Journal of Energy

VOLUME 73 Number 4 | 2024

journal homepage: http://journalofenergy.com/

Regional Solar Irradiance Forecasting Using Multi-Camera Sky Imagery and Machine Learning Models

Alen Jakoplić, Dubravko Franković, Tomislav Plavšić, Branka Dobraš

***Summary* — With the increasing integration of photovoltaic (PV) systems into power grids, accurate short-term solar irradiance forecasting is essential for efficient energy management. This paper presents a machine learning model developed using a synthetic dataset designed to analyze the potential of multicamera sky imaging for regional solar irradiance forecasting. The dataset, generated in a controlled simulation environment, captures cloud dynamics and solar irradiance at multiple locations within a region. The proposed model utilizes sky images from multiple virtual cameras strategically positioned to provide spatially distributed observations. By combining image-based features with historical irradiance measurements, the model shows improved forecasting accuracy compared to single-camera approaches. The results indicate that multi-camera systems better capture the spatial variability of cloud cover and allow the model to predict solar irradiance for locations without installed cameras. This research highlights the potential of multi-camera configurations for regional forecasting and provides valuable insights for grid operators and energy planners. The results support the adoption of distributed sky imaging networks as a practical approach to improve solar irradiance predictions and ultimately contribute to the stability and reliability of solarpowered energy systems through improved forecast accuracy.**

***Keywords* — Solar irradiance forecasting, photovoltaic systems, multi-camera sky imaging, renewable energy integration.**

I. Introduction

Photovoltaic power plants (PVPPs) are among the most widely used renewable energy power plants [1]. Their popularity stems from their ability to harness solar energy, an abundant and inexhaustible resource while generating electricity with minimal environmental impact. Due to their lower environmental impact, research has recently focused on further improving solar cells in terms of efficiency, production costs, and durability. The constant advances in photovoltaic technology have led to higher efficiency, longer lifespan, and lower manufacturing costs, which have accelerated the use of PV systems worldwide. As a result, the share of PVPPs in the structure of production units in the energy sector is steadily increasing, making solar energy a cornerstone of the transition to cleaner energy systems [2].

The unpredictability of electricity generation from renewable energy sources, including solar energy, leads to voltage and frequency fluctuations within the power grid, causing difficulties in its management [3]. These fluctuations are caused by sudden changes in solar radiation due to cloud movements, atmospheric conditions, and other meteorological factors. Changes in the availability of renewable energy can occur within very short periods, often only a few minutes, during which other, conventional power plants cannot adjust their output quickly enough. The inherently slow response of conventional power plants, such as coal or gas-fired power plants, exacerbates the imbalance between electricity generation and consumer demand. When the balance between generated and consumed electricity is disturbed, deviations from nominal voltage and frequency values occur, resulting in reduced quality of electrical energy and potential damage to sensitive equipment [4], [5].

To mitigate the negative impact of renewable energy sources on the power grid, it is necessary to predict changes in the availability of these energy sources with a certain degree of accuracy. The highly dynamic nature of meteorological conditions makes accurate long-term cloud forecasting at a given location difficult [6]. Conventional meteorological models, while effective at larger scales, are often inadequate when applied to local cloud dynamics relevant to PV power prediction. A promising solution to this challenge is the short-term prediction of cloud cover at the observed location, typically 10 to 15 minutes in advance within a radius of 2000 meters. Narrow spatial and temporal scales enable more accurate prediction of cloud cover and thus better integration of solar energy into the grid and more effective planning of the operation of conventional power plants [7].

Implementing a reliable power generation forecasting system reduces the need for balancing power, i.e. the reserve power needed to compensate for deviations of renewable energy generation from the contracted schedule. More accurate forecasts consequently reduce the cost of integrating renewable energy sources into the electricity grid by minimizing the dependence on reserve power plants and ancillary services [8]. In addition, the reduction of production curtailments due to large fluctuations leads to higher efficiency of existing systems. This improved efficiency combined with lower integration costs not only benefits grid operators but also leads to lower electricity prices for consumers within the grid. In addition, improved forecasting capabilities support grid stability, reliability, and resilience, especially as the share of variable renewable energy sources continues to increase in modern power systems [9].

(Corresponding author: Alen Jakoplić)

Alen Jakoplić, Dubravko Franković, and Branka Dobraš are with the Faculty of Engineering, University of Rijeka, Rijeka, Croatia (e-mails: alen.jakoplic@riteh.uniri.hr, dubravko.frankovic@riteh.uniri.hr, branka.dobras@riteh.uniri.hr)

Tomislav Plavšić is with the Croatian Transmission System Operator (HOPS), Zagreb, Croatia (e-mail: tomislav.plavsic@hops.hr)

Alen Jakoplić, Dubravko Franković, Tomislav Plavšić, Branka Dobraš, Regional Solar Irradiance Forecasting Using Multi-Camera Sky Imagery and Machine Learning Models, Journal of Energy, vol. 73 Number 4 (2024), 26–33

https://doi.org/10.37798/2024734707

Building on the existing foundation, this research focuses on the development of a short-term prediction model for solar radiation using a synthetic database. The main objective is to investigate the potential of using multiple sky cameras at different locations to predict solar irradiance on a regional scale. By using data from multiple cameras, the model can detect cloud patterns, movements, and shadow propagation, all of which have a significant impact on the production of photovoltaic power plants. The use of multiple cameras provides the model with a more comprehensive understanding of atmospheric dynamics, as the combination of different viewing angles enables better recognition of cloud formation and movement trends.

The ability to forecast solar radiation regionally offers several advantages. It enables predictions for areas where no cameras are directly installed, extending the practical applications of the model beyond the monitored locations. This is particularly valuable for distributed PV systems, where individual installations may be spread over a larger geographical area. In addition, improved short-term forecasting supports better grid management, as operators can anticipate fluctuations in solar power generation and implement necessary balancing measures more effectively.

Ultimately, this study shows that it is possible to use multicamera configurations to improve short-term solar irradiance forecasting at a regional level. The knowledge gained from this study contributes to ongoing efforts to improve the integration of renewable energy sources into modern power systems and to support a more stable, efficient, and reliable use of solar energy.

II. Basics and Motivation for Regional Solar Forecasting Models

In recent years, with significant integration of PV power plants, mostly at lower voltage levels, accurate solar irradiance forecasting is becoming crucial for stable and secure power system operation. Solar irradiance forecasting is an input variable for two important power system operational planning processes:

- PV production forecasting
- Load forecasting

While direct PV production forecasting is mainly used for PV power plants connected to the transmission voltage levels, load forecasting algorithms have an indirect forecast of PV generation connected to the distribution voltage levels embedded as a part of the overall forecasting function. While the load forecasting function is usually only performed as part of day-ahead operational planning processes, the forecast of renewable energy production is also performed in intra–day operational planning processes, usually one hour ahead. Such an approach is justified due to the sudden changes in local weather forecasts, which can have a significant impact on the production of PV and wind power plants. Ultra-shortterm PV production forecasting 15 minutes ahead of real-time could further improve the power system operation efficiency and security, enabling the operators in control centers to take the necessary preventive operational measures just ahead of real-time. This way, the operating personnel still have time to optimize power system operation, while if those measures were curative and done after the changes in the PV production have occurred, there would be much less room for optimized operational actions.

Over the years, a large number of different methods and approaches have been developed to predict the production of PV systems [10]. These methods have evolved significantly due to the growing need for more accurate and reliable predictions to optimize grid operations and support the increasing share of solar energy in modern power systems. The categorization of these methods is generally based on the type of input data, approaches to data pre-processing, temporal frequency of data collection, spatial resolution, and temporal and spatial horizon [11]. In addition, factors such as the complexity of the model, the computational requirements, and the availability of historical data play a crucial role in determining the effectiveness of these forecasting methods.

An important aspect of forecasting the production of photovoltaic power plants lies in the analysis of local weather conditions, especially solar radiation, whose fluctuations directly affect the output power and allow an accurate prediction of future production. Solar radiation is the main factor influencing photovoltaic output, and its fluctuations are influenced by various atmospheric phenomena such as cloud cover, aerosols, and seasonal changes. Parameters such as wind speed, temperature, time of day, and relative humidity, on the other hand, have a much lower correlation with the production of photovoltaic systems. While these meteorological variables are useful for broader climatological analyses, they are often less relevant for short-term predictions. Such data are often subject to fluctuations due to various conditions, such as changes in wind direction and speed at different heights above the ground and the relative stability of temperature over a short period within a day [12]. Therefore, these parameters often require more complex models and algorithms for effective inclusion in forecasting models.

The importance of short-term prediction of solar irradiance cannot be neglected when it comes to predicting the output power of photovoltaic power plants. This type of prediction is particularly important during periods of high solar variability, e.g. on partly cloudy days, when rapid changes in irradiance can lead to significant fluctuations in PV output. The complexity of this task lies in the inherent randomness and non-linearity of solar radiation, which is particularly pronounced in changing weather conditions. Many scientific studies have emphasized the use of artificial neural networks (ANN) for such forecasting models due to their ability to adapt to complex and nonlinear patterns [13], [14]. These models use historical and real-time data to learn the intricate relationships between atmospheric conditions and solar radiation. Nevertheless, further refinement of these models in terms of accuracy and robustness is needed [15], especially in scenarios with highly dynamic cloud formations.

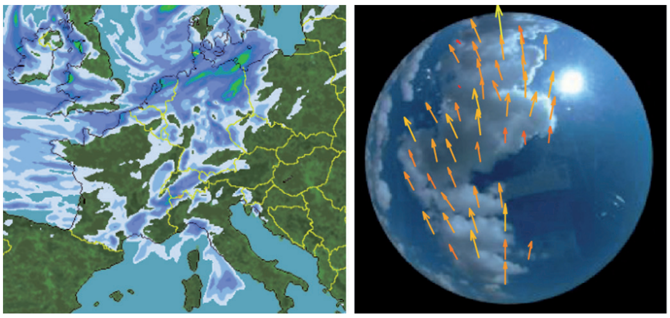

Fig. 1. Comparison of satellite and ground-based cloud motion vectors.

Forecasts of the output power of photovoltaic power plants are often based on satellite images and use models such as cloud motion vectors, as can be seen in Figure 1 on the left and right. These methods use historical cloud movement data to predict future irradiance patterns. However, the limitations of these models, such as the assumption of constant cloud shapes and sensitivity to local weather conditions, make them less accurate [16]. While satellite images are beneficial for large-scale analyses, they often lack the spatial and temporal resolution needed for accurate short-term predictions.

Accurate prediction of cloud changes over PVPPs requires detailed data on the state of clouds, including their amount, position,

Alen Jakoplić, Dubravko Franković, Tomislav Plavšić, Branka Dobraš, Regional Solar Irradiance Forecasting Using Multi-Camera Sky Imagery and Machine Learning Models, Journal of Energy, vol. 73 Number 4 (2024), 26–33
https://doi.org/10.37798/2024734707

and movement, which are usually obtained from satellite and radar imagery [17]. However, due to the limited temporal and spatial resolution of these images, they are often not suitable for short-term predictions [18]. Ground-based sky cameras equipped with wide-angle lenses and high-frequency imaging offer a promising alternative for capturing cloud dynamics in real-time. The development of new databases containing more detailed information from sky photography can significantly improve the accuracy of predictions, but high equipment costs limit their application [19], [20].

Convolutional neural networks (CNNs) are particularly well suited for detecting nonlinear relationships between input and output data in models for short-term prediction of photovoltaic power plant production. CNNs excel at processing visual information and can automatically extract relevant features from sky images without manual intervention. These networks can recognize patterns in photographs so that they can use these images as input data. Since photographs of the sky above a photovoltaic power plant are directly correlated with its output [21], convolutional neural networks can use these photographs to discover correlations between sky images and the output of photovoltaic power plants. By incorporating additional meteorological data, CNNs can achieve even higher forecasting accuracy, especially for short-term forecasts.

Based on the review of the available methods, it is concluded that the accuracy of the models generated by convolutional neural networks depends on the quality of the input data. The availability of high-quality sky images in combination with precise irradiance measurements plays a crucial role in the training and validation of the models. This underlines the importance of a detailed and comprehensive database that would enable better training of the neural network and thus a more accurate prediction of the production of photovoltaic power plants.

The development of specialized databases for short-term regional solar forecasting is essential for testing and validating new model architectures. General or single-camera datasets often fail to capture the spatial variability of cloud cover over larger areas, which is critical for understanding cloud dynamics and their impact on photovoltaic production. A customized dataset that accounts for different weather conditions and spatial configurations provides the necessary basis for improving model performance and exploring novel approaches for regional solar irradiance forecasting.

III. Development of a Multi-Camera Solar Forecasting Model

The development of a reliable model for short-term solar irradiance forecasting at a regional level requires an innovative approach that takes into account the complex interactions between atmospheric conditions and solar irradiance. In this study, a synthetic database with data from multiple wideangle cameras is used to capture cloud movements and their effects on solar radiation. This method enables the creation of models capable of predicting solar irradiance for locations without direct measurements by utilizing spatial relationships and cloud dynamics.

A. Synthetic Data Set Simulation Framework

The synthetic data base used in this study was created using the Unity development platform, which was chosen for its flexibility in generating realistic atmospheric scenarios and simulating dynamic cloud behavior. Unity's advanced 3D rendering capabilities and real-time simulation tools enable the accurate reproduction of sunlight behavior and cloud patterns under different weather conditions.

To ensure the fidelity of the simulations, the High Definition Render Pipeline (HDRP) was used. HDRP supports realistic light interactions, which is essential for modeling variations in solar radiation due to cloud cover, as can be seen in Figure 2. The simulation framework consists of several userdefined scripts that control environmental parameters such as sun position, cloud density, cloud movement, and temporal progression to replicate diurnal cycles.

Camera placement is calculated to cover key areas within the simulation. A script assigns coordinates to each camera, providing a variety of perspectives in the monitored region. The cameras capture sky images at regular intervals, providing a continuous stream of visual data that is essential for training the forecasting model. The image resolution is set to 64x64 efficiency. However, the framework allows customization for higher resolutions.

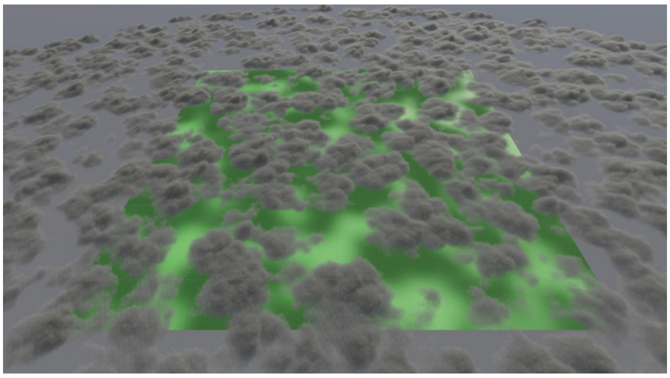

Fig. 2. Simulated cloud coverage over the target region generated using the Unity HDRP framework. The image illustrates the spatial distribution of clouds and the resulting shadows cast on the ground. The observed area is represented by the green square in the center, corresponding to a simulated surface of 50 × 50 kilometers.

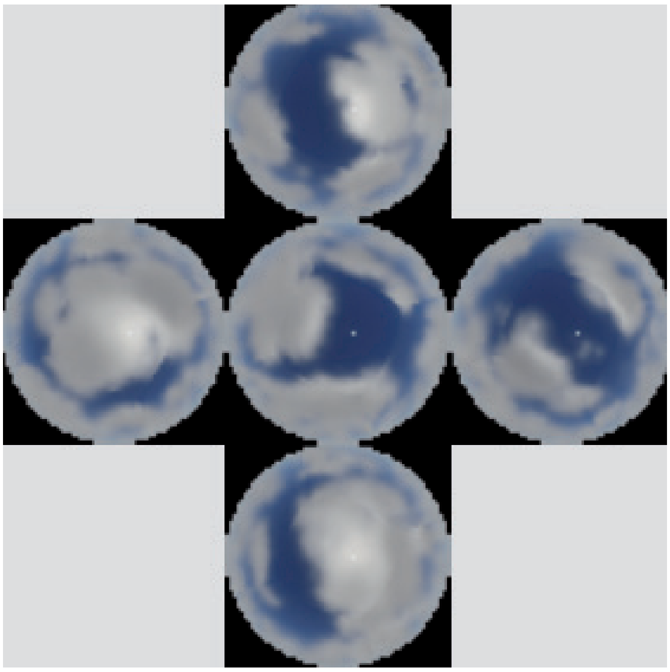

Fig. 3. Sky Images from Different Perspectives: Central Camera (0,0) with Surrounding Cameras at (-5000,0), (5000,0), (0,5000), and (0,-5000).

The simulation process was tested on an M3 MacBook Air generating a single day's data from five camera positions with images taken every five minutes took about two minutes. The resulting data set requires 8.1 MB of disk space, which underlines the scalability of the system for larger simulations. The temporal resolution

Alen Jakoplić, Dubravko Franković, Tomislav Plavšić, Branka Dobraš, Regional Solar Irradiance Forecasting Using Multi-Camera Sky Imagery and Machine Learning Models, Journal of Energy, vol. 73 Number 4 (2024), 26–33
https://doi.org/10.37798/2024734707

is set to five minutes and provides sufficient data granularity for short-term forecast models. Shorter intervals, e.g. one minute, can be configured if required to capture rapid changes in cloud cover.

The simulated area spans 50 x 50 km, with cameras strategically placed in the center and four surrounding cameras placed 5 km away in different directions to capture different cloud perspectives, as shown in Figure 3. Each camera is accompanied by light sensors that provide reference irradiance measurements to ensure that the image features match the irradiance data.

The database is publicly accessible via the Kaggle platform [22], facilitating data sharing and collaboration between researchers. The published dataset contains extensive metadata and a total size of 5.1 GB, providing a rich resource for model training and validation.

This simulation framework forms the basis for the development of advanced models for short-term solar irradiance forecasting. By incorporating multiple perspectives and different weather conditions, the model can more accurately predict cloud-related variations in solar irradiance, supporting more reliable and efficient photovoltaic power generation on a regional scale.

B. Architecture Of The Neural Network

The development and training of the neural network model for the short-term prediction of solar irradiance was performed with Google Colab, using an L4 GPU for efficient parallel processing. Google Colab provides a cloud-based environment that simplifies access to computational resources without the need for extensive local infrastructure. This platform was chosen for its flexibility, ease of collaboration, and support for GPU-accelerated machine learning workflows.

The synthetic dataset used to train the model includes 340 days of data, with measurements taken every five minutes between 6:00 am and 6:00 pm. For each of the five simulated camera locations, the database contains hemispheric sky images together with corresponding measurements of available solar radiation. This setup provides a diverse and dynamic dataset that reflects varying meteorological conditions on different days and at different times. As illustrated in Figure 4, the solar irradiance levels for five different locations are shown for two randomly selected days, providing insight into the temporal variations in solar radiation under different meteorological conditions.

Prior to training, the data was pre-processed to improve the performance of the model. The pre-processing steps included normalizing the irradiance values and temporal features to standardize the input distribution. Normalization is crucial to accelerate the convergence of the model and improve the stability of the training by ensuring that the input features have similar scales.

A custom sequence generator was implemented to prepare the data for training. This generator creates sequences of nine consecutive images for each camera, with corresponding irradiance measurements, timestamps, and camera coordinates. Each sequence represents a 45-minute time window selected based on the observed average time for cloud movement from the edge to the center of the sky image. The target value for the prediction is the irradiance at the center position (0.0) 15 minutes in the future, which corresponds to the third image ahead of the input sequence.

The decision to use a 45-minute input sequence with a 15-minute forecast horizon was guided by the dynamic characteristics of cloud movement within the synthetic data set. In the simulations, wind speeds vary daily and throughout the day. However, the 45-minute window has been shown to capture the predominant cloud movement patterns and allows the model to learn the relationship between cloud dynamics and solar irradiance variations.

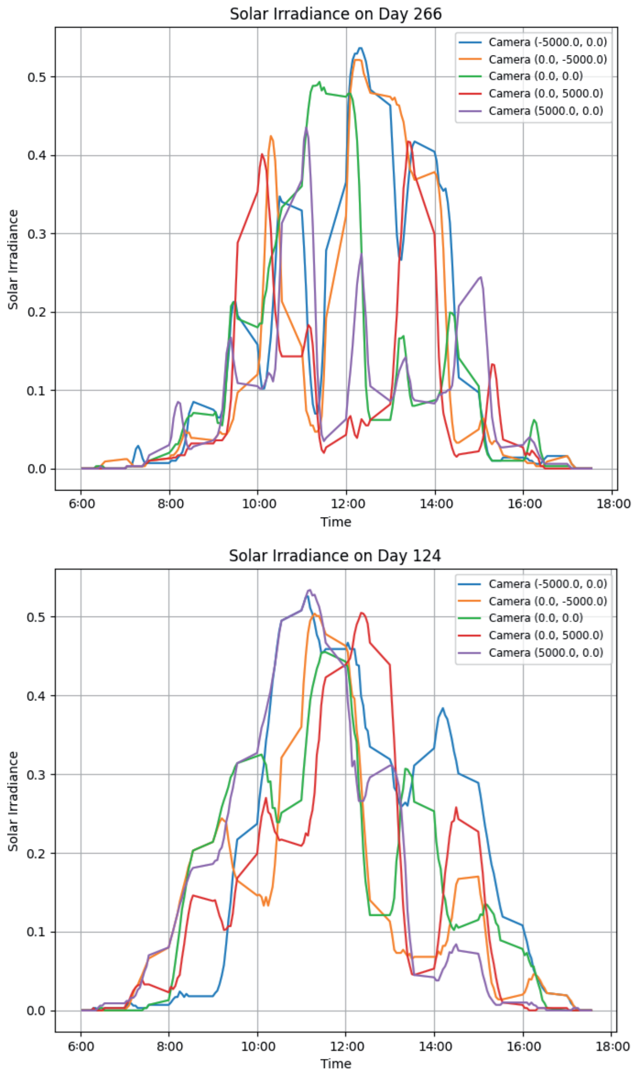

Fig. 4. Measured solar irradiance at five simulated locations on Day 266 (top) and Day 124 (bottom). Each curve represents irradiance values recorded by a dedicated irradiance module colocated with one of the five virtual sky cameras positioned at coordinates (0, 0), (±5000, 0), and (0, ±5000).

The selection of these parameters was aimed at enabling a meaningful performance analysis. Future research using real data, which is currently unavailable due to a lack of databases of sky images from multiple locations, will investigate the optimal length of input sequences and forecast horizons for regional solar forecasting.

The neural network model developed for this study was designed for short-term solar irradiance forecasting using data from multiple sky cameras together with numerical weather information. The architecture integrates convolutional neural networks (CNNs) for image processing and fully connected layers for numerical data, allowing the model to capture spatial cloud patterns and their relationship to solar radiation.

As illustrated in Figure 5, the model accepts two different inputs: The first input consists of time-sequenced images from five cameras, one central reference camera, and four peripheral cameras positioned around it. These images capture cloud movements and atmospheric conditions that influence solar irradiance. The second input comprises the corresponding meteorological param-

https://doi.org/10.37798/2024734707

eters, such as irradiance and cloud-related features, for the same time intervals.

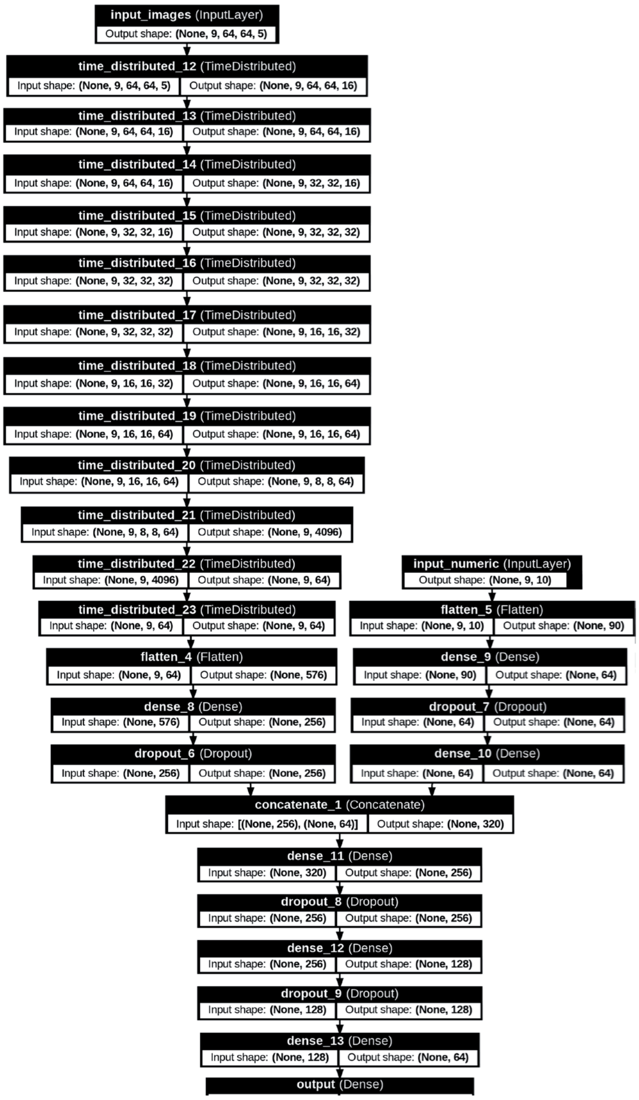

Fig. 5. Neural network architecture for regional solar forecasting.

The image processing branch uses a TimeDistributed Convolutional Neural Network (CNN) to extract spatial features from the sky images over the nine time steps. The network consists of:

- A series of three convolutional layers with filter sizes of 16, 32, and 64, each using a 3x3 kernel and ReLU activation
- Batch normalization layers after each convolutional layer
- Max-pooling layers to reduce spatial dimensions and retain essential features
- TimeDistributed flattening and dense layers to capture temporal dependencies
- A final dense layer with 256 units to combine the extracted features across all time steps

The numerical data stream processes a feature set consisting of the capture time and coordinates for each camera. This data stream passes through:

- An initial flattening layer
- Two dense layers with 64 units each and ReLU activation
- Dropout layers with a dropout rate of 20 to prevent overfitting

The outputs of the two branches are concatenated into a combined feature vector that is further processed by:

- A dense layer with 256 units and ReLU activation
- A dropout layer with a dropout rate of 30
- Additional dense layers with 128 and 64 units
- A final dense layer with a linear activation function to predict the future solar irradiance value

The model was compiled with the Adam optimizer, mean squared error (MSE) as the loss function, and a custom R^2 metric for performance evaluation. The training was performed with a batch size of 32 over 50 epochs.

The dataset was split into training (80%), validation (10%), and test subsets (10%). During training, the model was exposed to a diverse range of cloud patterns and atmospheric conditions to improve its predictive capabilities for unseen data.

To improve generalization and prevent overfitting, the training process employed early stopping with a patience of 10 epochs and model checkpointing, ensuring the retention of the best-performing model based on validation loss. The model includes a total of 566,097 trainable parameters.

Figure 6 shows the training and validation loss curves along with the validation R2 score throughout the 50 training epochs for scenario c) Camera-Free Target Location. The plot indicates stable convergence of the training loss, while fluctuations in the validation R2 metric reflect the challenging nature of the scenario, where no direct camera data from the target location is available.

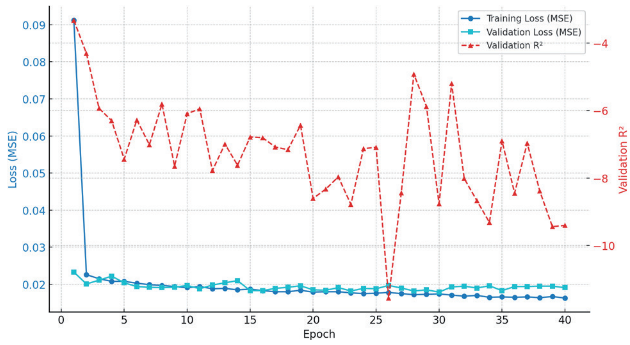

Fig. 6. Training and validation loss (MSE) and validation R2 score per epoch for scenario c) Camera-Free Target Location.

IV. Analysis of Model Accuracy in Regional Solar Forecasting

To evaluate the performance and feasibility of the developed model for regional solar forecasting, three different scenarios were designed, as illustrated in Figure 7. Each scenario explores different configurations of camera input data to evaluate the impact of different perspectives on forecasting accuracy and to investigate the potential of predicting solar irradiance for unmonitored locations.

In the first scenario (see Figure 7a), the model uses only the data from the central camera positioned at coordinates (0,0) to predict future solar irradiance at the same location. This scenario serves as a reference point against which the performance of the other scenarios can be compared. Since only a single camera is used, the model relies solely on the cloud movement patterns observed from this single viewpoint.

The second scenario (see Figure 7b) adds additional information by including the images from the four peripheral cameras at (5000,0), (-5000,0), (0,5000), and (0,-5000) meters in addition to

Alen Jakoplić, Dubravko Franković, Tomislav Plavšić, Branka Dobraš, Regional Solar Irradiance Forecasting Using Multi-Camera Sky Imagery and Machine Learning Models, Journal of Energy, vol. 73 Number 4 (2024), 26–33
https://doi.org/10.37798/2024734707

the data from the central camera. This configuration aims to determine whether the inclusion of multiple viewpoints can improve prediction accuracy. The peripheral cameras provide spatial context by capturing cloud motion from multiple vantage points, allowing the model to better understand the dynamics of cloud formation, breakup, and movement in the region.

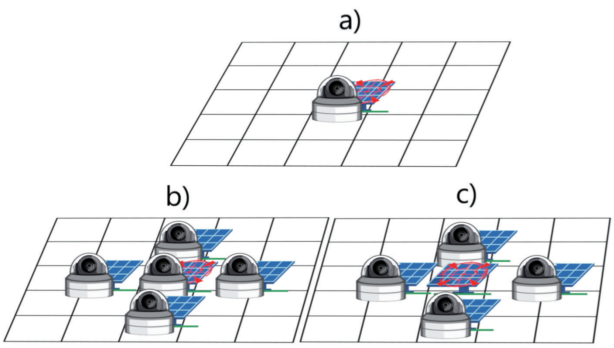

Fig. 7. Evaluation scenarios for model performance. a) Single-Camera Scenario, b) Multi-Camera Scenario, c) Camera-Free Target Location.

In the third and final scenario (see Figure 7c), the model uses only the data from the four peripheral cameras to predict solar irradiance at the central location (0,0). This scenario is particularly important as it tests the model's ability to estimate solar irradiance for a location where there is no direct monitoring device. The success of this approach would show that the model can provide forecasts for locations without installed cameras and thus support solar forecasting on a regional scale.

The performance in these three scenarios provides insight into the benefits of using multiple cameras for short-term solar irradiance prediction. It also evaluates the model's ability to generalize spatial relationships between cloud structures and irradiance patterns, providing valuable information for optimizing the placement of cameras in real-world applications.

Figures 8, 9 and 10 show the results of the solar radiation forecasts for all three evaluation scenarios. Each figure compares the actual irradiance available at the target location with the values predicted by the model. These visualizations provide a clear understanding of how the model performs under different input configurations and illustrate the impact of including multiple cameras on prediction accuracy.

The results show that the model performs satisfactorily in all three scenarios and effectively captures the relationship between cloud motion and solar irradiance variations. However, the differences in the performance metrics reveal the additional benefits of using multiple camera perspectives.

In the single-camera scenario (Scenario A), using only the data from the central camera, the model achieved a R2 score of 0.85. This scenario, illustrated in Figure 8, serves as a baseline for comparison and indicates that the model is capable of learning and predicting irradiation patterns to a reasonable extent even with a single camera. However, the limitations of using only one camera are obvious, as it provides a limited perspective on cloud motion. Without additional viewpoints, the model lacks a comprehensive spatial context, leading to occasional discrepancies between the predicted and actual values.

In contrast, the multi-camera scenario (Scenario B), which includes the images from all five cameras, significantly improves the prediction accuracy and leads to a R2 score of 0.87 as shown in Figure 9. The additional spatial context provided by the peripheral cameras allows the model to better understand cloud dynamics, shadow propagation, and irradiance variations. This improvement demonstrates the benefit of multi-camera configurations as they allow the model to generalize more effectively across different weather conditions and improve short-term forecasting capabilities.

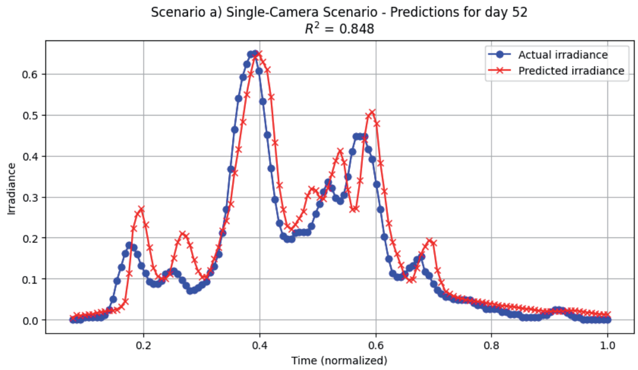

Fig. 8. Predicted and actual solar irradiance values for the single-camera cenario (Scenario A) on Day 52. Only data from the central camera at position (0, 0) were used for prediction.

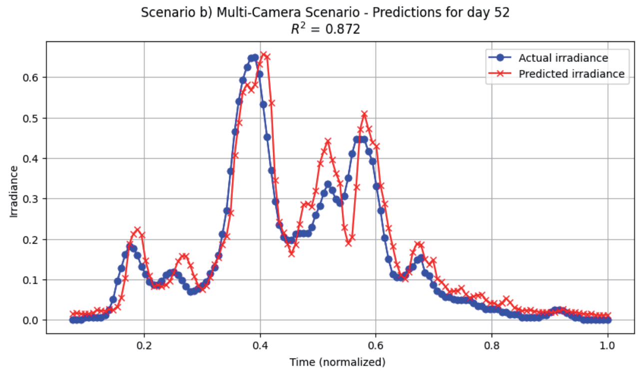

Fig. 9. Predicted and actual solar irradiance values for the multi-camera scenario (Scenario B) on Day 52. The model utilizes image data from five spatially distributed cameras to generate predictions.

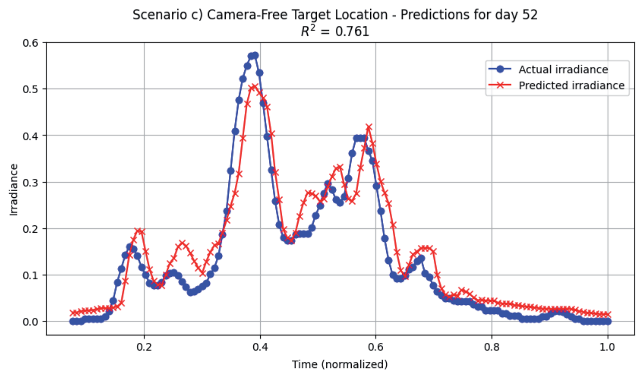

Fig. 10. Predicted and actual solar irradiance values for the camera-free target location scenario (Scenario C) on Day 52. In this setup, the model predicts irradiance at the central location (0, 0) using only images from the four peripheral cameras.

The most critical test is the camera-free target location scenario (Scenario C), where the model predicts solar irradiance at the central location without using direct images from that point. In this case, the model relies solely on the four peripheral cameras to determine the irradiance at the target location. Despite the absence of a direct monitoring device, the model still achieves a

Alen Jakoplić, Dubravko Franković, Tomislav Plavšić, Branka Dobraš, Regional Solar Irradiance Forecasting Using Multi-Camera Sky Imagery and Machine Learning Models, Journal of Energy, vol. 73 Number 4 (2024), 26–33
https://doi.org/10.37798/2024734707

R2 score of 0.76 as is evident from Figure 10., confirming that spatially distributed observations can successfully estimate irradiance for unmonitored areas. Although the performance is slightly lower compared to Scenario A and Scenario B, the results indicate that regional solar forecasting is feasible even if cameras are not installed at every location of interest.

Table I summarizes the R2 performance metrics for all three scenarios shown in Figures 8–10. The single-camera scenario (a) serves as a baseline, while the multi-camera configuration (b) shows improved accuracy due to spatially enriched input. The third scenario (c) still achieves satisfactory prediction performance despite the exclusive use of peripheral cameras, which supports the applicability of the model in cases where direct sky images of the target location are not available.

Table I

Prediction Performance across Different Forecasting Scenarios.

Scenario	Description	R^2 Score
a) Single-Camera Scenario	Only central camera	0.848
b) Multi-Camera Scenario	Central + four peripheral cameras	0.872
c) Camera-Free Target Location	Only four peripheral cameras	0.761

It is important to recognize the limitations associated with using a purely synthetic dataset for this study. While the Unity simulation framework allows for controlled experiments and rapid generation of different scenarios, it inevitably simplifies the complexity of real atmospheric physics. Factors such as unpredictable, rapid weather changes beyond the simulated patterns, variations in aerosol concentration, subtle cloud formations (e.g. thin cirrus clouds), and potential sensor noise or calibration issues with physical cameras are not fully captured by the current synthetic data. While the presented results show the potential and feasibility of the multi-camera approach for regional forecasting, the achieved performance metrics (R2 values) should be interpreted as an upper-bound estimate under idealized conditions. Future work will focus on the validation of this model using real data. It is planned to set up a physical network of sky cameras in the target region to collect real images and irradiance measurements. This realworld data set will be crucial to thoroughly evaluate the practical performance, robustness, and generalizability of the model. It will allow the necessary fine-tuning and adjustment to account for the inherent stochasticity and complexity of actual atmospheric conditions.

In addition to evaluating the forecasting accuracy, the practical feasibility of deploying a multi-camera system must also consider the cost of individual units. The proposed camera modules were designed using low-cost, off-the-shelf components with the goal of supporting scalable deployment. Table II summarizes the hardware components and their associated costs, with the total price of one complete unit remaining below C60. This cost-effective configuration, based on opensource platforms and simple solar power integration, makes the system suitable for distributed implementations in both research and real-world settings. The affordability and modularity of the setup support its application in community monitoring, smart grid demonstrations, and large-scale regional deployments.

Table Ii

Component Cost Breakdown for f Single Sky Camera Unit (2023).

Component	Description	Price [€]
Raspberry Pi Camera	Wide-angle camera compatible with Raspberry Pi	18.32
Raspberry Pi Zero W	Mini computer (single-board)	27.50
Small Solar Panel	5V, 0.125 W (45 × 25 mm)	1.59
INA219 Module	Current and voltage measurement module	7.30
3D-Printed Housing	Protective enclosure	0.89
Mounting Clamp	For camera mounting	1.31
Total		**56.91**

V. Conclusion

The growing adoption of photovoltaic (PV) systems in modern energy grids presents both opportunities and challenges, particularly with regard to the variability of solar energy production. Accurate short-term forecasting of solar irradiance plays a crucial role in ensuring stable and efficient grid operation. This paper demonstrates the potential of using a synthetic database in combination with a machine learning model to overcome these challenges by analyzing cloud dynamics using multiple sky cameras.

The synthetic database developed in this study provides a controlled environment for testing different model configurations, weather conditions, and data structures. This flexibility allows systematic experimentation with different parameters to analyze the potential for regional solar irradiance forecasting. By using images from multiple cameras strategically distributed across the monitored region, the model can detect patterns in cloud movement and shadow dynamics. The results show that the use of multiple cameras significantly improves forecasting accuracy compared to a single-camera setup.

The performance evaluation across the three defined scenarios confirms that multiple perspectives contribute to more accurate predictions, not only for locations with installed cameras but also for locations without direct visual input. The model has successfully demonstrated that it is able to predict solar irradiance at the central target location using only the data from the peripheral cameras. This finding highlights the potential of a distributed camera network to support regional solar forecasting without the need for a dense sensor installation.

The results show that using multiple cameras gives the model a more detailed understanding of cloud dynamics, which in turn improves forecast accuracy. By capturing cloud movement from different angles, the model gains insight into shadow behavior and irradiance variations. This capability is particularly useful for forecasting production at locations without direct camera measurements.

In addition, the synthetic database supports the simulation of camera placements in real regions, enabling the development of optimal configurations for specific locations. This approach makes it possible to train the model in a synthetic environment and then reconcile it with real data. As the synthetic database can be created much faster than collecting real data (in minutes rather than days) it provides a practical and efficient solution for testing different configurations. Once the model has been trained in this flexible environment, it captures the dependencies between cloud movement and shadow formation, reducing the amount of real data needed to fine-tune it to the actual meteorological conditions at the selected location.

Alen Jakoplić, Dubravko Franković, Tomislav Plavšić, Branka Dobraš, Regional Solar Irradiance Forecasting Using Multi-Camera Sky Imagery and Machine Learning Models, Journal of Energy, vol. 73 Number 4 (2024), 26–33
https://doi.org/10.37798/2024734707

This research underlines the importance of flexible, dataefficient approaches to renewable energy forecasting. The method presented not only demonstrates the feasibility of regional predictions of solar irradiance but also highlights the potential of multi-camera systems for accurate forecasting over larger areas. By enabling predictions on a regional scale, this work provides a solid foundation for future advances in solar energy forecasting and its practical implementation in power grid operations.

Future research will focus on deploying the camera modules in real-world environments and collecting observational data to validate the model's forecasting capabilities under actual atmospheric conditions. The ultimate goal is to enable reliable, region-wide solar irradiance forecasting based on a distributed network of low-cost sky imagers. The synthetic framework and results presented in this study provide a robust foundation for this transition and serve as a critical preparatory step toward real-world implementation.

References

[1] A. Qazi, F. Hussain,N.A. Rahim, G. Hardaker, D. Alghazzawi, K. Shaban, & K. Haruna, Towards sustainable energy: a systematic review of renewable energy sources, technologies, and public opinions, IEEE access, 7, 63837-63851, 2019.

[2] Murdock, Hannah E., et al., Renewables 2021-global status report, 2021.

[3] Kumar, Varun, A. S. Pandey, and S. K. Sinha., Grid integration and power quality issues of wind and solar energy system: A review, International conference on emerging trends in electrical electronics & sustainable energy systems (ICETEESES). IEEE, 2016.

[4] Impram, S., Nese, S. V., & Oral, B., Challenges of renewable energy penetration on power system flexibility: A survey, Energy Strategy Reviews, 31, 100539, 2020.

[5] Infield, D., & Freris, L., Renewable energy in power systems, John Wiley & Sons, 2020.

[6] Anvari, M. et al., Short term fluctuations of wind and solar power systems. New Journal of Physics 18, 063027, 2016., New Journal of Physics 18, 063027, 2016.

[7] Akhter, M. N., Mekhilef, S., Mokhlis, H., & Mohamed Shah, N., Review on forecasting of photovoltaic power generation based on machine learning and metaheuristic techniques, IET Renewable Power Generation, 13(7), 1009-1023, 2019.

[8] G. e. Boyle, Renewable electricity and the grid: The challenge of variability, London, UK: Earthscan Publications Ltd., 2012.

[9] S. Jenniches, Assessing the regional economic impacts of renewable energy sources – A literature review, Renewable and Sustainable Energy Reviews, Volume 93, Pages 35-51, ISSN 1364-0321, 2018.

[10] Chaturvedi, D. K., & Isha, I., Solar power forecasting: A review, International Journal of Computer Applications, 145(6), 28-50, 2016.

[11] Yang, B., Zhu, T., Cao, P., Guo, Z., Zeng, C., Li, D., ... & Yu, T., Classification and summarization of solar irradiance and power forecasting methods: A thorough review, CSEE Journal of Power and Energy Systems, 2021.

[12] U. &. W. Z. Munawar, A framework of using machine learning approaches for short-term solar power forecasting, Journal of Electrical Engineering & Technology, 15(2), 561-569, 2020.

[13] Wang, F., Mi, Z., Su, S. and Zhao, H., Short-term solar irradiance forecasting model based on artificial neural network using statistical feature parameters, Energies, 5(5), pp.1355-1370, 2012.

[14] Sivaneasan, B., Yu, C.Y. and Goh, K.P., Solar forecasting using ANN with fuzzy logic pre-processing, Energy procedia, 143, pp.727-732, 2017.

[15] Wentz, V.H., Maciel, J.N., Gimenez Ledesma, J.J. and Ando Junior, O.H., Solar Irradiance Forecasting to Short-Term PV Power: Accuracy Comparison of ANN and LSTM Models, Energies, 15(7), p.2457, 2022.

[16] Oh, M., Kim, C.K., Kim, B., Yun, C., Kang, Y.H. and Kim, H.G., Spatiotemporal optimization for short-term solar forecasting based on satellite imagery, Energies, 14(8), p.2216, 2021.

[17] Cheng, L., Zang, H., Wei, Z., Ding, T., Xu, R., & Sun, G., Short-term solar power prediction learning directly from satellite images with regions of interest, IEEE Transactions on Sustainable Energy, 13(1), 629-639, 2021.

[18] Miller, S.D., Rogers, M.A., Haynes, J.M., Sengupta, M. and Heidinger, A.K., Short-term solar irradiance forecasting via satellite/model coupling, Solar Energy, 168, pp.102-117, 2018.

[19] Ryu, A., Ito, M., Ishii, H., & Hayashi, Y., Preliminary analysis of short-term solar irradiance forecasting by using total-sky imager and convolutional neural network, 2019 IEEE PES GTD Grand International Conference and Exposition Asia, 2019.

[20] Feng, C., Zhang, J., Zhang, W., & Hodge, B. M., Convolutional neural networks for intra-hour solar forecasting based on sky image sequences, Applied Energy, 310, 118438, 2022.

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

[22] Kaggle dataset, "Multi-Camera Solar Forecasting Dataset" [Online]. Available: https://www.kaggle.com/datasets/alenjakopli/ multicamera-solar-forecasting-dataset. [Accessed: Feb. 2025].

Alen Jakoplić, Dubravko Franković, Tomislav Plavšić, Branka Dobraš, Regional Solar Irradiance Forecasting Using Multi-Camera Sky Imagery and Machine Learning Models, Journal of Energy, vol. 73 Number 4 (2024), 26–33
https://doi.org/10.37798/2024734707