data

MDPI

Data Descriptor

Dataset for Machine Learning: Explicit All-Sky Image Features to Enhance Solar Irradiance Prediction

Joylan Nunes Maciel [1,2,3,*], Jorge Javier Gimenez Ledesma [1,3] and Oswaldo Hideo Ando Junior [1,3,4,*]

[1] Interdisciplinary Postgraduate Program in Energy & Sustainability (PPGIES), Federal University of Latin American Integration–UNILA, Paraná City 85867-000, Brazil; jorge.ledesma@unila.edu.br
[2] Postgraduate Program in Applied Physics (PPGFISA), Federal University of Latin American Integration–UNILA, Paraná City 85867-000, Brazil
[3] Research Group on Energy & Energy Sustainability (GPEnSE), Academic Unit of Cabo de Santo Agostinho (UACSA), Federal Rural University of Pernambuco (UFRPE), Cabo de Santo Agostinho 54518-430, Brazil
[4] Program in Energy Systems Engineering (PPGESE), Academic Unit of Cabo de Santo Agostinho (UACSA), Federal Rural University of Pernambuco (UFRPE), Cabo de Santo Agostinho 54518-430, Brazil
* Correspondence: joylan.maciel@unila.edu.br (J.N.M.); oswaldo.ando@ufrpe.br (O.H.A.J.)

Abstract: Prediction of solar irradiance is crucial for photovoltaic energy generation, as it helps mitigate intermittencies caused by atmospheric fluctuations such as clouds, wind, and temperature. Numerous studies have applied machine learning and deep learning techniques from artificial intelligence to address this challenge. Based on the recently proposed Hybrid Prediction Method (HPM), this paper presents an original and comprehensive dataset with nine attributes extracted from all-sky images developed using image processing techniques. This dataset and analysis of its attributes offer new avenues for research into solar irradiance forecasting. To ensure reproducibility, the data processing workflow and the standardized dataset have been meticulously detailed and made available to the scientific community to promote further research into prediction methods for photovoltaic energy generation.

Keywords: all-sky image features; solar irradiance prediction; image processing; photovoltaic energy forecasting

check for updates

Citation: Maciel, J.N.; Ledesma, J.J.G.; Ando Junior, O.H. Dataset for Machine Learning: Explicit All-Sky Image Features to Enhance Solar Irradiance Prediction. *Data* **2024**, *9*, 113. https://doi.org/10.3390/data9100113

Academic Editors: Hugo Morais, Rui Castro and Cindy Guzman

Received: 10 July 2024
Revised: 11 September 2024
Accepted: 17 September 2024
Published: 29 September 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The increasing demand for energy is being met by clean and so-called renewable sources, such as wind and photovoltaics. Several countries are investing heavily in the installation of photovoltaic solar energy. China stands out in the world ranking, as do Japan, the United States, Germany, and India, which are the five countries with the largest installed capacity [1].

The adoption of photovoltaic solar energy has driven research into methods for directly predicting solar energy generation, or even solar irradiance, which is the main component determining the energy generated [2]. The biggest obstacle that influences and promotes intermittency in the process of generating photovoltaic solar energy is variation in atmospheric climatic factors, especially clouds [3,4].

A growing number of studies on methods for predicting solar photovoltaic energy generation (PSPEG) have been published in the scientific literature [4], and mainly apply machine learning (ML) and deep learning (DL) [3,5–10]. The increase in computer processing power and the use of DL models allows the use of all-sky (180°) images for PSPEG [11–16]. In this sense, several datasets containing images have been built and made available to promote the development of predictive methods [17–21].

Research on PSPEG at the Federal University of Latin American Integration (UNILA) has developed scientific studies on short-term PSPEG [22–27]. Based on this research,

the study by [27] proposed and evaluated a new hybrid approach to predicting solar irradiance, which applies image processing (IP) and artificial intelligence machine learning techniques [28]. Named the Hybrid Prediction Method (HPM) [27], the proposal was developed from a dataset containing historical data with all-sky images and meteorological and solar irradiance information, and created with controlled quality and samples collected every minute for the full 3-year period (2014 to 2016) [17].

Based on the HPM proposition, this paper presents a comprehensive study of the process of developing a dataset containing a set of instantaneous sky characteristics obtained with image processing techniques applied to all-sky images, which can be used to develop direct and indirect prediction methods (solar irradiance) applicable to photovoltaic solar power generation. This study provides a unique dataset, developed from all-sky images, making it relevant to improving the accuracy of predictions of photovoltaic solar energy generation, which is fundamental for the efficient operation and integration of solar energy into electrical systems. The creation of a comprehensive dataset and the use of innovative AI techniques, based on a set of explicit characteristics, can overcome the limitations of traditional methods and provide more effective tools for researchers and professionals in the sector.

In Section 2, this paper presents the main concepts on the subject of PSPEG, emphasizing the HPM hybrid prediction approach. Section 3 presents dataset construction, description, and samples, with details of the design and implementation of metrics, including the tools, technologies and discusses the analysis of the descriptive statistics for each metric, as well as their results and limitations, paving the way for understanding and optimizing them. Section 4 presents the final considerations, as well as making the dataset available to leverage the development and optimization of PSPEG methods.

2. Theoretical Background

According to [4,29], there has been an increase in the accuracy and precision of photovoltaic solar energy generation prediction (PSPEG) methods, especially since 2017, with the emergence of machine learning and deep learning artificial intelligence methods [7,10,30–32]. Although there is no consensus on the classification of these methods, the main characteristics that distinguish them are the prediction horizon, the origin and types of data, classes of predictive algorithms, and error metrics adopted [4].

One study [27] proposed the Hybrid Prediction Method (HPM), a new hybrid approach to predicting solar irradiance that is applicable to photovoltaic solar power generation. The HPM uses image processing and machine learning techniques to make short-term solar irradiance predictions at horizons of 1, 5, 15, 30, and 60 min. This study was carried out using a dataset with a large volume of data [17] and controlled quality [17] consisting of standardized endogenous and exogenous data relating to meteorological information, solar irradiance, and all-sky images collected over a full three-year period (2014–2016) and sampled every 1 min. Table 1 shows data used in the HPM and captured in the city of Folsom, California, United States of America. Nighttime values between 8 pm and 6 am were not taken into account.

Table 1. Folsom dataset variables [17] and sample sizes by horizon used in HPM [27].

Dataset Variables	Type	Horizon (min)	Number of Samples			
			2014	2015	2016	Total
Year, Month, Day, Hour, Minute	**Date and Time**	1	250.579	263.023	261.961	775.563
		5	250.574	263.018	261.956	775.548
All-sky Images	Image	10	250.569	263.013	261.951	775.533
		15	250.564	263.008	261.946	775.518
Global Horizontal Irradiance (GHI)	Solar Irradiance	30	250.549	262.993	261.931	775.473
		60	250.519	262.963	261.901	775.383

Figure 1 provides examples of all-sky images and their respective global horizontal irradiance (GHI) captured every hour for two days, one of which was completely sunny (2 September 2024), and the other partially cloudy (28 June 2015). The displayed time is Coordinated Universal Time (UTC-07) local to the city of Folsom, CA, USA. These two days are used as an example to demonstrate the logic of the image processing and results of the features (measurements) extracted from all-sky images that make up the dataset prepared in this study.

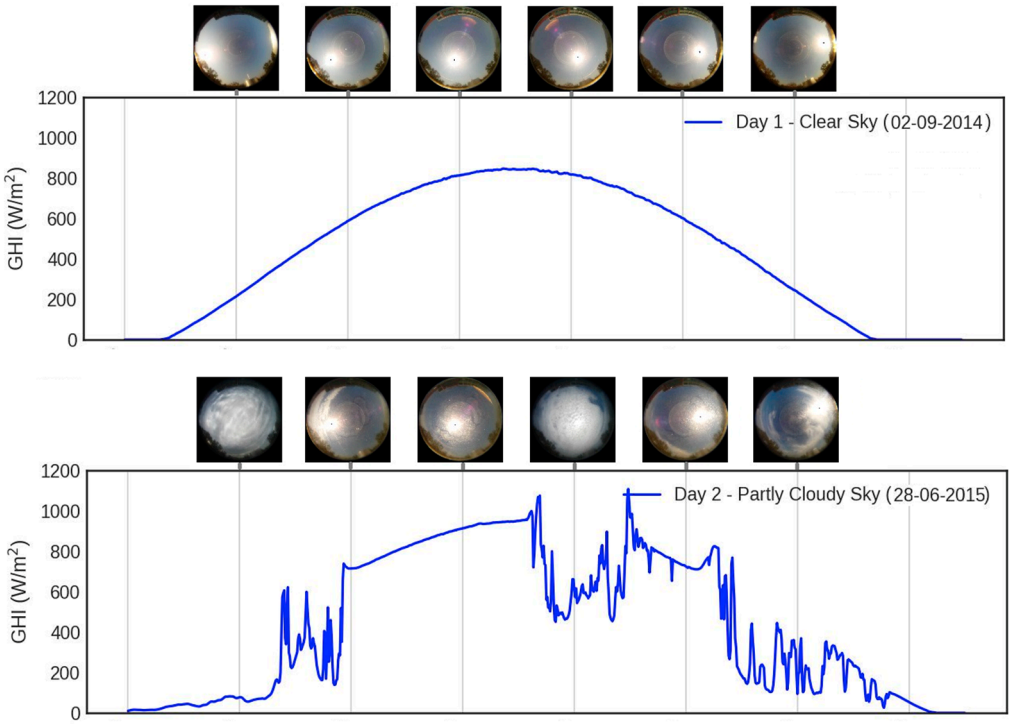


Figure 1. Samples of all-sky images and respective GHI values for two days. Adapted from [27].

The hybrid prediction approach proposed in [25] is depicted in Figure 2. A set of nine features (measurements) (f_1 to f_9) was calculated from each all-sky image using image processing techniques. In addition, another five features (f_{10} to f_{14}) relating to temporal data and the climatic season form part of the inputs calculated from the images. This set of fourteen features for each all-sky image was then used as input for short-term predictions in the HPM using machine learning models previously trained on historical data [25,27].

In this study, the HPM was used to create a dataset of fourteen features (f_1 to f_{14}). The focus was, therefore, on the explicit features image extraction stage, i.e., calculating the set of nine features (f_1 to f_9) extracted with image processing: *Clouds Movement*, *Clouds Coverage*, *Clouds Around Sun*, *Clear Sky GHI*, *Sun Luminance*, *Sun Luminance Adjusted*, *Sun Located,* and *White Pixel Ratio*. The sampling of the dataset built corresponded to the values of all the features for each image captured, every one minute, over a period of three years (2014 to 2016). In addition, the value of the target variable, global horizontal irradiance (GHI), corresponding to the all-sky image, is provided, and represents the GHI at the moment the image was captured.

The presented dataset, in addition to its application in predicting solar irradiance for photovoltaic generation, offers significant potential for other research, such as multistep forecasting for controlling the ramp rate in photovoltaic systems, which is essential for the efficient integration of solar energy into the grid, mitigating intermittency, and improving stability [33]. Also noteworthy is cloud movement forecasting, which contributes to short-term irradiance forecasting, and cloud base height estimation, which is relevant for monitoring in regions with high climate variability [34]. These applications reinforce the versatility of the dataset and its importance for the advancement of solar energy forecasting

methods. The next section presents the characteristics and implementation of the nine features extracted from the images.

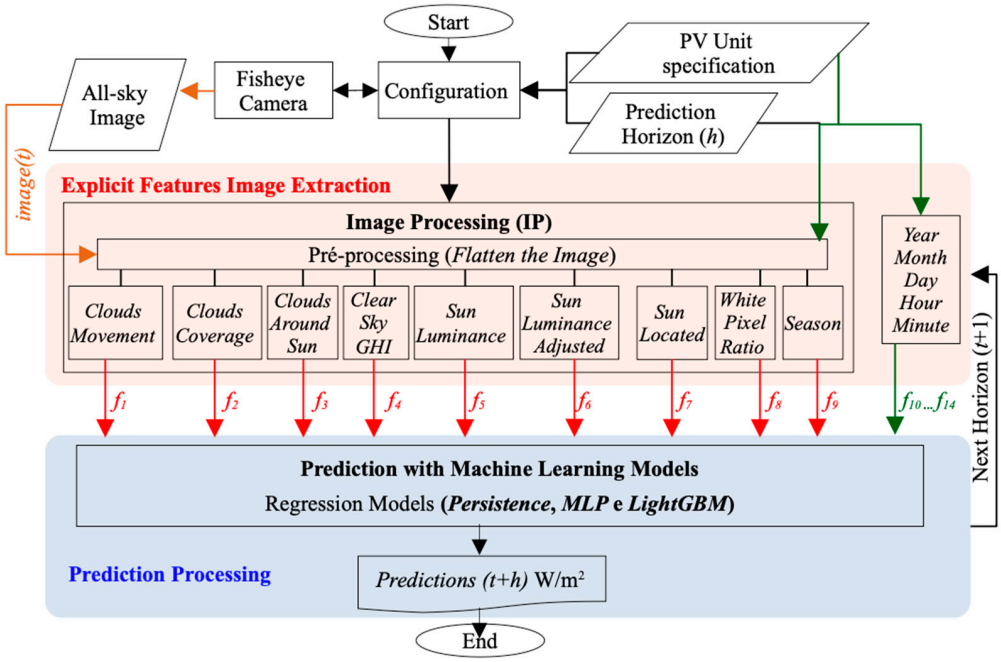


Figure 2. Hybrid Prediction Method architecture diagram. Adapted from [27].

3. Dataset Construction, Description, and Samples

All-sky images obtained from the dataset [17] were in red–green–blue (RGB) format [35], in which each point (pixel)'s color consisted of three components (channels), red, green, and blue, with a value between 0 and 255 each. The resolution of the figures was 96 dots per inch (DPI), with 256 colors (8 bits) and dimensions of 1536 × 1536 pixels (width and height).

Figure 3 shows these channels in the formation of the pixel and exemplifies the formation of the colors yellow (A), pink (B), and grey (C). From each of the all-sky images, the set of features f_1 to f_9 were computed sequentially using image processing techniques [35] using the Python 3.10 programming language [36]. In total, the dataset provided 764,867 records extracted from these all-sky images and respective solar irradiances from [17].

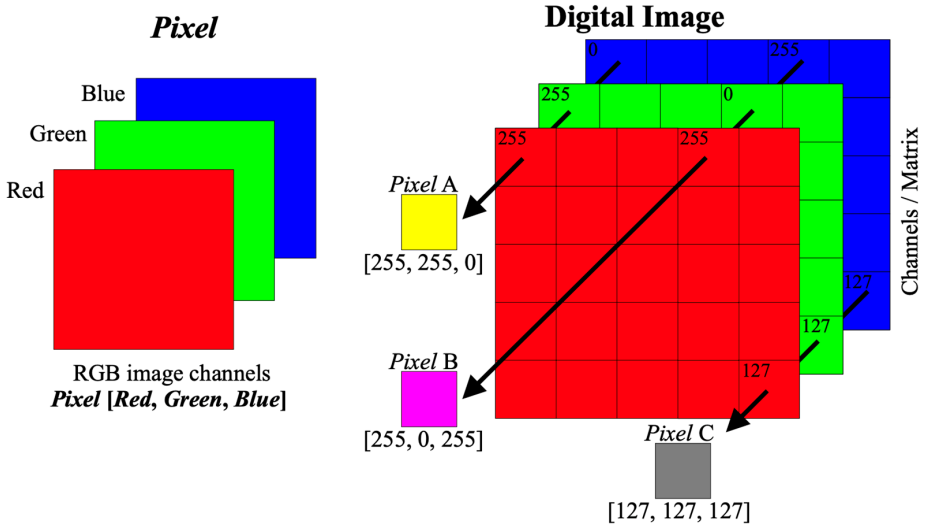


Figure 3. Channels and colors in a RGB image. Adapted from [27].

To demonstrate the image processing applied to each feature collected from all-sky images, two days with different weather conditions and GHI were selected, as shown in Figure 1. Considering all-sky image pre-processing, the following sections present processing details for the following features: *Preprocessing*, *Clouds Movement* (f_1), *Clouds Coverage* (f_2), *Clouds Around Sun* (f_3), *Clear Sky GHI* (f_4), *Sun Luminance* (f_5), *Sun Luminance Adjusted* (f_6), *Sun Located* (f_7), *White Pixels Rate* (f_8), and *Season* (f_9).

3.1. Image Preprocessing

This first stage had the function of flattening the all-sky image and correcting image distortion caused by the fisheye lens (180°). Using the approach provided in [37], the flattening procedure was carried out for a *linear* lens, *full-frame* image format, input image field of view (*ifov* = 180°), and output field of view perspective (*ofov* = 180°) [37]. Figure 4 shows original examples of images resulting from preprocessing (defished), which were used to calculate the features (f_1 to f_9).

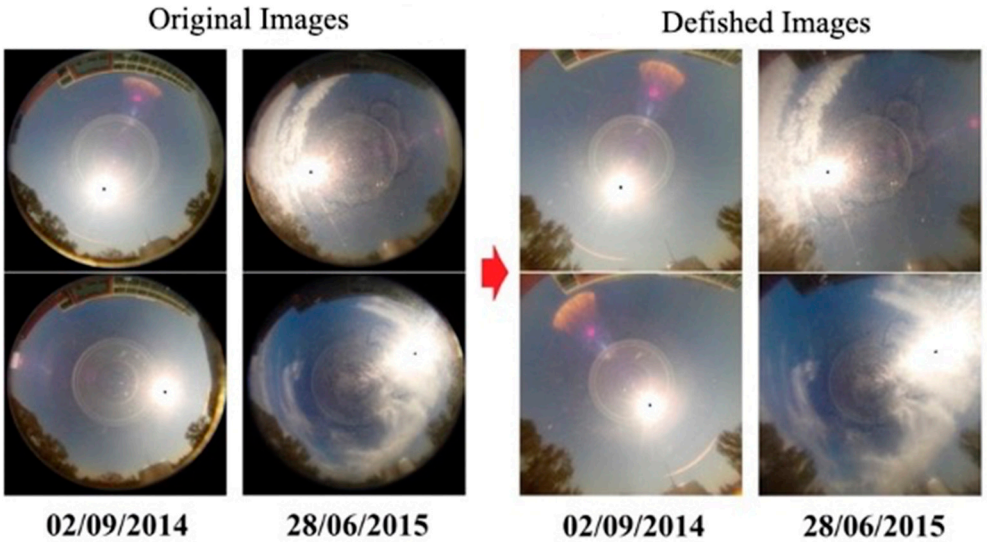


Figure 4. Channels and colors in an RGB image. Adapted from [25].

Finally, this stage also extracted the following information from each captured all-sky image: year, month, day, hour, minute, and weather season. The dataset generated provides this information in respectively named columns.

3.2. Clouds Movement

This feature quantifies the existence and intensity of cloud movement based on the difference between two sequentially captured images, considering the prediction horizon (h). Cloud movement at time t is calculated by the difference between $imagem(t)$ and $imagem(t-h)$. A simple and computationally efficient approach that quantifies the degree of difference between two images was developed and applied: the mean square error (MSE) [38]. Each pixel in corresponding coordinates between two images was compared individually and, at the end, the total difference between the two images was calculated using the MSE. Although it was not normalized in the dataset, it is recommended to normalize the manual MSE to the automatic MSE by AM-based algorithms. Figure 5 shows an example of the *Clouds Movement* over two full days.

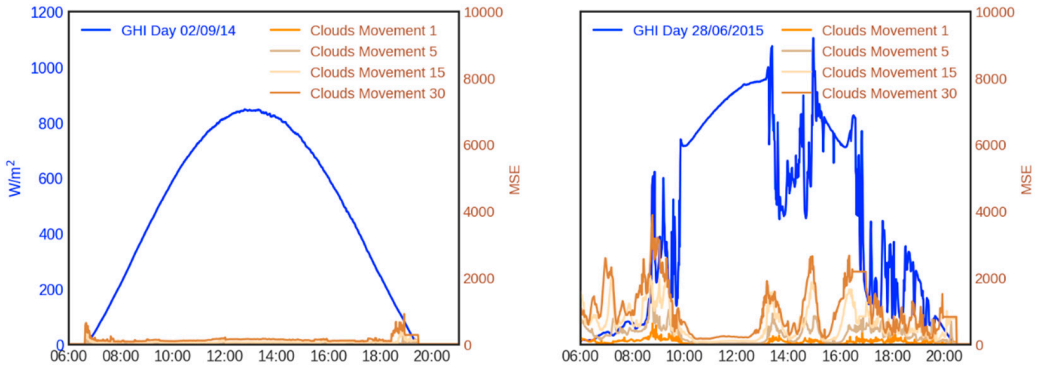


Figure 5. Clouds Movement samples. Adapted from [27].

In order to facilitate the use of the *Clouds Movement* feature in different prediction horizons, the value of this feature for each horizon was made available in the dataset generated in the "*cloudsMovementH*" columns, where *H* represents the prediction horizon in minutes (h = 1, 2, 3, 4, 5, 10, 15, 30, or 60). Therefore, the other columns, except for h, can be removed.

3.3. Clouds Coverage

This feature represents the percentage of clouds in the sky image at time t. Among the challenges and proposals for identifying clouds in the sky [34,39–44], the dataset provides results from the following algorithm, adopted in the HPM: (i) resizing the flat image to 30% of its original size (460 × 460 pixels) to optimize computational processing; (ii) converting the image to a single channel using the normalized red–blue ratio (RBR) algorithm proposed in [45]; and (iii) applying thresholding with the Otsu threshold algorithm [46]. Commonly used in similar problems [39,47–50], the Otsu algorithm calculates an adaptive intensity value for each image, separating pixels into clouds (foreground) and sky (background). This adaptive intensity is a constant that may not perform adequately in situations such as low solar elevation [46]; (iv) clouds percentage calculation (*CP*) of clouds in the sky at time t (CP_t) using Equation (1) is as follows:

$$\mathrm{CP_t} = 100 \times \frac{\sum \rho_{cloud}}{\sum \rho_{sky}} + cf \tag{1}$$

where ρ_{cloud} and ρ_{sky} correspond to the total number of pixels showing clouds and sky, respectively, and cf is an adjustment factor for objects that interfere with the sky area in the image, such as buildings or trees. In the HPM, $cf = 0.1$ (10%) was used, and in the dataset generated the "*cloudsCoverage*" column, which corresponds to the value of the *Clouds Coverage* feature. Figure 6 shows a sample of the *Clouds Coverage* over two full days.

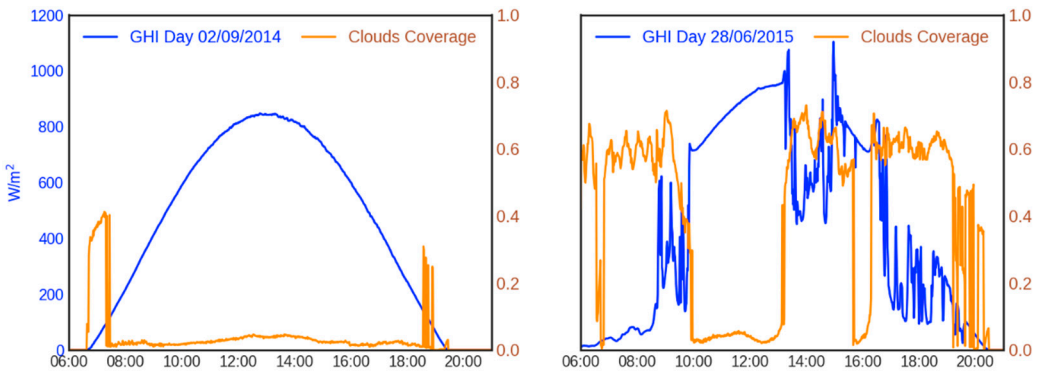


Figure 6. Clouds Coverage samples. Adapted from [27].

3.4. Clouds Around Sun

This feature gives the percentage of clouds relative to a region around (surrounding) the sun in the image at time t. After classifying each pixel between sky and cloud, the central coordinate of the sun is located with a threshold filter of 240 in the red channel of the image [51]. If a region with this threshold is located, the sun is located and an area using the red–blue ratio (RBR) algorithm of 300 × 300 pixels on the surrounding area is selected for segmentation and application of the Otsu threshold, allowing the percentage of clouds in the area surrounding the sun to be calculated. If the sun is not located, 100% of the area around the sun consists of clouds. It is important to note that the *Clouds Around Sun* feature is inconsistent at the beginning and end of the day, caused by the sun's reflection in the lens or known limitations of the RBR and Otsu threshold methods. In the generated dataset, values in the "*cloudsSunAround*" column correspond to the *Clouds Around Sun* feature, an example of which is shown in Figure 7.

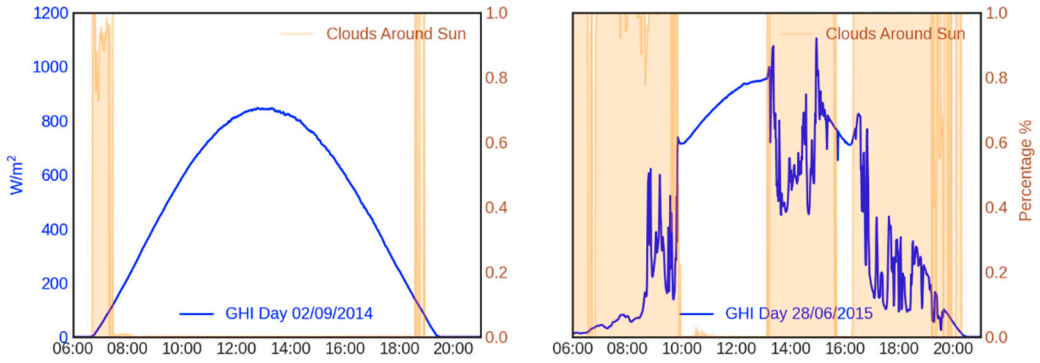


Figure 7. Clouds Around Sun samples. Adapted from [27].

3.5. Clear Sky GHI

This feature represents the global solar irradiance expected on days with completely clear skies, and is calculated using the Ineichen–Perez model [52] provided in the PVlib 0.9.1 programming package [53]. The generated dataset used the geographical coordinates of the city of Folsom, USA (latitude = 38.6779591 and longitude = −121.1760583) and its altitude (67 m) with other default values from PVlib. Figure 8 shows examples of the solar irradiance GHI (*Clear Sky GHI*) values predicted by the Ineichen–Perez model and their correlation (R^2) with the Folsom GHI. In the dataset generated, values in the "*clearSkyGHI*" column correspond to this feature.

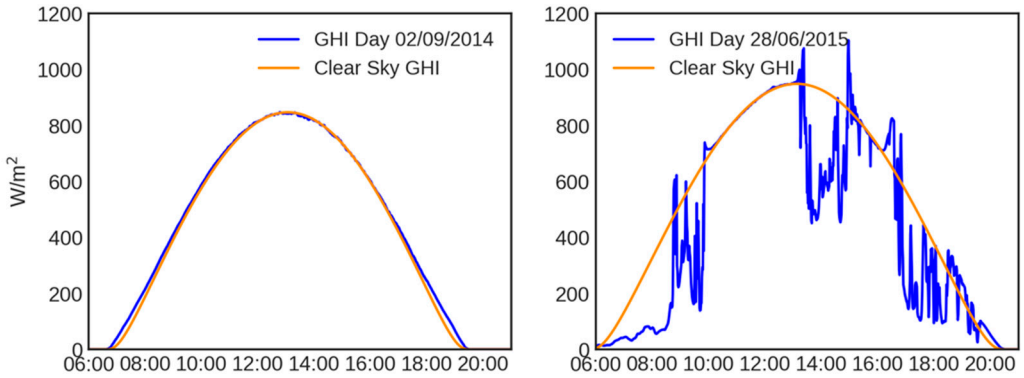


Figure 8. Clear Sky GHI samples. Adapted from [27].

3.6. Sun Luminance

This feature is based on the concept of the relative luminance of the pixel [54] in a given area of the image [27]. The HPM adopted the region around the sun as the Y luminance, using the ITU-R BT.709-6 standard [55]. Based on the study by [12], the luminance (Y) of

a pixel p is computed from the linear values of the red (R_{lin}), green (G_{lin}), and blue (B_{lin}) channels of the pixel in RGB images, using Equation (2) [55]:

$$Y = 0.2126 \times R_{lin} + 0.7152 \times G_{lin} + 0.0722 \times B_{lin} \quad (2)$$

where Y represents the relative luminance of pixel p and $R_{lin}, G_{lin},$ and B_{lin} are the linear values of the red, green, and blue color components of p. According to [27], the average luminance of the area around the sun (300 × 300 pixels) is computed, and when the sun is not located, the luminance of the central area of the image (900 × 900 pixels) is taken into account [27]. In the generated dataset, the values in the "*sunLuminance*" column correspond to the *Sun Luminance* feature. Results are shown in Figure 9.

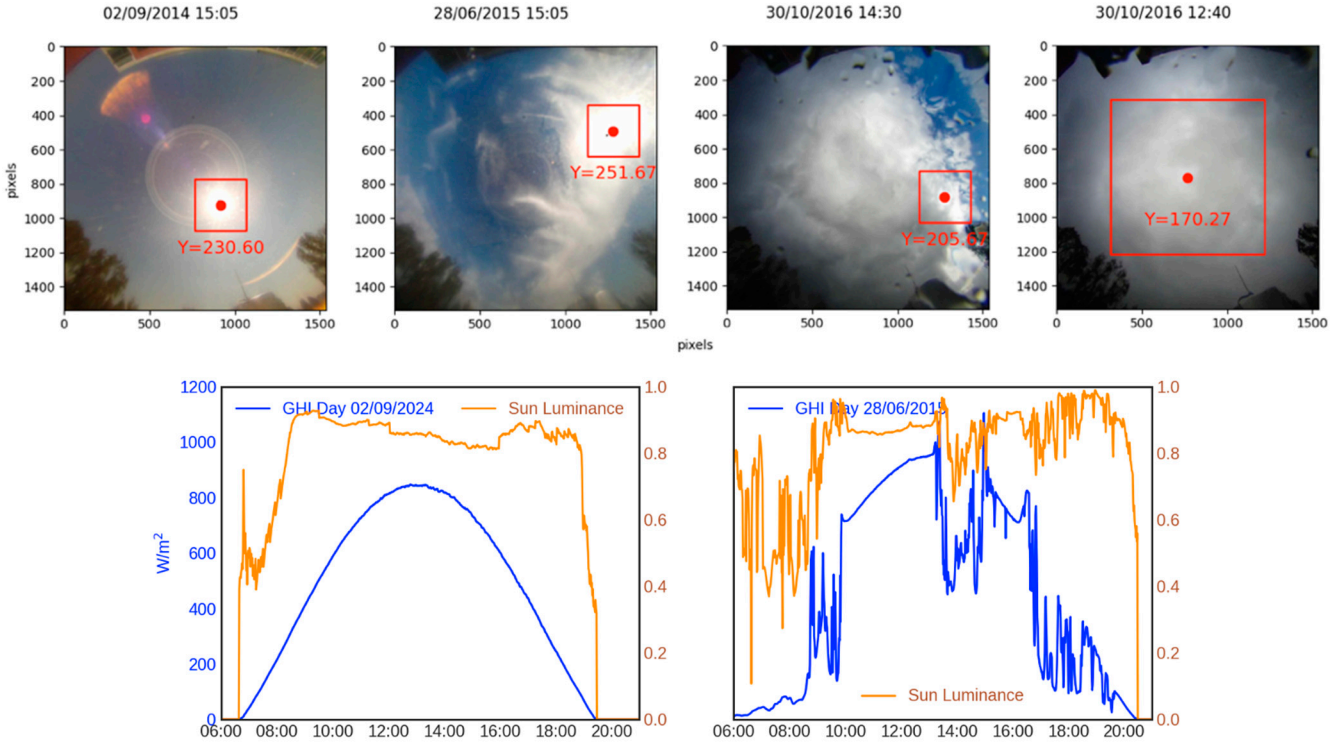


Figure 9. Sun Luminance image processing and samples. Source: [27].

3.7. Sun Luminance Adjusted

This feature is similar to *Sun Luminance* with post-processing applied to the Y value. Given that Y_{hm} is obtained from the image captured hour (h) and minute (m), an adjustment factor f_{hm} is applied based on the $clearSkyGHI_{hm}$ feature previously computed [27]. This factor is calculated by the ratio of the clear sky GHI at that hour (h) and minute (m) in terms of the maximum GHI expected for that day, according to Equation (3):

$$f_{hm} = \frac{clearSkyGHI_{hm}}{max(clearSkyGHI)} \quad (3)$$

where $clearSkyGHI_{hm}$ is the clear sky GHI irradiance value at time $h : m$ provided by the Ineichen–Perez synthetic model, and $max(clearSkyGHI)$ is the maximum daily clear sky irradiance value. Finally, considering the luminance Y value calculated for the image taken at hour h and minute m, the adjusted sun luminance value (Y_{adj}) is computed using the equation $Y_{adj} = Y_{hm} \times lf_{hm}$. Figure 10 shows examples of the *Sun Luminance Adjusted* feature; their values in the generated dataset are in the "*fitLuminanceCSI*" column.

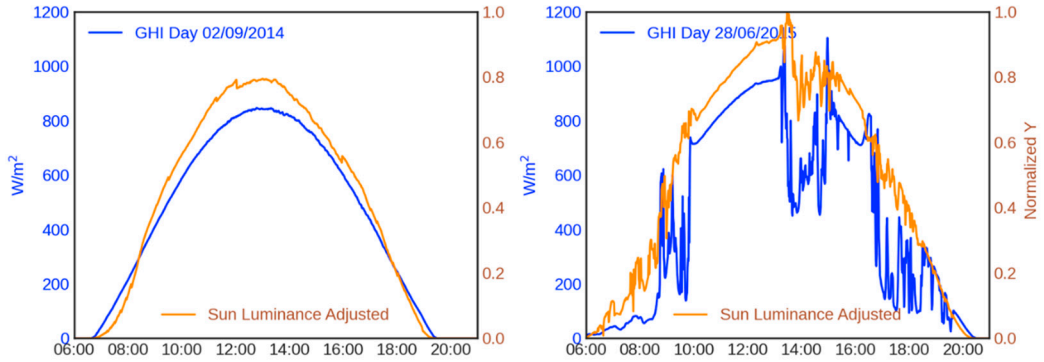


Figure 10. Sun Luminance Adjusted samples. Adapted from [27].

3.8. Sun Located

This feature returns a true (1) or false (0) value for the location of the sun in the image. Its processing consists of locating the brightest region in the image in the red channel, similar to processing used in the *Sun Luminance* and *Clouds Around Sun* features [27]. The values for this feature are in the "*sunLocated*" column of the generated dataset.

3.9. Season

This feature corresponds to the annual climatic season in which the all-sky image was captured. These values are represented by the *season* column in the generated dataset.

3.10. White Pixel Ratio

The *White Pixel Ratio* feature represents the percentage of white pixels in the sky area of the image. According to [27], a threshold of 240 was used to incorporate the grayish tones of denser clouds. Examples of this feature are shown in Figure 11, and contained in the "*whitePixelRatio*" column in the generated dataset.

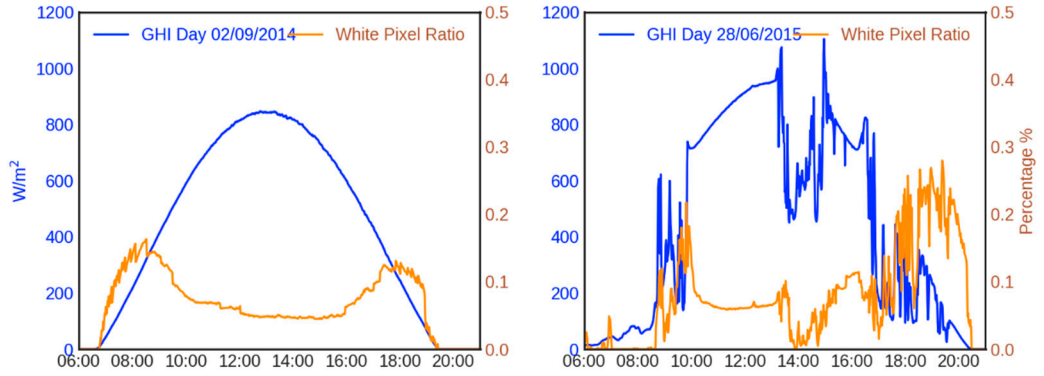


Figure 11. White Pixel Ratio samples. Adapted from [27].

Finally, the set of nine features explicitly extracted from each all-sky image (f_1 to f_9) and temporal data (f_{10} to f_{14}) can be used to predict global solar irradiance over a given horizon. In this sense, the dataset generated provides a choice of the following horizon options for the Clouds Movement feature: 1, 2, 3, 4, 5, 10, 15, 30, and 60 min.

4. Conclusions and Future Perspectives

The construction and availability of datasets for training machine learning models is a crucial practice in data science and artificial intelligence research. In the field of PSPEG, this paper details and provides an original dataset built from the application of the HPM method for the promotion and development of new studies on predictive models using machine learning and deep learning.

The dataset provided is original in the PSPEG literature. Its originality is due to the set of 14 features (characteristics) that can be jointly applied to the prediction of global solar irradiance. This is the first dataset in the scientific literature to provide a set of features explicitly extracted from all-sky images. In general, datasets in the literature provide meteorological data or images for use in PSPEG. It is important to note that the set of features from the dataset provided has shown promising accuracy results [27], although there is a dependence on all-sky images and the additional computational cost of pre-processing the images.

The study by [27] highlighted the possibility of improving the HPM, such as optimizing its features, analyzing and selecting more representative sets of features, and optimizing for specific prediction horizons, among others. Therefore, the description and availability of a dataset based on the application of the features provided by the HPM allows the development of new studies on PSPEG in the scientific community to be leveraged and facilitated.

The generated dataset is quality controlled, with standardized data, validated types, and no missing or inconsistent values. Therefore, the provided dataset contributes to the validation of machine learning models, increasing the reliability of scientific findings. In addition, it can be used as a benchmark for future research, as it was built from a standardized dataset of all-sky images. The provided dataset also allows fair comparison between different algorithms and approaches. This helps to identify the most effective techniques and to promote continuous improvements in the HPM, including various prediction horizons (1, 2, 3, 4, 5, 10, 15, 30, and 60 min).

Due to the original HPM approach applied to solar irradiance prediction, this dataset is also original, as it is the first dataset made available for global solar irradiance prediction and provides a standardized and explicit set of features extracted using all-sky image processing techniques. Therefore, the availability and detailing of the calculation of the features in the dataset generated promotes the development of new machine learning methods and algorithms, especially those based on time series.

Finally, the creation of datasets is a fundamental activity that contributes significantly to the advancement of science and technology. In the area of predicting solar irradiance and photovoltaic solar energy generation, these datasets drive innovation, improve the accuracy of artificial intelligence models, and facilitate decision-making at multiple levels, from mitigating intermittent energy generation to negotiating future energy supply and demand.

Author Contributions: Conceptualization: J.N.M., J.J.G.L. and O.H.A.J.; methodology: J.N.M. and O.H.A.J.; validation: J.N.M. and O.H.A.J.; investigation and simulation: J.N.M. and O.H.A.J.; writing—original draft preparation: J.N.M., J.J.G.L. and O.H.A.J.; writing—review and editing: J.N.M., J.J.G.L. and O.H.A.J.; project administration: O.H.A.J.; funding acquisition: O.H.A.J. All authors have read and agreed to the published version of the manuscript.

Funding: This paper was partially supported by the FACEPE agency (Fundação de Amparo a Pesquisa de Pernambuco) throughout the project, with references APQ-0616-9.25/21 and APQ-0642-9.25/22. O.H.A.J. was funded by the Brazilian National Council for Scientific and Technological Development (CNPq), grant numbers 407531/2018-1, 303293/2020-9, 405385/2022-6, 405350/2022-8, and 40666/2022-3, as well as the Program in Energy Systems Engineering (PPGESE) Academic Unit of Cabo de Santo Agostinho (UACSA), Federal Rural University of Pernambuco (UFRPE). J.N.M. was funded by the Federal University of Latin American Integration (UNILA).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset, data dictionary, and related codes are available in the Mendeley Data Repository (http://doi.org/10.17632/347p98c2rd.1) accessed on 18 September 2024.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. International Energy Agency: IEA. *Solar PV*; IEA: Paris, France, 2022.
2. Antonanzas-Torres, F.; Urraca, R.; Polo, J.; Perpiñán-Lamigueiro, O.; Escobar, R. Clear Sky Solar Irradiance Models: A Review of Seventy Models. *Renew. Sustain. Energy Rev.* **2019**, *107*, 374–387. [CrossRef]
3. Ying, C.; Wang, W.; Yu, J.; Li, Q.; Yu, D.; Liu, J. Deep Learning for Renewable Energy Forecasting: A Taxonomy, and Systematic Literature Review. *J. Clean. Prod.* **2023**, *384*, 135414. [CrossRef]
4. Maciel, J.N.; Ledesma, J.J.G.; Ando Junior, O.H. Forecasting Solar Power Output Generation: A Systematic Review with the Proknow-C. *IEEE Lat. Am. Trans.* **2021**, *19*, 612–624. [CrossRef]
5. Kotsiantis, S.B.; Zaharakis, I.D.; Pintelas, P.E. Machine Learning: A Review of Classification and Combining Techniques. *Artif. Intell. Rev.* **2006**, *26*, 159–190. [CrossRef]
6. Kumar, D.S.; Yagli, G.M.; Kashyap, M.; Srinivasan, D. Solar Irradiance Resource and Forecasting: A Comprehensive Review. *IET Renew. Power Gener.* **2020**, *14*, 1641–1656. [CrossRef]
7. Wang, H.; Lei, Z.; Zhang, X.; Zhou, B.; Peng, J. A Review of Deep Learning for Renewable Energy Forecasting. *Energy Convers. Manag.* **2019**, *198*, 111799. [CrossRef]
8. Obando, E.D.; Carvajal, S.X.; Pineda Agudelo, J. Solar Radiation Prediction Using Machine Learning Techniques: A Review. *IEEE Lat. Am. Trans.* **2019**, *17*, 684–697. [CrossRef]
9. Voyant, C.; Notton, G.; Kalogirou, S.; Nivet, M.L.; Paoli, C.; Motte, F.; Fouilloy, A. Machine Learning Methods for Solar Radiation Forecasting: A Review. *Renew. Energy* **2017**, *105*, 569–582. [CrossRef]
10. Kumari, P.; Toshniwal, D. Deep Learning Models for Solar Irradiance Forecasting: A Comprehensive Review. *J. Clean. Prod.* **2021**, *318*, 128566. [CrossRef]
11. Wang, F.; Xuan, Z.; Zhen, Z.; Li, Y.; Li, K.; Zhao, L.; Shafie-khah, M.; Catalão, J.P.S. A Minutely Solar Irradiance Forecasting Method Based on Real-Time Sky Image-Irradiance Mapping Model. *Energy Convers. Manag.* **2020**, *220*, 113075. [CrossRef]
12. Dev, S.; Savoy, F.M.; Lee, Y.H.; Winkler, S. Estimating Solar Irradiance Using Sky Imagers. *Atmos. Meas. Tech.* **2019**, *12*, 5417–5429. [CrossRef]
13. Yang, H.; Kurtz, B.; Nguyen, D.; Urquhart, B.; Chow, C.W.; Ghonima, M.; Kleissl, J. Solar Irradiance Forecasting Using a Ground-Based Sky Imager Developed at UC San Diego. *Solar Energy* **2014**, *103*, 502–524. [CrossRef]
14. Zhen, Z.; Zhang, X.; Mei, S.; Chang, X.; Chai, H.; Yin, R.; Wang, F. Ultra-short-term Irradiance Forecasting Model Based on Ground-based Cloud Image and Deep Learning Algorithm. *IET Renew. Power Gener.* **2022**, *16*, 2604–2616. [CrossRef]
15. Yang, H.; Wang, L.; Huang, C.; Luo, X. 3d-Cnn-Based Sky Image Feature Extraction for Short-Term Global Horizontal Irradiance Forecasting. *Water* **2021**, *13*, 1773. [CrossRef]
16. Zhang, R.; Feng, M.; Zhang, W.; Lu, S.; Wang, F. Forecast of Solar Energy Production-A Deep Learning Approach. In Proceedings of the 2018 IEEE International Conference on Big Knowledge (ICBK), Singapore, 17–18 November 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 73–82.
17. Pedro, H.T.C.; Larson, D.P.; Coimbra, C.F.M. A Comprehensive Dataset for the Accelerated Development and Benchmarking of Solar Forecasting Methods. *J. Renew. Sustain. Energy* **2019**, *11*, 036102. [CrossRef]
18. Williamson, S.; Businger, S.; Matthews, D. Development of a Solar Irradiance Dataset for Oahu, Hawai'i. *Renew. Energy* **2018**, *128*, 432–443. [CrossRef]
19. Ntavelis, E.; Remund, J.; Schmid, P. SkyCam: A Dataset of Sky Images and Their Irradiance Values. *arXiv* **2021**, arXiv:2105.02922. [CrossRef]
20. Moncada, A.; Richardson, W.; Vega-Avila, R. Deep Learning to Forecast Solar Irradiance Using a Six-Month UTSA SkyImager Dataset. *Energies* **2018**, *11*, 1988. [CrossRef]
21. Nie, Y.; Li, X.; Scott, A.; Sun, Y.; Venugopal, V.; Brandt, A. SKIPP'D: A SKy Images and Photovoltaic Power Generation Dataset for Short-Term Solar Forecasting. *Sol. Energy* **2023**, *255*, 171–179. [CrossRef]
22. Victor, H.; Joylan, N.; Gimenez, J.J.; Júnior, A.; Oswaldo, H.; Wentz, V.H.; Maciel, J.N.; Ledesma, J.J.G.; Ando Junior, O.H. Construção de Um Banco de Dados Para a Predição Do Potencial de Geração de Energia Solar Fotovoltaica. In Proceedings of the IX Encontro Anual de Iniciação Científica-EICTI, Foz do Iguaçu, PR, Brazil; 2020; p. 2020.
23. Ardila, V.M.S.; Maciel, J.N.; Ledesma, J.J.G.; Junior, O.H.A. Fuzzy Time Series Methods Applied to (In)Direct Short-Term Photovoltaic Power Forecasting. *Energies* **2022**, *15*, 845. [CrossRef]
24. Wentz, V.H.; Maciel, J.N.; Gimenez Ledesma, J.J.; Ando Junior, O.H. Solar Irradiance Forecasting to Short-Term PV Power: Accuracy Comparison of ANN and LSTM Models. *Energies* **2022**, *15*, 2457. [CrossRef]
25. Maciel, J.N. *Método Híbrido de Predição Da Irradiância Solar Com Processamento de Imagens e Inteligência Artificial Aplicável a Geração de Energia Solar Fotovoltaica*; Tese de Doutorado, Universidade Federal da Integração Latino-Americana (UNILA): Foz do Iguaçu, PR, Brazil, 2022.
26. Maciel, J.N.; Wentz, V.H.; Ledesma, J.J.G.; Junior, O.H.A. Analysis of Artificial Neural Networks for Forecasting Photovoltaic Energy Generation with Solar Irradiance. *Braz. Arch. Biol. Technol.* **2021**, *64*, e21210131. [CrossRef]
27. Maciel, J.N.; Ledesma, J.J.G.; Ando Junior, O.H. Hybrid Prediction Method of Solar Irradiance Applied to Short-Term Photovoltaic Energy Generation. *Renew. Sustain. Energy Rev.* **2024**, *192*, 114185. [CrossRef]
28. Russell, S.; Norvig, P. *Artificial Intelligence: A Modern Approach*, 3rd ed.; Pearson Education: London, UK, 2016; ISBN 9781292153964.

29. Blaga, R.; Sabadus, A.; Stefu, N.; Dughir, C.; Paulescu, M.; Badescu, V. A Current Perspective on the Accuracy of Incoming Solar Energy Forecasting. *Prog. Energy Combust. Sci.* **2019**, *70*, 119–144. [CrossRef]
30. Lin, F.; Zhang, Y.; Wang, J. Recent Advances in Intra-Hour Solar Forecasting: A Review of Ground-Based Sky Image Methods. *Int. J. Forecast.* **2022**, *39*, 244–265. [CrossRef]
31. Tina, G.M.; Ventura, C.; Ferlito, S.; De Vito, S. A State-of-Art-Review on Machine-Learning Based Methods for PV. *Appl. Sci.* **2021**, *11*, 7550. [CrossRef]
32. Rajagukguk, R.A.; Ramadhan, R.A.A.; Lee, H.-J. A Review on Deep Learning Models for Forecasting Time Series Data of Solar Irradiance and Photovoltaic Power. *Energies* **2020**, *13*, 6623. [CrossRef]
33. Antonanzas, J.; Osorio, N.; Escobar, R.; Urraca, R.; Martinez-de-Pison, F.J.; Antonanzas-Torres, F. Review of Photovoltaic Power Forecasting. *Sol. Energy* **2016**, *136*, 78–111. [CrossRef]
34. Chow, C.W.; Belongie, S.; Kleissl, J. Cloud Motion and Stability Estimation for Intra-Hour Solar Forecasting. *Sol. Energy* **2015**, *115*, 645–655. [CrossRef]
35. Gonzalez, R.C.; Woods, R.E. *Digital Image Processing*, 2nd ed.; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2002; Volume 1, ISBN 0201180758.
36. van Rossum, G. Python 3.9.0 Documentation. Available online: https://docs.python.org/3/ (accessed on 29 July 2022).
37. dos S. Pereira, E. Defisheye. Available online: https://github.com/duducosmos/defisheye (accessed on 29 July 2022).
38. Sara, U.; Akter, M.; Uddin, M.S. Image Quality Assessment through FSIM, SSIM, MSE and PSNR—A Comparative Study. *J. Comput. Commun.* **2019**, *7*, 8–18. [CrossRef]
39. Chauvin, R.; Nou, J.; Thil, S.; Traoré, A.; Grieu, S. Cloud Detection Methodology Based on a Sky-Imaging System. *Energy Procedia* **2015**, *69*, 1970–1980. [CrossRef]
40. Nou, J.; Chauvin, R.; Eynard, J.; Thil, S.; Grieu, S. Towards the Intrahour Forecasting of Direct Normal Irradiance Using Sky-Imaging Data. *Heliyon* **2018**, *4*, e00598. [CrossRef]
41. Dissawa, L.H.; Godaliyadda, R.I.; Ekanayake, P.B.; Agalgaonkar, A.P.; Robinson, D.; Ekanayake, J.B.; Perera, S. Sky Image-Based Localized, Short-Term Solar Irradiance Forecasting for Multiple PV Sites via Cloud Motion Tracking. *Int. J. Photoenergy* **2021**, *2021*, 9973010. [CrossRef]
42. Urquhart, B.; Ghonima, M.; Nguyen, D.; Kurtz, B.; Chow, C.W.; Kleissl, J. Sky-Imaging Systems for Short-Term Forecasting. In *Solar Energy Forecasting and Resource Assessment*; Solar Energy Forecasting and Resource Assessment; Elsevier Inc.: San Diego, CA, USA, 2013; pp. 195–232. ISBN 9780123971777.
43. Yang, J.; Min, Q.; Lu, W.; Ma, Y.; Yao, W.; Lu, T.; Du, J.; Liu, G. A Total Sky Cloud Detection Method Using Real Clear Sky Background. *Atmos. Meas. Tech.* **2016**, *9*, 587–597. [CrossRef]
44. Chow, C.W.; Urquhart, B.; Lave, M.; Dominguez, A.; Kleissl, J.; Shields, J.; Washom, B. Intra-Hour Forecasting with a Total Sky Imager at the UC San Diego Solar Energy Testbed. In *Solar Energy*; Elsevier Ltd.: San Diego, CA, USA, 2011; Volume 85, pp. 2881–2893.
45. Shields, J.E.; Johnson, R.W.; Koehler, T.L. Automated Whole Sky Imaging Systems for Cloud Field Assessment. In Proceedings of the Fourth Symposium on Global Change Studies of the American Meteorological Society, Boston, MA, USA, 17–22 January 1993.
46. Otsu, N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Trans. Syst. Man. Cybern.* **1979**, *9*, 62–66. [CrossRef]
47. Li, Q.; Lu, W.; Yang, J. A Hybrid Thresholding Algorithm for Cloud Detection on Ground-Based Color Images. *J. Atmos. Ocean. Technol.* **2011**, *28*, 1286–1296. [CrossRef]
48. Li, X.; Lu, Z.; Zhou, Q.; Xu, Z. A Cloud Detection Algorithm with Reduction of Sunlight Interference in Ground-Based Sky Images. *Atmosphere* **2019**, *10*, 640. [CrossRef]
49. Kazantzidis, A.; Tzoumanikas, P.; Bais, A.F.; Fotopoulos, S.; Economou, G. Cloud Detection and Classification with the Use of Whole-Sky Ground-Based Images. *Atmos. Res.* **2012**, *113*, 80–88. [CrossRef]
50. Rajagukguk, R.A.; Kamil, R.; Lee, H.-J. A Deep Learning Model to Forecast Solar Irradiance Using a Sky Camera. *Appl. Sci.* **2021**, *11*, 5049. [CrossRef]
51. Savoy, F.M.; Dev, S.; Lee, Y.H.; Winkler, S. Geo-Referencing and Stereo Calibration of Ground-Based Whole Sky Imagers Using the Sun Trajectory. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 7473–7476.
52. Ineichen, P.; Perez, R. A New Airmass Independent Formulation for the Linke Turbidity Coefficient. *Sol. Energy* **2002**, *73*, 151–157. [CrossRef]
53. Holmgren, W.F.; Hansen, C.W.; Mikofski, M.A. Pvlib Python: A Python Package for Modeling Solar Energy Systems. *J. Open Source Softw.* **2018**, *3*, 884. [CrossRef]
54. Szeliski, R. *Computer Vision: Algorithms and Applications (Texts in Computer Science)*, 2nd ed.; Springer: Washington, DC, USA, 2022; Volume 1, ISBN 978-3030343712.
55. International Telecommunication Union. *ITU Parameter Values for the HDTV Standards for Production and International Programme Exchange*; International Telecommunication Union: Geneva, Switzerland, 2015.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.