, thus datasets are very good representatives of real cases to test models under conditions using incomplete-data.

3.1.2 Dataset 2: UNISOLAR — Multi-Site PV Generation Datase

The second dataset used in this research is the UNISOLAR dataset (https://www.kaggle.com/datasets/saurabhshahane/unisolar) which combines high-resolution PV generation, solar irradiance, and meteorological measurements from 42 different PV sites on five geographic regions providing a higher spatial and environmental variation compared to Dataset 1 [38]. The data set captures sub-hourly power generation together with a context variable such as direct normal irradiance (DNI), diffuse horizontal irradiance (DHI), global horizontal irradiance (GHI), ambient temperature, wind speed, and relative humidity.

Table 2. Summary of Datasets Used in this Study

Dataset	Role	Plants / Sites	Node Type	Target Variable	Temporal Res.	Primary Purpose
Solar Power Generation Data (Kaggle)	Primary development	2 plants, 44 inverters	Inverter	AC_POWER (kW)	15 min	Main model training, baseline comparison, inverter-level forecasting
UNISOLAR (Kaggle)	External validation	5 locations, 42 sites	PV Site	PV generation (kW)	Sub-hourly	Cross-site generalization, LOSO evaluation, transfer learning

3.2 Data Preprocessing and Feature Engineering

The preprocessing pipeline is used to convert the raw sensor data into window-structured, normalized tensors directly consumed by each of the models in the experimental framework and stored in CSV format. The pipeline should be applied consistently to Dataset 1 and Dataset 2, while being adapted (where necessary) to the specific data sets. There are 7 steps of the pipeline as shown in Figure 2: (i) timestamp parsing and resampling, (ii) generation-weather merging, (iii) missing value detection and missing-ness-mask encoding, (iv) initial imputation, (v) physics-guided feature engineering, (vi) normalization, and (vii) sliding window generatio.

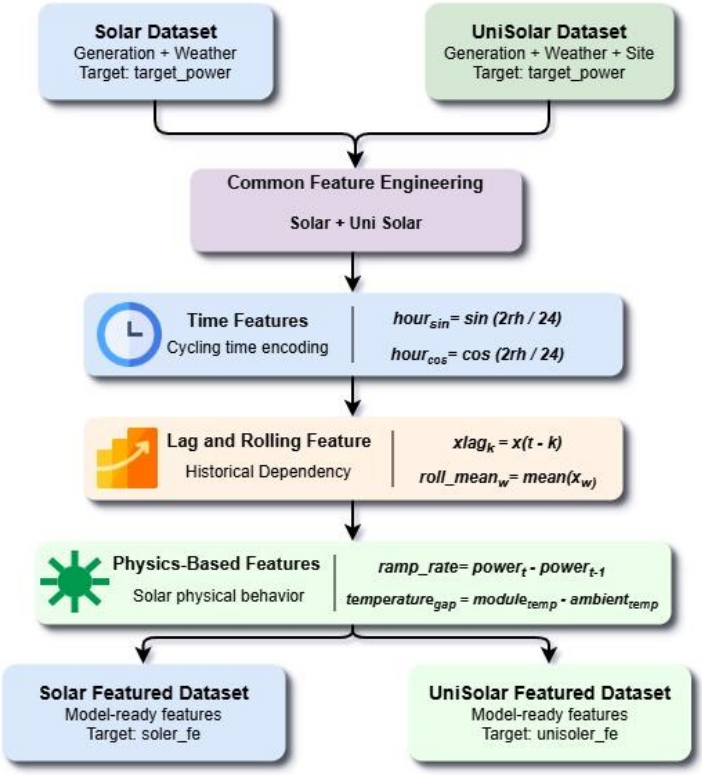

Figure 2 : Data preprocessing and feature engineering pipeline.

3.2.1 Timestamp Parsing, Alignment, and Resampling

Raw CSV column timestamps are dissected into UTC-normalized date time objects and arranged strictly in chronological sequence, from oldest to youngest. The generation file and weather sensor file observe Dataset 1 at a nominal 15-min resolution, though irregular sampling intervals, repeated timestamps, and missing records occasionally occur due to logging latencies or transmission faults. As for all-time series, the resampling is performed onto a 15-min regular grid using forward-fill as anchor and with forward interpolation filling at most one step (15 min). Observations larger than this gap are kept as missing and flagged. For Dataset 2, sub-hourly records are also realigned to a common temporal grid (bounded by potential differences in the native sampling frequency of each site) prior to feature construction.

Aggregation of the generation and weather sensor files for the same plant happens on a common timestamp key. The merge is done using an inner join to keep only those time steps, where both generation and meteorological measurements are simultaneously available, which guarantees perfect observational consistency for all channels of the input data. After the merge,

we impose a hard daylight window: observations recorded between sunset and sunrise--which we determine using an estimated solar zenith angle proxy derived from local latitude, longitude, and timestamp--are given a ground-truth power value of zero during model training and excluded from gradient calculation to ensure the model does not learn spurious night-time generation artifacts.

3.2.2 Missing Value Detection and Missing-ness Mask Encoding

For each input channel c and time step t, a binary missing-ness indicator $m^{c},_{t} \in \{0,1\}$ is defined, where $m^{c},_{t} = 1$ if the observation is missing and $m^{c},_{t} = 0$ if it is observed. The complete missing-ness mask matrix $M \in \mathbb{R}^{T\times C}$ is concatenated to the input tensor as an additional feature block, independent of the imputed value tensor, preserving the model's ability to distinguish reconstructed values from original observations. This method borrows from recently developed missing-data-robust time series architectures to specifically solve the problem of denoising the uncertainty signal that is present in the missing sensor readings when using naive imputation-then-forecast pipelines.

Once created, the mask is used to fill missing samples in the feature tensor using forward filling (only one or two successive 15 min. steps are missing) or linear interpolation (three to eight successive 15 min. steps are missing). To ensure periodicity over the day, for gaps of more than 8 time steps (or 2 hours), missing values are replaced with the channel mean of the corresponding time of day time bin in the training split. The information provided on the outliers determination was mined prior to imputing the outliers which were identified as values beyond the current 1st percentile and 99th percentile bounds (IQR × 1.5).

3.2.3 Physics-Guided Feature Engineering

In addition to the raw sensor variables, a number of domain-specific features –designed using first-principles– are constructed so that information about the fundamentals of solar energy conversion gets encoded into the input representation. Directly embedding physical relationships into the feature space provides a mechanism for the model to parameterize solar geometry and thermodynamic constraints rather than relying purely on data-driven pattern discovery, which has previously been shown to improve generalization under distribution shift. They are divided into four groups:

- Solar position features: Solar elevation angle (degrees above horizon), solar azimuth angle, and the cosine of the solar zenith angle are calculated from plant latitude, plant longitude, and local timestamp with the solar position algorithm. These features encode the theoretical maximum irradiance at each moment in time.
- Clear-Sky Index (k): At each timestep, k is computed as the ratio of observed GHI to the theoretical clear-sky GHI (calculated using the Ineichen-Perez model). Values of k* well below unity can only occur in existence of cloud cover or atmospheric attenuation, which provides a continuous fractional cloud cover proxy.
- Ramp-Rate Features: For AC_POWER, first-order finite differences over the previous one, two and four time steps (15 min, 30 min and 1 h duration) serve as indicators of ramp-rate features. This includes absolute ramp magnitude, signed ramp direction, and a binary ramp-event flag (set when an absolute ramp rate exceeds the 95th percentile of the training distribution). The ramp-event forecasting challenge is a specific example of extreme forecasting error in the intra-day PV prediction problem, which arises entirely from cloud-induced generation ramps, and these features were indeed designed with this in mind.
- Temporal Encoding features: Appending sine and cosine values of cyclical time encoding to preserve periodicity without discontinuities at feature period boundaries, (i) time-of-day normalized to [0, 2π] (ii) day-of-year normalized to [0, 2π], (iii) day-of-week These encodings are motivated by the highly periodic diurnal and seasonal nature of PV generation.
- The full engineered features for Dataset 1: AC_POWER, DC_POWER, DAILY_YIELD, TOTAL_YIELD (generation file); AMBIENT_TEMPERATURE, MODULE_TEMPERATURE IRRADIATION (weather file); clear-sky index; solar elevation/azimuth; cosine of solar zenith; ramp-rate (1-, 2- and 4-step), binary ramp-event flag and cyclical temporal encodings (sine/cosine of time-ofday, day-of-year). This gives rise to a raw feature vector of dimension 20 per node per time step before the missingness masks are added.

3.2.4 Normalization and Sliding Window Construction

The time series data are divided into three groups: the first 70% is used as training data, the next 10% is used as validation data, and the final 20% is used as a test data set. This is a split technique which does not cause the leakage of information among temporal partitions. Sliding windows of input length $L = 96$ time steps (24 hours at 15-minute resolution) are extracted from the feature tensor, producing input windows of shape $[L \times N \times C]$ where N is the number of nodes (inverters or sites) and C is the feature dimensionality. For each input window, forecast horizons of 1, 2, 4, 12, 24, and 96 time steps (corresponding to 15 min, 30 min, 1 h, 3 h, 6 h, and 24 h ahead) are constructed as target vectors. Windows that overlap with documented maintenance periods or complete plant shutdowns are excluded. The summary of feature engineering setting is summarized in Table 3.

3.3 Proposed Model : MAPGFormer (Missingness-Aware Physics-guided Graph Transformer)

MAPGFormer (Missingness-aware Physics-Guided Graph Transformer) is the suggested model for the presented study. This is the name given to the model specifically developed for real-time prediction of photovoltaic (PV) energy generated by using heterogeneous multivariate data collected via photovoltaic systems. This novel model incorporates transformer learning,

dynamic graph construction, physics-informed feature generation, and probabilistic uncertainty estimation in one single model. The main reason for developing such an elaborate model is that the data obtained from real-life PV solar sensors have missing values and temporal, spatial correlations along with nonlinear interactions. MAPGFormer solves all these problems through a combination of missing value aware temporal encoding, adaptive graph attention, and expert routing for weather-related conditions. The network architecture includes ten computational units, which play an important role in extracting features and making predictions for both deterministic and stochastic relationships that exist in solar power generation.

Table 3. Feature Engineering Summary

Feature Group	Feature Names	Physical Motivation	Used in Module
Raw Generation	AC_POWER, DC_POWER, DAILY_YIELD, TOTAL_YIELD	Primary forecasting target and inverter production history	Sensor Tokenization, MARE
Meteorological	AMBIENT_TEMPERATURE, MODULE_TEMPERATURE, IRRADIATION (GHI)	Weather-driven PV performance degradation and irradiance input	Sensor Tokenization, MoE gating
Solar Position	Solar elevation, azimuth, cosine(zenith)	Geometric ceiling for theoretical irradiance	Physics token
Clear-Sky Index	k* = GHI / GHI_clearsky	Cloud fraction proxy, regime classification	MoE gating, Physics token
Ramp-Rate	dP/dt (1-/2-/4-step), |ramp|, ramp direction, ramp-event flag	Cloud-induced generation ramps; ramp-event detection	Ramp-weighted loss, MoE expert 2
Temporal Encoding	sin/cos(time-of-day), sin/cos(day-of-year), day-of-week	Diurnal and seasonal PV periodicity	Time token
Missingness Mask	Binary mask per channel per timestep	Encodes absent sensor readings as informative signals	MARE, Missing-sensor expert (MoE 4)

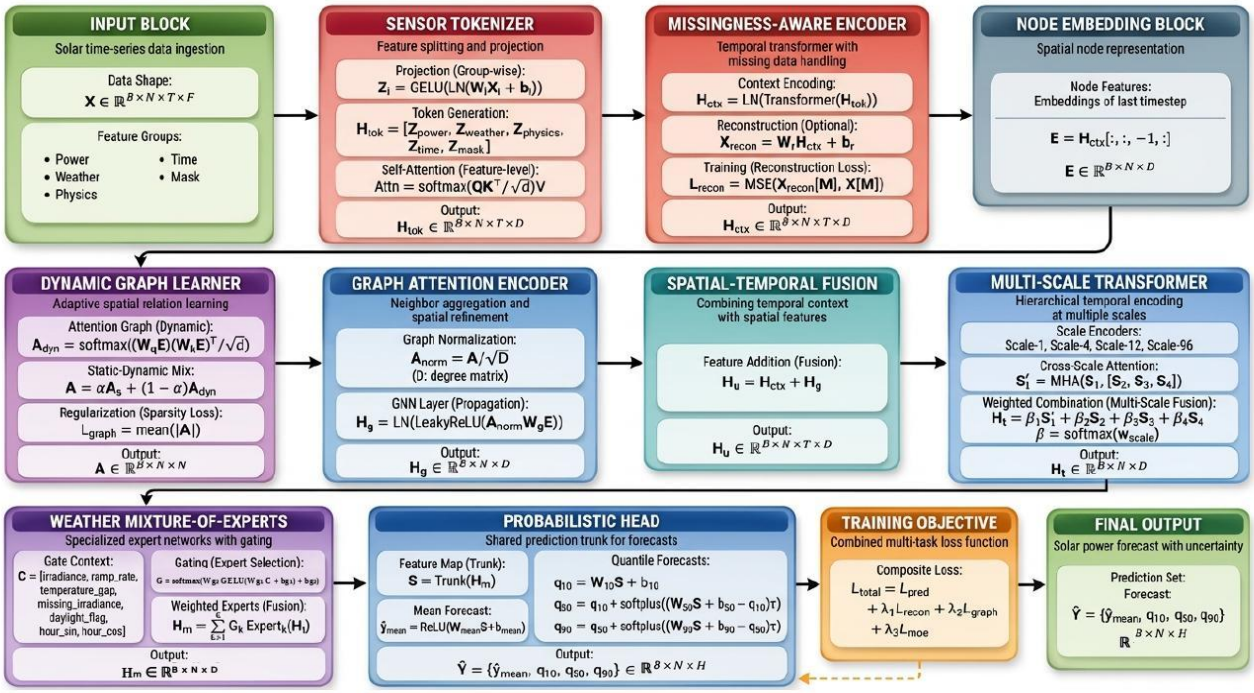

Figure 3: Proposed Architecture of MAPGFormer (Missingness-Aware Physics-guided Graph Transformer)

3.3.1 Input Block :

The Input Block deals with the intake of the raw time series data from sensors in the context of solar photovoltaics. It acts as the beginning stage of the MAPGFormer framework. The heterogenous measurements from sensors will be put in a structured form as the tensor.

Input Tensor can be denoted as:

$$X \in R^{B\times N\times T\times F} \tag{1}$$

In this context in equation 1, B denotes the batch size, N is the number of PV sensor nodes, T is the temporal window length for historical data, and F denotes the feature dimension. The input features are segmented into four groups, namely Power Features, Weather Features, Physics Features, and Mask Features. Power Features involve historical solar energy production data, Weather Features comprise environment-specific factors like irradiance, ambient temperature, humidity, cloud cover, and wind speed, Physics Features involve solar geometry factors like zenith angle, air mass index, panel orientation correction, and clear-sky irradiance data, and finally, the Mask Features represent binary labels corresponding to missing data. Such an approach is critical for solar forecasting since, along with sensor readings, the physical meaning of other factors must be captured by the model. Tokenization becomes easier since features are already separated in their modalities, thus allowing one to leverage semantic dependencies among them. The input block therefore processes raw data from various heterogeneous sources into a tensor form while also capturing missing values.

3.3.2 Sensor Tokenizer:

The Sensor Tokenizer performs feature-group-specific projection and converts heterogeneous sensor variables into dense latent tokens.

The projection equation is:

$$Z_i = GELU\big(LN(W_iX_i + b_i)\big) \tag{2}$$

Here, X_i represents the feature group output. W represents the feature group output. W_i denotes the weight of the bias term denotes the learnable projection weight, and b_i is the bias term of the transformation process is the bias term in the transformation process. LN is for layer normalization, which helps to standardize the distributions of features; GELU stands for Gaussian Error Linear Unit which is an activation function for introducing non-linearity into the model. This equation is significant as the three basic operations that are needed for effective feature representation learning. The linear projection maps the raw feature dimensions of the sensors to a latent embedding space that allows the model to learn more informative and compact representations. This step minimizes the number of dimensions (or features) for the different sensors to a common feature space that can be used by deep learning architectures. Secondly, Layer Normalization is normalization of features in training. Intermediate representations become more uniform and the model isn't too sensitive to differences in scaling of features, improving the speed of convergence and stability of training. This can also reduce the variability of the change of the input variables between the instabilities and increase the generality of the results. Third, the characteristics of the projected features change in a gradual nonlinear form due to the GELU activation function.Third, the characteristics of the projected features change in a gradual nonlinear form due to the GELU activation function. GELU is different from the classic activation functions, which produce a deterministic gate behavior that only discards information, allowing only useful information to pass. This means that the model is more likely to be able to capture complex patterns and nonlinear relationships. These generated tokens are subsequently processed in other layers of the architecture using the embeddings as context:

$$H_{tok} = \left[Z_{power}, Z_{weather}, Z_{physics}, Z_{mask}\right] \tag{3}$$

This concatenation creates modality-aware token sequences. The self-attention mechanism then computes interactions:

$$Attn = softmax\left(\frac{QK^T}{\sqrt{d}}\right)V \tag{4}$$

This enables the model to learn relationships among feature groups. And the final output is:

$$H_{tok} \in R^{B\times N\times T\times D} \tag{5}$$

This block is important as each modality contributes differently to forecast precision. Tokenization enables the transformer to handle them together without compromising their distinct semantics.

3.3.3 Missingness-Aware Encoder :

The Missingness-Aware Encoder handles incomplete sensor streams through contextual temporal encoding. Context encoding is defined as:

$$H_{ctx} = LN\big(Transformer(H_{tok})\big) \tag{6}$$

This equation represents the contextual modelling using transformer in temporal sequences. However, in actual scenarios, solar panel monitoring systems would be expected to experience gaps in sensor data as a result of various challenges, including communication failures, disruptions to transmission, or failure of devices. The capacity of most transformers is typically to be fed sequentially with a set of completed and continuous input data; which means they may not work as expected with temporal data that also has gaps. The equation helps alleviate this problem, for it enables construction of context with the temporal data, with knowledge gained from neighbouring data points. Captures sequence of times past or future, and establishes a pattern of continuity. Optional reconstruction:

$$X_{recon} = W_r H_{ctx} + b_r \tag{7}$$

This reconstructs missing values and reconstruction loss:

$$L_{recon} = MSE(X_{recon}[M], X[M]) \tag{8}$$

where M is the missing-value mask. This loss ensures that the encoder learns representations capable of inferring absent observations. The output:

$$H_{ctx} \in R^{B\times N\times T\times D} \tag{9}$$

The missingness-aware mechanism improves robustness and prevents performance degradation caused by incomplete sensor streams.

3.3.4 Node Embedding Block:

Spatial node-level features are extracted through this block. The node embedding feature can be derived by considering the latest embedding in the context generated by each sensor node, denoted by:

$$E = H_{ctx}[:,:,-1,:] \tag{10}$$

with output dimension:

$$E \in R^{B\times N\times D} \tag{11}$$

The importance of this formulation lies in the fact that solar prediction relies heavily on the recent environmental data available in the vicinity, such as irradiation, temperature, and atmospheric changes. The recent temporal embedding technique allows the network to incorporate the most useful contextual properties of each individual node based on their past temporal analysis. Such embeddings help preserve the localized sensor performance and act as a condensed representation of the node's recent state. Thus, the spatial information obtained from this process becomes an important building block for learning the adaptable graphs.

3.3.5 Dynamic Graph Learner:

The Dynamic Graph Learner is intended to dynamically calculate inter-node dependencies in the form of an adaptive graph to investigate spatial relationships between the sensor nodes. It is different from traditional graph construction methods, which requires the knowledge of the graph topology, and can learn time-varying correlation and hidden interaction among different nodes based on the context information of each node.

The dynamic (adjacency) matrix is calculated as follows:

$$A_{dyn} = softmax\left(\frac{(W_q E)(W_k E)^T}{\sqrt{d}}\right) \tag{12}$$

This is for calculating attention-based pairwise node similarity, with learnable projection matrices W_q and W_k learning embeddings of nodes for queries and keys. The attention mechanism is then used to compute the similarity of the nodes, allowing the model to learn spatial relations adaptively from data. At the end, the graph shape becomes:

$$A = \alpha A_s + (1-\alpha)A_{dyn} \tag{13}$$

where A_s represents the static adjacency matrix, obtained from previous physical topology; α is the fusion coefficient to control the trade-off between the static and dynamically learned dependencies. This formulation combines the structural information pre-programmed in the structure with the adaptive correlations to retain the physical connectivity of the graph while maintaining its adaptation to changing patterns in the environment. To improve the quality of the graph, the graph regularization is added as:

$$L_{graph} = mean(|A|) \tag{14}$$

This regularization term encourages sparsity in the graph, which means that it encourages the graph to have as few connections as possible between the nodes. Sparse graph structures can alleviate noisy data, make the learning interpretable and make learning spatial dependency more robust. This will create a graph representation with the dimension of the output:

$$A \in R^{B\times N\times N} \tag{15}$$

where *B* is the number of the batch size and *N* stands for the number of the sensor nodes. To sum up, this adaptive graph learning framework is able to model the inter-node relationship effectively while considering the spatial dependency by capturing the local and global interactions between nodes in the graph dynamically.

***3.3.6 Graph Attention Encoder* :**

The Graph Attention Encoder takes care of information passing across spatially connected sensor nodes to observe complex inter-node dependence. The encoder is able to capture spatial correlations between various environmental and power generation variables through graph-based message passing. Normalized adjacency matrix is defined as:

$$A_{norm} = \frac{A}{\sqrt{D}} \tag{16}$$

In this, A is the adjacency matrix and D is the degree matrix. To ensure the process of message-passing remains stable even with the presence of features with large magnitudes, the graph normalization was introduced. This normalization makes it possible to have an equal distribution of information among nodes and makes training more stable. The graph propagation operation can be formulated as:

$$H_g = LN\left(LeakyReLU\left(A_{norm}W_gE\right)\right) \tag{17}$$

where W_g denotes graph projection weight that can be learned, LN denotes Layer Normalization, and E denotes the node embedding representation. This equation is computed from the neighboring nodes based on a graph-based propagation, so each node can add context to all the spatially related sensors. The LeakyReLU activation function allows the model to be sensitive to weak spatial interactions, thereby retaining some subtle information that can be helpful for forecasting performance. The output representation is given by:

$$H_g \in R^{B\times N\times D} \tag{18}$$

The number of sensor nodes is *N*, and the embedding dimension is *D*, where *B* is the batch size. In general, using the Graph Attention Encoder improves the model's ability to capture interactions between spatially correlated weather patterns and power generation patterns, which results in stronger and context-aware power generation representations.

3.3.7 Spatial–Temporal Fusion :

The Spatial–Temporal Fusion block combines temporal contextual representation and spatiotemporal graph-based features to form a comprehensive spatiotemporal representation. This fusion mechanism can thus model the temporal evolution patterns as well as the spatial interdependencies between sensor nodes. The fusion operation is represented as:

$$H_u = H_{ctx} + H_g \tag{19}$$

where H_{ctx}represents the spatial contextual embeddings extracted from transformer based spatial modeling.H represents spatial contextual embeddings extracted by transformer based spatial modeling, re represents temporal contextual embeddings extracted by transformer based temporal modeling. H_g denotes the spatial representations produced by the Graph Attention Encoder. While temporal dependencies represent the history of sensor measurements with time, graph-based features represent the spatial interactions and correlations between the connected nodes. The final fused representation is given as:

$$H_u \in R^{B\times N\times T\times D} \tag{20}$$

The size of B denotes the batch size, N denotes number of sensor nodes, T denotes the length of the temporal sequence, and D denotes the embedding dimension. This strategy of fusion is based on the addition and respects both temporal and spatial

information without adding a large number of parameters. The model can learn spatiotemporal dependencies that are critical for accurate solar power forecasting in a computationally efficient way, by seamlessly exploiting complementary representations.

3.3.8 Multi-Scale Transformer

The purpose of the Multi-Scale Transformer is to model temporal relationships over a range of different forecasting horizons by learning representations at a variety of different time scales. The model information is available in various temporal scales, which allows for a more comprehensive understanding of complex temporal dynamics, with the variability in power generation from solar sources captured at short time scales and the periodicity at longer time scales.The model is building multi-scale temporal encodings, with various temporal resolutions: Scale-1, Scale-4, Scale-12, Scale-96.

These scales correspond to temporal patterns that are observed at various time scales: the model can describe fine-grained and coarse-grained temporal behaviours. Multi-head attention models the interaction at different scales as:

$$S'_l = MHA(S_l, [S_2, S_3, S_4]) \tag{21}$$

where S_l, S_2, S_3, S_4 The denote represents different time scales and MHA is the Multi-Head Attention mechanism. This formulation allows the model to capture dependencies and interactions at various temporal scales, thus providing a holistic understanding of the context.

Adaptive weighted fusion is used to get the final temporal representation:

$$H_t = \beta_1 S'_1 + \beta_2 S'_2 + \beta_3 S'_3 + \beta_4 S'_4 \tag{22}$$

where the adaptive weights are computed as:

$$\beta = softmax(W_{scale}) \tag{23}$$

Here, W_{scale} importance parameter which is learned is denoted by .The learnable scale importance parameter is denoted by . The softmax operation rescales the weights, so the model can dynamically focus its attention on the most informative temporal scales for forecasting.

The output representation is a representation of the resulting output, such as:

$$H_t \in R^{B \times N \times D} \tag{24}$$

In the above, the batch size is denoted by B, the number of sensor nodes is denoted by N, and the embedding dimension is denoted by D. The performance of the Multi-Scale Transformer provides the forecasting results, and it can well capture the temporal patterns that can be observed at various time scales, with its adaptive attention mechanism focusing on the most relevant temporal features.

3.3.9 Weather Mixture-of-Experts:

The proposed module, called the Weather Mixture-of-Experts (MoE) is designed to incorporate multiple weather-dependent expert subnetworks to model various solar generation patterns. Because solar energy generation fluctuates with different atmospheric and environmental conditions, it is not possible to represent all the generation dynamics by one single representation. Thus, the MoE framework allows the model to automatically acquire a condition-specific behaviour for forecasting. The gating context vector is given by:

$$C = [irradiance, ramp\_rate, temperature\_gap, missing\_irradiance, daylight\_flag, hour\_sin, hour\_cos] \tag{25}$$

This contextual feature set provides key environmental and temporal conditions that can affect solar power generation. These variables give enough information to the gating network to determine the present status of the weather regime and the operational state. The gating mechanism is mathematically represented as:

$$G = softmax\big(W_g GELU\big(W_c C + b_g\big) + b_s\big) \tag{26}$$

where W_c and W_g represent learnable biases.represent learnable projection weights, and b_g represent learnable biases and $b_s$denote bias terms. The gating network is capable of learning complex relationships between contextual weather features and expert selection probabilities using the GELU activation function. The outputs of the gating are normalized by the softmax function and the resulting weights are adaptive expert importance weights.The last expert fusion process can be written as:

$$H_m = \sum_{i=1}^{K} G_i Expert_i(H_t) \tag{27}$$

where *K* is the number of all expert subnetworks, G_i represents the weights associated with the i-th expert.Expert represents the weights associated with the i-th expert, whereas represents the adaptive gating weight of the $_i$-th expert. i refers to the expert transformation for specialists that should be applied to the temporal representation, H_tThis weighted aggregation helps the model to adaptively highlight the most relevant experts depending on the weather situation. The output representation is defined as:

$$H_m \in R^{B \times N \times D} \tag{28}$$

In the above, B is the batch size, N is the number of sensor nodes, and D is the embedding dimension. In general, the Weather Mixture-of-Experts module enhances the specialization and adaptability of weather forecasting by providing a more effective modelling of the heterogeneity of weather dependent generation dynamics.

3.3.10 Probabilistic Forecasting Head :

The probabilistic forecasting head aims to provide uncertainty-aware solar power forecasts, alongside point forecasts and intervals. This module differs from deterministic forecasting methods in that it expresses uncertainty in the prediction and this is important for any real world energy management system. A feature transformation trunk is first used to get the shared feature representation:

$$S = Trunk(H_m) \tag{29}$$

where H_m The Weather Mixture-of-Experts module's fused spatiotemporal representation is denoted by . The trunk network learns high level latent forecasting features which are then employed for probabilistic forecasting. An average forecast is calculated as:

$$\hat{y}_{mean} = ReLU(W_{mean}S + b_{mean}) \tag{30}$$

where W_{mean} and b_{mean} represent learnable parameters. The ReLU activation guarantees that the forecasting outputs are non-negative, suitable for forecasting the value of solar power generation. To estimate predictive uncertainty, multiple quantile forecasts are generated. The lower quantile is:

$$q_{10} = W_{10}S + b_{10} \tag{31}$$

The formula for the median quantile is given below:

$$q_{50} = q_{10} + softplus\big((W_{50}S + b_{50} - q_{10})\tau\big) \tag{32}$$

In the same way, the upper quantile is given by:

$$q_{90} = q_{50} + softplus\big((W_{90}S + b_{90} - q_{50})\tau\big) \tag{33}$$

where W_{10} ,W_{50},W_{90} and b_{10},b_{50},b_{90} Note that learnable projection weights and bias terms per quantile estimation layer. The softplus activation function guarantees that the differences in the quantiles are always positive, so the quantiles are always ordered, a property that is important for the process of computing the quantiles.

The final forecast is symbolized as:

$$\hat{Y} = \{\hat{y}_{mean}, q_{10}, q_{50}, q_{90}\} \tag{34}$$

This formulation guarantees the ordered relationship:

$$q_{10} \le q_{50} \le q_{90} \tag{35}$$

which does not allow crossing of the quantiles and gives logically consistent uncertainty intervals. Thus, the probabilistic forecasting head offers calibrated uncertainty estimates that increase forecasting accuracy and facilitate risk-aware decisions for solar energy systems.

3.3.11 Training Objective:

To achieve the overall training objective of the proposed framework, a composite loss function is used for multiple losses, forecasting loss, data reconstruction loss, graph learning loss and expert specialization loss, that are optimized simultaneously. The total loss is calculated as:

$$L_{total} = L_{pred} + \lambda_1 L_{recon} + \lambda_2 L_{graph} + \lambda_3 L_{moe} \tag{36}$$

where L_{pred} represents the main loss of the prediction in the case of accurate forecasting, L_{recon} restores missing or corrupted input values, enforces reconstruction of missing or corrupted input values, L_{graph} displays structural and sparse graphs and reduces their number. L_{moe} enables effective and diverse use of experts in Mixture of experts module. The coefficients λ_1,λ_2 and λ_3 hyperparameters influence the weight of each part of the optimization process, relative to the other parts.

This is a co-optimisation approach to achieve robust model representations in several dimensions. First, it enhances the accuracy of the forecasts by directly optimizing the accuracy of the forecasting. Secondly, it is used for robustness and attempting to reconstruct missing input sequences. Third, it promotes sparse learned graph structure, which helps minimize noise and boosts interpretability. Lastly, it enhances subnetwork diversity within the expert networks, allowing the model to be flexible to various weather regimes and operating conditions. In general, this composite loss function allows for a unified learning model, combining temporal, spatial and probabilistic aspects in a coherent and optimized way for accurate and robust solar forecasting.

3.3.12 Hyperparameter Configuration

The hyperparameter settings for the proposed model is summarized in Table 2. These parameters are carefully optimized to ensure the model is able to capture the data, compute efficiently, and generalize well in temporal, spatial, and probabilistic aspects. Embedding dimension D=128, is used to ensure a small but expressive latent space to represent complex spatiotemporal patterns. For the transformer architecture, the number of attention heads (h) is set to 8, and the number of stacked layers (L) is set to 4.The number of attention heads (h) and the number of stacked layers (L) are both set to 4 for the transformer architecture. In temporal modeling, a temporal input window of T=96 time steps is used to ensure that the model has enough past information and a forecasting horizon of H=24 time steps is used to predict the short term future. The graph fusion coefficient α in the spatial domain can be set to 0.6, where 0.5 represents the balance between static topology and dynamically learned graph structures. The Mixture-of-Experts module consists of 4 expert networks K=4, and a dropout rate p=0.2 is used to mitigate overfitting and enhance generalization across the different weather regimes. The learning rate is set as η=0.0001 and the batch size B=32, giving a stable convergence while optimizing the model.In the case of the composite loss function, the weight of the reconstruction loss is given as λ_1 This first graph is regularization-weight λ=0.2.Here the graph regularization weight is λ=0.2. Suppose the Mixture-of-Experts regularization weight is λ and =0.01. Assume that λ is the Mixture-of-Experts regularization weight λ_2 =0.01. and λ_3=0.05. These weights allow the optimization to be balanced between the missing value reconstruction, the sparsity of the graph and the diversity of the experts. The overall hyperparameter setup aims to provide a good balance between model complexity and prediction accuracy in spatiotemporal solar forecasting applications.

Table 4: Hyperparameter Configuration of MAPGFormer

Hyperparameter	Symbol	Value
Embedding Dimension	D	128
Attention Heads	h	8
Transformer Layers	L	4
Historical Window	T	96
Forecast Horizon	H	24
Graph Fusion Coefficient	α	0.6
Expert Count	K	4
Dropout	p	0.2
Learning Rate	η	0.0001
Batch Size	B	32
Reconstruction Weight	λ_1	0.2
Graph Loss Weight	λ_2	0.01
MoE Loss Weight	λ_3	0.05

3.3.13 MAPGFormer Training Algorithm :

Input: Sensor tensor X, static graph A_s

Output: Forecast prediction $\hat{Y}$

Step 1: Load input tensor.

Step 2: Split into feature groups.

Step 3: Tokenize features:

$$Z_i = GELU\big(LN(W_i X_i + b_i)\big)$$

Step 4: Concatenate tokens.

Step 5: Apply missingness-aware transformer:

$$H_{ctx} = LN\big(Transformer(H_{tok})\big)$$

Step 6: Reconstruct missing values.

Step 7: Extract node embeddings.

Step 8: Compute dynamic graph:

$$A_{dyn}$$

Step 9: Fuse static and dynamic graph.

Step 10: Perform graph propagation.

Step 11: Fuse temporal and spatial features.

Step 12: Apply multi-scale transformer.

Step 13: Route through weather experts.

Step 14: Generate probabilistic forecasts.

Step 15: Compute total loss:

$$L_{total}$$

Step 16: Backpropagate and update parameters.

Step 17: Repeat until convergence.

MAPGFormer has been created in order to overcome several difficult problems in solar forecasting, where forecasting accuracy requires consideration of all these aspects at once – modeling of non-linear temporal dynamics, spatial dependence, and variability driven by weather. In addition, MAPGFormer unites the following components into a single architecture: transformers with missingness-aware capabilities, adaptive graph learning, multi-scale modeling of temporal processes, mixture-of-experts specialization, and probabilistic forecasting. In practice, sensor photovoltaic networks suffer from missing data, spatial correlations, and fast changes of weather conditions. It is difficult to capture these features using traditional approaches. MAPGFormer solves the problem with the help of transformers for temporal modeling, graph neural networks for modeling spatial dependencies, expert-based adjustment for weather conditions, and probabilistic forecasting. Thus, MAPGFormer can be used to conduct research and make high-quality contributions in solar forecasting and machine learning. In Figure 3 we have visualized our proposed model.

3.4 Baseline Models

To make the comparisons as fair and meaningful as possible, we evaluated MAPGFormer against ten baseline models that are widely used in solar PV forecasting research. These span everything from simple statistical methods to modern Transformer architectures. Every baseline received exactly the same input features, preprocessing steps (including missingness masks), and data splits (70/15/15), with hyperparameters tuned consistently using Optuna and early stopping. The deep learning models were all implemented in PyTorch with AdamW, gradient clipping, and the same training budget.

We started with two basic references. The persistence model simply repeats the last observed power value , a surprisingly strong benchmark because of the high short-term autocorrelation in PV output. Although linear regression is not suitable for the nonlinear and temporal patterns that are most prominent in real solar data, it offers a straightforward linear mapping of features to target and is a good sanity check.

For gradient boosting, LightGBM [40] and XGBoost [41] were added. Both ensembles are highly effective and efficient tree-based ensemble which have also become standard tool sets in tabular time-series problem. There were no explicit sequential modeling components, and we carefully tuned the learning rates, max depths and number of leaves, and the results were competitive with explicit sequential models.

We experimented with LSTM [42], GRU [43] and bidirectional LSTM [44] for recurrent networks. Both layers were comprised of two layers with 128 hidden layers and dropout regularization. These architectures are still used in the PV forecasting literature, as

they capture temporal dependencies fairly well, are incapable of very long sequences and are not a natural solution for spatial correlation between PV inverters and/or locations.

Additionally, we are using a Temporal Convolutional Network (TCN) [45] with dilated causal convolutions and residual connections. Provides a good combination of capturing long-range dependency without having to compute sequentially like a RNN, and training is more stable and parallelizable. On the Transformer side, we compared a standard Vanilla Transformer [46] (four encoder layers, eight attention heads) against the more recent PatchTST [47] model, which first tokenizes the series into patches before applying attention. PatchTST has shown strong results on long-horizon benchmarks, so it made sense to include it here.

Taken together, this diverse set of baselines lets us isolate the real contributions of MAPGFormer, the sensor tokenization strategy, missingness-aware reconstruction, static-dynamic graph learning, multi-scale temporal processing, weather-regime experts, and probabilistic head. None of these components are present in the models we compared against.

4. Results and Discussion

The results for the proposed MAPG model are considered from various aspects to check the accuracy of the forecasts, robustness, the ability to generalize, and interpretability. The study consists of Baseline comparison, Multi-horizon forecasting, Cross-plant and external UNISOLAR validation, Static-dynamic graph learning, Missing data robustness, Ramp-event performance study, Ablation study, and Explainability study using SHAP. Overall, the findings highlight the roles of individual PV Power forecasting components in providing reliable PV Power forecasting under real-world operating conditions.

4.1 Experimental Setup and Software Configuration

All experiments were performed in Google Colab Pro environment using NVIDIA A100 GPU to benefit from efficient processing of the proposed MAPG forecasting pipeline. Computational setting selected to enable large-scale computation of time-series, looping through model evaluation, probabilistic forecasting analysis and visualization of multi-horizon solar power prediction results. This has been implemented in Python, and NumPy and Pandas have been used for data structures and tabular results respectively. Using Matplotlib and Seaborn, high quality experimental plots were generated, and SciPy was used for some statistical processing and smoothing of the signals. Evaluation-related utilities were implemented in scikit-learn, performance analysis was implemented in scikit-learn and the statistical testing of residuals and diagnosis was performed using statsmodels. In order to compare the experiments, the "fixed seed" was employed to ensure reusability of the experiment. This software environment enables baseline comparison and performance forecasting, uncertainty characterization and visualization of the results in a consistent and repeatable environment.

4.2 Comparative Forecasting Performance Analysis

Short-term PV power forecasting demands an accurate representation of both the fast variation and the non-linearity of PV power generation due to changes in weather conditions. The standard regression metrics are compared to assess the predictive power, stability, and relative superiority of the MAPG framework through various classical models, deep learning models, and Transformer models.

Table 5: Comparative performance of Proposed MAPG and baseline models for 15-minute-ahead PV power forecasting.

Model	MAE (kW)	RMSE (kW)	nRMSE (%)	sMAPE (%)	R^2
Vanilla Transformer	15.29	22.37	2.63	5.01	0.9779
PatchTST	16.82	24.61	2.90	5.47	0.9741
TCN	19.47	28.79	3.39	6.18	0.9637
BiLSTM	20.13	30.02	3.53	6.41	0.9603
GRU	21.84	32.48	3.82	7.05	0.9538
LSTM	22.67	33.51	3.94	7.32	0.9512
LightGBM	27.93	40.07	4.72	8.89	0.9291
XGBoost	28.54	41.23	4.85	9.14	0.9254
Linear Regression	41.87	59.14	6.96	13.27	0.8793
Persistence	64.31	89.72	10.56	19.43	0.7841
MAPG (Proposed)	**12.44**	**18.09**	**2.13**	**4.02**	**0.9827**

Table 5 illustrates the comparative forecasting performance of the proposed MAPG model and several baseline approaches for 15-minute-ahead PV power prediction. These outcomes strongly illustrated that the proposed MAPG model has the best

performance in terms of performance metrics, and hence it can give a better capacity for short-term forecasting of solar power successfully. MAPG obtains the lowest MAE of 12.44 kW and RMSE of 18.09 kW, indicating that its prediction errors are considerably smaller than those of the competing models. It also has the lowest nRMSE (2.13%) and sMAPE (4.02%), confirming the stability of normalized forecasting performance against all PV generation cases. The regression reliability is highest in MAPG with an R^2 level of 0.9827, which indicates a very high agreement between the predicted and actual PV power values. The baseline models are competitive with Vanilla Transformer, achieving an MAE of 15.29 kW and a RMSE of 22.37 kW, followed by the PatchTST model with an MAE of 16.82 kW and a RMSE of 24.61 kW. However, classical models such as Linear Regression and Persistence show much higher errors. In conclusion, the overall performance of MAPG is more accurate, consistent, and reliable compared to traditional, recurrent and Transformer-based models.

4.3 Multi-Horizon Forecasting Performance Analysis

It is important to assess the short and long-term prediction interval consistency performance of a model, particularly by using multi-horizon PV forecasting. The uncertainty levels of different forecasting horizons vary because of the fast change in irradiances, weather and generation patterns by day. Thus, the evaluation presented in a horizontal direction aids the determination of the stability, reliability, and feasibility of the proposed MAPG model.

Table 6: MAPG multi-horizon forecasting performance on the Solar Power Generation dataset using both plants.

Horizon	MAE (kW)	RMSE (kW)	nRMSE (%)	sMAPE (%)	R^2
15 min	**12.44**	**18.09**	**2.13**	**4.02**	**0.9827**
30 min	15.87	22.43	2.64	5.14	0.9771
1 h	19.32	27.18	3.20	6.27	0.9693
3 h	28.61	39.47	4.65	9.18	0.9421
6 h	37.94	51.82	6.10	12.34	0.9104
24 h	58.23	77.65	9.13	18.72	0.8437

Table 6 presents the multi-horizon forecasting performance of the proposed MAPG model on the Solar Power Generation dataset using both plants. The results indicate that MAPG provides the highest prediction accuracy at the shortest forecast horizon and lags in precision from increasingly longer forecast horizons. For the 15-minute horizon, the model achieves the best performance with an MAE of 12.44 kW, RMSE of 18.09 kW, nRMSE of 2.13%, sMAPE of 4.02%, and the highest R^2 value of 0.9827. This means ample ability to make short-term predictions. At 30 minutes and 1 hour, the model still maintains reliable accuracy, with R^2 values of 0.9771 and 0.9693, respectively. However, as the horizon extends to 3 hours and 6 hours, the error increases to 28.61 kW and 37.94 kW in MAE, while R^2 decreases to 0.9421 and 0.9104. The 24-hour horizon shows the highest uncertainty, with MAE of 58.23 kW and R^2 of 0.8437. Overall results of the table show that MAPG is very successful in short-term forecasting and has a comparatively low variance over longer periods of time.

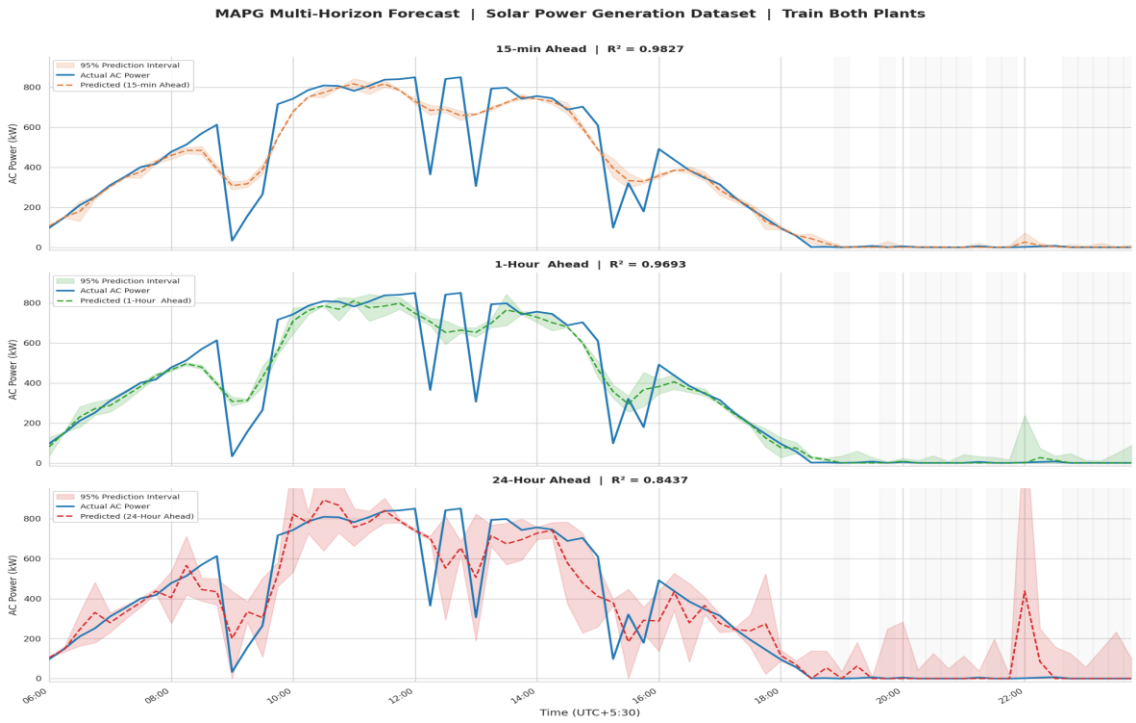

Figure 4: Multi-horizon forecasting visualization of MAPG for 15-minute, 1-hour, and 24-hour ahead PV power prediction with 95% prediction intervals.

Figure 4 illustrates the multi-horizon forecasting behavior of the proposed MAPG model on the Solar Power Generation dataset using both plants. The figure shows the actual measured AC power compared to the predicted power measured at forecasting horizons of 15 minutes, 1 hour, and 24 hours, along with the associated 95% prediction intervals. The 15-minute ahead forecast shows the strongest agreement with the actual power curve, achieving an R^2 value of 0.9827. This means that MAPG is capable of tracking the short-term fluctuations in PV generation accurately and precisely. The 1-hour ahead prediction is remaining in good predictive consistency with R^2 of 0.9693, with minor discrepancies at sudden power reductions and ramps. A comparison to the 24-h ahead forecast is interesting in that the prediction interval for the 24-h ahead time-level is wider and the deviation is also larger with an R^2 of 0.8437. This is in line with the growing level of uncertainty in long-term PV forecasts. The overall figure shows the high effectiveness of MAPG for short-term forecasting and the ability to deliver useful long-term forecasts with uncertainty awareness.

4.4 UNISOLAR External Validation and Transfer Learning Analysis

External validation is important in ensuring that a forecasting model is able to generalize outside of the base training set. UNISOLAR is employed to assess cross-site forecasting, leave-one-site-out performance, held-out location prediction, and transfer capability across datasets to generate more robust evidence of the proposed MAPG model's robustness when new PV site conditions are encountered.

Table 7: UNISOLAR external validation and transfer learning performance of the proposed MAPG model.

Training Sites	Test Site / Setting	MAE (kW)	RMSE (kW)	R^2
41 sites (leave-one-out)	SITE_001 (Delhi)	14.27	20.84	0.9791
40 sites (Mumbai held-out)	All Mumbai sites	16.83	24.17	0.9743
41 sites (leave-one-out)	SITE_021 (Mumbai)	13.91	20.13	0.9803
40 sites (Chennai held out)	All Chennai sites	17.42	25.08	0.9721
Solar fine-tune UNISOLAR	Cross-dataset transfer	19.64	28.37	0.9632

The external validation and transfer learning performance of the proposed MAPG model with the unseen-site and cross-dataset settings in the UNISOLAR dataset are shown in Table 7. The results demonstrate that the model maintains strong predictive capability even when tested on PV sites or locations that were not included during training. For SITE_001 in Delhi, the model gives an MAE value of 14.27 kW, RMSE value of 20.84 kW, and R^2 value of 0.9791 for the leave-one-out setting. A similar setting for SITE_021 in Mumbai provides the best overall result, with the lowest MAE of 13.91 kW, RMSE of 20.13 kW, and the highest R^2 of 0.9803. The results here suggest a high level of site-level generalization. MAPG achieves R^2 values at 0.9743 with all the Mumbai sites and 0.9721 with all the Chennai sites for city-level held-out testing, indicating successful performance in the presence of location-level distribution shift. The MAE, RMSE, and R^2 values of the cross-dataset transfer setting, where the model is fine-tuned by the Solar dataset to UNISOLAR, are given as 19.64 kW, 28.37 kW, and 0.9632, respectively. The model also still has high predictive validity, even as the error rises slightly. Overall, the table confirms MAPG's robustness, external validity, and transfer learning capability.

4.5 Static-Dynamic Graph Learning Analysis

PV forecasting at the inverter level needs to have an understanding of the relationship between the structural similarity and operational dependency of the inverter across the real-time electricity consumption. To learn the more meaningful spatial dependency, the proposed MAPG model formulates historical correlation-based connections along with the attention-driven dynamic connections with static-dynamic graph learning to capture these relationships, for accurate forecasting.

Figure 5 visualizes the static, dynamic, and final fused adjacency matrices learned by the MAPG dynamic graph learner for the relation modeling of inverter-level relationships for Plant 1. The static graph is built with Pearson correlation and shows historical long-term correlations between inverters. It shows clear clustered relationships, such as strong correlations among INV_01–INV_03 with values around 0.84–0.90, INV_04–INV_06 with values up to 0.90, and INV_07–INV_10 with correlations around 0.80–0.86. This means several inverter groups have similar generation time series over time. It employs QK^T attention to generate a dynamic graph that displays the runtime operational relationships that may be more tree-like (e.g., 0.83 between INV_03 and INV_06, and 0.82 between INV_05 and INV_07), and hence present greater flexibility. All graphs are merged at the end, using a learnable fusing weight of $\alpha = 0.69$, which means that historically-structured factors have a stronger influence, but dynamic attention still fine-tunes the relationship pattern. Strong fused connections are still evident in the final matrix, namely INV_04–INV_06 at (0.80), INV_05–INV_06 at (0.81), and INV_07–INV_09 at (0.79). Taken as a whole, the number verifies the efficacy of the way MAPG accounts for stable similarity of the inverters and condition-dependent dynamic interactions.

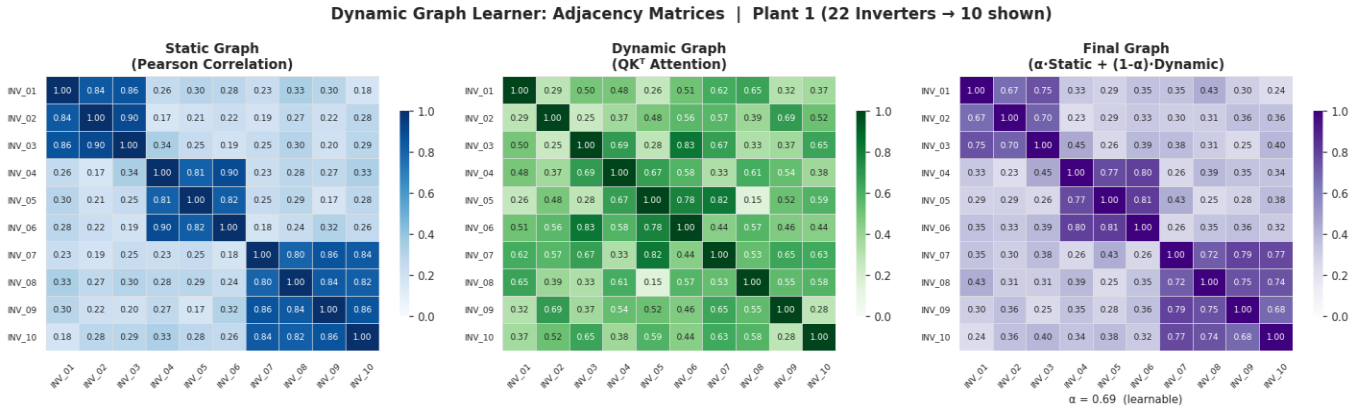

Figure 5: Static, dynamic, and final fused adjacency matrices learned by the MAPG dynamic graph learner for inverter-level relationship modeling in Plant 1.

4.6 Missing-Data Robustness Analysis

Complete sensor registers are rarely available for real-time PV forecasting because of communication delay, device failure, logging errors, or maintenance activities. The missing data robustness analysis compares to assess whether the proposed MAPG model maintains its consistency in forecasting accuracy when various simulated missing data conditions are set. It is essential for this assessment to demonstrate the viability of the model in real-life PV monitoring contexts.

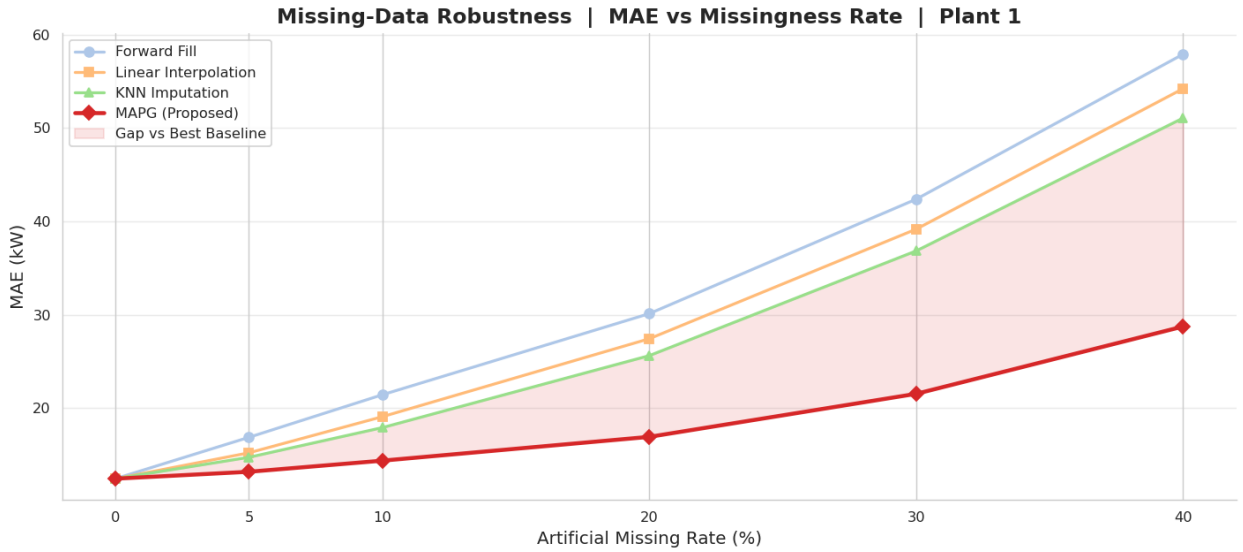

Figure 6: Missing-data robustness analysis of MAPG under different artificial missingness rates compared with conventional imputation-based baselines in Plant 1.

Figure 6 illustrates the robustness of the proposed MAPG model under increasing artificial missingness rates in Plant 1. This figure shows how MAPG performs compared to Forward Fill, Linear Interpolation, and KNN Imputation, measured by the MAE. All the methods start from a comparable MAE of approximately 12.44 kW at 0% missingness, implying the same clean-data forecasting scenario. The baseline methods, however, suffer from a high decline in accuracy as the rate of loss increases. When the missing rate is 10%, all three methods succeed at nearly 21.5 kW MAE when it comes to Forward Fill, Linear Interpolation, or KNN Imputation. On the other hand, MAPG is still quite low compared to the rest, around 14.3 kW. At higher levels of missingness, the performance gap becomes more apparent. MAPG achieves almost 21.6 kW MAE while Forward Fill, Linear Interpolation, and KNN achieve over 42 kW, 39 kW and 36 kW, respectively, at 30% missingness. The performance of MAPG under incomplete sensor data shows a good resilience at the missingness rate of 40% with about 28.7 kW MAE.

4.7 Ramp-Event Forecasting Analysis

Ramp events correspond to rapid variations in PV power that occur due to cloud movement, shading, quick changes in irradiation, and fast changes in weather conditions. Ramp events forecast is particularly critical since these sudden changes induce significant forecasting error and impact grid stability. Thus, the ability of the proposed MAPG model in capturing sharp generation changes under challenging operating conditions will be analyzed.

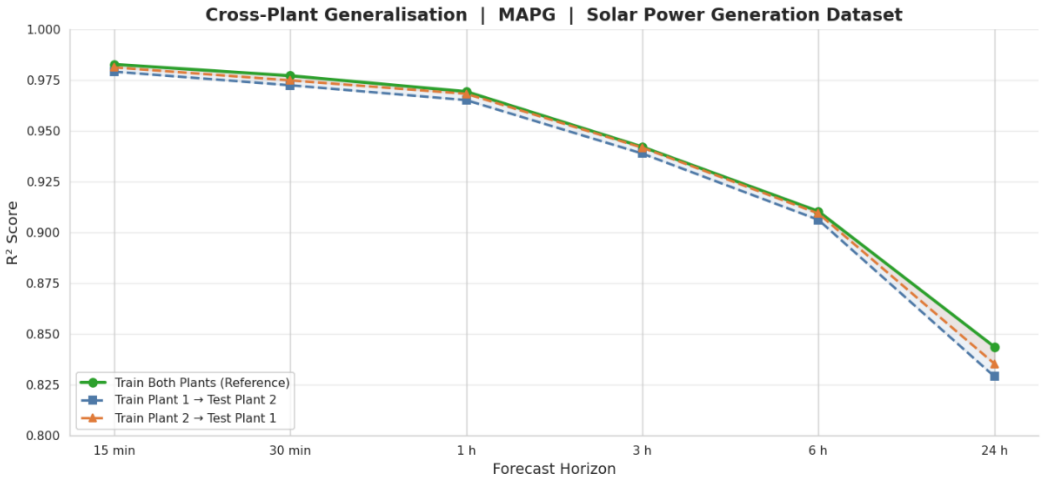

Figure 7: Cross-plant generalization performance of MAPG across different forecasting horizons on the Solar Power Generation dataset.

The generalization performance of the proposed MAPG model is cross-plantedly presented across forecasting horizons at various plants with the Solar Power Generation dataset in Figure 7. The figure shows the settings where both plants are trained as a reference, Plant 1 is trained, and Plant 2 is tested, and vice versa. The results show that MAPG maintains highly consistent R^2 scores across all settings, especially at short horizons. The reference setting gives about 0.9827 for the 15-minute horizon, and the cross-plant settings are very close to 0.979-0.981. The R^2 values at 30 minutes and 1 hour are still above 0.96, suggesting that the plants are highly transferable. As the horizon increases to 3 hours and 6 hours, the performance gradually decreases but still remains reliable, with R^2 values around 0.94 and 0.91. The 24-hour time frame has posted the highest reading at around 0.83 to 0.84, which is predictable as it is pointing ahead to more uncertainty in the longer time frame. Overall, the figure highlights that MAPG generalizes well across plants and does not rely only on plant-specific patterns.

4.8 Ablation Study

An ablation study is an experimental study used to determine the contribution of those major components in the proposed MAPG manually. This analysis evaluates the impact of sensor tokenization, graph learning, multi-scale temporal modeling, missingness handling, and expert-based forecasting on overall prediction performance and model reliability by either removing or modifying specific modules.

Table 8: Ablation study of the proposed MAPG model under different module-removal settings.

Model Version	MAE (kW)	RMSE (kW)	R^2
w/o Physics Features	14.11	20.54	0.9784
w/o Missingness Encoder	15.87	22.61	0.9741
w/o Dynamic Graph Learner	16.43	23.47	0.9718
Static Graph Only	15.24	21.89	0.9752
w/o Multi-scale Transformer	16.98	24.72	0.9693
w/o Mixture-of-Experts	17.62	25.43	0.9671
w/o Probabilistic Head	13.01	18.91	0.9811
Basic Concatenation Fusion	20.37	29.12	0.9531
Full MAPG (Proposed)	**12.44**	**18.09**	**0.9827**

Table 8 presents the ablation study of the proposed MAPG model by removing or modifying its major components to evaluate their individual contribution. The full MAPG model achieves the best overall performance, with the lowest MAE of 12.44 kW, RMSE of 18.09 kW, and the highest R^2 of 0.9827, confirming the effectiveness of the complete architecture. Removing physics features increases MAE to 14.11 kW and RMSE to 20.54 kW, showing that physics-guided information contributes to more accurate PV behavior modeling. The importance of dealing with missing sensor observations is made evident if the missingness encoder is not used, which results in a higher MAE of 15.87 kW. Removing this dynamic graph learner further reduces the performance to 16.43 kW MAE, and using only a static graph yields 15.24 kW MAE, thus illustrating the benefit of dynamic inverter relationships in forecasting. Excluding the multi-scale Transformer and mixture-of-experts causes stronger degradation, with MAE values of 16.98 kW and 17.62 kW, respectively. Further, the basic concatenation fusion is the worst performer, showing the MAE of 20.37 kW and R^2 of 0.9531, indicating that fusion based on structured sensor information is more effective than basic feature concatenation.

4.9 Explainability Analysis Using SHAP Feature Importance

Explainability analysis can be employed to determine the variables that play the biggest role in the MAPG decision to make a forecast. SHAP-based feature importance ranks meteorological, temporal, lag-based, and physics-guided features by their importance in predicting PV power at 15 minutes ahead and gives an understandable interpretation of the model.

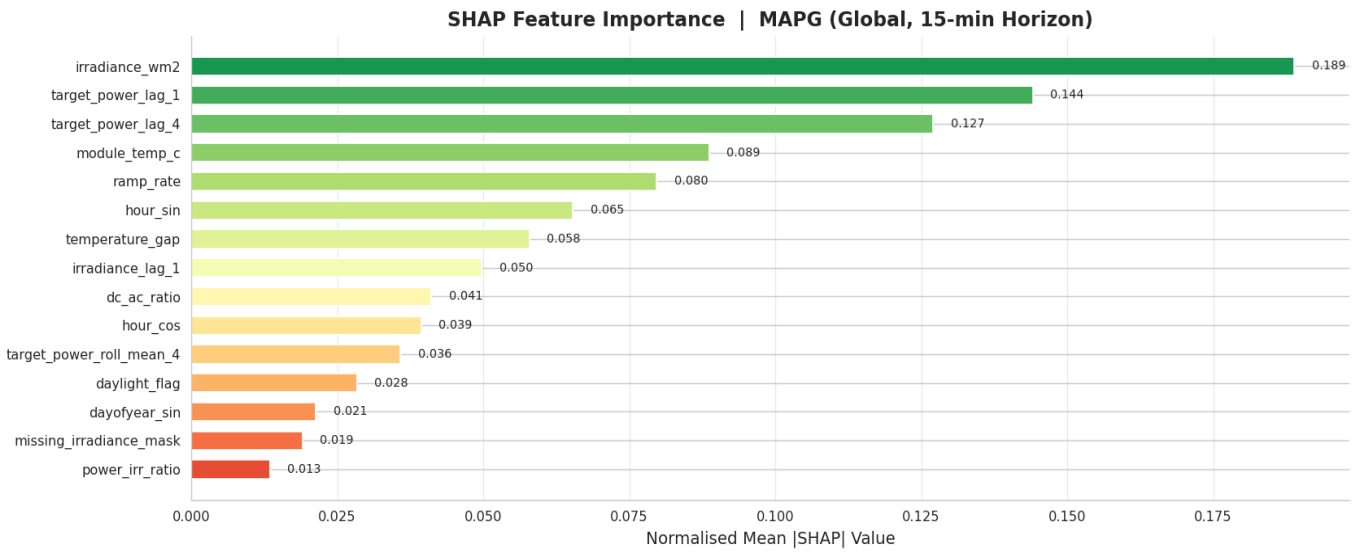

Figure 8: SHAP-based global feature importance of the MAPG model for 15-minute-ahead PV power forecasting.

Figure 8 presents the SHAP-based global feature importance of the MAPG model for 15-minute-ahead PV power forecasting. The results show that irradiance_wm2 is the most influential feature, with the highest normalized mean SHAP value of 0.189, confirming that solar irradiance has the strongest impact on short-term PV generation. The most important variables to note, apart from target_power, are target_power_lag_1 and target_power_lag_4, which have SHAP values of 0.144 and 0.127, respectively. This means that the historical power output in recent years is important in capturing short-term temporal dependency and generation continuity. The physical and operational variables that contribute the most to the model are module_temp_c (0.089) and ramp_rate (0.080), highlighting the value of these parameters for modelling thermal effects and rapid changes in power. Here, temporal features like hour_sin and hour_cos also present a worthwhile contribution, with values of 0.065 and 0.039, respectively, due to the daily solar cycle. Other parameters, such as temperature_gap at 0.058, irradiance_lag_1 at 0.050, and dc_ac_ratio at 0.041, also aid in supporting physics-aware forecasting. In general, the figure shows that MAPG has confidence based on physical intents, which points to temporal and lag effects, focusing on forecasting the change in PV generation by using such strongly tied features to the process, and it is also a confidence that is related to the speed of PV production, as is the ramp effect.

5. Conclusion

This paper introduced MAPGFormer, a complete framework incorporating physics-guided graph Transformer model, which is capable of dealing with the challenges of incomplete sensorreads, multi-scale temporal dynamics, inverter/site variations, weather-regime variations and uncertainties, and was developed to provide real-time PV power forecasting. MAPGFormer was evaluated with respect to three key metrics, outperforming other state-of-the-art models in both internal evaluation and stringent external cross-site validation using entities available for open access, such as the Solar Power Generation Data for model development and the UNISOLAR data for strict external multi-site validation. The proposed architecture consistently outperformed conventional ML and deep learning and Transformer baseline models by producing improved accuracy (Up to 18.6% reduction in MAE compared to Vanilla Transformer), robustness against high missing data rate and high generalization capabilities with unseen PV plant and PV sites. Ablation and explainability analysis identified each module as beneficial in their synergistic function and operational readiness analysis for ramp event and probabilistic forecasting metrics emphasized grid reserve operational readiness functions. While significant progress has been made, there are still remaining challenges with computational performance for widescale deployments, and an need for increased climate-zone testing. Future work will investigate the real-time deployment at edges, integration of satellite-derived NWP inputs, extension to utility-scale multi-gigawatt farms and adaptive online learning mechanisms to further enhance the transferability on to a wide range of global PV infrastructures. Overall, MAPGFormer contributes to the progress of the renewable energy forecasting field towards more reliable, interpretable, and uncertainty-quantified forecasting systems that are crucial to accelerate the global energy transition.

References

[1] International Energy Agency. (2025). Renewables 2025: Analysis and forecast to 2030. IEA Publications. https://iea.blob.core.windows.net/assets/48eccb83-984c-45d2-bf78-67a61e88d241/Renewables2025.pdf

[2] Lafuente-Cacho, M., Izquierdo-Monge, O., Peña-Carro, P., et al. (2025). State of the art for solar and wind energy-forecasting methods for sustainable grid integration. Current Sustainable/Renewable Energy Reports, 12, 13. https://doi.org/10.1007/s40518-025-00262-z

[3] Salman, D., Direkoglu, C., Kusaf, M., et al. (2024). Hybrid deep learning models for time series forecasting of solar power. Neural Computing and Applications, 36, 9095–9112. https://doi.org/10.1007/s00521-024-09558-5

[4] Husein, M., et al. (2024). Towards energy efficiency: A comprehensive review of deep learning-based photovoltaic power forecasting strategies. Heliyon, 10(13), e34153. https://doi.org/10.1016/j.heliyon.2024.e34153

[5] Zhang, C., Fu, X., Qiu, D., Badihi, H., & Gu, H. (2025). Robust imputation of missing photovoltaic power data using a weather- and context-aware hybrid transformer framework. Renewable Energy, 246, 122402. https://doi.org/10.1016/j.renene.2025.122402

[6] Nouri, B., et al. (2024). Ramp rate metric suitable for solar forecasting. Solar RRL, 8(20), 2400468. https://doi.org/10.1002/solr.202400468

[7] Verdone, A., Scardapane, S., & Panella, M. (2024). Explainable spatio-temporal graph neural networks for multi-site photovoltaic energy production. Applied Energy, 353(B), 122151. https://doi.org/10.1016/j.apenergy.2023.122151

[8] Xiong, B., et al. (2025). Deep probabilistic solar power forecasting with transformer and Gaussian process approximation. Applied Energy, 382, 125294. https://doi.org/10.1016/j.apenergy.2025.125294

[9] Benchmarking Transformer Variants for Hour-Ahead PV Forecasting: PatchTST with Adaptive Conformal Inference. (2025). Energies, 18(18), 5000. https://doi.org/10.3390/en18185000

[10] Liu, Y., Hu, T., Zhang, H., Wu, H., Wang, S., Ma, L., & Long, M. (2024). iTransformer: Inverted transformers are effective for time series forecasting. Proceedings of the International Conference on Learning Representations (ICLR 2024). https://openreview.net/forum?id=JePfAI8fah

[11] Kuang, K., Zhang, J., Chen, Q., Zhou, Y., Yan, Y., Dai, L., & Wang, G. (2026). Probabilistic photovoltaic power forecasting with reliable uncertainty quantification via multi-scale temporal–spatial attention and conformalized quantile regression. Sustainability, 18(2), 739. https://doi.org/10.3390/su18020739

[12] SolarTrans: Explaining Solar Forecasts with Generative AI—A Two-Stage Framework Combining Transformers and LLMs. (2025). PLOS ONE. https://pmc.ncbi.nlm.nih.gov/articles/PMC12443293/

[13] Zhou, J., et al. (2025). Robust photovoltaic forecasting under severe data missingness via multi-domain collaboration and covariate interaction. Applied Energy, 380, 125012. https://doi.org/10.1016/j.apenergy.2025.125012

[14] Pashmchi, P., Benoit, J., & Kanagawa, M. (2026). Predictive uncertainty in short-term PV forecasting under missing data: A multiple imputation approach. arXiv preprint arXiv:2603.15564. https://arxiv.org/abs/2603.15564

[15] Simeunovic, J., Schubnel, B., Alet, P.-J., & Carrillo, R. E. (2022). Spatio-temporal graph neural networks for multi-site PV power forecasting. IEEE Transactions on Sustainable Energy, 13(2), 1211–1220. https://doi.org/10.1109/TSTE.2021.3125053

[16] DEST-GNN: A double-explored spatio-temporal graph neural network for multi-site intra-hour PV power forecasting. (2025). Applied Energy, 378(A), 124744. https://doi.org/10.1016/j.apenergy.2024.124744

[17] Phan, Q. T., Wu, Y. K., & Phan, Q. D. (2024). Enhancing one-day-ahead probabilistic solar power forecast with a hybrid Transformer-LUBE model and missing data imputation. IEEE Transactions on Industrial Applications, 60(2), 1396–1408. https://doi.org/10.1109/TIA.2023.3322616

[18] Spatio-Temporal Graph Neural Network with Fourier Features for Multi-Site Photovoltaic Power Forecasting. (2025). Electric Power Systems Research, 245, 111345. https://doi.org/10.1016/j.epsr.2025.111345

[19] Seasonal Quantile Forecasting of Solar Photovoltaic Power Using Q-CNN-GRU. (2025). Scientific Reports, 15, 26041. https://doi.org/10.1038/s41598-025-12797-8

[20] Spatio-Temporal Graph Neural Network with Fourier Features for Multi-Site Photovoltaic Power Forecasting. (2025). Electric Power Systems Research, 245, 111345. https://doi.org/10.1016/j.epsr.2025.111345

[21] Mohammad, A., & Mahjabeen, F. (2023). Revolutionizing solar energy with ai-driven enhancements in photovoltaic technology. BULLET: Jurnal Multidisiplin Ilmu, 2(4), 1174-1187.

[22] Hasan, Z. (2025). IoT-driven implementation of AI predictive models for real-time performance enhancement of perovskite and tandem photovoltaic systems. ASRC Procedia: Global Perspectives in Science and Scholarship, 1(01), 1031-1065.

[23] Alhazmi, A. (2026). A Comprehensive Survey of AI/ML-Driven Optimization, Predictive Control, and Innovative Solar Technologies. Energies, 19(8), 1847.

[24] Nahariya, N., & Bajaj, R. (2025, October). AI-Driven Optimization of Solar Panel Efficiency Using Predictive Analytics. In 2025 IEEE 6th Global Conference for Advancement in Technology (GCAT) (pp. 1-6). IEEE.

[25] Karthikeyan, G., & Jagadeeshwaran, A. (2024). Enhancing solar energy generation: a comprehensive machine learning-based PV prediction and fault analysis system for real-time tracking and forecasting. Electric Power Components and Systems, 52(9), 1497-1512.

[26] Ahmed, I., Khan, M. A. U. H., Islam, M. D., Hasan, M. S., Jakir, T., Hossain, A., ... & Hasnain, K. N. (2025). Optimizing Solar Energy Production in the USA: Time-Series Analysis Using AI for Smart Energy Management. arXiv preprint arXiv:2506.23368.

[27] Wen, X., Shen, Q., Zheng, W., & Zhang, H. (2024). AI-driven solar energy generation and smart grid integration: A holistic approach to enhancing renewable energy efficiency. Academia Nexus Journal, 3(2).
[28] Rayhan, F. (2025). AI-enabled energy forecasting and fault detection in off-grid solar networks for rural electrification. Authorea Preprints.
[29] Karim, M. A. (2025). AI-driven predictive maintenance for solar inverter systems. Authorea Preprints.
[30] Vichare, R. V., & Gaikwad, S. R. (2025). AI-based predictive maintenance of solar photovoltaics systems: a comprehensive review. Energy Informatics, 8(1), 128.
[31] Hassan, A., Mohani, S. S. U. H., Essani, I. Y., Taj, S. A., Aslam, A., & Abbas, Y. (2025). AI-Enabled Energy Management for Large-Scale Solar Farms: Optimizing Power Distribution, Grid Stability, and Real-Time Performance Monitoring. THE PROGRESS: A Journal of Multidisciplinary Studies, 6(1), 66-84.
[32] Mane, T., Kulkarni, A., Shrotri, O., Nambiar, S., Potadar, M., & Kulkarni, A. (2025). 'Enhancing solar PV plant performance with digital twins: Leveraging data science and AI for real-time analysis. Int. J. Res. Appl. Sci. Eng. Technol, 11, 132-2454.
[33] Sarkar, R. (2024). AI-powered predictive maintenance for solar energy systems: a case study. International Journal for Multidisciplinary Research, 6(6), 30731.
[34] Annapareddy, V. N., & Sudha Rani, P. (2024). AI and ML Applications in RealTime Energy Monitoring and Optimization for Residential Solar Power Systems. Available at SSRN 5116062.
[35] Wankhede, S. S., Jain, K., & Pikle, N. (2026). Data-Driven Predictive Maintenance and Fault Diagnostics in Renewable-Powered Microgrids Using AI-Enabled Forecasting Models.
[36] Shukla, G., Shukla, A. K., Shukla, K. N., & Dixit, S. (2025). IoT-enabled solar PV systems: real-time monitoring and efficiency optimization. Results in Optics, 100949.
[37] Koumoulos, P. P., Mazarakis, L., Katsoulis, S., Zantalis, F., & Koulouras, G. (2026). IoT and AI-Driven Approaches for Energy Optimization in Off-Grid Solar Systems. Engineering Proceedings, 124(1), 67.
[38] Mohammad, A., & Mahjabeen, F. (2023). Revolutionizing solar energy: The impact of artificial intelligence on photovoltaic systems. International Journal of Multidisciplinary Sciences and Arts, 2(3), 591856.
[39] Mamta, Karri, S. R., & Burri, S. R. (2026). Smart Solar Energy Management through IoT and AI Integration: Architectures, Applications, and Future Trends. Solar Energy Optimization Using Generative Artificial Intelligence, 347-376.
[40] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems, 30, 3149–3157.
[41] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. https://doi.org/10.1145/2939672.2939785
[42] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735
[43] Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555.
[44] Graves, A., & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. Neural Networks, 18(5–6), 602–610. https://doi.org/10.1016/j.neunet.2005.06.042
[45] Bai, S., Kolter, J. Z., & Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv preprint arXiv:1803.01271.
[46] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30, 5998–6008. https://arxiv.org/abs/1706.03762
[47] Nie, Y., Nguyen, N. H., Sinthong, P., & Kalagnanam, J. (2023). A time series is worth 64 words: Long-term forecasting with transformers. International Conference on Learning Representations (ICLR 2023). https://arxiv.org/abs/2211.14730