Available online www.worldscientificnews.com

World Scientific News

An International Scientific Journal

WSN 212 (2026) 1-31 EISSN 2392-2192

Attention Convolutional Neural Network Model Estimation of Solar Energy Potential Using Radiance Sky Imagery

Richard Iherorochi Nneji[1], Joyce Uchechi Nneji[2], Chibueze Favour Aririguzo[3], Anunobi Miracle Ugomma[4], Confidence Chigozirim Olumba[5], Benjamin Chiemeka Opara[6], Gladys Chinyere Olumba[7], Grace Ugochi Nneji[1, 2, *]

[1]Intelligent Computing Lab, Hace LLC, North Carolina, USA

[2]College of Law, Chengdu University of Technology, China

[3]Department of Civil Engineering, Lumkalt Energy Limited, Rivers, Nigeria

[4]Science Laboratory Technology, University of Nigeria, Nsukka, Nigeria

[5]Computer Science, Federal College of Education Obudu, Cross River State, Nigeria

Educational Management (Political Science), University of Benin, Benin, Nigeria

[6]College of Ecology, Chengdu University of Technology, China

[2]OBU Computing, Chengdu University of Technology, China

*Author for Correspondence: grace.nneji@zycdut.edu.cn

https://doi.org/10.65770/VVBQ3503

ABSTRACT

The rapid growth of renewable energy adoption has heightened the need for accurate solar energy prediction to ensure grid stability, particularly in regions with high solar penetration. However, traditional forecasting methods relying on historical meteorological data often fail to address short term fluctuations caused by dynamic cloud movements, limiting real time adaptability. To overcome this challenge, this study proposes a deep learning framework integrating convolutional neural networks (CNNs) with attention mechanisms to predict photovoltaic (PV) output from radiance sky images.

(Received 10 December 2025; Accepted 15 January 2026; Date of Publication 4 February 2026)

Two datasets capturing diverse sky conditions were used to evaluate three architectures: a baseline CNN, CNN with Squeeze-and-Excitation (SE) Attention, and CNN with Spatial Attention. The CNN with SE-Attention model significantly outperformed baseline models, reducing prediction errors and improving explanatory power, as validated by metrics including RMSE, MAE, and R^2. Gradient-weighted Class Activation Mapping (GradCAM) further demonstrated the model's ability to prioritize meteorologically critical regions, such as cloud edges and solar disk areas, with distinct attention patterns for sunny and cloudy scenarios. The framework's practical utility was enhanced through deployment in an interactive web-based Graphical User Interface, enabling real-time solar potential simulations for energy operators. By combining attention mechanisms with interpretable design, this work advances short-term solar forecasting accuracy while providing actionable insights for grid management. Future research directions include multi-modal data fusion and hybrid transformer-CNN architectures to improve robustness across diverse climatic conditions.

Keywords: Solar energy prediction, Deep learning, Attention mechanisms, Grad-Cam, Sky image analysis.

1. INTRODUCTION

In recent years, with the rapid growth of global energy demand and the improvement of environmental protection awareness, the application of clean energy, especially solar energy, has become the focus of global attention. With its renewable and clean nature, solar energy is seen as an important part of sustainable energy in the future [1]. Accurate solar power forecasting is critical to maintaining grid stability, especially in areas with high solar penetration, where grids need to respond in real time to fluctuations in solar radiation due to cloud cover or weather changes to avoid power supply disruptions due to sudden drop in sunlight [2].

However, because solar power generation is dependent on the intensity and duration of sunlight, rapid changes in weather conditions make traditional forecasting methods challenging. Most of the traditional forecasting methods rely on historical meteorological data, but such methods are difficult to adapt to the rapidly changing weather conditions in time, especially the dynamic changes of cloud cover in a short period of time, resulting in limited real-time and flexible forecasting [3]. In order to address this challenge, deep learning technology has made significant progress in the field of image recognition and pattern recognition in recent years, making it possible to analyze weather conditions and predict solar radiation in real time using radiated sky images [4].

Through real-time analysis of sky images, deep learning models can automatically identify cloud type, thickness, coverage, and other information, so as to provide dynamic and accurate data support for solar potential prediction. This image-based forecasting method has shown significant advantages over traditional methods, especially in response to short-term weather changes [5]. As illustrated in Figure 1, regions with higher solar penetration, such as California and New England, exhibit significantly higher error costs. Although the current correlation between error costs and solar penetration levels is relatively weak, the potential negative impacts of forecasting errors are expected to become more pronounced as the proportion of solar energy in the energy mix continues to increase [6]. Nowadays, ground-based sky imagery and deep learning models have gradually emerged as effective ways to address these short-term fluctuations, reducing uncertainty by providing fast short-term forecasts.

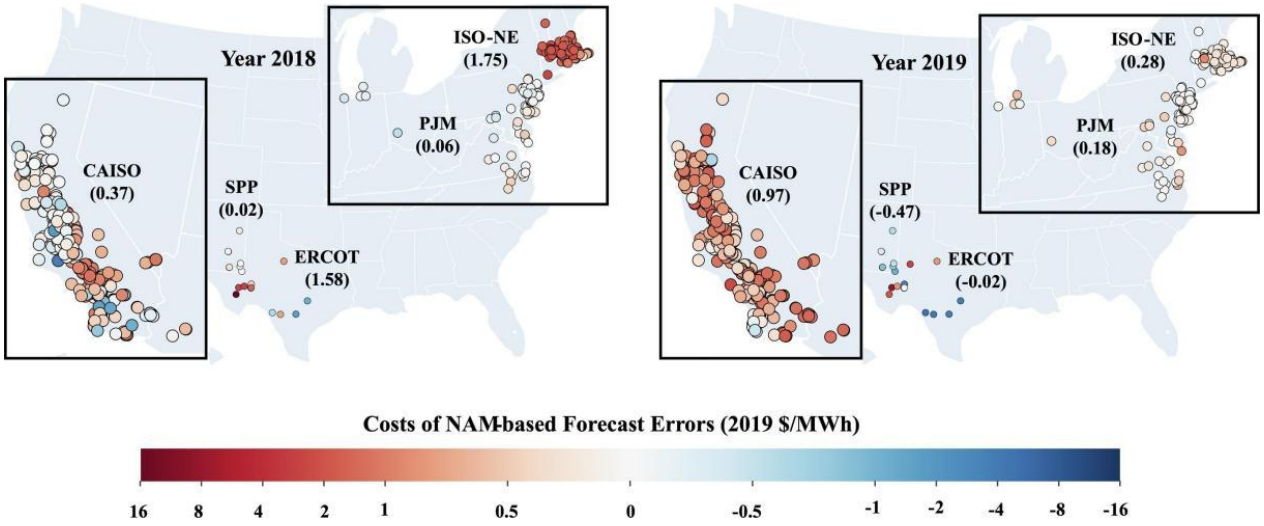


Figure 1. Costs of NAM Forecasts Errors in 2018 and 2019 by Wang et al. [6]

The improvement of model performance is inseparable from the support of high-quality and diverse datasets, which is one of the main challenges facing the field at present. To ensure the reliability of the model, a high-quality dataset representative of a wide range of atmospheric conditions is essential. Although a growing number of open-source sky imagery datasets are becoming available, these datasets vary in coverage, resolution, and quality control, and these factors are critical to training models with high generalization capabilities. Researchers are actively promoting the accessibility of data and the standardization of solar prediction datasets, which can help improve model generalization and support further research in energy meteorology and atmospheric sciences [7].

Predicting solar energy potential by combining radiated sky imagery and deep learning technology, which can not only significantly improve the accuracy of prediction, but also help the grid achieve more efficient power management and dispatch, thereby ensuring the stability and reliability of power supply. This image-based prediction method has strong real-time, low cost, and high scalability, which is expected to bring new application prospects and technological innovation to the field of solar power generation.

With the rapid development of renewable energy across the globe, accurate prediction of solar power generation potential is critical for smart grids and energy management. To achieve this, many studies employ image-based deep learning models to analyze sky images for solar energy prediction. However, traditional methods often struggle to provide real-time, accurate forecasts under rapidly changing weather conditions, such as moving clouds [8]. Therefore, this study proposes a model combining Convolutional Neural Network (CNN) and attention mechanisms to automatically extract and enhance spatial features from irradiated sky images. By integrating these features, the fore casting model aims to correlate sky radiance data with photovoltaic (PV) output, providing efficient and accurate solar energy potential predictions under diverse weather conditions.

The main goal of this research is to develop a deep learning model combining CNN with attention mechanisms for real-time prediction of solar energy potential from sky radiation images. To achieve this overall goal, this study focuses on a series of specific tasks, including data processing, model design and optimization, evaluation metric selection, and performance validation. First, this paper constructs a diverse sky image dataset covering different meteorological conditions, including sunny and cloudy days. To enhance robustness and reduce time-dependent biases, data preprocessing includes grouping by day blocks, shuffling, image normalization, and tensor conversion of PV output. Batch processing is applied to optimize training efficiency. Secondly, the paper designs a CNN-based architecture with integrated attention mechanisms to extract spatial features, such as cloud density and distribution patterns, and enhance the focus on relevant regions. These design choices enable the model to perform real-time and accurate predictions, particularly for short-term photovoltaic output under rapidly changing weather conditions. The model's performance will be assessed using key evaluation metrics, including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Coefficient of Determination (R^2), and Explained Variance Score (EVS). These metrics will quantify prediction accuracy and generalization ability. Furthermore, comparative experiments with traditional persistence models will demonstrate the effectiveness of the nowcasting model, particularly under dynamic weather conditions.

This section introduces a CNN-based model integrated with attention mechanisms for real-time solar energy prediction, emphasizing its potential to enhance energy management and the key stakeholders who will benefit from advancements in accurate predictions of solar energy.

To facilitate a clearer understanding of the research's overall architecture, Figure 2 is a schematic representation depicting the key components of the research.

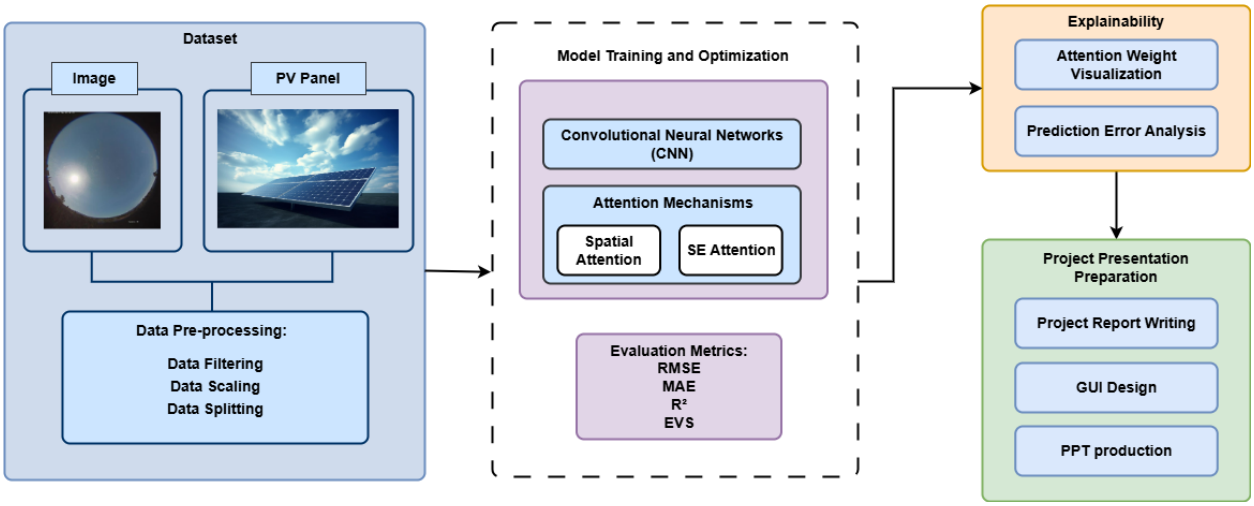


Figure 2. Flow chart of research overview.

In recent years, with the rapid development of deep learning technology, the CNNs have achieved significant success in the field of image processing and have been increasingly applied to complex tasks such as weather prediction and energy management. However, standard CNN models may face challenges when dealing with complex atmospheric conditions and dynamic weather patterns, such as limited adaptability to spatial variations and reduced robustness under rapidly changing conditions. To address these limitations, this paper introduces a CNN-based model enhanced with attention mechanisms.

The CNN layers extract spatial features such as cloud density and light intensity from sky images, while the attention modules emphasize relevant regions, improving the model's ability to handle dynamic weather scenarios and enabling accurate real-time prediction of solar energy potential.

The potential contributions and practical implications of this research include:

- Accurate prediction of solar power potential by making solar energy systems smarter and more reliable by analyzing cloud distribution and light intensity in radiated sky images to provide round-the-clock real-time predictions.

- Optimize the management and deployment of solar energy in the grid through real-time prediction, reduce energy waste, and improve the stability of power supply.

- Provide real-time predictive support for smart grids, photovoltaic power generation systems, distributed energy management systems, etc., and enhance the adaptability of solar energy systems.

- Improve forecasting economics and scalability by reducing backup energy requirements and reduce overall operating costs through efficient data processing and predictive models.

- Provide technical support for green energy transition by promoting social trust and use of clean energy through efficient and convenient prediction tools.

2. LITERATURE REVIEW

The application of deep learning in solar radiation prediction is becoming more and more extensive, and several research teams have proposed different model structures to improve the accuracy and timeliness of prediction. Feng and Zhang [9] proposed the SolarNet, a 20-layer deep CNN specifically designed for hourly prediction of global horizontal irradiance (GHI). Alani et al. [2] further researched the application of CNNs by developing a hybrid CNN-Multi-Layer Perceptron (MLP) model, which are used to extract spatial features from images, and MLP networks are used to explore complex relationships between image information, GHI, and different weather variables. Papatheofanous et al. [10] introduced a CNN-based Image Regression Module (IRM) for estimating short-term solar radiation from sky images.

In addition to CNN structures, more and more studies are trying to introduce time series information into predictive models to capture the dynamic changes in cloud cover and radiation. Paletta et al. [11] compared the effects of four deep learning models: CNN, CNN-Long Short-Term Memory (LSTM), 3D-CNN, and ConvLSTM, and the experimental results showed that recurrent neural networks containing time series information (LSTM, 3D-CNN, and ConvLSTM) outperformed traditional CNN models in prediction accuracy, especially 3D-CNN and ConvLSTM, performed well in short-term prediction skill indicators. Zhang et al. [12] also evaluated three deep learning models: MLP, CNN, and LSTM, for photovoltaic nowcasting using sky images, finding that the LSTM-based model significantly outperformed others, achieving a 21% improvement in RMSE skill score over the persistence baseline in predicting one-minute-ahead solar power output.

In this section, three main approaches to daylight forecasting are presented: traditional methods, machine learning methods, and deep learning methods. Among the deep learning methods, this section describes the application of CNNs and attention mechanisms, which have shown great potential to improve the accuracy and efficiency of daylight forecasting in particular.

2.1. Traditional Method of Solar Radiation Prediction

Traditional solar radiation prediction methods mainly rely on empirical models, statistical methods, and physical simulations, and since the 80s of the 20th centuries, time series methods have become the early focus of PV forecasting research, with countries such as Europe, the United States, and Japan leading the development of PV technology. Sidrach-de-Cardona and Lopez [13] of the University of Malaga, Spain, were among the first scholars to apply multiple linear regression models to predict the energy output of independent photovoltaic systems. Chowdhury and Salfur [14] used autoregressive moving averages (ARMA) and autoregressive moving averages (ARMA) to predict the energy output of photovoltaic systems. ARMA) and Autoregressive Integral Moving Average (ARIMA) models to study the energy output of photovoltaic systems, further developing this field. Subsequently, Hassanzadeh et al. [15] proposed an ARMA model for hourly PV generation forecasting in collaboration with NV Energy. Besides, to address large-scale solar forecasting, Numerical Weather Prediction (NWP) systems solve fluid dynamics and radiative transfer equations, though their computational complexity limits temporal resolution. Jimenez et al. [16] researched a WRF-Solar model, which is an operational NWP model optimized for solar forecasting, integrating aerosol-radiation feedbacks and high-temporal-resolution outputs to address industry needs, which is shown in Figure 3.

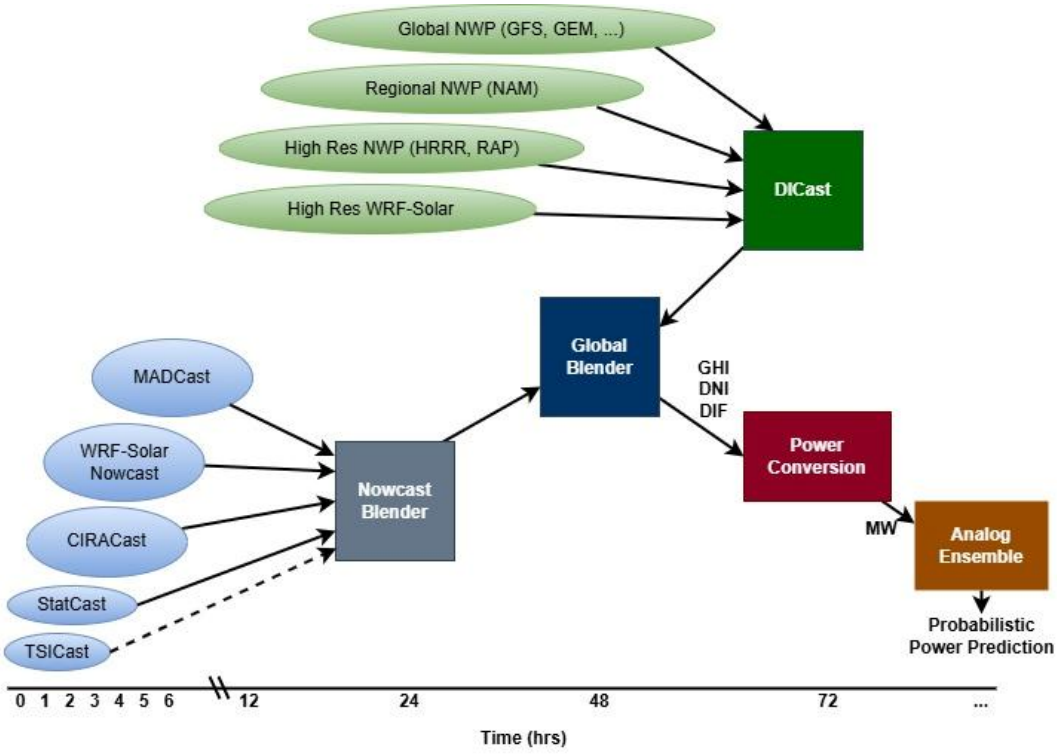


Figure 3. Suncast solar power forecasting system [16].

2.2. Machine Learning Method of Solar Radiation Prediction

Machine learning has greatly improved the ability to predict solar radiation, which solves the limitation that it is difficult for traditional power systems to accurately predict unconventional power generation [17]. Machine learning models enable accurate predictions by using historical solar data to identify and exploit complex relationships between inputs and outputs. This process begins with preprocessing of large datasets to ensure that the model is trained on relevant, high-quality data. Machine learning models in this field typically employ one of three approaches: models based on meteorological and geographic parameters, time-series models using historical solar irradiance data, and hybrid models with exogenous variables [18]. Figure 4 illustrates a comprehensive framework for forecasting PV generation using machine learning techniques, emphasizing the integration of data collection, data preparation, model training, and predictive validation. The most commonly used machine algorithms for solar irradiance prediction include K-nearest neighbor (KNN), support vector machine (SVM), decision tree (DT), and random forest (RF). However, despite its advantages, machine learning models also have their limitations, such as relying on large amounts of high-quality training data, being prone to overfitting, and requiring large amounts of computational resources [19]. These challenges need to be further explored, especially as the use of photovoltaic power generation systems continues to increase, and the need for advanced data preprocessing techniques to manage large and complex data sets increases.

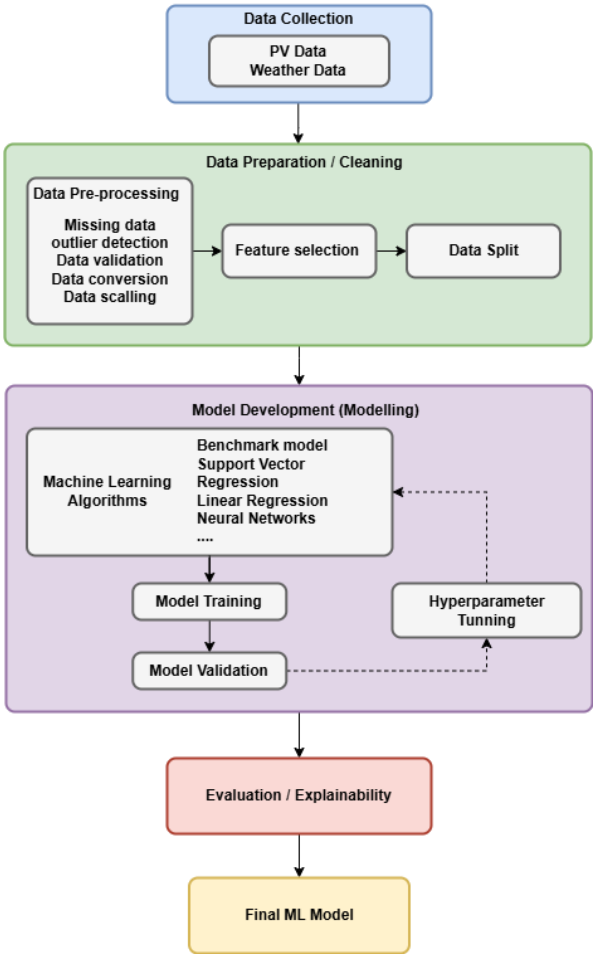


Figure 4. ML Solar PV Power [17].

2.3. Deep Learning Method of Solar Radiation Prediction

It has been proven that deep learning methods, especially convolutional neural networks (CNNs), can effectively extract spatial features from sky images for solar irradiance prediction.

2.3.1. CNN Models

As shown in Figure 5, a convolutional neural network consists of a convolutional layer, an aggregation layer, and a fully connected layer. The convolutional layer uses adaptive filters to extract spatial features, the aggregation layer reduces feature sampling to improve efficiency, and the fully connected layer maps features to output. During training, the kernel parameters were optimized using backpropagation and gradient descent, which allowed the CNN to learn patterns from low-level to high-level structures hierarchically [20]. This adaptability makes CNNs very effective for image-based tasks, such as sun prediction by analyzing sky images.

Researchers are actively working on neural networks, with a particular focus on CNNs and their hybrid models. In integrated systems, these models often improve prediction accuracy, especially in the case of short-term forecasts [21]. While researchers have made important contributions to the development of solar irradiance prediction, CNNs are inherently incapable of processing temporal information. Because of this limitation, further research is needed on how to combine CNNs with other models and methods that can process temporal data so that forecasts can be based on both temporal variation and spatial analysis.

2.3.2. Attention Mechanism

The integration of attention mechanisms into deep learning models has significantly enhanced the accuracy of solar radiation prediction by enabling models to selectively focus on critical input data. Wang and Zhang [22] proposed a CNN-Attention model for solar irradiance prediction, which standardizes and normalizes input features before processing them through a convolutional layer with ReLU activation and L2 regularization, which depicted in the Figure 6.

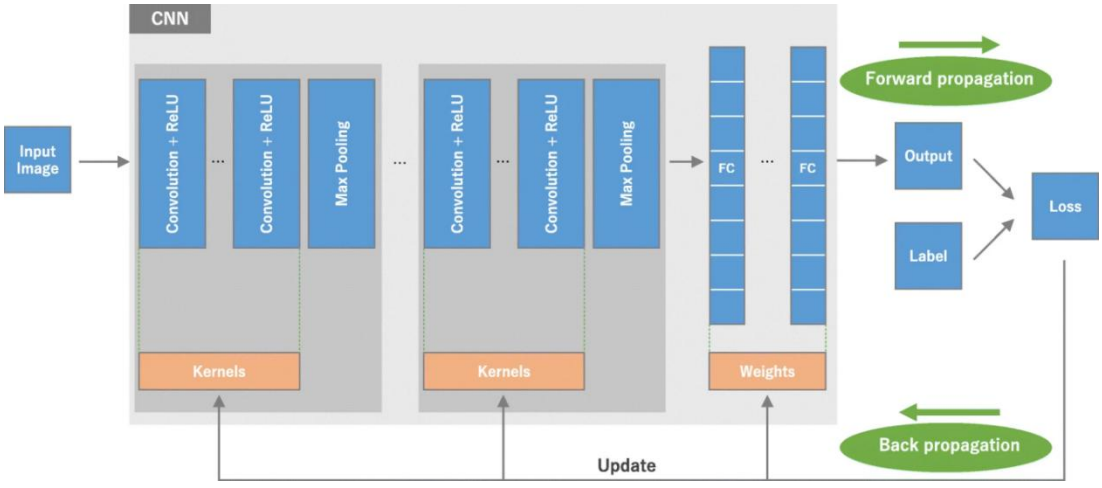


Figure 5. Convolutional Neural Network Overview by Yamashita et al. [20]

The attention mechanism integrates feature maps, ensuring vital features are considered, followed by another convolutional layer for enhanced feature extraction. The model outputs solar irradiance predictions via a linear activation Conv2D layer, trained on a comprehensive dataset from 48 scenarios for accurate estimation.

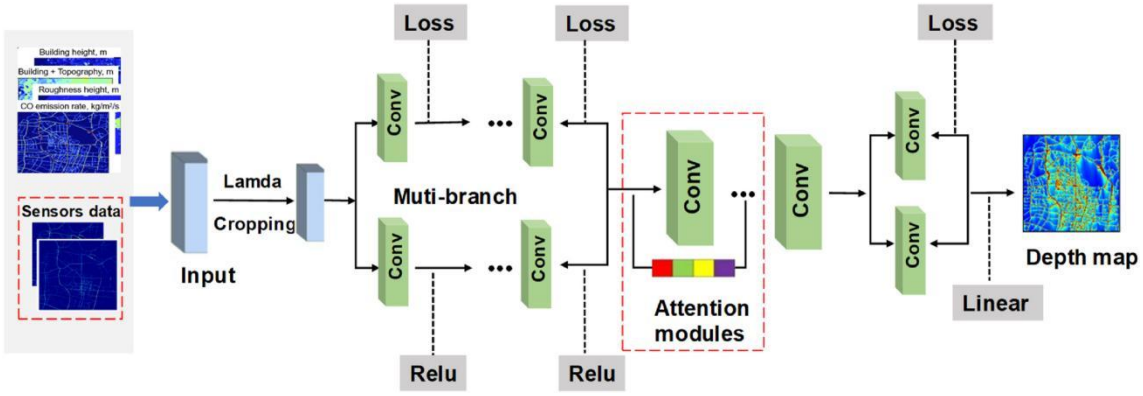


Figure 6. Architecture of CNN-Attention model by Wang and Zhang. [22]

Jonathan et al. [23] introduced an attention-embedded Convolutional Neural Network (ATT_CNN) model for solar irradiance forecasting using sky image sequences, achieving superior accuracy. Qu et al. [24] proposed integrating attention into CNN-LSTM models to capture both short-term and long-term temporal changes in time series data, further improving prediction accuracy. This capability is crucial for optimizing photovoltaic system performance by enhancing responsiveness to changing weather conditions. While attention mechanisms improve model interpretability and prediction precision, they also introduce complexity and increase computational demands. Despite these challenges, their benefits in solar radiation forecasting make them a valuable addition to predictive tools for managing and optimizing solar energy resources. Table 1 summarizes the datasets used and the results obtained for the different studies:

Table 1. Summary of Related Works.

Author	Datasets	Methods & Models	Limitations	Results
Feng and Zhang [9]	SRRL	CNN	Weather impact and single input	nRMSE: 8.85% FSS: 25.14%
Alani et al. [2]	GEP	CNN-MLP	Limited to single time and site	RMSE:13.05W/m²-49.16W/m²
Papatheofanous et al. [10]	CA	CNN	Need real-time irradiance forecasting application	RMSE: 10.44W/m²

Paletta et al. [11]	SIRTA	CNN, CNN-LSTM, 3D-CNN, ConvLSTM	Limited effectiveness of past image sequence training	10-min ahead forecast skill: 20.4%
Zhang et al. [12]	Hemispherical HDR sky images	MLP, CNN, and LSTM	Single site, single camera, and single photovoltaic panel	RMSE: 21%
Jonathan et al. [23]	SRRL	Attention-embedded CNN	Rely solely on sky image sequences	RMSE: 62.75W/m² MBE: 2.71W/m² FSS: 38.81%
Qu et al. [24]	Alice Springs photovoltaic power system	Attention-based CNN-LSTM	Long prediction range is limited	nRMSE: 6.34%

3. METHODOLOGY

This section introduces the core methods of solar forecasting, including the following key aspects: proposed attention-based CNN architecture design, sky image data collection, and data prepossessing steps. The approach aims to address the challenges of short-term solar forecasting by establishing a systematic workflow from data processing to predictive modeling.

3.1. Proposed Model Structure

This section introduces the proposed deep learning model, which focuses on improving the accuracy of solar energy prediction through attention-enhanced feature extraction. The development of the model begins with the construction of a basic CNN for sky image processing, and then use separately two attention mechanisms, extruded excitation (SE) attention and spatial attention, to improve weather feature learning. Finally, a comparative analysis is carried out to optimize the prediction performance and interpretability of the proposed model in solar forecasting applications.

3.1.1. CNN Model

The CNNs are the backbone of this study's deep learning architecture. CNNs are highly effective in processing grid-structured data like images, thanks to their convolutional layers, which extract spatial features through kernel operations. A kernel K, represented as a small matrix of weights, slides over the input image I to compute feature maps F:

$$F(x,y) = \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} K(i,y) \cdot I(x+i, y+i) \qquad (1)$$

Here, m and n represent the kernel dimensions, and (x,y) denote the spatial coordinates of the output feature map. CNNs are particularly adept at capturing local patterns such as edges, textures, and shapes, making them ideal for extracting cloud and radiation-related features from sky images.

3.1.2. Attention CNN Model

Attention mechanisms are incorporated into the CNN framework to enhance its ability to focus on the most critical regions of the input. This research employs Squeeze-and-Excitation (SE) Attention; nevertheless, for a better justification why we have focused on SE-Attention, the research did an ablation study on the other type of attention mechanism called spatial attention.

Spatial Attention focuses on "where" the model should look in the image by emphasizing spatially important regions. The attention map is computed as:

$$A_8 = sigmoid(Conv2D([F_{avg}, F_{max}])) \quad (2)$$

Where and represent the average-pooling and max-pooling operations applied along the channel dimension of feature maps F. Although this mechanism is effective in identifying important spatial regions, it is still insufficient for solar forecasting. For example, it treats all channels in the same way, ignoring their unique spectral signatures that are essential to distinguish between cloud types and solar features. In addition, the clustering operations it performs magnify local artifacts, causing larger weather patterns that span multiple regions to be ignored. These limitations prompted this study to focus more on SE-Attention mechanism in order to better capture the channel relationships that are critical for accurate solar forecasting.

With the limitation of spatial attention mechanism, this research has focus on squeeze and excitation (SE) attention. SE Attention focuses on "what" features to emphasize by adaptively recalibrating channel-wise feature responses. This involves a two-step process:

- Squeeze: Global average pooling reduces each feature map to a single value.

- Excitation: Fully connected layers apply non-linear transformations to model inter-dependencies between channels:

$$s = \sigma(W_2 \cdot ReLU(W_1 \cdot z)) \quad (3)$$

Here, is the squeezed vector, and are trainable weights, and σ denotes the sigmoid activation.

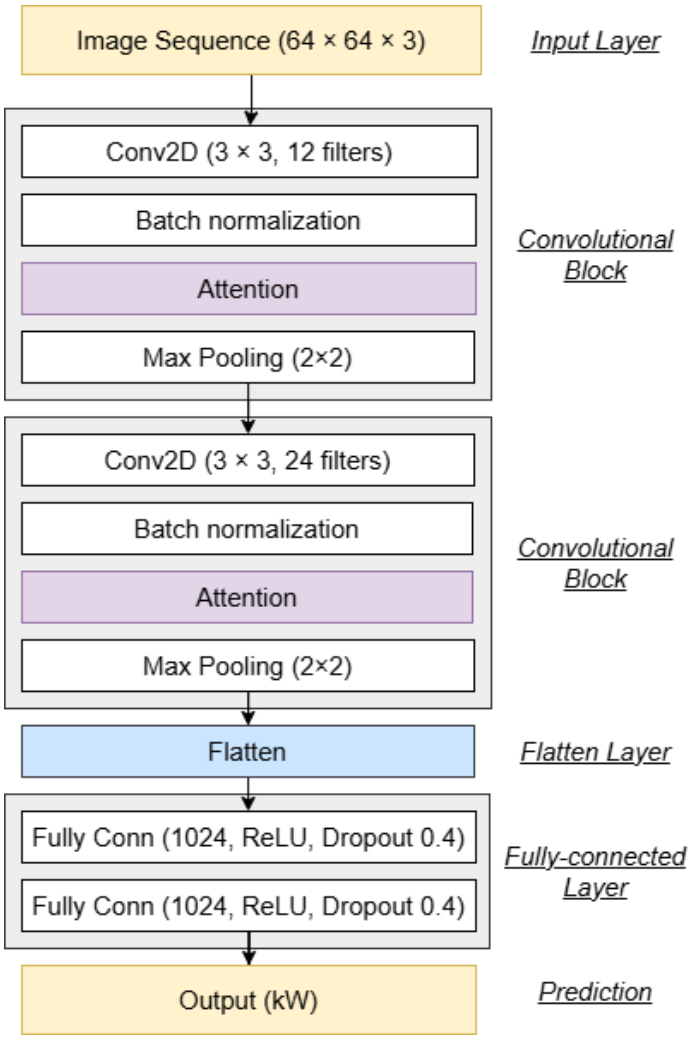


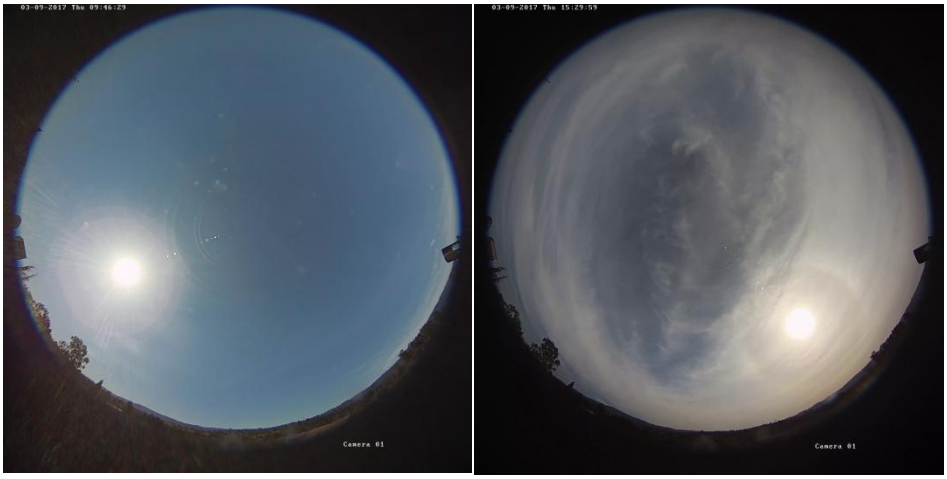
Figure 7. Proposed model architecture.

The architecture of the proposed model, illustrated in Figure 7, combines CNN with attention mechanisms to predict solar energy potential from radiance sky images. The input is a 64×64×3 image representing the radiance distribution of the sky. The model consists of two convolutional blocks, each comprising a convolutional layer, batch normalization, an attention module, and max pooling. These blocks progressively extract and refine spatial features, focusing on the most relevant regions through attention mechanisms. The extracted features are flattened and passed through two fully connected layers, each with 1024 neurons and dropout regularization. The final output layer predicts the solar energy potential as a single numerical value, making this model suitable for regression tasks. The model is trained with the Mean Squared Error (MSE) loss function and the Adam optimizer, with a learning rate of 3×10^{-6}, to ensure efficient parameter updating and stable convergence. To avoid overfitting, a dropout rate of 0.4 was used in the fully connected layers, while batch normalization stabilized the learning of the convolutional layer. In addition, the experiment used 10-fold cross-validation to verify performance to create a robust evaluation framework that improves the generalization ability of the model.

3.2. Dataset Collection

The first dataset, called Sky Images and Photovoltatic Power Generation Dataset (SKIPPD), was developed by Stanford University's Environment Assessment and Optimization (EAO) Group. The SKIPPD dataset includes sky imagery and photovoltaic data, making it suitable for short-term solar forecasting. The sky image was recorded by a 6 MP 360-degree fisheye camera at video resolution at 2048x2048 pixels, running at 20 frames per second. The recorded video then generates an image in JPG format at a 1-minute sample rate. Figure 8 shows examples of sky images in different weather conditions. In addition, raw photovoltaic (PV) data was recorded for a photovoltaic installation located approximately 125 meters away from the camera. These PV data are also recorded at one-minute intervals, which coincides with the frequency at which sky images are captured, allowing for accurate correlation between visual data and output power [25].

The second dataset, also collected on the 6Stanford campus, contains 2048 x 2048 pixels ultra-high-resolution video images of the sky recorded at 20 frames per second and every 60 seconds using photographic images and solar measurements of the same resolution. This continuous imaging method preserves complete spatial details and records changing cloud covers, allowing for a detailed study of the effects of the atmosphere on solar energy production [26].



(a) sunny (b) cloudy

Figure 8. Sample of the sky images in different weather conditions; (a) is sunny day and (b) is cloudy day.

In this study, the benchmark dataset used has undergone necessary image preprocessing steps, which include resizing the image frames and filtering out duplicate images caused by sporadic anomalies in the OpenCV video capture functionality. Additionally, the sky photos and PV power generation data in the dataset are organized into aligned pairs to ensure consistency and alignment of the data. Figure 9 illustrates the distribution of photovoltaic power generation in both the development and test sets, and a detailed profile of PV power generation over 20 specific days in the test set.

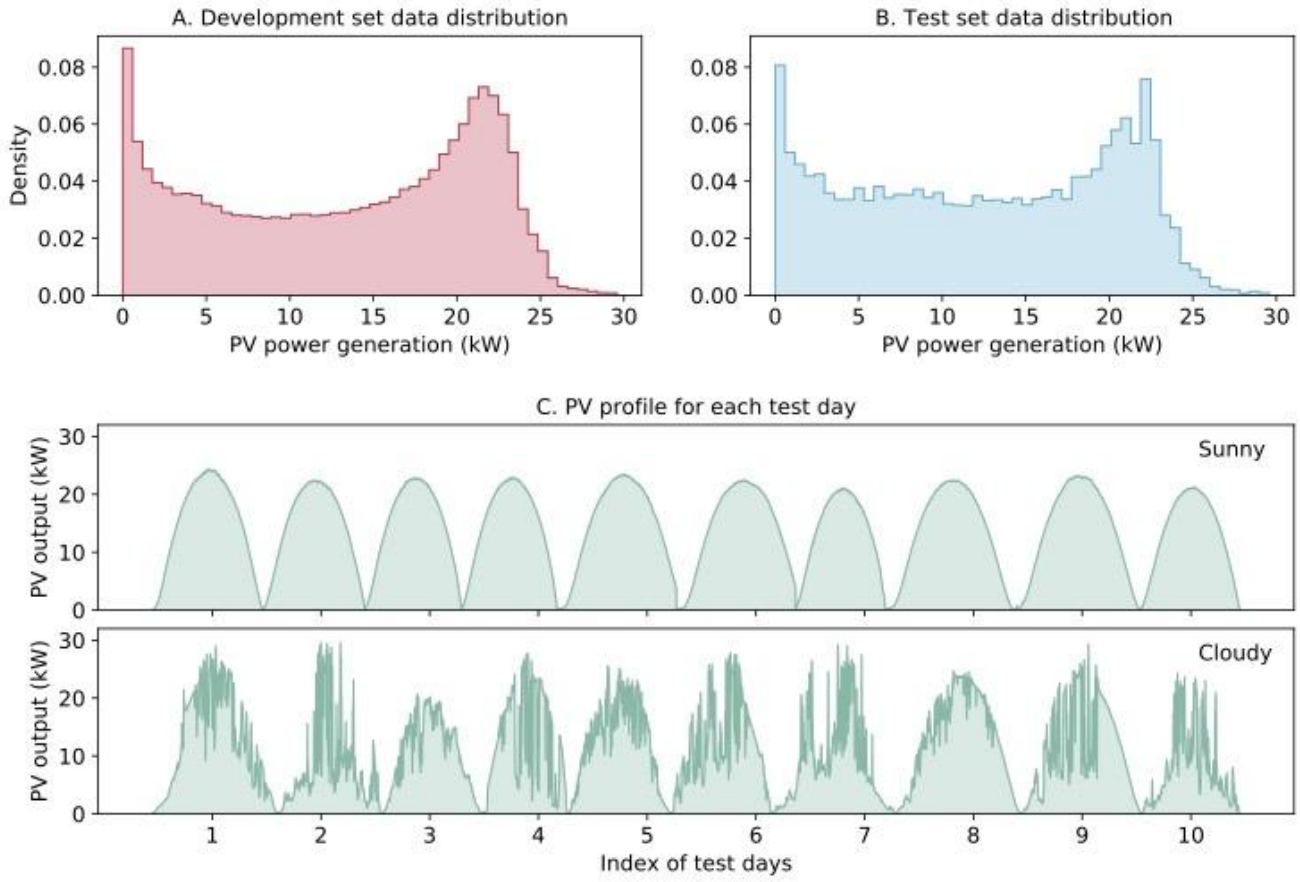


Figure 9. The PV power generation data distribution: A is the development set data distribution, B is the test set data distribution, C is PV profile for each test day [25]

3.3. Data Preprocessing

The data preprocessing phase establishes a robust foundation for model training by systematically transforming raw sky images into standardized inputs while ensuring representative data partitioning. This stage addresses two critical requirements for solar forecasting models: consistent input dimensions and temporally coherent evaluation splits that reflect real-world operational conditions.

3.3.1. Data Resizing

Sky images undergo spatial and numerical standardization to meet model input specifications. Original images are rescaled to a uniform resolution of 64×64 pixels through aspect ratio-preserving transformations, followed by pixel value normalization to the [0,1] range using linear scaling. This standardized processing offers two advantages: the ability to batch image data in different weather conditions and the ability to reduce brightness differences due to changes in solar altitude and weather fluctuations. During the pre-processing process, we paid special attention to maintaining the color fidelity of the RGB three-channel, which is essential for the subsequent attention mechanism to effectively identify the texture features of the clouds.

3.3.2. Data Partitioning

In the data partitioning section, the observations are clustered into daily chunks of the full solar cycle, which are randomly shuffled while preserving the intra-day chronological order. The hierarchical k-fold cross-validation scheme allocates the entire day block into different folds, ensuring that each validation set contains unique weather states that are proportional to the weather conditions that occur in the entire dataset. The above steps can be used to maintain time consistency during the performance evaluation process so that the model can be exposed to different cloud mobility scenarios during the training iteration process. The partitioning logic used in the study, combined with index remapping capabilities, allows for both full-resolution datasets and computationally optimized subsets without compromising the consistency of the evaluation.

3.3.3. Evaluation Metrics

The evaluation metrics of the proposed model are designed to rigorously assess its performance in predicting solar energy potential from radiance sky images. The evaluation process involves the use of a separate test dataset, appropriate preprocessing, performance metrics, and visual analysis to validate the model's effectiveness and generalizability. The test dataset consists of 14,003 samples, including images and corresponding PV output values. These samples are preprocessed by normalizing image data to the range [0, 1] and ensuring consistency in data types. To analyze the model's performance under varying weather conditions, the test data is categorized into sunny and cloudy days based on known dates. This allows for a more granular evaluation of the model's behavior in different scenarios.

The testing process begins by evaluating the performance of each of the 10 models trained during cross-validation on the test dataset. For each model, predictions are generated and compared against the ground truth PV outputs to calculate key evaluation metrics. The ensemble approach is applied by averaging the predictions of all models, resulting in a final prediction that leverages the strengths of multiple models to enhance robustness. To quantify the model's performance, several metrics are employed, including Root RMSE, MAE, R^2, and EVS. RMSE measures the average magnitude of the prediction error and is computed as follows, where n is the number of data, is the predicted value, and is the measured value:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \qquad (4)$$

MAE calculates the average absolute error between predictions and ground truth values, expressed as:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \qquad (5)$$

The coefficient of determination (R^2) evaluates how well the predictions explain the variance in the ground truth data:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2} \qquad (6)$$

Finally, the Explained Variance Score (EVS) measures the proportion of variance captured by the model:

$$EVS = 1 - \frac{Var(\hat{y}-y)}{Var(y)} \quad (7)$$

3.4. Experimental Setup and Technology

To ensure the effective development and evaluation of our deep learning models for solar energy prediction, the research has carefully selected a robust technology stack that balances computational efficiency, development flexibility, and reproducibility. The chosen tools and hardware components are specifically optimized for handling image-based deep learning tasks, particularly the processing of high-resolution sky images and the implementation of attention mechanisms. Table 2 below outlines the complete technology stack that will be utilized in this research, including both software frameworks and hardware specifications.

Table 2. Summary of Technology for the Research.

Software	Framework	TensorFlow
	Language	Python
	Libraries	NumPy, Keras, Matplolib, itertools, h5py
Hardware	Central Processing Unit (CPU)	Gen Intel(R) Core(TM) i7-11800H @ 2.30GHz 2.30 GHz
	Graphic Processing Unit (GPU)	NVIDIA GeForce RTX 3050

4. EXPERIMENTAL RESULT AND ANALYSIS

This section presents the practical development and outcomes of the solar energy prediction system. It covers three main aspects: the design and implementation of the deep learning models, the explainability analysis of model decisions, and the deployment of an interactive Graphical User Interface (GUI) for real-world application. Together, these components demonstrate the research's technical execution, interpretability, and practical usability.

4.1. Design and Implementation

This section details the design, development, and implementation of a solar potential prediction model. Three model variants, benchmark CNN, CNN with spatial attention, and CNN with SE attention, were evaluated to find the best performing model. In addition, we also visually analyzed the prediction results of the best model to verify its practical application value.

To ensure fair comparison, all models were trained and evaluated under the same experimental conditions. Each architecture uses 10-fold cross-validation, combined with the division of time-series block data, which not only retains the time dependence but also ensures the balanced distribution of the data at each fold. The parameters were reinitialized at the start of training, using the Adam optimizer, with the mean square error (MSE) as the optimization target. To prevent overfitting, we employ an early stop mechanism with a 5-epoch patience window and automatically save the optimal model weights based on validation losses. During training, training and validation losses for each iteration are recorded for detailed convergence analysis.

The architectural differences between the three models are mainly in the attention mechanism while maintaining the same underlying convolutional layer. The baseline CNN is used as a reference model without any attention module, while the spatial attention variant dynamically weights the characteristics of a specific region through the spatial importance map of learning. In contrast, the SE attention model uses channel-level feature recalibration to enhance discriminative feature learning by squeezing and excitation.

Performance evaluation was conducted through a multi-dimensional assessment framework, including convergence behavior analysis via training/validation loss, optimal model selection per fold based on validation performance, with comprehensive results across both training and test sets presented in Table 4.

Table 4. The overall results.

		Train loss	validation loss	RMSE	MAE	R^2	EVS
Dataset 1	CNN	2.20	2.00	2.482	1.524	0.895	0.902
	CNN+Spatial-Attention	1.46	1.50	2.347	1.390	0.906	0.912
	CNN+SE-Attention	1.70	1.72	2.231	1.311	0.908	0.912
Dataset 2	CNN+SE-Attention	4.05	4.01	3.927	2.792	0.655	0.655

As shown in Table 4, the baseline CNN shows higher training loss than validation loss, suggesting mild underfitting. In contrast, attention models demonstrate closely matched losses, with SE at 1.70 versus 1.72 and Spatial at 1.46 versus 1.50, indicating strong generalization. The SE-Attention's smaller gap highlights its stability, likely due to effective channel-wise feature weighting. Moreover, the CNN + SE-Attention model outperforms both the baseline CNN and the CNN + Spatial Attention variants across all metrics, including RMSE, MAE, R^2. The improvement in RMSE and MAE highlights the effectiveness of SE-Attention in enhancing prediction accuracy, particularly under dynamic weather conditions. Furthermore, the EVS metrics demonstrate the SE-Attention model's superior ability to explain and capture variance in the data.

While SE-Attention achieved the best results on Dataset 1, RMSE 2.231 and MAE 1.311, its performance degraded significantly when applied to Dataset 2, RMSE 3.029 and MAE 1.709, despite maintaining similar R^2 and EVS scores. This suggests that while SE-Attention generalizes well within Dataset 1's distribution, it struggles with Dataset 2's characteristics, likely due to domain differences in the video-extracted images. The performance metrics of first dataset, including sunny and cloudy conditions, are summarized in Table 5.

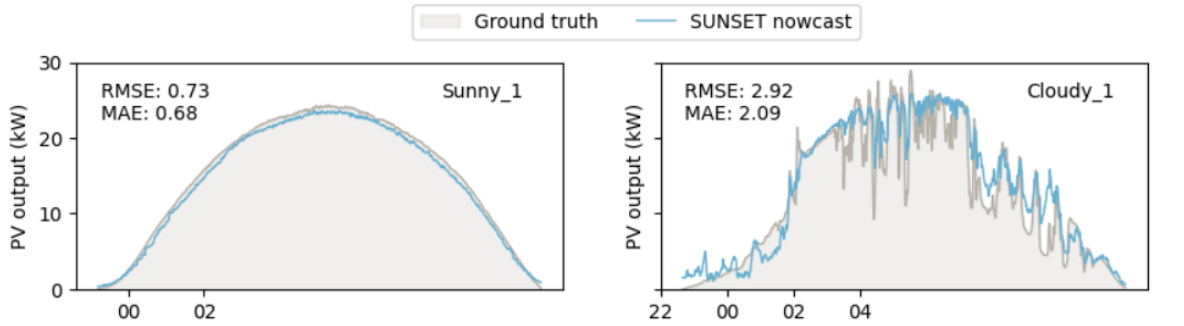
Table 5. Dataset 1 results including sunny and cloudy days.

		RMSE	MAE	R^2	EVS
Sunny	CNN	0.751	0.61	0.999	0.993
	CNN +spatial-Attention	0.675	0.547	0.992	0.996
	CNN+SE-Attention	0.511	0.414	0.995	0.997
Cloudy	CNN	3.425	2.434	0.793	0.805
	CNN+spatial-Attention	3.246	2.229	0.814	0.823
	CNN+SE-Attention	3.238	2.204	0.815	0.822

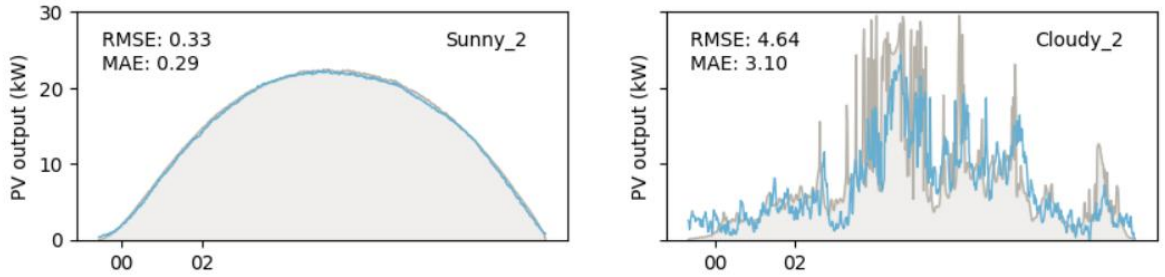
To further validate the best performance CNN + SE-Attention model in the first dataset, its predictions were visualized for representative sunny and cloudy days. Figure 10 illustrates the predicted versus actual PV outputs under these two distinct weather conditions. In sunny scenarios, the model consistently achieves high accuracy in sunny conditions. Across all ten repetitions, the RMSE values for sunny days range from 0.23 in Sunny_5 to 0.76 in Sunny_3, with a mean of 0.49 and standard deviation of 0.19. Similarly, MAE spans from 0.17 in Sunny_5 to 0.71 in Sunny_3, averaging 0.43 with a standard deviation of 0.19. Notably, Sunny_5 yields the lowest errors, RMSE 0.23 and MAE 0.17, whereas Sunny_3 in Repetition 3 shows slightly elevated values, RMSE 0.76 and MAE 0.71. Despite minor fluctuations, all repetitions maintain RMSE below 0.76 and MAE below 0.71, demonstrating robust performance under stable atmospheric conditions.

In contrast, for cloudy scenarios, the model reveals higher variability, reflecting the complexity of dynamic cloud cover. RMSE spans from 1.44 in Cloudy_8 to 4.70 in Cloudy_4, with a mean of 3.15 and standard deviation of 1.18. MAE ranges from 1.17 in Cloudy_8 to 3.31 in Cloudy_4, averaging 2.26 with a standard deviation of 0.85. Early repetitions, like Cloudy_2 with RMSE 4.64 and MAE 3.10, exhibit significant errors, but later iterations show progressive stabilization. For instance, Cloudy_8 achieves the lowest errors, RMSE 1.44 and MAE 1.17, and Cloudy_10 maintains RMSE 2.23 and MAE 1.50.

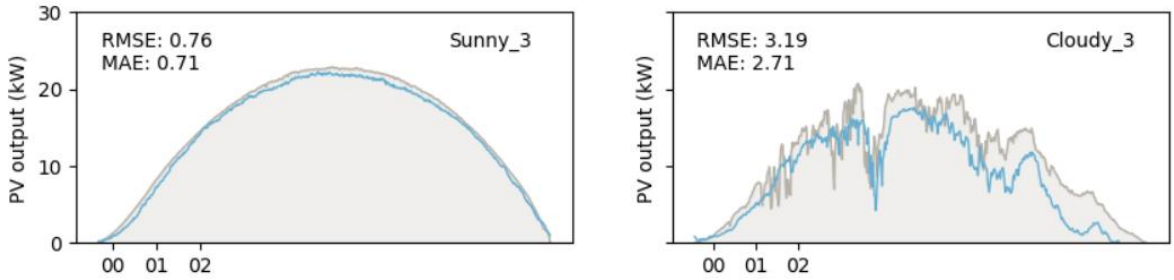
The dual visualization in Figures 10 shows that sunny predictions remain stable with minimal variability, confirming the model's reliability in ideal conditions. Besides, cloudy predictions initially fluctuate but gradually stabilize, with RMSE decreasing by 52% from Repetition 4 to Repetition 10 and MAE improving by 55%. While short-term variability exists in challenging conditions, aggregated performance across repetitions ensures reliable generalization for real-world applications.



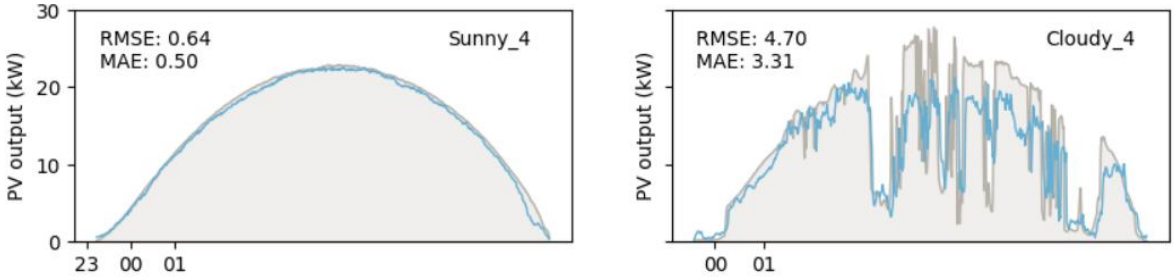
(a) Repetition 1 model



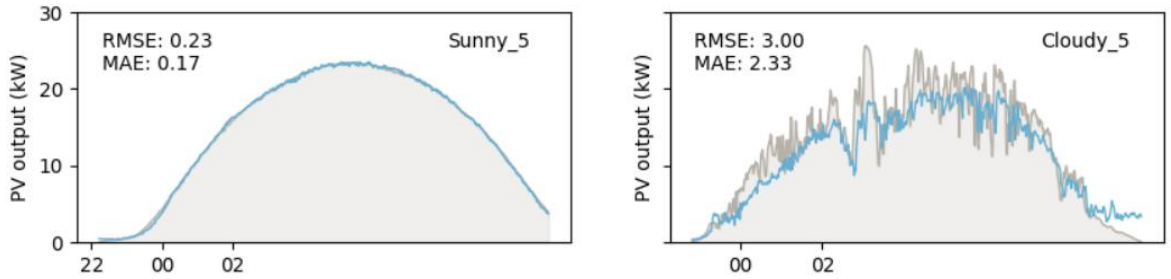
(b) Repetition 2 model



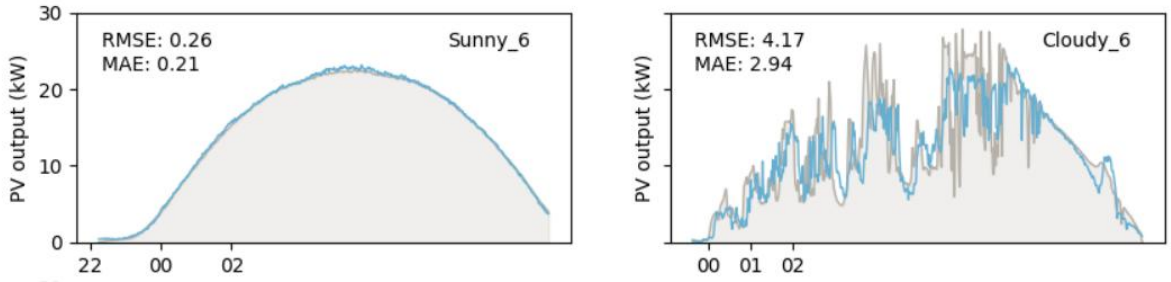
(c) Repetition 3 model



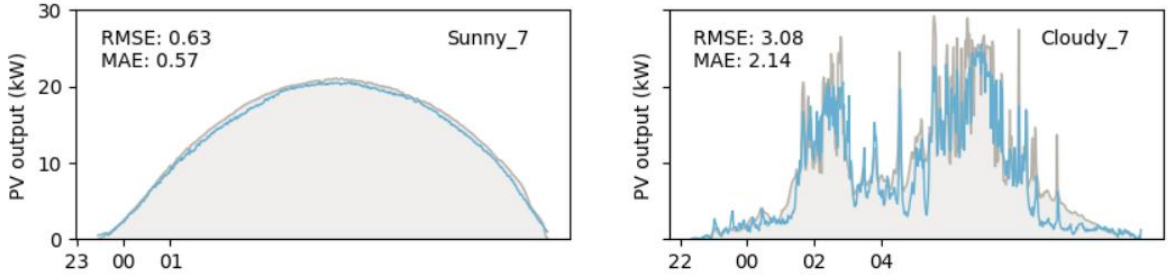
(d) Repetition 4 model



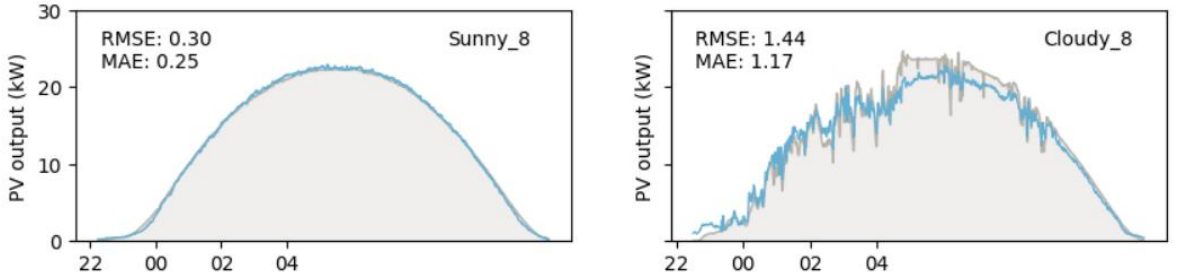
(e) Repetition 5 model



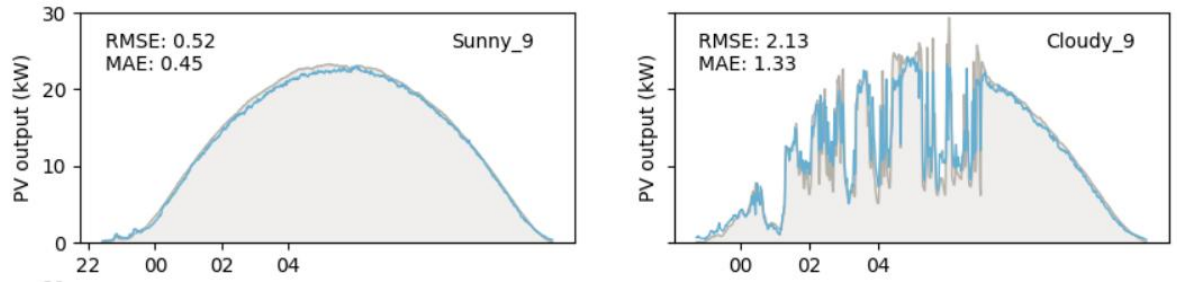
(f) Repetition 6 model



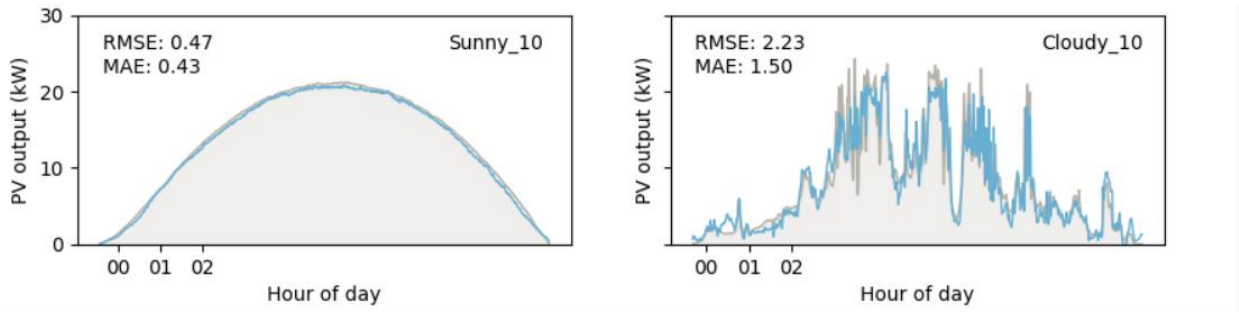
(g) Repetition 7 model



(h) Repetition 8 model



(i) Repetition 9 model



(j) Repetition 10 model

Figure 10. Visualization of nowcast predictions of CNN+SE-Attention model, (a) - (j) shows 10 repetitions.

This study quantifies the uncertainty during the training process and verifies the reliability of the practical application of the model in Figure 11. The left figure reveals the sensitivity of model training to data division through the RMSE distribution of each fold in ten cross-validations, while the right figure proves the stability of model performance through the mean distribution of ten full cross-validations. Random fluctuations in a single training are quantified through dual visual design and provide a key basis for evaluating the robustness of the model to cope with changes in data distribution in real scenarios.

4.2. Model Explainability

To interpret the decision-making process of the best performance CNN+SE-Attention model in the first dataset, the research employs Gradient-weighted Class Activation Mapping (Grad-CAM), a visualization technique that highlights the regions of input images most influential for the model's predictions. To improve the efficiency of model interpretability analysis, the research strategically reduced the size of the first dataset to 20% of its original capacity. The implementation incorporates robust error handling to ensure reliable heatmap generation even when gradient vanishing issues occur, while introducing elliptical sky masking to better align with the physical imaging characteristics. Moreover, the analysis compares model attention patterns, including sunny and cloudy conditions, by examining four representative sample categories: best-performing cases, median-error cases, worst-performing cases, and randomly selected samples.

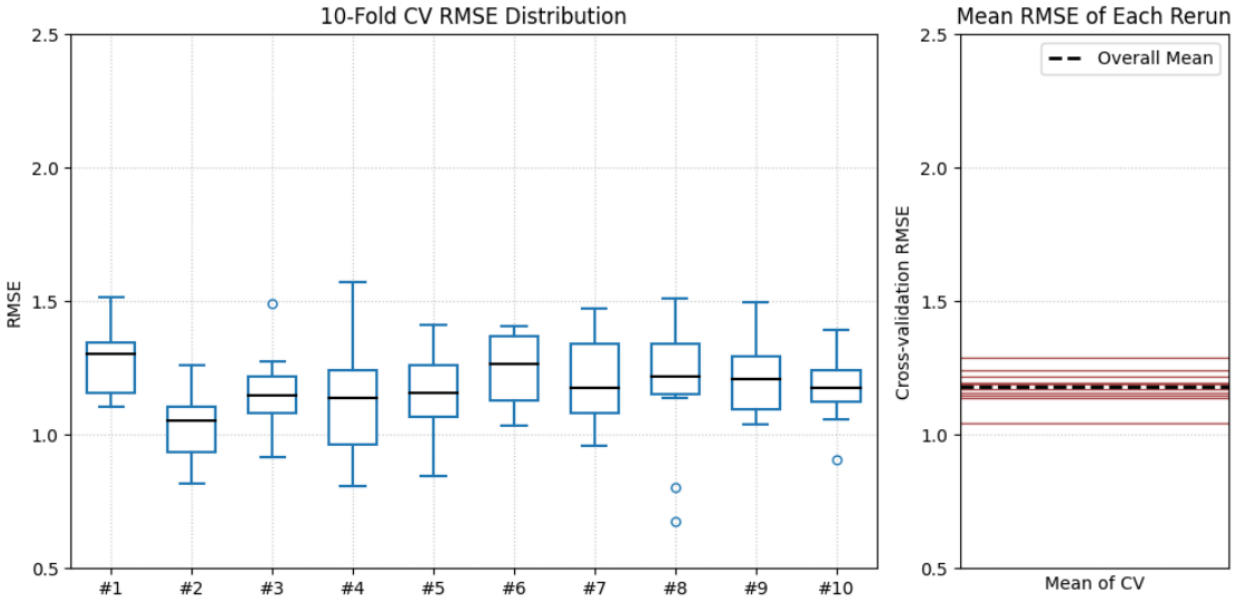
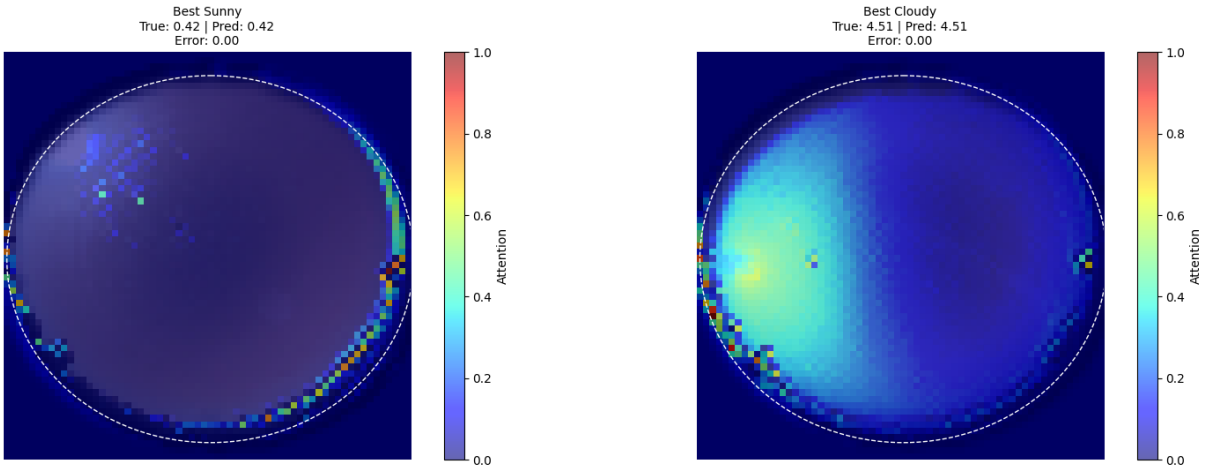


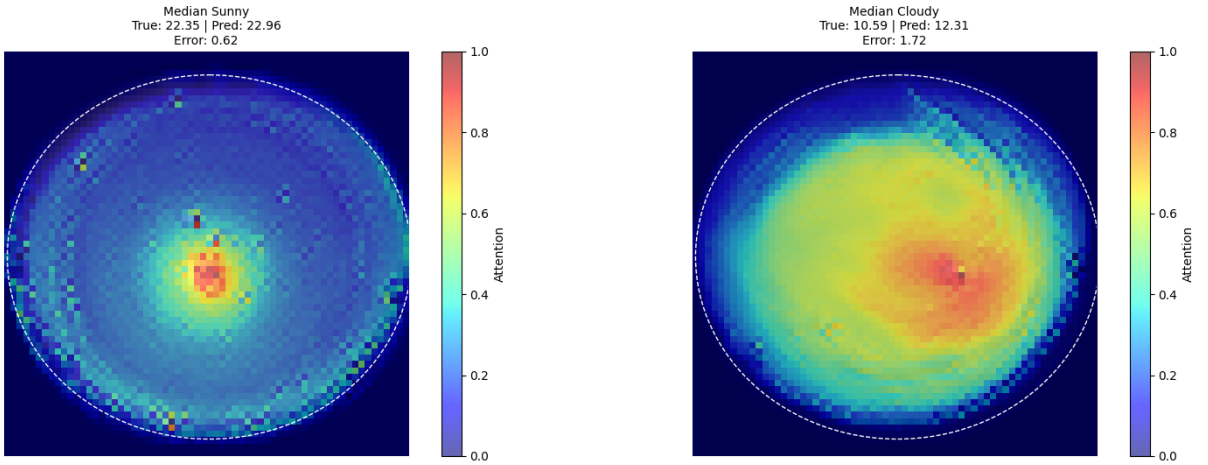
Figure 11. Model performance evaluation: Left panel shows the RMSE distribution across all 10 fold cross validation runs, demonstrating the variability in individual fold performance; the right panel displays the distribution of mean RMSE values for each complete 10 fold CV repetition.

The visualization results, as shown in Figure 12, reveal a strong correlation between prediction accuracy and attention localization precision. For both sunny and cloudy conditions, the best-performing samples, error =0.00, demonstrate optimal attention focusing within the elliptical sky region. For median-error samples, the sunny sample, error =0.62, maintains relatively concentrated attention in the central sky area, though with slight radial dispersion, while the cloudy sample, error =1.72, shows spiral-shaped attention distribution that extends beyond key meteorological features. This structural difference in attention patterns corresponds to the nearly threefold increase in prediction error for cloudy conditions. However, the worst-performing samples reveal catastrophic attention misalignment. The sunny case with error =3.15 displays fragmented attention clusters in non-sky regions, while the cloudy sample with error =22.12 exhibits completely scattered activation patterns with no discernible focus on cloud structures. The magnitude of error escalation suggests an abrupt breakdown in feature extraction rather than gradual performance decay. Random samples demonstrate intermediate characteristics, with attention maps showing partial sky-region coverage combined with erratic activation.

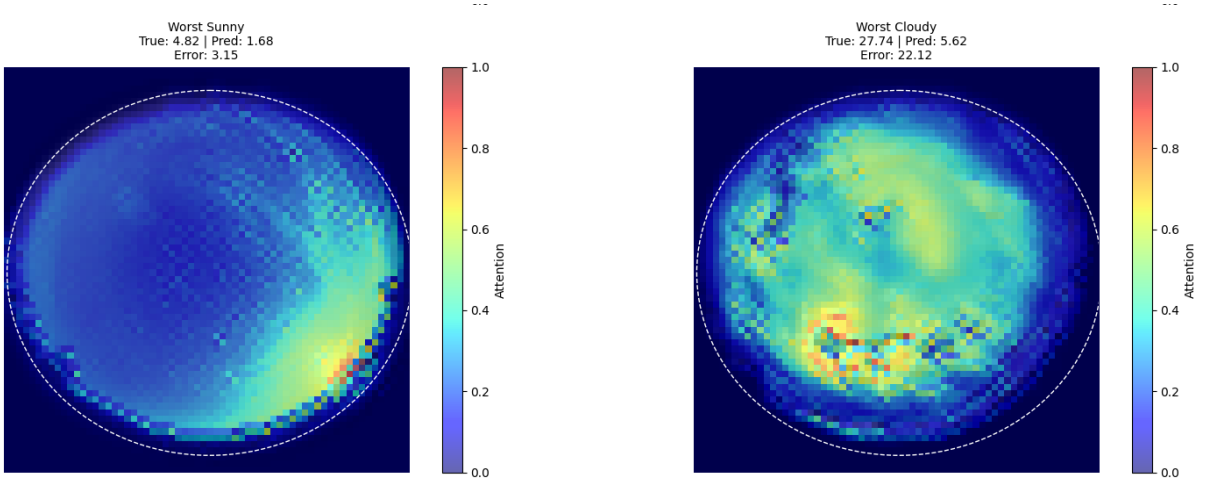
These observations collectively indicate that prediction accuracy is highly sensitive to the geometric organization of attention. Optimal performance requires tight spatial clustering of attention within physically relevant regions, while even minor dispersion correlates with measurable error increases. The most severe errors occur when attention loses its spatial coherence entirely, suggesting the model's decision-making process depends fundamentally on maintaining proper attention localization.



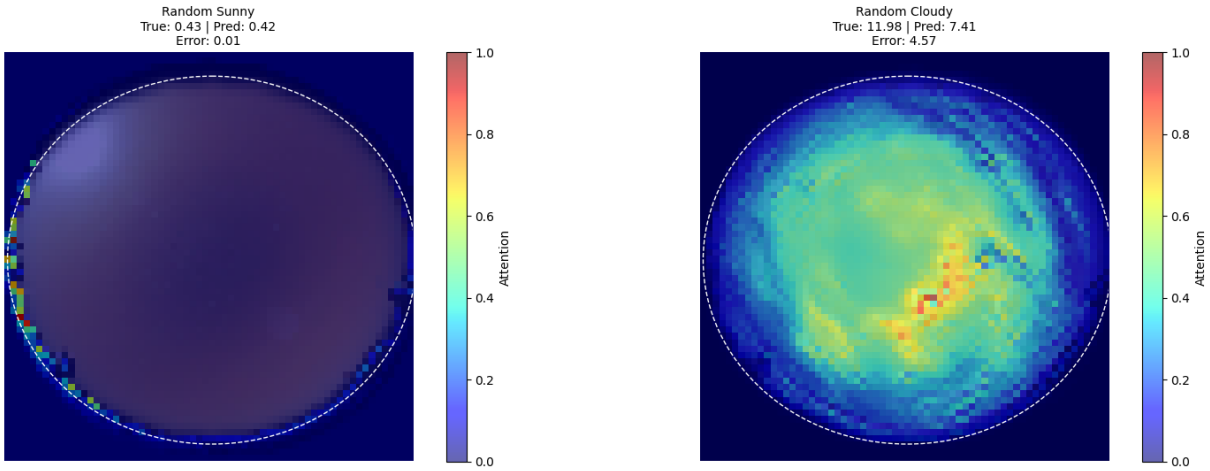
(a)Best sunny and best cloudy samples



(b)Median sunny and cloudy samples



(c)Worst sunny and cloudy samples



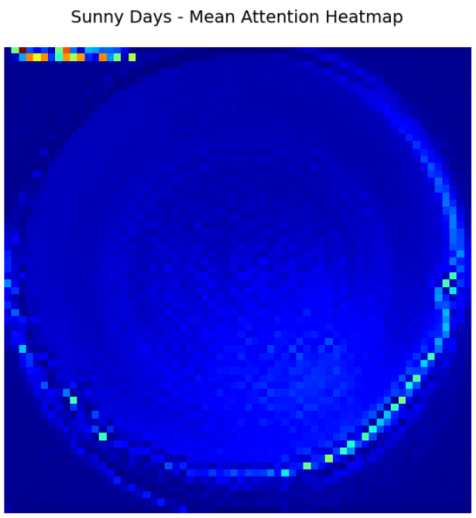
(d)Random sunny and cloudy samples

Figure 12. The visualization of four categories samples, (a) - (d) shows best, median, worst, and random samples.

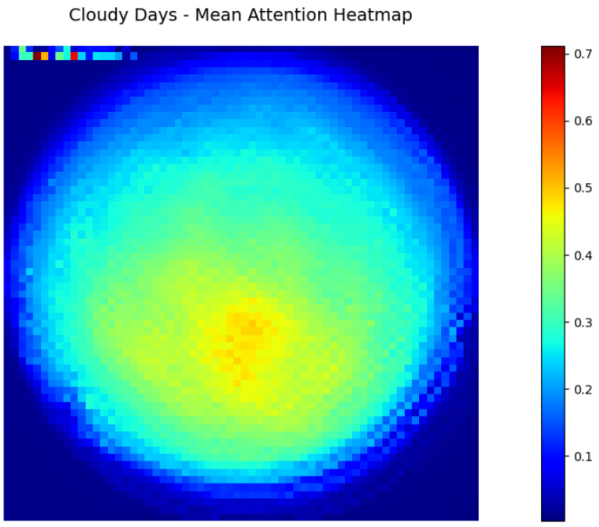
Building upon Grad-CAM's localized explanations, we systematically analyze attention patterns through a multi-stage comparative approach. The analysis includes aggregate heatmaps for sunny and cloudy conditions, pixel-wise statistical testing to identify significant differences, and regional comparisons through partitioned image analysis, which displayed in Figure 13.

The Regional Attention Intensity Comparison reveals three key findings. First, all five regions demonstrate extreme statistical significance of $p<0.001$, confirming fundamental differences in how the model processes sunny versus cloudy conditions. Second, sunny conditions consistently show near-zero attention in peripheral regions from Top-Left or Top-Right ≈ 0.0, while maintaining moderate focus on central areas. In contrast, cloudy conditions exhibit substantially heightened attention across all regions, particularly in the Bottom-Right quadrant where scattering effects are most pronounced.

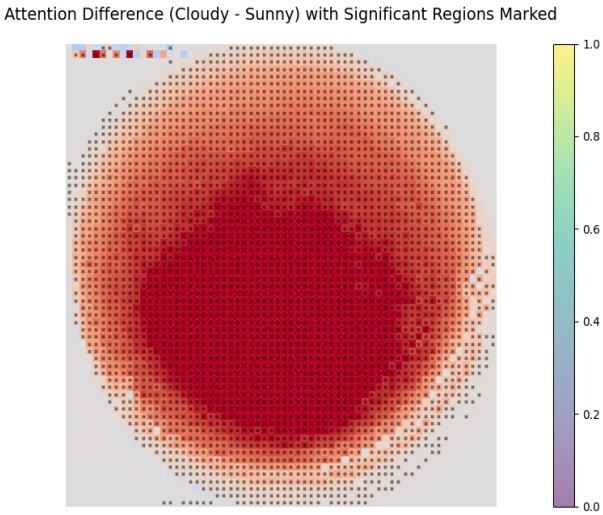
Most notably, the Center region maintains stable attention levels regardless of weather conditions, serving as an invariant reference point for the model's decision-making. This spatial pattern reflects meteorologically sound reasoning, which suppress noise in clear skies while actively monitoring cloud formations during overcast conditions. The Bottom-Right quadrant's particularly strong response to cloudy conditions suggests the model has learned to prioritize regions where atmospheric scattering signatures are typically most visible.



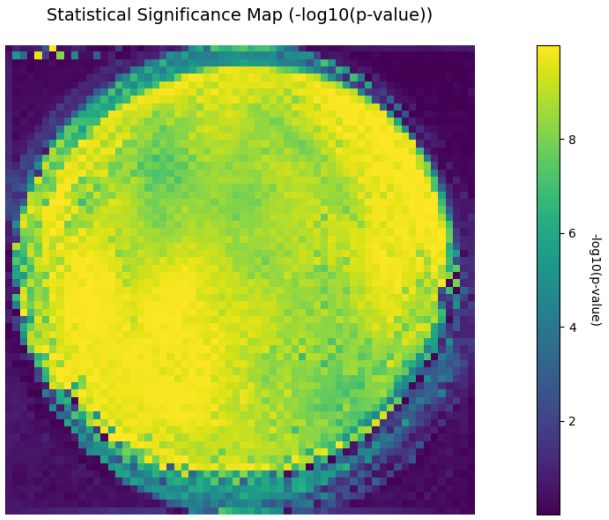
(a) Sunny mean attention



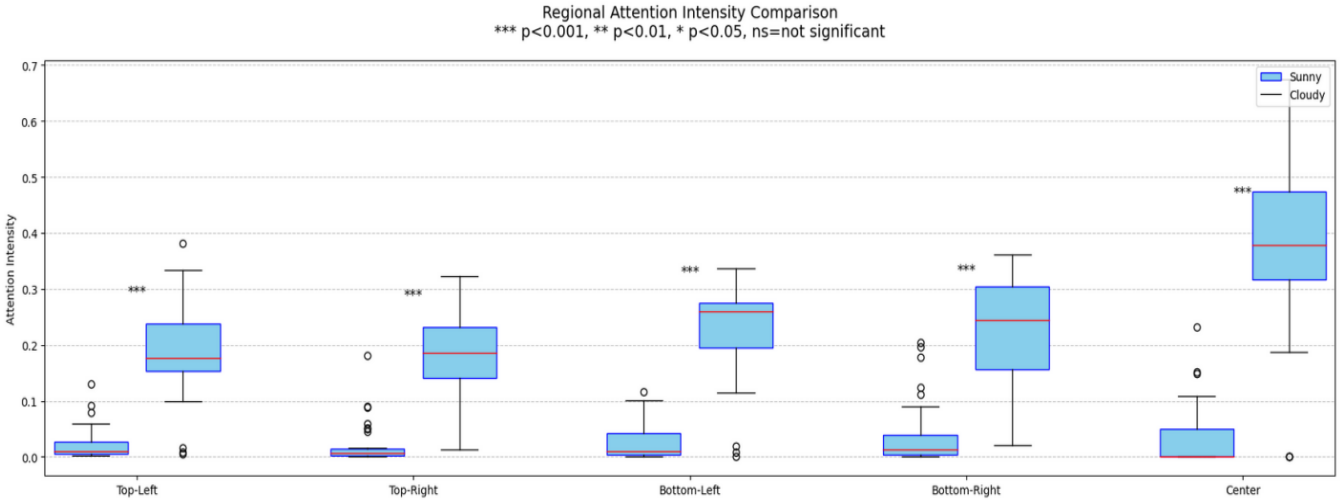
(b) Cloudy mean attention



(c) Attention Difference



(d) Significance map



(e)Regional Attention Intensity Comparison.

Figure 13. The comparison of attention pattern heatmaps, (a) - (e) shows sunny and cloudy days' mean attention heatmaps, attention difference (Cloudy - Sunny), significance map, and regional attention intensity comparison.

4.3. Model Visualization – GUI Design

The developed web application integrates the trained model through a Flask + TensorFlow backend with Bootstrap + jQuery (v1.11.1) frontend, featuring three core functional modules:

4.3.1. Main Pages

To enhance the usability and accessibility of the solar energy prediction system, the research developed an interactive web-based GUI that enables real-time forecasting and model explainability. The platform consists of three main pages, which shown in Figure 14, 15, 16: Home Page, providing an overview of the research and navigation; Dataset Display Page, allowing users to explore the sky image datasets used for training; and Solar Prediction Model Page, which offers core functionalities including solar irradiance prediction and model explainability visualization using Grad-CAM. This GUI bridges the gap between deep learning research and practical solar energy applications, making the CNN + Attention model accessible to both technical and non-technical users.

4.3.2. Core functionality test

The predictive model was first validated by testing its ability to successfully generate correct evaluation metrics (RMSE, MAE, R^2, EVS) and Grad-CAM explainability visualizations. As demonstrated in Figure 17 and 18, the system properly processed uploaded test datasets, including times_test.npy time series files, displaying all required prediction results and attention heatmaps in the web interface.

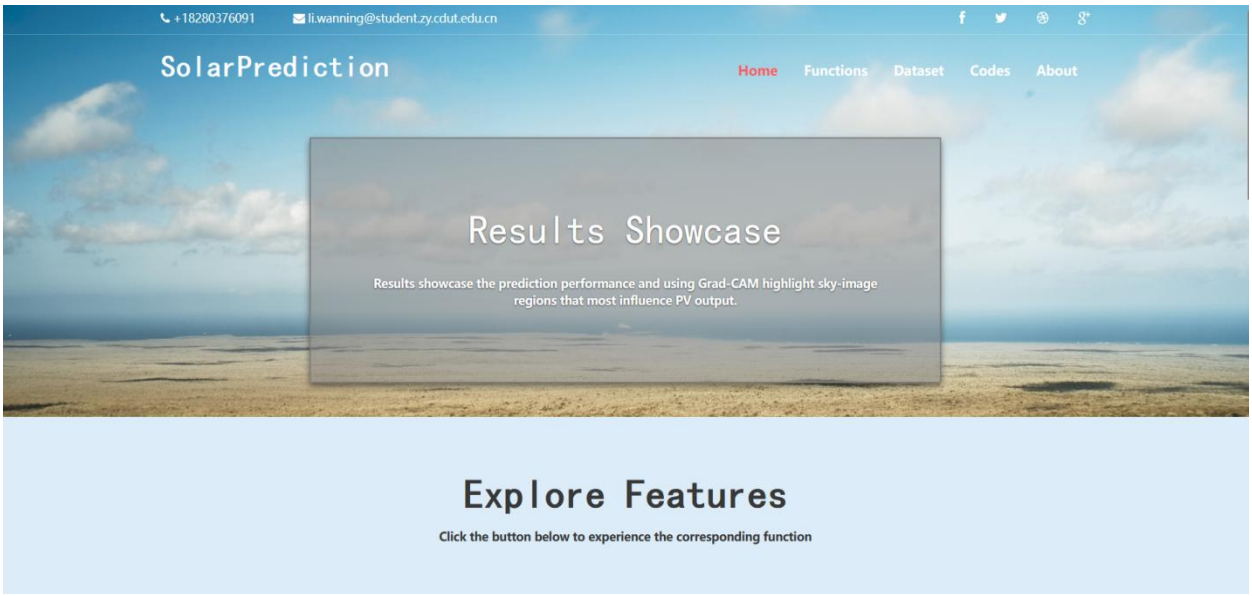


Figure 14. Home Page.

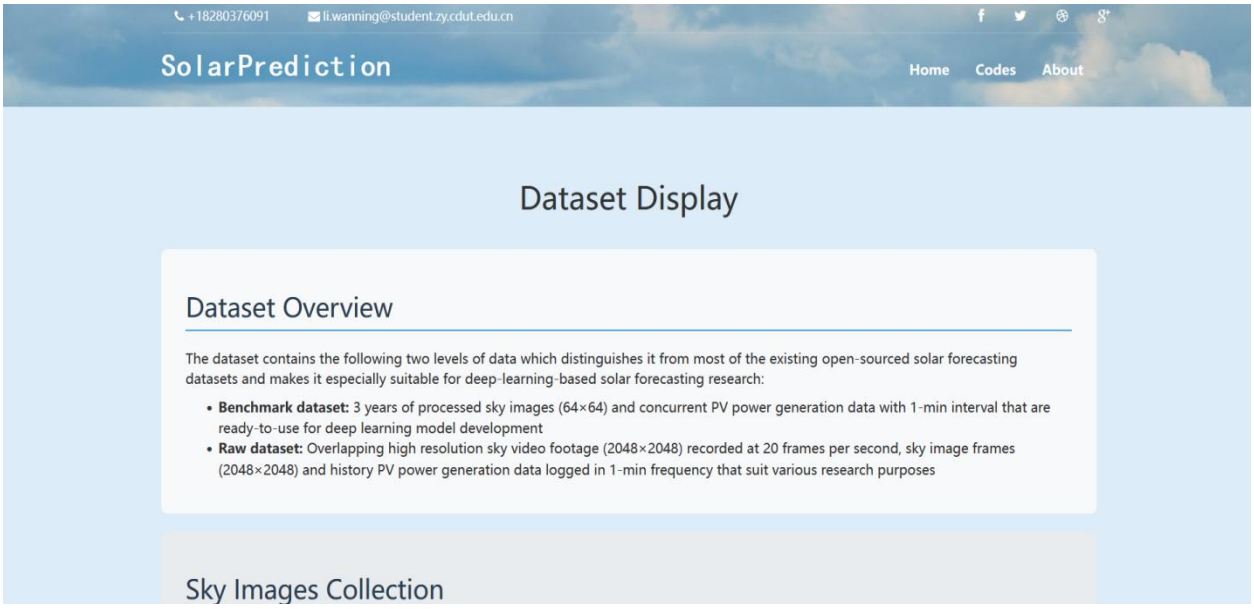


Figure 15. Dataset Display Page.

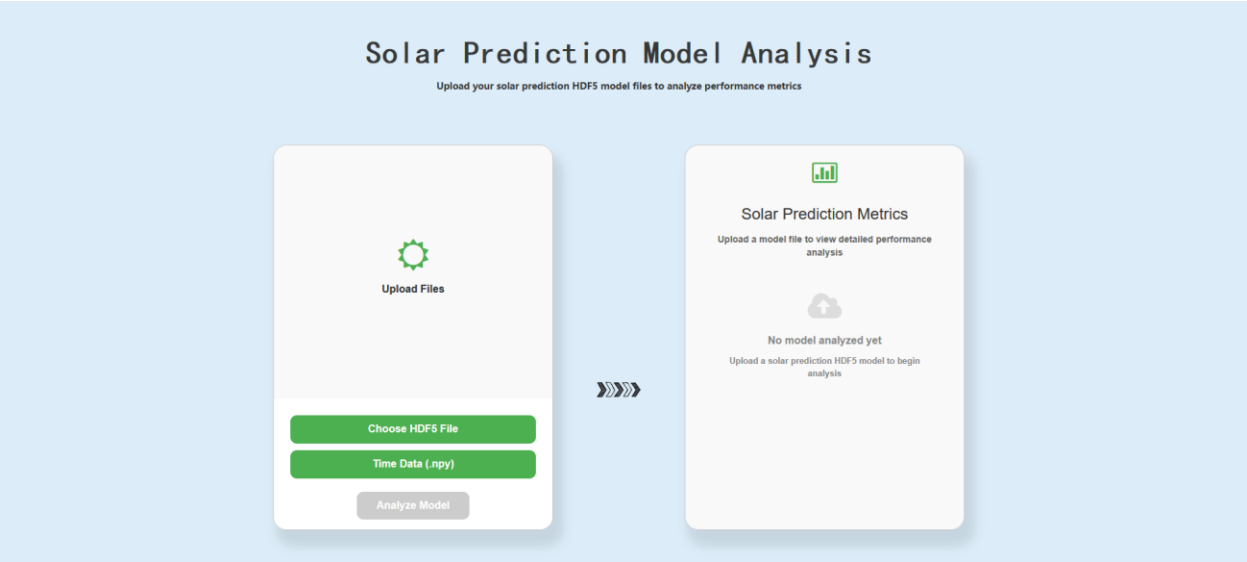


Figure 16. Solar Prediction Model Page.

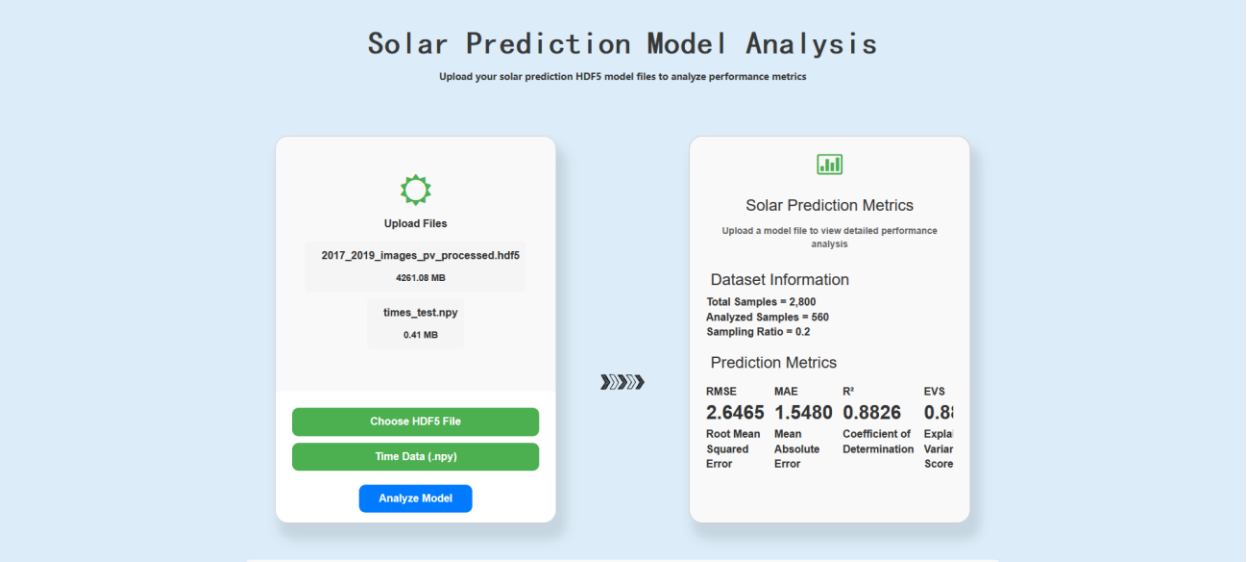


Figure 17. Solar Prediction Metrics Result in GUI.

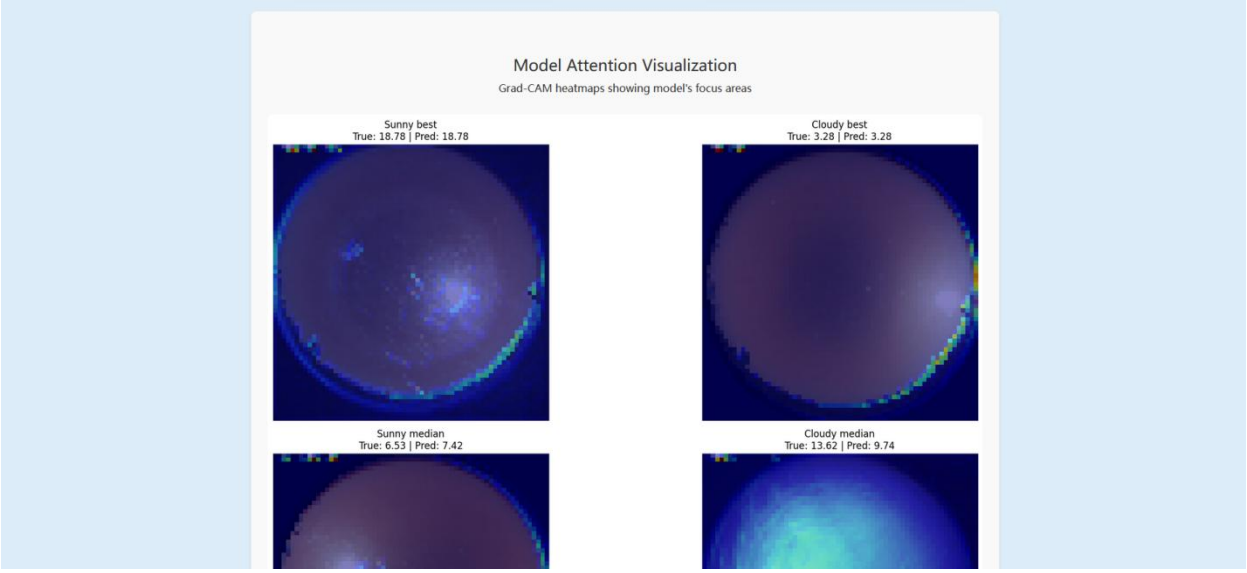


Figure 18. Model Explainability results in GUI.

5. CONCLUSION

This paper investigates the prediction of solar energy potential from radiance sky images using deep learning, systematically evaluating three convolutional architectures: CNN, CNN+SE-Attention, and CNN + Spatial-Attention, on two sky image datasets. Experimental results demonstrated that the CNN+SE-Attention model achieved optimal performance, with the performance of first dataset better compared to the second, highlighting its suitability for solar irradiance mapping under varying sky conditions. For the best SE-Attention model in the first dataset with the best performance, the study further uses Grad-CAM visualization to verify the interpretability of the model, revealing different attention patterns in sunny and cloudy scenarios and prioritizing key sky regions in the prediction process. By integrating these findings into an interactive GUI, the research bridges theoretical advancements in attention mechanisms with practical applications, offering a deployable tool for solar energy forecasting. This work establishes a framework for sky image-based renewable energy prediction, emphasizing model transparency, meteorological relevance, and user accessibility.

However, there are still limitations to the current research. Model performance is still limited by the diversity of datasets, especially for rainy weather and rapidly changing cloud cover scenarios. The current training data is mainly from a single weather station at a fixed location, lacking a balanced distribution of different seasons and weather patterns. In addition, due to the limitation of experimental equipment, the captured sky images have low resolution and are greatly affected by lighting conditions, so exposure anomalies are prone to occur during sunrise and sunset. The model is slow to respond to sudden weather changes, and there is a delay in prediction results. In terms of hardware deployment, the existing model requires GPU support with high computing power, which is difficult to directly apply to the embedded monitoring system of solar power plants.

Future research will address these limitations by optimizing data acquisition schemes to obtain more comprehensive images of the sky using multi-angle camera arrays. Secondly, a lightweight model based on mobile devices is developed, and knowledge distillation and quantization techniques are used to reduce the computational requirements. The timeliness of short-term forecasts can also be improved by introducing time series analysis methods, and a more reliable correction mechanism can be established by combining historical power generation data. In addition, a complete test platform will be built to evaluate the long-term performance of the model in the actual solar power plant, and the adaptation of the model to different weather conditions will be enhanced through transfer learning technology. These improvements will significantly increase the utility of the forecasting system in the field of renewable energy.

References

[1] Pérez-Rodríguez, S. A. et al., "Metaheuristic Algorithms for Solar Radiation Prediction: A Systematic Analysis," IEEE Access, vol. 12, pp. 100134-100148, Jul. 2024, doi: 10.1109/ACCESS.2024.3429073.

[2] El Alani, O.et al., “Short term solar irradiance forecasting using sky images hybrid CNN-MLP,” Energy Reports, vol. 7, pp. 888–900, doi: https://doi.org/10.1016/j.egyr.2021.07.053

[3] Martinez Lopez V. A. et al. , "Using sky-classification to improve the short-term prediction of irradiance with sky images and convolutional neural networks," Solar Energy, vol. 269, pp. 112320, Jan. 2024, doi: 10.1016/j.solener.2024.112320.

[4] Paletta. Q. et al., "Advances in solar forecasting: Computer vision with deep learning," *Advances in Applied Energy*, vol. 11, 100150, Aug. 2023, doi: 10.1016/j.adapen.2023.100150.

[5] Ajith. M and Martínez-Ramón. M, "Deep learning algorithms for very short-term solar irradiance forecasting: A survey," Renewable and Sustainable Energy Reviews, vol. 182, 113362, 2023, doi: 10.1016/j.rser.2023.113362.

[6] Wang. Y. et al., “The cost of day-ahead solar forecasting errors in the United States,” Solar Energy, vol. 231, pp. 846-856, Jan. 2022, doi: 10.1016/j.solener.2021.12.012

[7] Nie. Y. et al., "Open-source sky image datasets for solar forecasting with deep learning: A comprehensive survey," Renewable and Sustainable Energy Reviews, vol. 189, 113977, 2024, doi: 10.1016/j.rser.2023.113977.

[8] Gao. H and Liu. M., "Short-Term Solar Irradiance Prediction From Sky Images With a Clear Sky Model," Australian National University, Canberra, Australia, 2024.

[9] Feng. C and Zhang. J, "SolarNet: A sky image-based deep convolutional neural network for intra-hour solar forecasting," Solar Energy, vol. 204, pp. 71-78, Apr. 2020, doi: 10.1016/j.solener.2020.03.083.

[10] Papatheofanous. E. A. et al., "Deep Learning-Based Image Regression for Short-Term Solar Irradiance Forecasting on the Edge," Electronics, vol. 11, no. 22, 3794, Nov. 2022, doi: 10.3390/electronics11223794.

[11] Paletta. Q. et al., "Benchmarking of deep learning irradiance forecasting models from sky images: An in-depth analysis," Solar Energy, vol. 224, pp. 855–867, 2021, doi: 10.1016/j.solener.2021.05.056.

[12] Zhang. J. et al., "Deep photovoltaic nowcasting," Solar Energy, vol. 176, pp. 267–276, Oct. 2018, doi: 10.1016/j.solener.2018.10.024.

[13] Sidrach-de-Cardona. M and Lopez. L. M., "A simple model for sizing stand alone photovoltaic systems," Solar Energy Materials and Solar Cells, pp. 199-214, 1997.

[14] Chowdhury. B. H. and Salfur. R., "Forecasting sub-hourly solar irradiance for prediction of photovoltaic output," IEEE, pp. 171-176, 1987.

[15] Hassanzadeh. M. et al., "Practical Approach for Sub-Hourly and Hourly Prediction of PV Power Output," North American Power Symposium, 2010.

[16] Jimenez. P. A. et al., "WRF-SOLAR Description and Clear-Sky Assessment of an Augmented NWP Model for Solar Power Prediction," AMERICAN METEOROLOGICAL SOCIETY, Sep. 2015, doi: 10.1175/BAMS-D-14-00279.1

[17] Rachna and Singh. A. K., "Prediction of Photovoltaic Power Generation using Machine Learning - A Review," IEEE, 2023.

[18] Voyant. C. et al., "Machine Learning methods for solar radiation forecasting: a review," Renewable Energy, 2016, doi: 10.1016/j.renene.2016.12.095.

[19] Gaboitaolelwe. J. et al., "Machine Learning Based Solar Photovoltaic Power Forecasting: A Review and Comparison," IEEE, vol. 11, Apr. 2023.

[20] Yamashita. R. et al., "Convolutional neural networks: an overview and application in radiology," Insights into Imaging, vol. 9, no.611-629, 2018, doi: 10.1007/s13244-018-0639-9.

[21] Ahmed. R. et al., "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, no. 109792, 2020, doi:10.1016/j.rser.2020.109792.

[22] Jonathan A. L., ".A radiant shift: Attention-embedded CNNs for accurate solar irradiance forecasting and prediction from sky images," Renewable Energy, Vol. 234, 121133, 2024, doi: 10.1016/j.renene.2024.121133.

[23] Wang. S. and Zhang. Y., " An attention-based CNN model integrating observational and simulation data for high-resolution spatial estimation of urban air quality," Atmospheric Environment, Vol. 340, 120921, 2025, doi: 10.1016/j.atmosenv.2024.120921.

[24] Qu. J. et al., "Day-ahead hourly photovoltaic power forecasting using attentionbased CNN-LSTM neural network embedded with multiple relevant and target variables prediction pattern," Energy, vol. 232, no.120996, 2021, doi: 10.1016/j.energy.2021.120996.

[25] Nie. Y. et al., "SKIPP'D: A SKy Images and Photovoltaic Power Generation Dataset for short-term solar forecasting," Solar Energy, vol.255, pp.171-179, 2023, doi: 10.1016/j.solener.2023.03.043.

[26] Sun. Y. et al.," Solar PV output prediction from video streams using convolutional neural networks," Energy & Environmental science, 11, pp.1811-1818, 2018, doi: 10.1039/c7ee03420b.