POLITECNICO DI TORINO
Repository ISTITUZIONALE

Minute-Scale Solar Irradiance Nowcasting with Infrared Sky Imagery

Original
Minute-Scale Solar Irradiance Nowcasting with Infrared Sky Imagery / Bortolotti, S., Rege Cambrin, D., Garza, P., Leva, S., Ogliari, E.. - In: IEEE ACCESS. - ISSN 2169-3536. - ELETTRONICO. - (2026). [10.1109/ACCESS.2026.3746430]

Availability:
This version is available at: 11583/3017516 since: 2026-11-04T13:34:19Z

Publisher:
IEEE

Published
DOI:10.1109/ACCESS.2026.3746430

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

11 November 2026

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2026.3746430

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2024.0429000

Minute-Scale Solar Irradiance Nowcasting with Infrared Sky Imagery

Simone Bortolotti[1], Daniele Rege Cambrin[2], Paolo Garza[2], Sonia Leva[1], and Emanuele Ogliari[1]

[1]Politecnico di Milano, Milan, Italy (e-mail: name.surname@polimi.it)
[2]Politecnico di Torino, Turin, Italy (e-mail: name.surname@polito.it)

Corresponding author: Daniele Rege Cambrin (e-mail: daniele.regecambrin@polito.it).

Simone Bortolotti and Daniele Rege Cambrin contributed equally to this work.

ABSTRACT Accurate one-minute-ahead solar irradiance forecasts are essential for real-time photovoltaic control and grid balancing. While persistence is strong on stable days, it tends to degrade during the most critical and highly variable periods. This work presents the Simple Irradiance-forecasting Model, which predicts Beam Horizontal Irradiance from sequences of long-wavelength infrared all-sky images and co-located meteorological measurements. The model combines a Convolutional Long Short-Term Memory encoder to capture cloud-motion dynamics with a ConvNeXt backbone, a modern CNN architecture, and a regression head conditioned on the solar elevation angle. To improve operational stability, the network predicts Beam Horizontal Irradiance, which is then converted into one-minute increments of Global Horizontal Irradiance through a physics-based postprocessing scheme that enforces admissible irradiance bounds. The proposed approach outperforms persistence on both a full dataset and a high-variability subset, achieving root mean square errors of 49.51 (12.15% forecast skill) and 69.71 (15.20% forecast skill), respectively. Evaluation on an additional dataset with lower Beam Horizontal Irradiance variability further demonstrates the robustness of the proposed pipeline, achieving performance comparable to persistence under conditions with limited scope for forecast improvement. Ablation results show that postprocessing and temporal encoding are critical, and case-study days confirm lower cumulative error under rapidly changing cloud conditions.

INDEX TERMS Deep Learning, Infrared Imagery, Nowcasting, Renewable Energy, Sky Camera, Solar Irradiance

I. INTRODUCTION

PhotoVoltaic (PV) generation is inherently variable due to cloud dynamics, atmospheric conditions, and diurnal cycles of the sun. At high penetration levels, this variability poses a challenge to grid operators, who must maintain real-time balance between supply and demand while respecting the operational constraints of dispatchable units and storage systems [1]. Minute-scale solar irradiance forecasts are particularly valuable for intra-hour scheduling, ramp-rate control, and inverter-level curtailment, where even small forecast errors can translate into unnecessary reserves or increased wear on flexible assets [2]. For these applications, nowcasting horizons of 1–10 minutes are critical, but also the most challenging, because cloud fields evolve rapidly and exhibit strong nonlinear dynamics.

Traditional short-term solar forecasting approaches rely on time series models, Numerical Weather Prediction (NWP) [3], or sky-imaging systems [2]. Ground-based all-sky imagers provide hemispheric views of the cloud field and have emerged as a key technology for solar nowcasting, enabling data-driven models to learn spatiotemporal patterns that precede irradiance ramps, which are not detectable with only a timeseries of irradiance. However, most existing approaches depend on visible-spectrum cameras and handcrafted features or CNNs applied frame by frame or with simple temporal aggregation [4], [5]. Visible imagers suffer from strong solar glare near the sun disk and operate reliably only under suitable illumination [6]. In contrast, Long-Wavelength Infra-Red (LWIR) imagery exhibits a stronger thermal contrast between clouds and the background sky, rendering solar glare negligible. Moreover, it operates reliably in low-light conditions [7]. Despite these advantages, the use of LWIR all-sky imagery for minute-scale irradiance nowcasting remains relatively underexplored. Furthermore, although Deep Learning (DL) models based on all-sky imagery have shown promising performance, most existing research has focused

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2026.3746430

on forecasting horizons of 5–15 minutes or longer, leaving the one-minute-ahead horizon relatively underexplored. At this timescale, irradiance dynamics are dominated by rapid cloud motion, and the persistence model becomes extremely difficult to outperform, particularly under stable atmospheric conditions. In addition, many existing approaches rely on complex architectures or multimodal data fusion, which can limit their applicability in real-time or resource-constrained environments. As a result, it remains unclear to what extent accurate and reliable one-minute-ahead forecasts can be achieved using lightweight and deployable models, and which modeling choices are most effective at this timescale [8]. Addressing this challenge requires capturing both the spatial structure of clouds and their temporal evolution, a task that cannot be effectively handled by traditional statistical or persistence-based methods. DL approaches, and in particular Convolutional Long Short-Term Memory (ConvLSTM) based architectures, provide a natural solution by enabling the learning of spatio-temporal representations directly from sequences of sky images [9]. This makes them especially well-suited for modeling the high-frequency variability inherent to ultra-short-term solar forecasting.

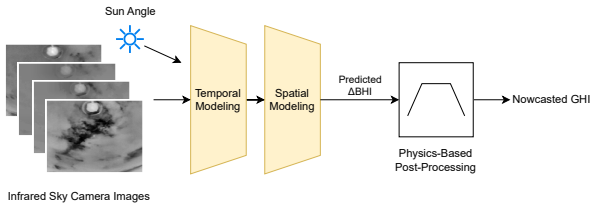


FIGURE 1. **Visual overview of the proposed forecasting framework. The system processes sequences of infrared sky images and solar angles to model cloud dynamics.**

To address this problem, this work develops a DL framework for one-minute-ahead solar irradiance nowcasting based on LWIR all-sky imagery, coupled with a physically based postprocessing pipeline, as summarized in Figure 1. Rather than directly forecasting the final irradiance output, the proposed approach targets the variation in Beam Horizontal Irradiance (BHI), which are subsequently transformed into stable and physically consistent GHI nowcasts through operational irradiance constraints. The study utilizes data collected from an LWIR all-sky imager co-located with a high-precision meteorological station. Hemispheric infrared sky images were captured every 60 seconds, along with one-minute irradiance measurements and auxiliary meteorological variables, to predict the next-minute irradiance. To exploit the thermal imagery while respecting physical constraints, the proposed Simple Irradiance-forecasting Model (SIM) combines a ConvLSTM [10] encoder for short-range temporal modeling, a ConvNeXt [11] backbone for spatial feature extraction, and a postprocessing scheme that operates on BHI and irradiance increments rather than absolute Global Horizontal Irradiance (GHI) values.

The main contributions of this work, targeting accurate and deployable one-minute-ahead solar irradiance nowcasting, are as follows:

- An LWIR-based nowcasting framework with a compact ConvLSTM–ConvNeXt architecture deployed on co-located infrared imaging and high-precision pyranometric measurements.
- A physically motivated target and postprocessing strategy that predicts BHI, ensures physically plausible results, and converts outputs to GHI.
- An extensive ablation study on architecture components providing quantitative evidence of the benefits of each design choice.
- A comparison against 3D-CNN and Vision Transformer baselines trained under an identical pipeline, showing that SIM outperforms a similarly-sized Transformer and a substantially lighter 3D-CNN, achieving the best absolute accuracy among the tested architectures.
- An evaluation on held-out data with markedly lower irradiance variability, confirming the robustness of the framework outside the training-period conditions.

The remainder of the paper is structured as follows. Section II presents the relevant literature. Section III describes the real-world dataset, including a 24-day lower-variability holdout, the proposed forecasting framework, and the evaluation methodology. The proposed method is validated against a persistence baseline and against alternative deep learning architectures. Section IV presents the experimental results. Section V discusses the implications and limitations, while Section VI concludes the paper and outlines future work.

II. RELATED WORKS

This section reviews the relevant literature on solar irradiance forecasting, with an emphasis on sky-imaging systems, DL architectures for spatiotemporal modeling, and physically based postprocessing approaches.

A. DEEP LEARNING FOR SKY IMAGE-BASED FORECASTING

Statistical and machine learning (ML) methods based purely on irradiance or power measurements capture temporal autocorrelation but lack explicit information about cloud motion, which is the dominant driver of rapid irradiance fluctuations at minute scales [2], [12]. NWP models, while indispensable for hourly to day-ahead horizons, are too coarse in space and time and often fail to resolve local cloud structures relevant for one-minute forecasts [2], [13]. DL models have become the state-of-the-art for solar irradiance nowcasting, consistently outperforming traditional naive, statistical, and physical models [14]. The advent of deep CNNs has enabled the end-to-end learning of cloud-to-irradiance mappings with sky imagery. Early CNN-based approaches processed single sky images or short sequences using 2D architectures, achieving improvements over persistence and classical methods for horizons up to 10 minutes. However, these models often treat temporal dependencies superficially by concatenating

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2026.3746430

frames as additional input channels or averaging predictions from individual frames, rather than explicitly modeling cloud motion dynamics [15], [16].

Recent work has explored recurrent and spatiotemporal architectures to better capture cloud evolution. ConvLSTM networks extend standard LSTMs by replacing fully connected layers with convolutional operations, preserving spatial structure while modeling temporal dependencies [17], [18]. ConvLSTM has been successfully applied to precipitation nowcasting and video prediction tasks, demonstrating strong performance on spatiotemporal sequences [10]. For solar forecasting, ConvLSTM provides a computationally efficient alternative to 3D CNNs and attention-based transformers, particularly when the relevant temporal context is short [19], [20]. Transformer-based architectures with self-attention mechanisms have also been proposed for solar nowcasting, leveraging their ability to capture long-range dependencies in image sequences [21], [22]. While effective for longer horizons and larger spatial domains, transformers require substantial computational resources and large training datasets to avoid overfitting [23], [24]. For one-minute horizons with sequences of 3–5 images, the added complexity of transformers may not be justified compared to more efficient recurrent architectures; we evaluate this claim empirically in Section IV.

Additionally, recent work has demonstrated that DL architectures can substantially enhance solar nowcasting performance by jointly processing image sequences and auxiliary meteorological data, while using hybrid architectures that combine convolutional feature extraction with temporal modeling [16], [25], [26]. However, the majority of nowcasting research addresses horizons of 15 minutes or more [27]. Only a few recent studies have specifically addressed the one-minute-ahead problem, revealing the unique challenges associated with this timescale. At this resolution, the persistence model becomes a surprisingly strong benchmark, especially during stable, clear-sky, or overcast conditions [28], placing high demands on the performance of any advanced model. Many recent models exploit multi-modal fusion to capture the complex, high-frequency dynamics present at the one-minute scale [29], [30]. Our work builds on these insights but diverges by focusing on the efficacy of a simpler, physics-based architecture with a highly optimized processing pipeline.

B. PHYSICAL CONSTRAINTS AND POSTPROCESSING

Most DL forecasting models produce raw predictions that may violate physical constraints—such as negative irradiance, values exceeding the clear-sky envelope, or unrealistic ramp rates. Quality control and postprocessing schemes are essential for operational deployment. Common approaches include simple clipping to physically valid ranges, statistical postprocessing (e.g., quantile mapping), and hybrid methods that blend model predictions with persistence or clear-sky estimates [31], [32]. Preprocessing and postprocessing are essential for improving model accuracy. Standard image preprocessing steps include fish-eye distortion correction and high dynamic range (HDR) adjustments [33], as well as the removal of atmospheric and radiation-related artifacts [34]. On the output side, postprocessing techniques are often used to correct for systematic model biases [35] and enforce physical constraints on the predictions. Careful quality control of the target variable itself is also paramount, as pyranometer measurement errors can severely degrade model performance [36].

Recent work has emphasized the benefits of predicting irradiance increments or ramp events rather than absolute values, particularly for short horizons where persistence is strong [37]. Increment-based targets mitigate bias accumulation in sequential forecasts and naturally account for the high temporal correlation of irradiance. However, translating model outputs back to absolute irradiance requires careful handling of boundary conditions and error propagation.

Another emerging direction is the use of physically decomposed targets. Rather than directly predicting GHI, some methods forecast Direct Normal Irradiance (DNI) and Diffuse Horizontal Irradiance (DHI) separately [18], exploiting their distinct variability characteristics: DNI exhibits rapid sub-minute fluctuations driven by cloud edges and optical thickness, while DHI varies more slowly due to atmospheric scattering [38]. For one-minute horizons, assuming near-constant DHI and predicting only DNI changes can reduce model complexity and improve robustness. Most research related to PV systems focuses on predicting GHI, as it directly relates to PV panel output [39]. Conversely, studies related to Concentrating Solar Power (CSP) often predict DNI, which is the relevant component for CSP systems [40], [41]. This distinction is crucial, as it implies that one component may be inherently more or less predictable than another under specific conditions. Our work proposes a novel methodology in which BHI is predicted as an intermediate step to enhance the final GHI forecast. $BHI = DNI \cos\theta$, where θ is the sun angle.

III. MATERIAL AND METHODS

This section details the data acquisition process, the dataset characteristics, the preprocessing and target definition procedures, the proposed forecasting model, and the evaluation framework used in this study.

A. DATA ACQUISITION AND SITE

The data for this study were collected at the SolarTechLAB facility at the Politecnico di Milano, Italy (45.50°N, 9.16°E, 140m a.s.l.). Milan features a humid subtropical climate (Köppen classification: Cfa), with hot, humid summers and cold, foggy winters. The instrumentation includes an all-sky imager (Figure 2) and a high-precision meteorological station (Figure 3).

1) All-Sky Imager

A Reuniwatt Sky InSight imager was used to capture hemispheric sky images. This device uses a LWIR camera to record

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2026.3746430

images with a 640x480 pixel resolution at a 60-second interval. The infrared spectrum is advantageous as it minimizes solar glare and operates continuously day and night. Key specifications are listed in Table 1.

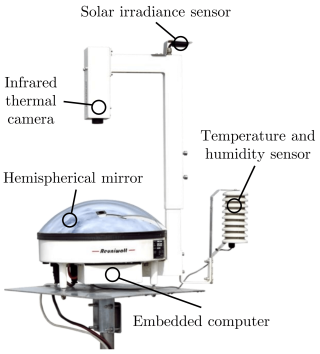


FIGURE 2. Instrumentation of the Reuniwatt Sky InSight imager station. The device captures hemispheric sky images using a LWIR camera.

TABLE 1. Technical specifications of the All-Sky Imager

Parameter	Specification
Operating Conditions	
Operation	Day and night, identical accuracy
Field of view	360° × 180°
Resolution	640×480 pixels
Image format	8bit-JPEG2000
Imaging interval	60 seconds
Environmental Conditions	
Operating temperature	−20 °C to +60 °C
IP rating	IP65
Power and Communications	
Power consumption	85 W
Power supply	110–220 VAC, 50–60 Hz
Communications	TCP/IP Ethernet 100 Mbit/s
Controller	Embedded PC
Power backup	Uninterrupted power supply
Mechanical Specifications	
Dimensions	350×615×800 mm
Weight	10 kg
Sensors	
Irradiance reference cell	Global Horizontal Irradiance (GHI)
Environmental parameters	Ambient temperature, relative humidity

2) Meteorological Station

Co-located with the imager, a meteorological station recorded key environmental variables. Solar irradiance components were measured using secondary-standard pyranometers (ISO 9060), providing GHI and DHI. The station logged data at a 10-second frequency, which was then averaged to one-minute intervals to synchronize with the sky imager. Sensor specifications are provided in Table 2.

B. DATASET

The primary dataset was collected over 54 days, from August 1 to September 30, 2023. To ensure data quality and relevance

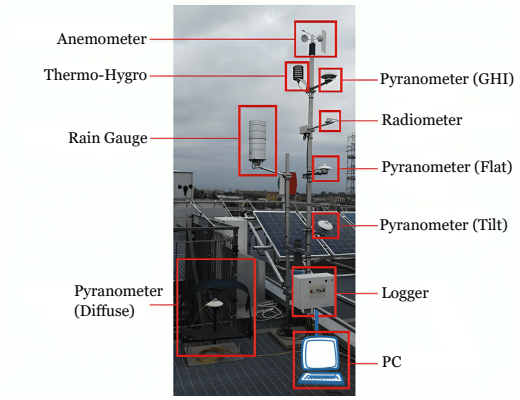


FIGURE 3. High-precision meteorological station co-located with the sky imager, used to record key environmental variables. Instrumentation includes secondary-standard pyranometers (ISO 9060) for GHI and DHI, radiometer, anemometer, rain gauge, thermo-hygrometer, and auxiliary devices for data logging and synchronization. The station logged data at a 10-second frequency and synchronized to one-minute intervals.

TABLE 2. Specifications of the solar irradiance pyranometer

Parameter	Specification
Standard	Secondary standard (ISO 9060)
Measurement Range	$< 2000\ \mathrm{W/m^2}$
Spectral Range	300–3000 nm
Achievable Daily Uncertainty	$< 2\,\%$
Directional Response	$< \pm 5.4\ \mathrm{W/m^2}$
Thermal Drift	$< 2\,\%$

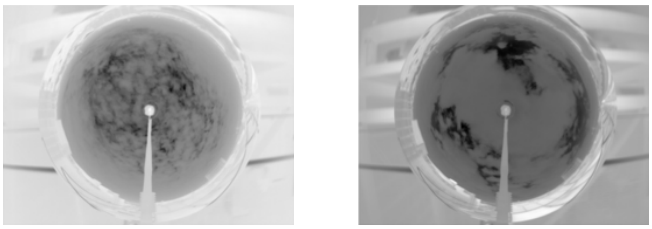
for PV applications, data between 06:30 and 16:30 were retained, excluding periods of low solar elevation where image quality is degraded. The dataset consists of 6480 sequences of consecutive single-channel infrared images (as shown in Figure 4) and the sun elevation angle at the last time step. This angle is included as a static numerical input for the sequence, as it changes negligibly over the four-minute interval, enabling physical correlation with clear-sky irradiance. The dataset comprises a total of 29,400 one-minute resolution single data points, each containing four consecutive images and corresponding meteorological data. To assess robustness beyond the primary training period, an additional 24-day holdout set was collected in October 2023 using the same acquisition and preprocessing pipeline, yielding 596 valid sequences per day. This period is characterized by substantially lower irradiance levels and, consequently, lower absolute variability than the August-September dataset. The October data were reserved exclusively for final evaluation and were not used at any stage of model development or hyperparameter selection.

C. TASK DEFINITION

The task we aim to solve can be framed as a forecasting task. Given a timeseries of images $I = \{I_0, I_1, I_2, I_3\}$ and time-synchronized meteorological measurements $X = \{x_0, x_1, x_2, x_3\}$, the objective is creating a model that given I and X can predict the *GHI* at the next timestep 5. Preliminary evaluation suggested that a time series of length 4 was

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2026.3746430



(a) (b)

FIGURE 4. Examples of sky images captured by the LWIR camera, illustrating cloud coverage under different atmospheric conditions.

optimal, as shown in Appendix A.

D. TARGET DEFINITION

The initial target was GHI, but due to inconsistencies in clear-sky assumptions and aerosol interference, this proved unstable. Instead, the BHI was defined as the target, computed as:

$$BHI = GHI - DHI = DNI \cos\theta$$

where *DNI* is the Direct Normal Irradiance, *DHI* is the Diffuse Horizontal Irradiance and θ is the sun angle.

To mitigate noise spikes from measurement artifacts, a rule-based filter was applied. Spikes in the GHI signal were identified by comparing them against a secondary GHI measurement and the site's PV power output. If a spike was absent in these correlated signals (Spearman correlation > 0.97 and 0.79, respectively), the anomalous value was corrected using last-observation-carried-forward imputation

E. FORECASTING MODEL: SIM ARCHITECTURE

The SIM is a DL architecture that handles the time dimension and then regresses the target using ConvLSTM and ConvNeXt as shown in Figure 5.

ConvLSTM provides a light and effective alternative to transformers. While the attention mechanism is effective for long sequences, it is more computationally intensive, and our time series are short (4 points). ConvLSTM retains all the advantages of LSTMs, while also preserving the spatial dimension, which is crucial for modeling the evolution of an image over time.

ConvNeXt is instead selected as a feature extractor, as it retains the computational advantages of CNNs and is less prone to overfitting than ViTs while incorporating all recent advancements, such as GELU activation, depth-wise separable convolutions, and patchify layers inspired by transformers.

The overall network works in the following way:

- **Sequence Encoding:** The network receives as input a sequence of four grayscale LWIR sky images of size 128×128. A ConvLSTM layer processes the temporal sequence while preserving the spatial structure, extracting temporal dynamics of cloud motion. The resulting spatiotemporal representation is composed of 16 feature maps of size 128×128.

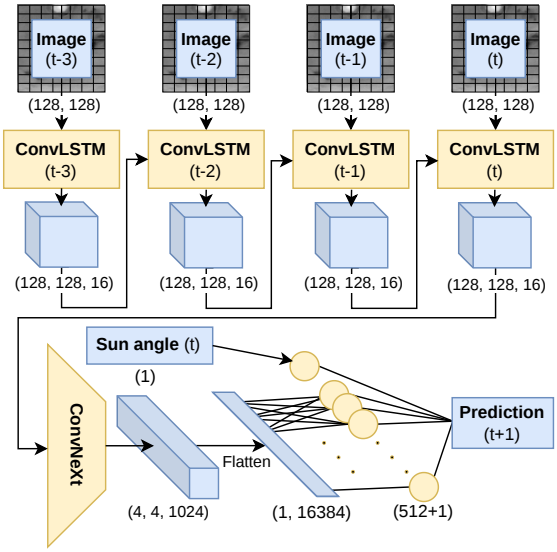


FIGURE 5. Proposed SIM architecture, where a ConvLSTM encoder processes four consecutive 128 × 128 LWIR sky images, a ConvNeXt backbone refines the resulting spatiotemporal features, and a fully connected head conditioned on the current solar elevation angle outputs the one-step-ahead BHI prediction.

- **Spatial Feature Extraction:** The ConvNeXt-Base network refines this representation, downsampling to a $4 \times 4 \times 1024$ feature map.
- **Prediction Head:** The flattened ConvNeXt output is projected to a 512-dimensional latent vector. The normalized solar elevation angle is concatenated before the final regression layer produces a one-step BHI forecast.

F. COMPARISON ARCHITECTURES

To contextualize SIM's performance against alternative spatiotemporal architectures, two additional baselines were trained and evaluated under the same data splits, preprocessing, postprocessing, and training protocol.

3D-CNN. The four-frame sequence is treated as a spatiotemporal volume and processed by four Conv3D–BatchNorm3D–GELU–MaxPool3D stages with channel widths of 32–64–128–256, progressively integrating spatial and temporal information. The temporal dimension is collapsed by the final stage, after which global average pooling yields a 256-dimensional feature vector. This feature is combined with the solar elevation angle and passed to the same regression head used by SIM. The resulting architecture contains approximately 1.1M parameters and provides a lightweight convolutional alternative to recurrent temporal modeling.

Vision Transformer (ViT). Following recent transformer-based solar nowcasting approaches [22], each frame is independently encoded by a pretrained ViT-B/16 backbone. Grayscale images are replicated across three channels and resized to 224×224, producing one CLS token per frame.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2026.3746430

A lightweight two-layer temporal transformer then attends across the four tokens to capture temporal dependencies and ordering. The resulting representations are mean-pooled into a 768-dimensional summary vector, concatenated with a projected solar elevation feature, and passed through a regression head equivalent to that used by SIM. With the pretrained backbone, the model contains approximately 87.6M parameters, comparable in scale to SIM's 96.0M, enabling a parameter-scale comparison between attention-based and convolutional-recurrent temporal modeling.

G. IMAGE PREPROCESSING

The preprocessing pipeline applied to the all-sky images consisted of the following steps:

1) Cropping the raw 640×480 images into a square format.
2) Applying a circular mask to remove non-informative background outside the sky dome.
3) Estimating the Sun position from the solar elevation and azimuth angles and extracting a 192×192 region centered on the solar disk, empirically determined to cover cloud motion within four minutes, as extensively described below.
4) Enhancing local contrast via Contrast-Limited Adaptive Histogram Equalization (CLAHE), with parameters `tileGridSize = (4,4)` and `clipLimit = 4.0`, improving cloud-edge visibility.
5) Resizing images to 128×128 pixels to reduce computational load.
6) Normalizing intensity values using z-score standardization across the dataset.

The Sun-centered crop was computed by projecting the solar position into image coordinates using the measured solar elevation angle θ_{elev} and azimuth angle θ_{azim}. The radial distance from the image center was defined as:

$$r = \left(1 - \frac{\theta_{elev}}{90}\right) \frac{W}{2} \left(1.025 + 0.020(|h - 11|)\right) \tag{1}$$

where W is the image width and h is the acquisition hour. The term $(1.025 + 0.020(|h - 11|))$ was empirically introduced to compensate for fisheye lens distortion and perspective effects.

The solar azimuth angle, originally expressed in degrees, was corrected using an empirical angular offset to compensate for camera orientation and projection alignment, and was subsequently converted into radians to obtain the angular coordinate:

$$\alpha = (\theta_{azim} + 183.8)\frac{\pi}{180} \tag{2}$$

The Sun position in image coordinates was then obtained as:

$$x_{sun} = \frac{W}{2} + r\sin(\alpha) + 4 \tag{3}$$

$$y_{sun} = \frac{H}{2} - r\cos(\alpha) - 13 \tag{4}$$

where H is the image height, while the offsets $(+4, -13)$ were empirically calibrated to align the projected coordinates with the actual solar disk position in the camera reference system.

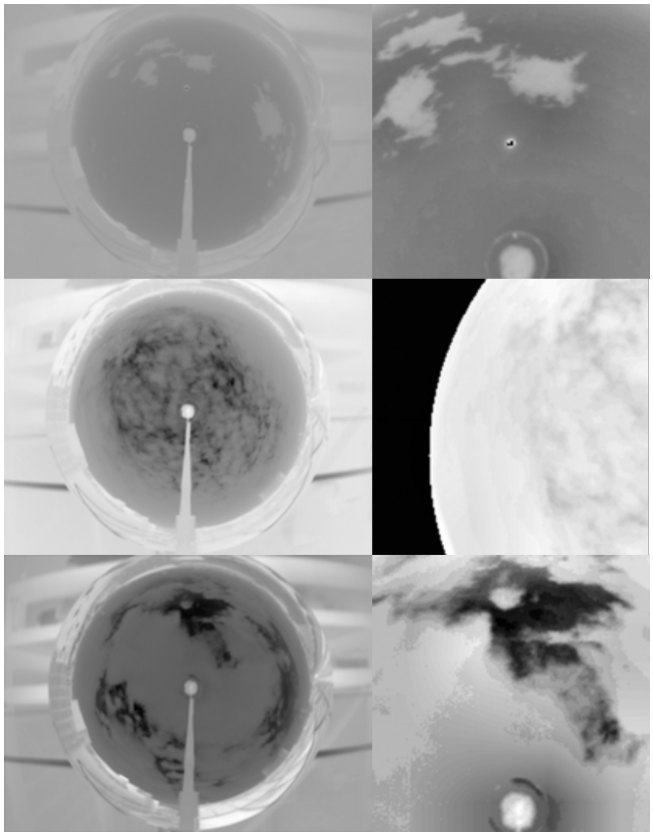


FIGURE 6. Representative LWIR all-sky images before (left) and after (right) the preprocessing pipeline. Three examples under different sky conditions are shown.

Finally, a fixed-size crop centered on (x_{sun}, y_{sun}) was extracted:

$$I_{crop} = I\left[x_{sun} - \frac{S}{2} : x_{sun} + \frac{S}{2},\, y_{sun} - \frac{S}{2} : y_{sun} + \frac{S}{2}\right] \tag{5}$$

where $S = 192$ pixels denotes the crop size. This procedure ensured that the Sun was consistently positioned near the image center, allowing the forecasting model to focus on cloud dynamics surrounding the solar disk, which are the most relevant features for very short-term irradiance variability.

The projection coefficients and positional offsets were empirically calibrated for the specific optical configuration and installation geometry of the employed all-sky imager. Consequently, the exact numerical values are not expected to generalize across different camera systems, lenses, or geographical installations. However, the proposed preprocessing methodology itself remains fully transferable, as it only requires a simple calibration procedure to align the projected solar coordinates with the observed Sun position in the image plane. No handcrafted cloud descriptors or manually engineered image features were extracted. Instead, the preprocessed Sun-centered image crops were directly used as input tensors to the forecasting network, allowing the convolutional

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2026.3746430

layers to automatically learn relevant spatial and temporal representations from the normalized pixel intensities.

The final image input at each timestep can therefore be represented as:

$$X_t \in \mathbb{R}^{128\times128\times1} \tag{6}$$

where each pixel corresponds to a normalized grayscale LWIR radiance value after preprocessing.

H. FORECAST POSTPROCESSING

To better support one-minute grid applications, the forecasting task is expressed in terms of irradiance increments rather than absolute values, i.e., the model output is transformed into $\Delta GHI_t = GHI_t - GHI_{t-1}$ instead of GHI_t itself, in line with prior work showing benefits of relative irradiance targets for short-term forecasting [37]. The network predicts absolute BHI, which is then converted into ΔGHI_t through a four-step postprocessing procedure.

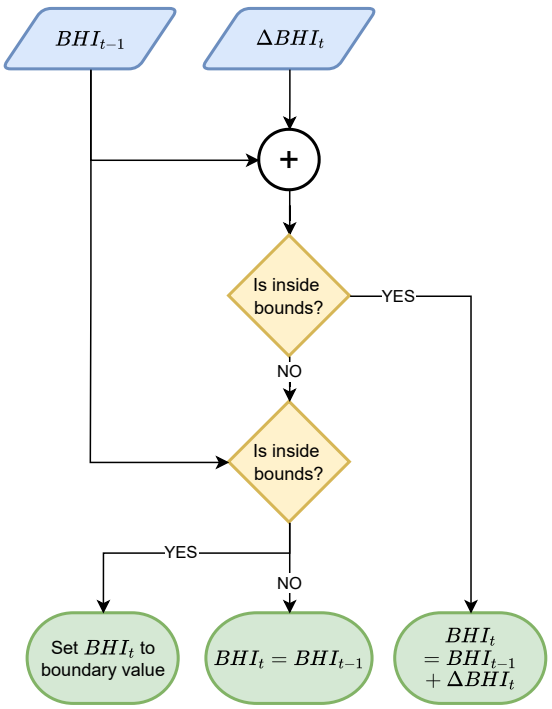


FIGURE 7. Operational-bounds postprocessing of one-step-ahead BHI forecasts, where the predicted increment ΔBHI_t is added to the previous observation BHI_{t-1}, the candidate value is checked against physically derived irradiance limits, and BHI_t is either accepted, projected to the closest admissible boundary, or reverted to persistence when both the previous and candidate values lie outside the bounds.

First, the predicted BHI values are denormalized (inverse z-score) to recover physical units. Second, the modelled BHI increment is obtained as $\Delta BHI_t = \hat{BHI}_t - \hat{BHI}_{t-1}$. Third, an operational-bounds algorithm decides whether to keep this increment or replace it with a more conservative alternative. The operational bounds are intentionally narrower than the theoretical physical irradiance range. Their objective is not only to reject physically implausible predictions but also to suppress small fluctuations occurring in operating regimes of limited practical relevance for one-minute PV nowcasting. Near the clear-sky limit, small predicted BHI variations are unlikely to correspond to meaningful cloud occlusions over the solar disk. Similarly, under conditions of almost complete cloud coverage, small BHI variations have limited operational impact on PV generation. Consequently, the postprocessing focuses on the irradiance range where accurate minute-ahead predictions are most beneficial. Given the previous true BHI, BHI_{t-1}, a candidate value $\tilde{BHI}_t = BHI_{t-1} + \Delta BHI_t$ is computed and checked against the admissible range. If $\tilde{BHI}_t$ moves from inside to outside the bounds, the forecast is projected to the crossed boundary; if both BHI_{t-1} and $\tilde{BHI}_t$ lie outside the bounds, the increment is set to zero, corresponding to the persistence benchmark ($GHI_{t+1} = GHI_t$). Otherwise, the model increment is accepted unchanged. This design adheres to common quality-control practices for solar-radiation time series, where physical and range-based limits are employed to reject unrealistic values. The procedure is depicted in Figure 7.

Finally, the accepted ΔBHI_t is converted into ΔGHI_t using the relation $GHI_t = BHI_t + DHI_t$ and the approximation $\Delta DHI_t \approx 0$, which is reasonable at the one-minute horizon because BHI exhibits much stronger sub-minute variability than the diffuse component [42], [43]. In this formulation, a zero increment corresponds to persistence, while non-zero increments reflect model-detected changes; exploiting the strong correlation between successive predictions in this way reduces bias accumulation and limits unrealistically large ramps. A visual example of the algorithm is illustrated in Figure 8.

I. ABLATION MODELS

To understand the impact of different design choices, we tested 6 different configurations to ablate the various components:

- SIM (the proposed framework): the input images are processed through ConvLSTM and ConvNeXt, and the exogenous sun elevation is provided in the fully connected layers. The output is the BHI, which is postprocessed to obtain the final GHI.
- SIM without Postprocessing (OP): the framework is equal to SIM, but without the postprocessing step.
- SIM without Boundaries (NB): it omits the high/low boundary constraints in the postprocessing stage. Compared to SIM, it isolates the effect of boundaries on the overall performance.
- SIM with GHI Target (GHI): the framework is equal to SIM, but the network is outputting the GHI directly
- NOEXOG: the SIM framework without sun elevation information.
- NOLSTM: the SIM framework without the ConvLSTM layers to model the time.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2026.3746430

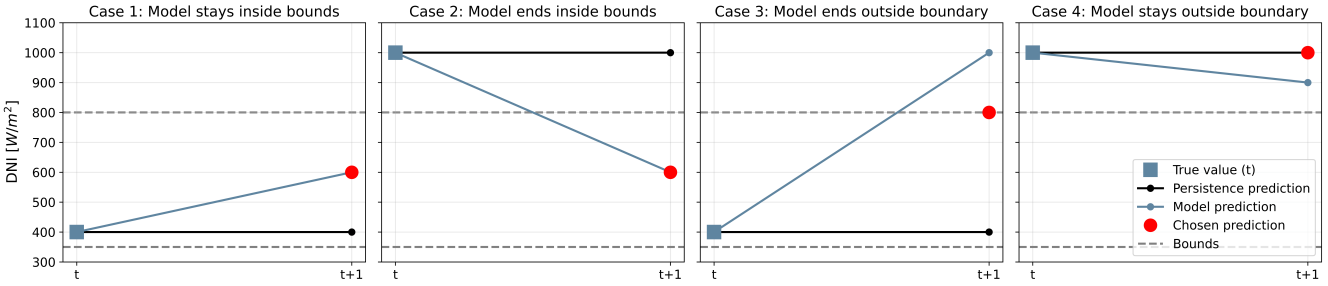


FIGURE 8. **Postprocessing algorithm behavior under four operational scenarios. Each panel shows the previous true BHI value at time t (blue square), the persistence prediction (black line at $t+1$), the raw model prediction (blue line to $t+1$), and the final chosen prediction (red circle). The horizontal dashed lines represent the operational bounds. Case 1: Model prediction stays within bounds and is accepted. Case 2: Model prediction ends within bounds from an internal starting point and is accepted. Case 3: Model prediction crosses from inside to outside the bounds and is clipped to the boundary value. Case 4: Both current and predicted values lie outside bounds, so the prediction reverts to persistence (zero increment).**

J. EVALUATION

The SIM model was compared to a Persistence benchmark, defined as $GHI_{t+1} = GHI_t$, which is a strong baseline for one-minute forecasting horizons. The performance was evaluated using the Root Mean Square Error (RMSE), Forecast Skill Score (FS), and R^2 score (R^2).

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2} \tag{7}$$

$$FS = 1 - \mathrm{RMSE}_m/\mathrm{RMSE}_p \tag{8}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{9}$$

where $\hat{y}$ is the model prediction, y is the ground truth value, $\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$ is the mean of the ground truth data, RMSE_m is the model RMSE, while RMSE_p is the persistence one. To complement forecast-accuracy metrics, model size (number of trainable parameters) is reported for the comparison architectures in Section IV as a proxy for deployability in resource-constrained, real-time settings.

IV. RESULTS

This section reports the experimental setup, the quantitative results, and the ablation studies.

A. EXPERIMENTAL SETUP

The dataset was split using a Leave-One-Out Cross-Validation (LOO) framework across all 54 days to obtain around 600 time series in each test set. Days were categorized into clear, variable, rainy, and other to ensure balanced train-validation splits. Rainy days (5 of 54) contain water droplets on the all-sky camera mirror, compromising image acquisition in part of the records. Clear-sky days (18 of 54) with negligible cloud presence, where GHI dropped below the theoretical clear-sky maximum for less than 30 minutes within the 10-hour window. Variable days (13 of 54) refer to the days with a standard deviation exceeding the 75th percentile of the entire daily dataset GHI. Days not falling in any of the previous categories fall into the other category.

The model was trained according to the hyperparameter configuration reported in Table 3. Training minimizes a squared error loss reweighted by inverse point-wise forecast skill relative to a persistence baseline, emphasizing poorly predicted regions. The network processes four grayscale sky images of size 128×128 using a ConvLSTM layer (16 channels, 3×3 kernel), followed by a ConvNeXt-Base backbone. The extracted features are flattened and passed through a fully connected layer with 512 neurons, where an exogenous variable is concatenated to form a 513-dimensional feature vector used for final regression.

TABLE 3. **Training hyperparameters used for training SIM.**

Parameter	Value
Optimizer	AdamW
Epochs	40
Batch size	32
Initial learning rate	1×10^{-4}
Learning-rate scheduler	ReduceLROnPlateau
Learning-rate reduction	Factor = 0.1
Scheduler patience	3 epochs
Weight decay	5×10^{-5}
Early stopping	Patience = 10 epochs
Loss function	Weighted MSE
Gradient clipping	Max norm = 1.0
Training precision	Mixed precision (16-bit)
Training strategy	Distributed Data Parallel (DDP)

The operational bounds of the postprocessing step are $[0.135 \cdot GHI_{clear}, 0.9 \cdot GHI_{clear}]$, where GHI_{clear} is the clearsky GHI. These coefficients were empirically selected on a validation subset as an engineering compromise between suppressing noisy predictions near the extremes of the irradiance range and preserving the operating region most relevant for PV applications. Although the exact numerical values are heuristic and may require recalibration for different sites or sensors, the proposed methodology is not expected to be highly sensitive to moderate variations of these coefficients because the bounds remain anchored to physically meaningful clear-sky irradiance estimates.

The 3D-CNN and ViT baselines (Section III-F) were trained using the identical LOO-CV protocol, optimizer,

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2026.3746430

TABLE 4. Results on the two sets for persistence and the proposed model

	FS (%) ↑	R^2 (%) ↑	RMSE ↓
Persistence	-	94.21	56.36
SIM	**12.15**	**95.53**	**49.51**

(a) Full Set

	FS (%) ↑	R^2 (%) ↑	RMSE ↓
Persistence	-	90.12	82.21
SIM	**15.20**	**92.90**	**69.71**

(b) Variable Set

scheduler, and preprocessing/postprocessing pipeline as SIM, ensuring a controlled comparison.

B. QUANTITATIVE RESULTS

The proposed SIM architecture was evaluated against the persistence baseline across the complete dataset (Full Set) and a subset specifically containing days with high irradiance variability (Variable Set). As detailed in Table 4, the SIM model consistently outperforms the persistence baseline in both scenarios.

On the Full Set, the SIM model achieved an RMSE of 49.51 W/m^2, representing a 12.15% improvement in Forecast Skill (FS) over the persistence model (56.36 W/m^2). The coefficient of determination (R^2) increased from 94.21% to 95.53%, indicating that the model successfully captures the majority of irradiance dynamics, even when averaged across stable clear-sky days where persistence is naturally strong.

The model's robustness is most evident in the Variable Set, which isolates days with high cloud-induced volatility (standard deviation $>$ 75th percentile). Here, the difficulty of the forecasting task increases significantly, as reflected by the baseline RMSE rising to 82.21 W/m^2. In this challenging regime, the SIM model maintained a strong performance, reducing the RMSE to 69.71 W/m^2. This corresponds to a Forecast Skill of 15.20%, confirming that the architecture effectively learns to anticipate rapid cloud occlusions rather than simply relying on the previous time-step's value.

C. ABLATION STUDIES

To validate the contribution of each methodological component, we conducted an extensive ablation study. Table 5 summarizes the performance of six model configurations, highlighting the necessity of the proposed pipeline.

TABLE 5. Ablation of different components of SIM

Model	FS (%) ↑	R^2 (%) ↑	RMSE ↓
SIM	**12.15**	**95.53**	**49.51**
OP	-9.35	93.07	61.63
NB	6.06	94.70	52.95
GHI	8.22	95.12	51.73
NOEXOG	11.26	95.40	50.02
NOLSTM	4.70	94.70	53.71

a: Impact of Postprocessing and Constraints

The most critical component proved to be the postprocessing strategy. The OP model (SIM without postprocessing) performed significantly worse than even the persistence baseline, with a negative FS of -9.35% and an RMSE of 61.63 W/m^2. This collapse in performance highlights the difficulty of regressing absolute BHI values directly in a DL framework without explicit differential guidance. By enforcing physical plausibility through boundary constraints, the NB model (No Boundaries) recovered a positive skill (FS = 6.06%), yet it remained inferior to the full SIM model. This demonstrates that while differencing (ΔGHI) is essential for stability, the explicit operational bounds are crucial for filtering unrealistic predictions and preventing large error propagation. The ablation results indicate limited sensitivity to the exact choice of coefficients. Comparable performance is obtained across a range of reasonable values, provided they remain consistent with the underlying physical constraints. Consequently, deployment in different geographical locations is expected to require only lightweight recalibration on a small validation set.

b: Target Selection: GHI vs. BHI

The comparison between SIM (BHI target) and GHI (direct GHI target) validates the choice of BHI as the intermediate variable for prediction. The GHI model achieved an FS of 8.22%, which is competitive but inferior to the 12.15% achieved by SIM. Predicting BHI enables the model to focus on the direct beam component, which is physically correlated with the cloud occlusions visible in the sky images, thereby separating it from the diffuse component that is less distinct in the visual data.

c: Architecture and Exogenous Inputs

The contribution of the ConvLSTM temporal processing is evident when comparing SIM with NOLSTM. Removing the recurrent layers resulted in a decrease in the FS from 12.15% to 4.70%. This sharp decline confirms that treating the input as a sequence to model cloud motion dynamics is far superior to static image processing. Finally, the NOEXOG model, which excludes solar elevation data, resulted in a slight performance decrease (FS = 11.26%). While the visual data implicitly contains solar position information (via the centered crop), explicitly providing the elevation angle helps the dense layers scale the final irradiance output more effectively, yielding a marginal but consistent gain.

D. COMPARISON WITH OTHER ARCHITECTURES

Table 6 compares SIM against the 3D-CNN and ViT baselines described in Section III-F. At a parameter budget comparable to SIM, the ViT baseline underperforms on both the Full Set and the Variable Set, indicating that, for four-frame sequences at this resolution, the inductive biases of the convolutional-recurrent design are more effective than self-attention alone. The 3D-CNN achieves competitive but still consistently lower Forecast Skill, confirming that explicit joint spatio-temporal

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2026.3746430

convolution is a reasonable lightweight alternative but does not match the combination of ConvLSTM temporal encoding and a pretrained spatial backbone used in SIM. Overall, SIM offers the best accuracy for its parameter budget among the tested architectures.

TABLE 6. Comparison of SIM against 3D-CNN and ViT baselines, trained under an identical protocol.

Model	FS (%) ↑	RMSE ↓	Params
Full Set			
SIM	**12.15**	**49.51**	96.0M
ViT	10.95	50.19	87.6M
3D-CNN	9.62	50.94	1.1M
Variable Set			
SIM	**15.20**	**69.71**	96.0M
ViT	13.08	71.46	87.6M
3D-CNN	12.16	72.21	1.1M

E. CASE-STUDY DAYS

This section analyzes the models on three representative days: August 2nd and 26th, 2023, two of the most variable days in the dataset, and August 29th, 2023, a cloudier and less variable day included for contrast. We report the ground-truth irradiance, the corresponding absolute cumulative error for persistence and SIM, and a scatter of predicted against measured irradiance, to better simulate a real use case on a daily basis.

1) Irradiance Variability Patterns

Both case study days exhibit highly complex irradiance patterns characterized by rapid fluctuations and numerous spikes, reflecting challenging meteorological conditions typical of variable cloud coverage. On August 26th, 2023, the irradiance (Figure 9 d) demonstrates a relatively smooth morning ramp-up from approximately 200 W/m^2 at 07:00 to peak values exceeding 900 W/m² around midday, with significant intermittent variations throughout. The afternoon period shows a dramatic decline with irradiance dropping to below 200 W/m^2 by 14:00, indicating substantial cloud interference. August 2nd (Figure 9 a) presents an even more volatile profile, with peak irradiances reaching approximately 1,200 W/m^2 and exhibiting pronounced oscillations between 400-1,000 W/m² during the 10:00-13:00 window. Such rapid variations pose significant challenges for forecasting models, as these conditions represent scenarios where persistence assumptions break down due to non-stationary cloud dynamics. August 29th (Figure 9 g) presents a comparatively less variable profile than the other two days, though still marked by frequent small-scale spikes and extended cloudy periods. Between 9:30 and 10:30, irradiance remains persistently below the lower operational bound, a regime in which the postprocessing step defers entirely to persistence rather than the model's raw prediction.

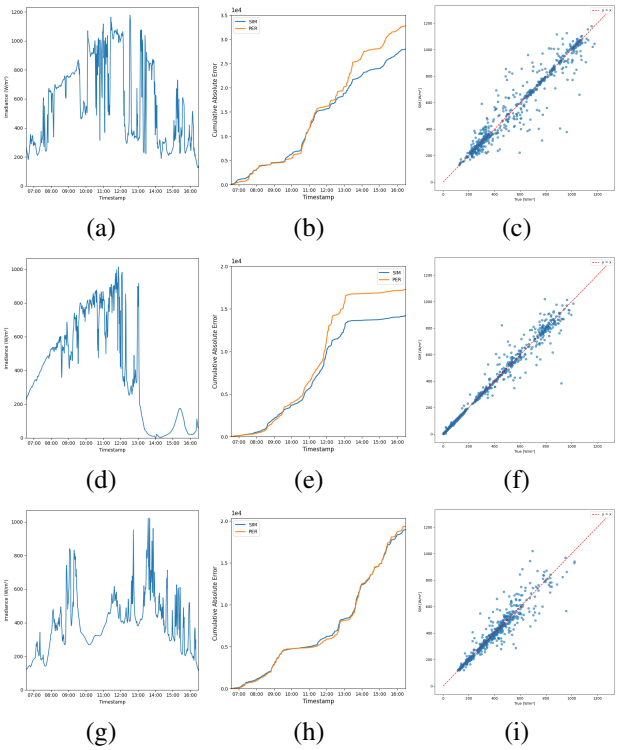


FIGURE 9. Representative test days (2, 26 and 29 August 2023). Rows correspond to days, while columns show (left) measured GHI, (middle) cumulative absolute forecasting error, and (right) predicted versus measured irradiance.

2) Model Performance Analysis

The cumulative absolute error trajectories reveal that the SIM model consistently outperforms the persistence (PER) model under these variable conditions. On August 26th (Figure 9 e), both models demonstrate similar error accumulation until approximately 11:00, after which the SIM model maintains a lower cumulative error trajectory, ending with a 17% lower error by 16:00. The divergence is even more pronounced on August 2nd (Figure 9 b), where the highly erratic irradiance profile causes the persistence model's error to accumulate rapidly. In contrast, the SIM model demonstrates superior stability, resulting in a cumulative error approximately 15% lower. This consistent margin indicates that the SIM model is better equipped to handle rapid fluctuations in irradiance than the persistence approach. On August 29th (Figure 9 h), the discrepancy between the two error metrics is particularly informative. The model achieves an RMSE-based Forecast Skill of 13.93%, yet the reduction in cumulative absolute error is markedly smaller, at just over 2%. Because RMSE penalizes errors quadratically, this pattern suggests that SIM incurs a larger number of small-to-moderate errors while substantially reducing the occurrence of larger errors.

3) Comparative Insights

The performance of the SIM model across the case study days highlights its robustness against different meteorological conditions. While the persistence model's error accumulates

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2026.3746430

sharply during periods of high volatility—such as the 10:00-13:00 window on August 2nd—the SIM model mitigates this effect, likely due to its ability to incorporate non-linear dynamics that simple persistence fails to capture. The fact that SIM outperforms PER even on the slightly less chaotic August 26th suggests that its advantage is not limited to extreme events but is structurally present in variable weather conditions. This aligns with broader literature suggesting that advanced forecasting models offer the most significant gains over persistence during periods of dynamic cloud cover, where the assumption of static conditions is fundamentally violated. Finally, August 29th further illustrates the post-processing mechanism in a real operating scenario: during the sustained low-irradiance period between 9:30 and 10:30, the operational-bounds check causes the pipeline to revert to persistence throughout, rather than propagating an unreliable prediction.

F. GENERALIZATION TO A REDUCED-VARIABILITY PERIOD

To assess robustness beyond the training period, SIM was evaluated on an independent 24-day holdout from October 2023 (Section III), characterized by lower solar elevation and substantially reduced irradiance variability relative to August–September. On this set, persistence itself achieves a low RMSE of 24.46 W/m^2, reflecting the limited scope for improvement when conditions are already stable. SIM achieves a comparable RMSE of 24.49 W/m^2 (FS -0.12%), indicating that the model neither degrades nor meaningfully improves on persistence in this regime. This result is consistent with expectations: under low-variability conditions, persistence is close to the theoretical error floor, and the primary value of SIM demonstrated in Table 4 is realized specifically when cloud-driven variability is high. The evaluation therefore supports the robustness rather than the accuracy gain of the proposed pipeline outside its original training distribution.

V. DISCUSSION

The proposed SIM delivers robust one-minute-ahead BHI and GHI forecasts from LWIR all-sky images and co-located meteorological measurements at a single site. On the Full Set, SIM achieves a Forecast Skill of 12.15% and reduces the RMSE from 56.36W/m^2 to 49.51W/m^2 relative to persistence, with a corresponding increase in R^2 from 94.21% to 95.53%. These gains become even more pronounced on the Variable Set, where the task is intrinsically harder: SIM reduces the RMSE from 82.21W/m^2 to 69.71W/m^2 and achieves 15.20% Forecast Skill over persistence, indicating that it can anticipate ramps associated with rapidly evolving cloud fields rather than merely tracking the last observation.

The case-study analysis on August 2nd and 26th, 2023, further clarifies the operational value of the proposed framework. On both highly variable days, SIM consistently accumulates less absolute error than the persistence baseline, with end-of-day cumulative error reductions of approximately 15–17%. This behavior is particularly relevant for PV plant operators and grid managers, as large irradiance ramps directly translate into power fluctuations. The ability to modestly but systematically reduce short-horizon errors can alleviate reserve needs and support smoother dispatch scheduling under dynamic meteorological conditions.

A. FRAMEWORK CONTRIBUTIONS

The ablation experiments highlight that the strong performance of SIM does not stem from a single component, but rather from the combined effect of the target definition, the temporal architecture, and the physically constrained post-processing. Removing the postprocessing step (OP) leads to a collapse in performance, with negative Forecast Skill and higher RMSE than persistence, underscoring the difficulty of learning absolute BHI values without explicit control on physically implausible values. Re-introducing differencing while omitting the operational bounds (NB) recovers part of the skill but remains clearly inferior to the full SIM configuration, showing that both the increment formulation and the enforcement of site-specific bounds are necessary to prevent error accumulation and unrealistic ramps in one-minute grids.

The comparison between the BHI-based SIM and the GHI-target model confirms the rationale for predicting BHI as an intermediate variable. Predicting direct irradiance enables the network to directly leverage the strong correlation between cloud occlusions in the LWIR images and the beam component, while deferring the contribution of the diffuse component to the postprocessing step. Although the GHI-target variant still improves over persistence, it remains behind SIM in all metrics, suggesting that decoupling direct and diffuse contributions at training time is beneficial when the input modality is an all-sky camera. This is consistent with broader evidence that relative or component-wise targets (e.g., clear-sky indices or irradiance increments) can simplify the learning of short-term variability by removing large-scale diurnal trends from the regression task.

From an architectural perspective, the ConvLSTM encoder is essential for extracting cloud motion information across the four-frame sequence. The NOLSTM variant, which treats the stack of images as a static volume, exhibits a marked decline in performance and a substantial increase in RMSE, indicating that explicit temporal modeling is required even for such short horizons. Conversely, removing the exogenous solar elevation input (NOEXOG) only slightly degrades all metrics, indicating that the centered crop already encodes much of the sun-position information; supplying the elevation angle as an explicit feature nonetheless reduces small systematic biases across different solar geometries.

The proposed postprocessing strategy can be interpreted as a simple yet effective bridge between purely data-driven models and physically informed forecasting. By operating on BHI increments, constraining forecasts within operationally meaningful bounds, and reverting to persistence when the model proposes changes outside physically admissible ranges, the algorithm acts as a guardrail that regularizes the learned predictor without requiring a full physical model of cloud radiative effects. Similar ideas have been explored in

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2026.3746430

postprocessing of solar and PV forecasts, where statistical or physics-based calibration layers are used to enforce realism, reduce bias, and improve reliability of point or probabilistic predictions [44]–[46]. In this context, the results suggest that even a relatively compact neural network can deliver competitive skill if coupled with a carefully designed, physically consistent postprocessing pipeline.

B. LIMITATIONS

Despite these encouraging results, several limitations must be acknowledged. First, the study is restricted to a single site in Milan and the dataset spans a single season transition (summer to early autumn) in a single year, which limits conclusions about generalization across full seasonal cycles, climates, and sensor configurations. Additional experiments on multi-season, multi-site datasets would be required to validate the robustness of SIM under different cloud regimes and aerosol loads. Second, the forecasting horizon is limited to one minute ahead; while this horizon is highly relevant for ramp-rate mitigation and fast control, many operational applications also require forecasts up to tens of minutes, which will likely demand coupling SIM with modules that predict future sky states. Finally, the present work focuses on deterministic point forecasts of irradiance; probabilistic extensions would provide system operators with richer information about forecast uncertainty.

VI. CONCLUSION

This work presented SIM, a compact DL framework for one-minute-ahead solar irradiance forecasting from LWIR all-sky images and co-located meteorological measurements. The model combines an image-based neural architecture with a physically motivated postprocessing layer that operates on BHI increments and enforces operational bounds. On a 54-day dataset from the SolarTechLAB facility, SIM consistently outperformed a persistence baseline, achieving up to 15.20% Forecast Skill, outperforming both a similarly-sized Vision Transformer and a substantially lighter 3D-CNN baseline, and maintaining stable performance on an independent, lower-variability holdout month.

The ablation study showed that the combination of design choices—BHI as an intermediate target, increment-based postprocessing with physical constraints, temporal modeling via ConvLSTM, and exogenous solar-elevation input—each contributes to the overall skill. In particular, the postprocessing layer and BHI-based target proved critical for avoiding unrealistic ramps and stabilizing training, while the temporal encoder enabled the model to capture rapid cloud dynamics that static processing cannot resolve. These results indicate that carefully coupling a lightweight neural architecture with simple, physically informed constraints can deliver meaningful improvements over persistence at very short horizons.

Future work will extend SIM beyond the single-site, short-horizon, deterministic setting considered here, by extending evaluation across full seasonal cycles and multi-site datasets, integrating image-based nowcasting modules for 5–30-minute horizons, and developing probabilistic variants that provide uncertainty-aware forecasts for grid operation and PV plant control.

APPENDIX A PERFORMANCE BY TIME SERIES LENGTH

To determine the optimal number of historical frames for the forecasting model, we conducted a systematic evaluation, varying the time series length from 3 to 8 consecutive images on August 3, 2023. Each configuration was trained and evaluated using identical hyperparameters to ensure fair comparison.

TABLE 7. Forecast Skill by varying the number of frames in the time series. The optimal configuration (4 frames) is highlighted in bold.

Frames	3	**4**	5	6	7	8
FS (%)	28.71	**29.40**	27.25	26.56	26.02	26.89

The results presented in Table 7 show that a time series length of 4 frames achieves the highest Forecast Skill (FS) of 29.40%, representing the best balance between historical context and model complexity. Shorter sequences (3 frames) provide insufficient temporal information about cloud motion dynamics, resulting in reduced performance (FS = 28.71%). Conversely, longer sequences (5–8 frames) introduce redundant information and increase computational overhead without improving accuracy, with performance declining to 26.02% at 7 frames. Based on these findings, all experiments reported in this study employ a time series length of 4 frames, corresponding to 4 minutes of historical observations at the 60-second imaging interval.

DATA AVAILABILITY

The data used in this study are not publicly available.

REFERENCES

[1] F. Fattori, N. Anglani, I. Staffell, and S. Pfenninger, "High solar photovoltaic penetration in the absence of substantial wind capacity: Storage requirements and effects on capacity adequacy," *Energy*, vol. 137, p. 193–208, Oct. 2017. [Online]. Available: http://dx.doi.org/10.1016/j.energy.2017.07.007

[2] Y. Chu, M. Li, C. F. Coimbra, D. Feng, and H. Wang, "Intra-hour irradiance forecasting techniques for solar power integration: A review," *iScience*, vol. 24, no. 10, p. 103136, Oct. 2021. [Online]. Available: http://dx.doi.org/10.1016/j.isci.2021.103136

[3] A. Nespoli, A. Niccolai, E. Ogliari, G. Perego, E. Collino, and D. Ronzio, "Machine learning techniques for solar irradiation nowcasting: Cloud type classification forecast through satellite data and imagery," *Applied Energy*, vol. 305, p. 117834, 2022.

[4] P. Kumari and D. Toshniwal, "Deep learning models for solar irradiance forecasting: A comprehensive review," *Journal of Cleaner Production*, vol. 318, p. 128566, Oct. 2021. [Online]. Available: http://dx.doi.org/10.1016/j.jclepro.2021.128566

[5] S.-A. Logothetis, V. Salamalikis, S. Wilbert, J. Remund, L. F. Zarzalejo, Y. Xie, B. Nouri, E. Ntavelis, J. Nou, N. Hendrikx, L. Visser, M. Sengupta, M. Pó, R. Chauvin, S. Grieu, N. Blum, W. van Sark, and A. Kazantzidis, "Benchmarking of solar irradiance nowcast performance derived from all-sky imagers," *Renewable Energy*, vol. 199, p. 246–261, Nov. 2022. [Online]. Available: http://dx.doi.org/10.1016/j.renene.2022.08.127

[6] A. Niccolai and A. Nespoli, "Sun position identification in sky images for nowcasting application," *Forecasting*, vol. 2, no. 4, pp. 488–504, 2020.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2026.3746430

[7] Y. He, B. Deng, H. Wang, L. Cheng, K. Zhou, S. Cai, and F. Ciampa, "Infrared machine vision and infrared thermography with deep learning: A review," *Infrared Physics & Technology*, vol. 116, p. 103754, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1350449521001262

[8] N. B. Nam, E. Ogliari, S. Leva, E. Pafumi, D. Alberti, and M. Q. Duong, "Comparative analysis of conformal prediction techniques and machine learning models for very short-term solar power forecasting," *Energy and AI*, vol. 21, p. 100573, 2025.

[9] N. B. Nguyen, E. Ogliari, S. Leva, and E. Pafumi, "A multi-source deep learning framework for very short-term solar irradiance nowcasting with satellite data and all-sky images," in *2025 IEEE PES 17th Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, 2025, pp. 1–6.

[10] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Advances in Neural Information Processing Systems*, vol. 28, 2015, pp. 802–810.

[11] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

[12] H.-M. Zuo, J. Qiu, Y.-H. Jia, Q. Wang, and F.-F. Li, "Ten-minute prediction of solar irradiance based on cloud detection and a long short-term memory (lstm) model," *Energy Reports*, vol. 8, p. 5146–5157, Nov. 2022. [Online]. Available: http://dx.doi.org/10.1016/j.egyr.2022.03.182

[13] R. H. Inman, H. T. Pedro, and C. F. Coimbra, "Solar forecasting methods for renewable energy integration," *Progress in Energy and Combustion Science*, vol. 39, no. 6, p. 535–576, Dec. 2013. [Online]. Available: http://dx.doi.org/10.1016/j.pecs.2013.06.002

[14] R. Ahmed, V. Sreeram, Y. Mishra, and M. D. Arif, "A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization," *Renewable and Sustainable Energy Reviews*, vol. 124, p. 109792, 2020.

[15] C. Feng and J. Zhang, "Solarnet: A deep convolutional neural network for solar forecasting via sky images," in *2020 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, 2020, pp. 1–5.

[16] Q. Paletta, G. Arbod, and J. Lasenby, "Benchmarking of deep learning irradiance forecasting models from sky images – an in-depth analysis," *Solar Energy*, vol. 224, p. 855–867, Aug. 2021. [Online]. Available: http://dx.doi.org/10.1016/j.solener.2021.05.056

[17] Q. Paletta, A. Hu, G. Arbod, and J. Lasenby, "Eclipse: Envisioning cloud induced perturbations in solar energy," *Applied Energy*, vol. 326, p. 119924, Nov. 2022. [Online]. Available: http://dx.doi.org/10.1016/j.apenergy.2022.119924

[18] S. Chaaraoui, S. Houben, and S. Meilinger, "Probabilistic end-to-end irradiance forecasting through pre-trained deep learning models using all-sky-images," *Advances in Science and Research*, vol. 20, p. 129–158, Jan. 2024. [Online]. Available: http://dx.doi.org/10.5194/asr-20-129-2024

[19] W. Kong, Y. Jia, Z. Y. Dong, K. Meng, and S. Chai, "Hybrid approaches based on deep whole-sky-image learning to photovoltaic generation forecasting," *Applied Energy*, vol. 280, p. 115875, 2020.

[20] A. Ravi and F. Karray, "Exploring convolutional recurrent architectures for anomaly detection in videos: a comparative study," *Discover Artificial Intelligence*, vol. 1, no. 1, Sep. 2021. [Online]. Available: http://dx.doi.org/10.1007/s44163-021-00004-2

[21] L. F. C. Varaschin and D. Silva, "Benchmarking deep learning-based methods for irradiance nowcasting with sky images," Preprint, 2025.

[22] O. Bayasgalan and A. Akisawa, "Nowcasting solar irradiance components using a vision transformer and multimodal data from all-sky images and meteorological observations," *Energies*, vol. 18, p. 2300, 2025.

[23] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," in *International Conference on Learning Representations*, 2021. [Online]. Available: https://openreview.net/forum?id=YicbFdNTTy

[24] A. P. Steiner, A. Kolesnikov, X. Zhai, R. Wightman, J. Uszkoreit, and L. Beyer, "How to train your vit? data, augmentation, and regularization in vision transformers," *Transactions on Machine Learning Research*, 2022. [Online]. Available: https://openreview.net/forum?id=4nPswr1KcP

[25] A. L. Jonathan, O. Bamisile, D. Cai, C. J. Ejiyi, J. J. N. Nkou, K. Victor, C. C. Ukwuoma, L. Wei, and Q. Huang, "A multimodal deep learning approach for very short-term solar forecasts using sky images and historical numerical data," *Renewable Energy*, vol. 255, p. 123774, Dec. 2025. [Online]. Available: http://dx.doi.org/10.1016/j.renene.2025.123774

[26] M. Husein, E. Gago, B. Hasan, and M. Pegalajar, "Towards energy efficiency: A comprehensive review of deep learning-based photovoltaic power forecasting strategies," *Heliyon*, vol. 10, no. 13, p. e33419, Jul. 2024. [Online]. Available: http://dx.doi.org/10.1016/j.heliyon.2024.e33419

[27] O. Gandhi *et al.*, "The value of solar forecasts and the cost of their errors: A review," *Renewable and Sustainable Energy Reviews*, vol. 189, p. 113915, 2024.

[28] U. K. Das *et al.*, "Forecasting of photovoltaic power generation and model optimization: A review," *Renewable and Sustainable Energy Reviews*, vol. 81, pp. 912–928, 2018.

[29] J. Zhang, R. Verschae, S. Nobuhara, and J.-F. Lalonde, "Deep photovoltaic nowcasting," *Solar Energy*, vol. 176, pp. 267–276, 2018.

[30] C. C. Ukwuoma *et al.*, "An attention fused sequence-to-sequence convolutional neural network for accurate solar irradiance forecasting and prediction using sky images," *Renewable Energy*, vol. 237, p. 121692, 2024.

[31] M. D. Bailey, D. W. Nychka, M. Sengupta, J. Yang, Y. Xie, A. Habte, and S. Bandyopadhyay, "Adapting quantile mapping to bias correct solar radiation data," *Solar Energy*, vol. 288, p. 113220, Mar. 2025. [Online]. Available: http://dx.doi.org/10.1016/j.solener.2024.113220

[32] K. Bakker, K. Whan, W. Knap, and M. Schmeits, "Comparison of statistical post-processing methods for probabilistic nwp forecasts of solar radiation," *Solar Energy*, vol. 191, p. 138–150, Oct. 2019. [Online]. Available: http://dx.doi.org/10.1016/j.solener.2019.08.044

[33] F. Lin, Y. Zhang, and J. Wang, "Recent advances in intra-hour solar forecasting: A review of ground-based sky image methods," *International Journal of Forecasting*, vol. 39, pp. 244–265, 2023.

[34] G. Terrén-Serrano and M. Martínez-Ramón, "Detection of clouds in multiple wind velocity fields using ground-based infrared sky images," *Knowledge-Based Systems*, vol. 274, p. 110628, 2023.

[35] R. Bessa *et al.*, "Towards improved understanding of the applicability of uncertainty forecasts in the electric power industry," *Energies*, vol. 10, p. 1402, 2017.

[36] R. Urraca *et al.*, "Quality control of global solar radiation data with satellite-based products," *Solar Energy*, vol. 158, pp. 49–62, 2017.

[37] J. E. Hammond, R. A. Lara Orozco, M. Baldea, and B. A. Korgel, "Short-term solar irradiance forecasting under data transmission constraints," *Renewable Energy*, vol. 233, p. 121058, 2024.

[38] J. Nou, R. Chauvin, J. Eynard, S. Thil, and S. Grieu, "Towards the intrahour forecasting of direct normal irradiance using sky-imaging data," *Heliyon*, vol. 4, no. 4, p. e00598, Apr. 2018. [Online]. Available: http://dx.doi.org/10.1016/j.heliyon.2018.e00598

[39] J. Wojtkiewicz, M. Hosseini, R. Gottumukkala, and T. L. Chambers, "Hour-ahead solar irradiance forecasting using multivariate gated recurrent units," *Energies*, vol. 12, p. 4055, 2019.

[40] B. Nouri *et al.*, "Real-time uncertainty specification of all sky imager derived irradiance nowcasts," *Remote Sensing*, vol. 11, p. 1059, 2019.

[41] P. Gregor, T. Zinner, F. Jakub, and B. Mayer, "Validation of a camera-based intra-hour irradiance nowcasting model using synthetic cloud data," *Atmos. Meas. Tech.*, vol. 16, pp. 3257–3271, 2023.

[42] G. M. Lohmann, A. H. Monahan, and D. Heinemann, "Local short-term variability in solar irradiance," *Atmospheric Chemistry and Physics*, vol. 16, no. 10, pp. 6365–6379, 2016. [Online]. Available: https://acp.copernicus.org/articles/16/6365/2016/

[43] F. Vignola, "Ghi correlations with dhi and dni and the effects of cloudiness on one-minute data," in *World renewable energy forum, PROCEEDINGS OF THE SOLAR CONFERENCE*, vol. 1. Boulder, Colo.: American Solar Energy Society;, 2012, pp. 621–626. [Online]. Available: https://www.tib.eu/de/suchen/id/BLCP

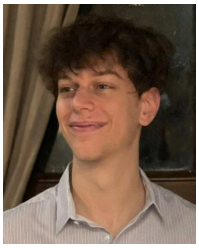
[44] B. Schulz, M. El Ayari, S. Lerch, and S. Baran, "Post-processing numerical weather prediction ensembles for probabilistic solar irradiance forecasting," *Solar Energy*, vol. 220, p. 1016–1031, May 2021. [Online]. Available: http://dx.doi.org/10.1016/j.solener.2021.03.023

[45] S. Baran, J. C. Marín, O. Cuevas, M. Díaz, M. Szabó, O. Nicolis, and M. Lakatos, "Machine-learning-based probabilistic forecasting of solar irradiance in chile," *Advances in Statistical Climatology, Meteorology and Oceanography*, vol. 11, no. 1, p. 89–105, Jun. 2025. [Online]. Available: http://dx.doi.org/10.5194/ascmo-11-89-2025

[46] M. J. Mayer, Ágnes Baran, S. Lerch, N. Horat, D. Yang, and S. Baran, "Post-processing of ensemble photovoltaic power forecasts with distributional and quantile regression methods," 2025. [Online]. Available: https://arxiv.org/abs/2508.15508

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2026.3746430



SIMONE BORTOLOTTI received a bachelor's degree in Computer Engineering from the University of Modena and Reggio Emilia in 2022, and a master's degree in Computer Science and Engineering from Politecnico di Milano in 2025.



DANIELE REGE CAMBRIN received a master's degree in Computer Engineering from Politecnico di Torino in 2022. He is a Ph.D. student in the same institution's Department of Control and Computer Engineering (DAUIN). His main research interests are machine learning applied to geospatial data and scalable algorithms.



PAOLO GARZA received the master's and Ph.D. degrees in Computer Engineering from Politecnico di Torino in 2001 and 2005, respectively. Since December 2018, he has been an associate professor at the Dipartimento di Automatica e Informatica, Politecnico di Torino. His research interests include data mining algorithms, big data analytics, and data science. He has worked on classification, clustering, itemset mining, and scalable algorithms.



SONIA LEVA. (Senior Member, IEEE) received a Ph.D. degree in electrical engineering from the Politecnico di Milano, Italy, in 2001. She is currently a Full Professor of electrical engineering with the Department of Energy, Politecnico di Milano. She is also the Director of the Solar Tech Laboratory (SolarTech^{LAB}) and the Laboratory of MicroGrids (MG^{2}LAB), Politecnico di Milano. She is a Senior Member of the IEEE Power and Energy Society and a member of the IEEE Working Group Distributed Resources Modeling and Analysis.



EMANUELE GIOVANNI CARLO OGLIARI (Member, IEEE) is an Associate Professor at the Department of Energy, Politecnico di Milano, Milan, Italy where he teaches Electrical Engineering and Photovoltaic-based Systems. He has a M.Sc. in electrical engineering, and he received the Ph.D. degree in electrical engineering from the Politecnico di Milano, Italy, in 2016 where he is staff member of the SolarTech^{LAB} and the Laboratory of MicroGrids (MG^{2}LAB).

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/