

Article

# Three Novel Artificial Neural Network Architectures Based on Convolutional Neural Networks for the Spatio-Temporal Processing of Solar Forecasting Data

Llinet Benavides Cesar , Miguel-Ángel Manso-Callejo \*  and Calimanut-Ionut Cira 

Departamento de Ingeniería Topográfica y Cartográfica, E.T.S.I. en Topografía Geodesia y Cartografía, Universidad Politécnica de Madrid, C/Mercator 2, 28031 Madrid, Spain; llinet.bcesar@upm.es (L.B.C.)

\* Correspondence: m.manso@upm.es

**Abstract:** In this work, three new convolutional neural network models—spatio-temporal convolutional neural network versions 1 and 2 (ST\_CNN\_v1 and ST\_CNN\_v2), and the spatio-temporal dilated convolutional neural network (ST\_Dilated\_CNN)—are proposed for solar forecasting and processing global horizontal irradiance (GHI) data enriched with meteorological and astronomical variables. A comparative analysis of the proposed models with two traditional benchmark models shows that the proposed ST\_Dilated\_CNN model outperforms the rest in capturing long-range dependencies, achieving a mean absolute error of 31.12 W/m<sup>2</sup>, a mean squared error of 54.07 W/m<sup>2</sup>, and a forecast skill of 37.21%. The statistical analysis carried out on the test set suggested highly significant differences in performance (*p*-values lower than 0.001 for all metrics in all the considered scenarios), with the model with the lowest variability in performance being ST\_CNN\_v2. The statistical tests applied confirmed the robustness and reliability of the proposed models under different conditions. In addition, this work highlights the significant influence of astronomical variables on prediction performance. The study also highlights the intricate relationship between the proposed models and meteorological and astronomical input characteristics, providing important insights into the field of solar prediction and reaffirming the need for further research into variability factors that affect the performance of models.

**Keywords:** solar forecast; dilated CNN; spatio-temporal; deep learning



**Citation:** Benavides Cesar, L.; Manso-Callejo, M.-Á.; Cira, C.-I. Three Novel Artificial Neural Network Architectures Based on Convolutional Neural Networks for the Spatio-Temporal Processing of Solar Forecasting Data. *Appl. Sci.* **2024**, *14*, 5955. <https://doi.org/10.3390/app14135955>

Academic Editors: César Berna-Escriche and Paula Bastida-Molina

Received: 13 May 2024  
Revised: 1 July 2024  
Accepted: 4 July 2024  
Published: 8 July 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 sustainable energy sources has led to a significant growth in the use of solar energy. In particular, the Spanish government, as set out in the National Integrated Energy and Climate Plan 2021–2030 (PNIEC), anticipates electricity generation with a total installed capacity of 157 GW by the year 2030, of which 37 GW is expected to come from solar PV [1]. For this reason, it is considered important to develop reliable predictive models for solar radiation [2]. One of the most important factors influencing the quality of solar forecasts is solar variability [3]. This variability is mainly a spatio-temporal phenomenon and, in recent years, researchers in this field have introduced the use of machine learning methods due to their capacity for processing large amounts of data. Machine learning methods have the capability of modelling this phenomenon more accurately [4], and have been extended to solar forecasting, showing significant improvements over traditional statistical methods [5].

In the deep learning field of machine learning, convolutional neural networks (CNNs) have exhibited remarkable capabilities in learning intricate patterns from spatially and temporally complex datasets, such as satellite and sky images [6,7] or graphs [8]. Among the various architectural modifications introduced to enhance the performance of CNNs [9–13], dilated convolutions [14] have emerged as a novel architectural refinement. They have

been used with univariate time series [15] for video-based crowd estimation [16], for speech emotion recognition [17], and for dense prediction [18], among other tasks.

Dilated convolutions enable networks to achieve wide receptive fields with minimal layers, maintaining input fields throughout the network while also ensuring computational efficiency [19]. Mishra et al. [15] used stacked dilated convolutional layers for load forecast with univariate time series and showed that the model surpassed other state-of-the-art models by capturing the local trends and seasonality. Chen et al. [20] used dilated convolutional layers to extract hidden time characteristics as part of a more complex multivariate time-series forecast model. The model was applied to traffic and engine datasets and obtained better performances than other baselines. In the same way, Liang et al. [21] used a stack of dilated convolutional layers with different values of dilation to capture temporal dependencies and to make predictions for the base time series. The convolutional layers were part of a more complex model that also presented an attention scheme. The model was used for the ultra-short-term spatio-temporal forecasting of renewable resources (wind speed and solar generation). Fan et al. [22] used a genetic-based attention network for photovoltaic power forecasting and utilized dilated convolution as the non-linear component to simplify the network structure and learn the spatial and temporal dependencies between historical weather data. The authors employed two dilated convolutional layers that concatenated with a Hadamard product used as an input for a more complex model.

In [23], the authors review the most widely used deep learning methods for solar radiation prediction. Recurrent networks are highlighted and, in particular, long short-term memory (LSTM), gated recurrent unit (GRU), and CNN networks combined with other networks or methods are mentioned. On the other hand, in [24], the authors concentrate the study on vision methods of deep learning for solar forecasting. CNNs stand out with their hybridization with other networks to improve results. All of these models offer major advantages due to their ability to work with sequential data, such as time series. They are able to learn complex patterns from the data more easily. They can handle large volumes of data. They are able to process data from different sources. Although they have some disadvantages, these models are often referred to as a 'black box' because it is often not clear which factors are used to predict the outcome; nonetheless, on-going research seeks to better explain this. Training time for the models is costly because of the time involved, but the increasingly high accuracy achieved by them makes their use widespread in the field of solar radiation forecast.

This study shows the benefits of using advanced artificial neural networks for solar forecasting (ST\_Dilated\_CNN considerably outperformed conventional methodologies) and highlights the appropriateness of applying a spatio-temporal approach to GHI data (considering the multidimensional influences of meteorological and astronomical variables). Dilated CNNs can preserve the resolution and dimensionality of the data at the output layer while providing a larger receptive field, are more computationally efficient, and have less memory consumption. This is because of the dilation of the layers rather than pooling of the layers, so the order of the data is preserved. For example, the structure of the convolution helps to preserve the order of the data in 1D dilated causal convolutions where the prediction of the output depends on previous inputs.

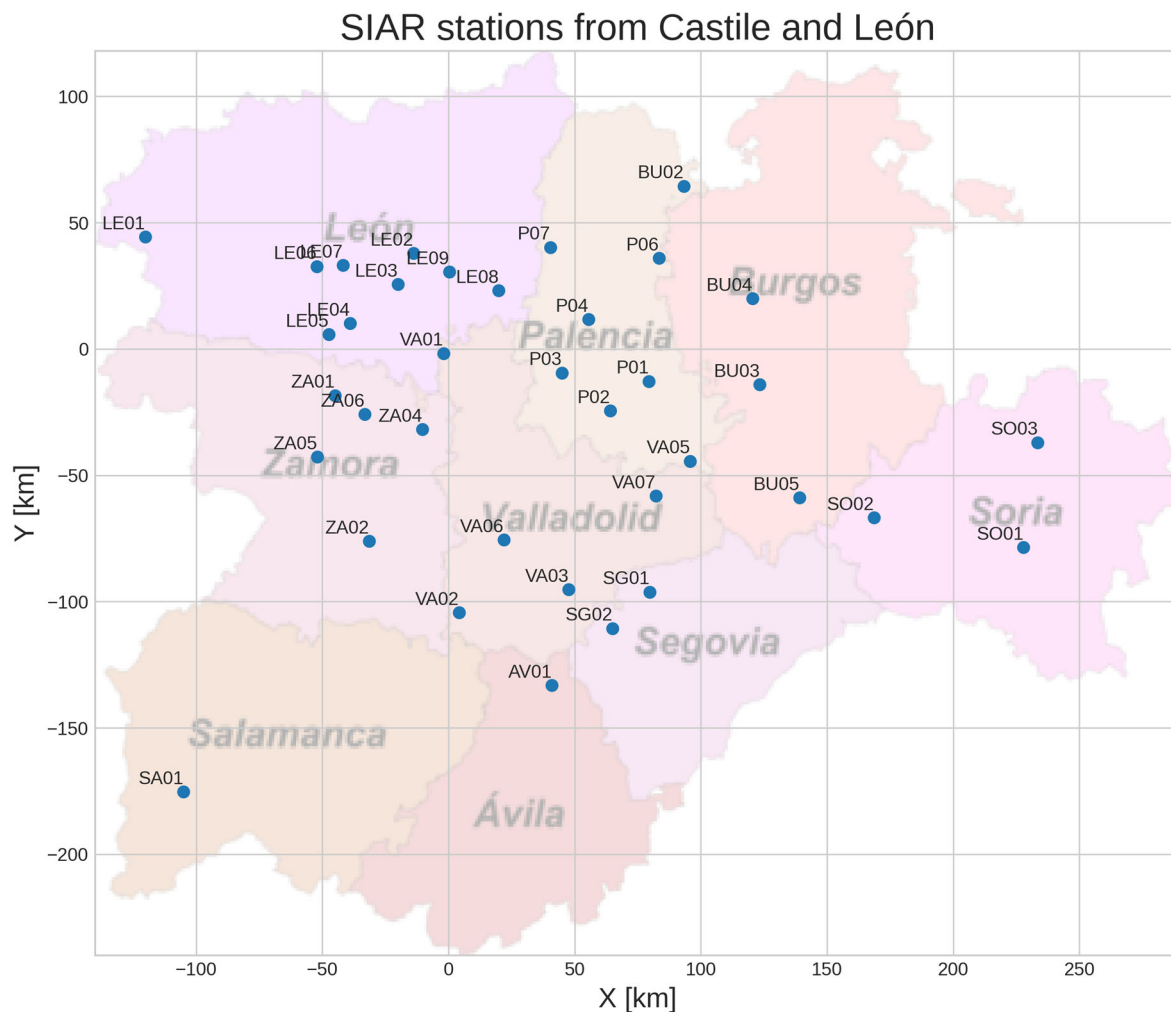
The main contributions of this work are as follows:

- The proposed network ST\_Dilated\_CNN succeeds in capturing representations of the complex spatio-temporal patterns present in solar data.
- The proposed architecture ST\_CNN\_v2 is able to simulate the dilated range, with the difference between the two models being less than one percentage point.
- The proposed models outperform the benchmark models, with a difference of up to three percentage points in forecast skill.

The paper continues with the Section 2, where the dataset used as well as the selection of the stations are described. Section 3 presents the proposed methods, while Section 4 presents the training methodology applied. The results and the associated discussion and analysis can be found in Section 5. The article ends with the Section 6.

## 2. Data

CyL-GHI—a global horizontal irradiance dataset from the region Castile and León in Spain—was used for the development of this study. CyL-GHI [25] is public, distributed under a CC-BY-SA-NC (“Creative Commons Attribution-ShareAlike-NonCommercial”) license, and can be downloaded from the Zenodo repository [26]. The data are provided in three csv (comma-separated value) files (“CyL\_GHI\_ast.csv”, “CyL\_meteo.csv”, and “CyL\_geo.csv”). The CyL-GHI dataset contains records from 1 January 2002 to 31 December 2019, with a time granularity of 30 min, from 37 stations located in Castile and León (Figure 1).



**Figure 1.** CyL-GHI dataset: global horizontal irradiance dataset from the region Castile and León in Spain. Source: “Figure 4. Distribution of the 37 stations selected for the CyL-GHI dataset after the exploratory data analysis process.” by Benavides et al. [25], licensed under Creative Commons Attribution 4.0 License.

The CyL\_GHI\_ast.csv file contains data on global horizontal irradiance, solar elevation angle, upper atmospheric radiation, and the azimuthal angle of the sun. The CyL\_meteo.csv file contains meteorological data on the variables of air temperature, humidity, wind speed, wind direction, and precipitation. The CyL\_geo.csv file contains the geographical latitude, longitude, and altitude data of the 37 stations, as well as the station codes and names. The data are for all stations for the period considered.

Seven stations (BU04, LE05, P03, P07, SG01, SO01, and ZA02) were evaluated to compare two conditions regarding the position of neighbors. Some stations were surrounded by neighbors, following the traditional approach, while others had no neighbors in ei-

ther direction. All stations had the same number of neighbors and were located within 50 to 100 km of the target station.

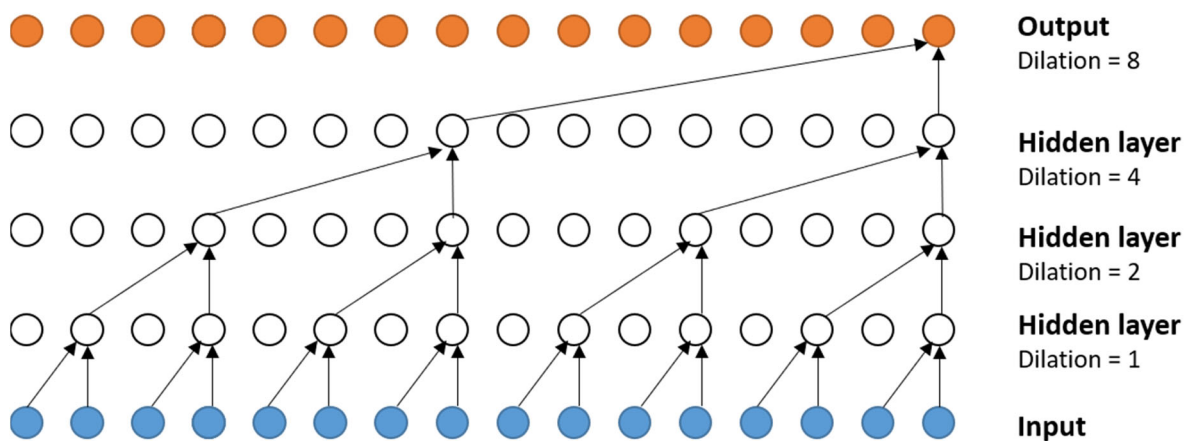
Other studies that have used data from this region have focused on stations that are surrounded by neighboring stations. Eschenbach et al. [27] used station VA01 to compare the impact of the spatial dispersion of neighboring stations on forecast skill by comparing stations from two pyranometer networks with different spatial coverage. Gutierrez-Corea et al. [28] used the data from station VA08 and its eight nearest neighbors using as a discriminant the database capacity to limit the number of neighbors. In both studies, it was ensured that the stations studied had neighboring stations in all directions. This is an optimal way of choosing stations because it is guaranteed that no matter where the clouds originate from, they will always have information from a neighbor in that direction. The selection of stations in this study aims to ensure that the models are prepared for a non-ideal case where they may not have information from neighbors in all directions.

### 3. Proposed Methods Based on Convolutional Neural Networks

Solar irradiance prediction is important for many applications, ranging from generation planning to grid management. In this context, we will make use of CNNs for spatio-temporal forecasting. Because CNNs are known for their ability to capture local patterns and learn hierarchical representations of data, these networks have shown their ability to predict time series and resolve complex spatial patterns [29,30]. By using these networks to predict solar radiation, we can exploit their talents to analyze how factors such as location, terrain distribution, and climate influence solar radiation prediction. CNNs have been evaluated for sequential data, and it has been shown that they can be equal or superior to recurrent networks, which are currently the most widely used for this type of data [31].

From a mathematical point of view, a convolution is an operation that transforms two functions into a third. This happens because each convolutional layer typically reduces the spatial dimensions (height and width) of the input. To avoid this reduction, the size of the convolutional kernel can be adjusted by applying techniques such as padding and stride adjustments.

One technique for expanding the kernel is dilated convolution, which includes “holes” between consecutive elements (as shown in Figure 2). The mentioned figure is an example of the structure of dilation layers; in the case of the presented model, the numbers of dilation layers are one, two, four, and eight. A dilated-based approach is introduced in [14] in the form a proposed context module comprising seven layers, applying  $3 \times 3$  convolutions with varying dilation factors: 1, 1, 2, 4, 8, 16, and 1. Each layer conducts  $3 \times 3 \times C$  convolutions followed by truncation and a final layer of  $1 \times 1 \times C$  convolutions.



**Figure 2.** A stack of dilated convolutional layers for dilations 1, 2, 4, and 8. Note: Dilation controls the distance between the neighboring elements that a neuron in the convolutional layer can see in its receptive field.

Another technique applied is padding, which consists of padding the output with zero values (to restore lost pixels). Causal padding is a special type of padding that works with conv1D layers and is used in sequence-to-sequence [32] models and time-series prediction: the stride is increased so that the output fields overlap less and there is less spatial dimension. This type of padding adds elements to the beginning of the data, which helps the prediction of the first elements in the time steps [33].

From this analysis, two models that fix the filter and increase the stride to increase the receptive field are produced in the form of spatio-temporal convolutional neural network version 1 and version 2 (described in Sections 3.1 and 3.2). Then, a method that enlarges the receptive field by stacking layers with different dilation factors is produced, named the spatio-temporal dilated convolutional neural network (in Section 3.3). All methods use ReLU (one of the more popular activation layers that increases the non-linear properties of the model and the entire network without affecting the reception field of the convolutional layer). The proposed artificial neural networks are described in Sections 3.1–3.3.

### 3.1. Spatio-Temporal Convolutional Neural Network Version 1 (ST\_CNN\_v1)

The first proposed model, called spatio-temporal convolutional neural network version 1 (ST-CNN-v1), is implemented using the Keras deep learning library [34]. The model architecture is defined in Figure 3.

ST-CNN-v1 consists of a one-dimensional convolutional layer with 64 filters, a kernel size of two, and stride of one. Next, a max-pooling layer was added for feature extraction, with a pooling size and stride of two. Afterwards, a flattened layer can be found, and its output is passed through a densely connected layer with a rectified linear unit (ReLU) [35] activation function. The final layer features a single unit modelling the predicted solar radiation value. The Adam [36] optimizer is used for minimizing the mean absolute error (MAE—measuring the average absolute difference between the predicted values and the actual values) loss function.

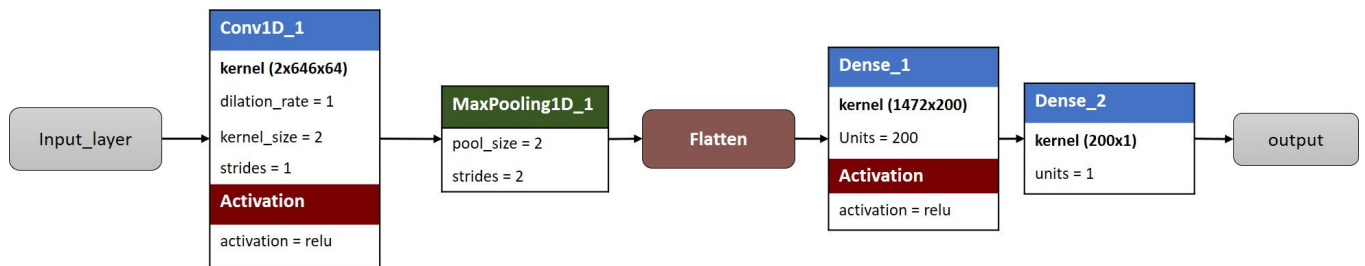


Figure 3. Description of the architecture of the ST\_CNN\_v1 model.

### 3.2. Spatio-Temporal Convolutional Neural Network Version 2 (ST\_CNN\_v2)

The second proposed model, named spatio-temporal convolutional neural network version 2 (ST\_CNN\_v2), uses learning structures specific to CNNs and can capture long-range dependencies within the input time series through configurations of filters and layer strides. The model architecture is defined in Figure 4.

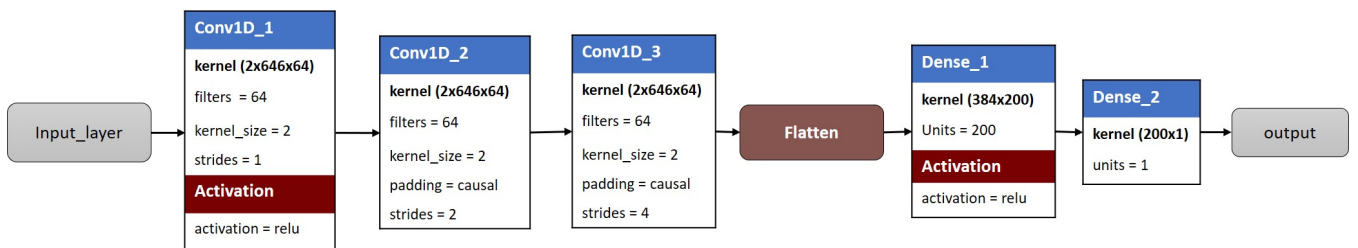


Figure 4. Description of the architecture of the ST\_CNN\_v2 model.

ST\_CNN\_v2 features multiple one-dimensional convolutional layers with a distinct stride size. All convolutional layers use a kernel size of two and the ReLU activation function, starting with the first one. Subsequent layers introduce causal padding and increasing stride to capture the different time scales of the input time series. The output of the flattened layer is then passed through two densely connected layers with ReLU activation, with the last one featuring a single unit to predict solar radiation. MAE is chosen again as the loss function and the Adam optimizer is used for training.

### 3.3. Spatio-Temporal Dilated Convolutional Neural Network (ST\_Dilated\_CNN)

The third proposed model, the spatio-temporal dilated convolutional neural network (ST\_Dilated\_CNN), employs a layer with more dilation factors than that proposed by [14]. The number is increased to 32 and finished with two dense layers. Moreover, it works with a kernel of two.

The model features the architecture shown in Figure 5. It starts with the input layer. Six one-dimensional convolutional layers with different expansion factors follow it. The dilation factors range from 1 to 32. This set of dilated layers allows the model to capture time dependencies on more than one scale of input data (time series). The model continues with a one-dimensional convolutional layer with 64 filters, followed by a flattening layer and two densely connected layers with the ReLU activation function. The last dense layer produces a single output, which represents the solar radiation to be predicted. MAE and Adam were used as the loss function and optimizer for model training.

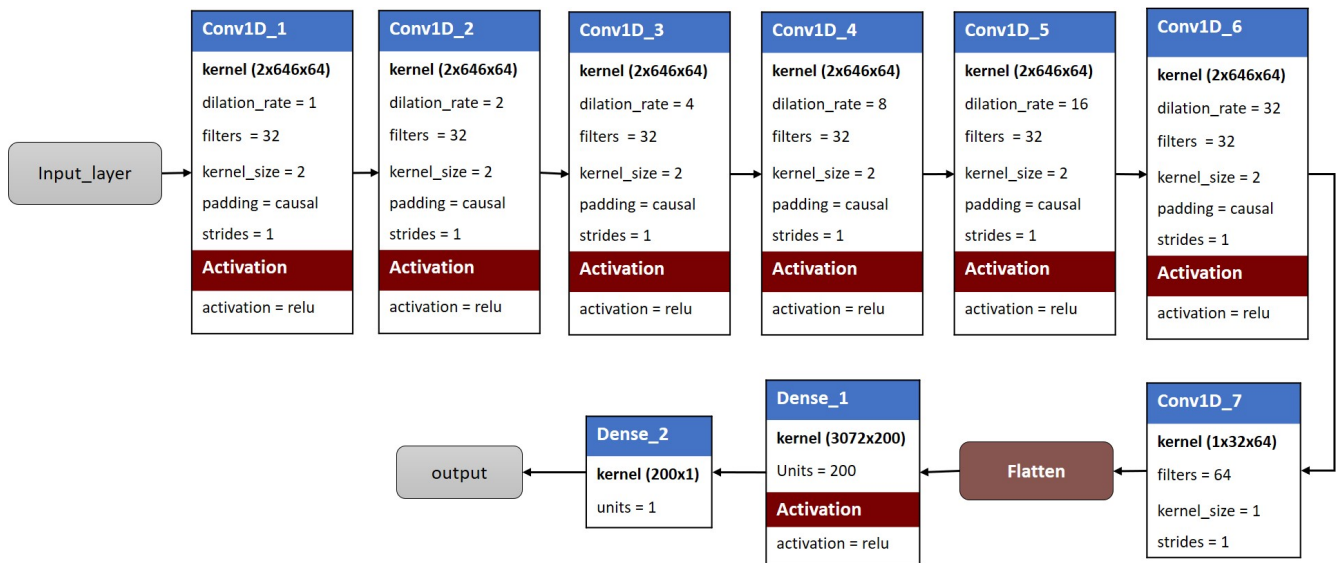


Figure 5. Description of the architecture of the ST\_Dilated\_CNN model.

## 4. Training Methodology

In this work, the predictive ability of deep learning models, which are known for their ability to extract spatio-temporal information from data, was evaluated. The experiments were divided into three scenarios based on the input characteristics of the models. Scenario 1 included only irradiance from the target and neighboring stations, Scenario 2 included astronomical data (solar elevation angle and azimuth angle) in addition to the irradiance used in Scenario 1, and the third scenario added meteorological variables (temperature, humidity, wind speed, wind direction, and precipitation for each station) in addition to the variables used in Scenario 2.

Considering that the temporal resolution of the data is half an hour, 48 lags representing 24 h in the past are used for each of the variables in order to predict the next instant of time, which in this case would be the next half hour (30 min). The data are prepared using the sliding window method and then flattened to enter to the model. This selection

of features ensures that the models have as much information about their environment as possible. Because low solar incidence angles tend to affect the quality of the measurements, only measurements for solar elevation angles greater than 5° were considered.

The methodology used is shown in Figure 6. Specific features tailored to each station’s particular scenario were chosen. The data for each fold of the dataset were then scaled and reshaped to prepare it for model training and fitting. Once the data were prepared, the model was trained and fitted. Error metrics were then calculated to assess model performance for each station. Finally, the results of these error metrics were combined to obtain an overall picture of the model performance for all selected stations.

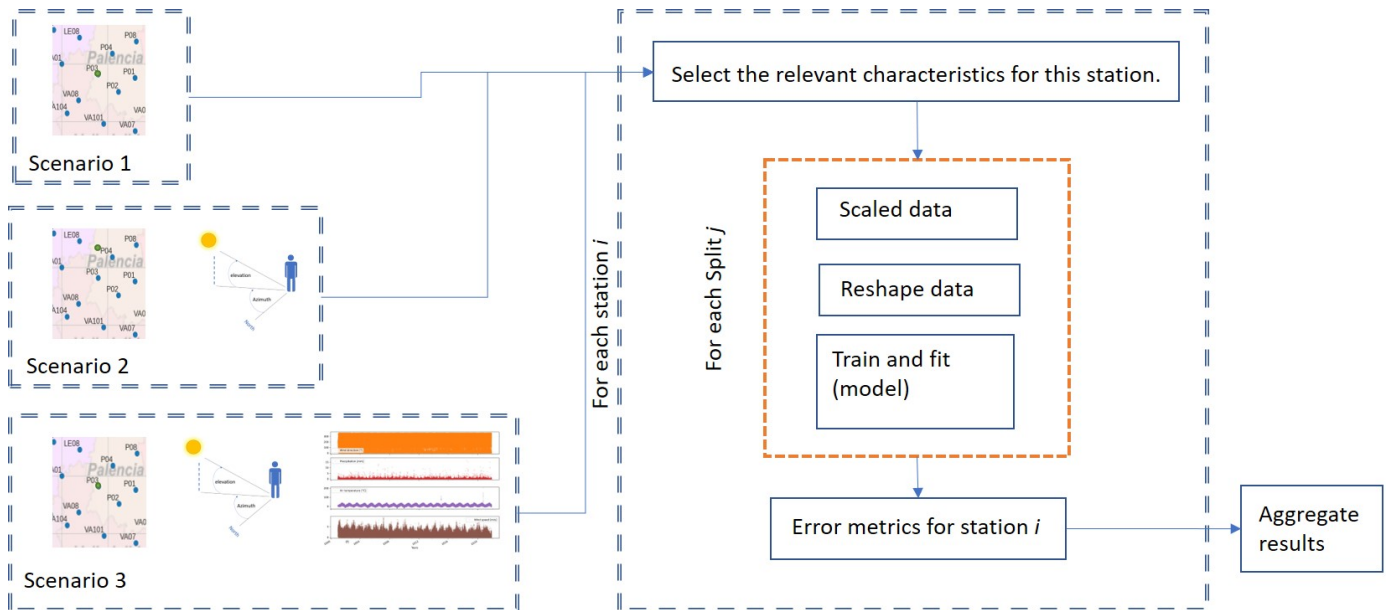


Figure 6. Diagram illustrating the training methodology applied.

The predictive model is defined as  $f(\cdot)$ , which aims to predict a target variable  $y$  based on a set of input variables  $X$ . The target variable  $y$  represents the spatio-temporal phenomenon of solar irradiance.

#### 4.1. Scenario 1

In Scenario 1, the predictive model considers the irradiance of the nearest neighbors of the target station as an exogenous variable (Figure 6). Mathematically, this can be represented with Equation (1).

$$f_1(y) = f(I_{lags}, I_{neighbors}) \tag{1}$$

Here,  $I_{lags}$  represents the irradiance values of the target station and  $I_{neighbors}$  the irradiances values of the nearest neighbors.

#### 4.2. Scenario 2

In Scenario 2, the predictive model includes not only the irradiance of the neighbors, but also the astronomical variables of both the neighbors and the target station (Figure 6). This can be represented as depicted in Equation (2).

$$f_2(y) = f(I_{lags}, I_{neighbors}, A_{target}, A_{neighbors}) \tag{2}$$

Here,  $A_{target}$  represents the astronomical variables of the target station and  $A_{neighbors}$  represents the astronomical variables of the neighbors.

### 4.3. Scenario 3

In Scenario 3, the predictive model extends Scenario 2 by including the meteorological variables of both the neighbors and the target station (Figure 6).

$$f_3(y) = f\left(I_{lags}, I_{neighbors}, A_{target}, A_{neighbors}, M_{target}, M_{neighbors}\right) \quad (3)$$

Here,  $M_{target}$  represents the meteorological variables of the target station and  $M_{neighbors}$  represents the meteorological variables of the neighboring stations.

In Equations (1)–(3),  $f(\cdot)$ , represents the predictive model, which could be a deep learning model or any other machine learning algorithm,  $I$  represents irradiance values,  $A$  represents astronomical variables, and  $M$  represents meteorological variables. Finally, the “neighbors” and “target” subscripts denote variables associated with the neighboring stations and the target station, respectively.

### 4.4. Baseline Models

Linear regression (LR) and Random Forest (RF) [37] models are frequently used as baseline models in solar forecast tasks [38–43].

Linear regression seeks to establish a linear relationship between the input variables and the output variable by employing a linear function to predict the values of the target variable based on the input characteristics; its interpretation is easily comprehensible.

Random Forest is an ensemble model that combines multiple decision trees to make predictions. Each tree is trained on a random sample of features and examples from the dataset. To make the prediction, the average of the tree predictions is calculated. Random Forest has proved its resilience against overfitting and its ability to effectively manage linear data and categorical features. In this study, these models serve as a benchmark for evaluating the performance of the models proposed in Section 3.

### Parameter Estimation

The dataset was split into two portions: training and testing. CyL-GHI data from the first 18 years were used for training, while the data from the final year were used for testing.

To identify the best parameters for the models, a 5-fold cross-validation was carried out. In the cross-validation, multiple iterations were performed, and the training set was divided into training and validation folds using the “TimeSeriesSplit” [44] method (which ensures that the time series maintain their time index when the cut is performed). During the preprocessing, the data were normalized between 0 and 1 using min–max normalization, which was then applied to both the training and validation sets.

The training consisted of several training operations for each model, with each iteration featuring a different combination of specified hyperparameters. The process considers several hyperparameters, such as the number of nodes (values of 10, 25, 50, 100), epochs (values of 20, 50, 100, 200), and batch size (values of 32, 64, 128, 256).

### 4.5. Statistical Analysis of the Performance of the Models

To further evaluate the proposed models, a data subsampling method was applied to allow the study of the consistency and variability of the models. This technique evaluates the performance of algorithms in the case of unseen or independent datasets. In the performance prediction of the models considered, models were only updated with new data, and previously trained data were not considered. By using the subsampling method, distortions in the dependency structure of a time series caused by the block bootstrap can be avoided [45]. This selected approach is useful for detecting patterns and ensuring that the model has not been overfitted to a specific section of the data. In this study, it was implemented by splitting the entire time series (in this particular case, using a separate year as a test set) into multiple segments, and each segment was used to validate the models independently.

Furthermore, to determine the statistical significance of the results obtained, the analysis of variance (ANOVA) test was applied to the metrics to compute the associated F-statistics and *p*-values. These statistical tools will ensure that the results obtained are not due to chance. A low *p*-value (lower than the significance threshold of 0.05) suggests that there are significant differences in the observed results that are unlikely to be due to chance, which reinforces confidence in the validity of the model, whilst a high value of the F-statistic indicates that the model provides a significant improvement in prediction compared to a model with no predictor variables. A *p*-value lower than 0.001 is considered highly statistically significant.

### 5. Results and Discussion

In this section, the effects of the three scenarios on the performance of the trained models are analyzed. We start by listing the metrics used, and then the experiments are detailed and separated by scenarios. In addition, a comparison between the scenarios is made.

Three performance metrics were used to assess the accuracy of the regression models, namely the mean absolute error (MAE),  $AE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n}$ ; the root mean squared error (RMSE),  $RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$ ; and the forecast skill score (FS),  $FS = 1 - \frac{RMSE_{model}}{RMSE_{persistence}}$ . All metrics were calculated using the actual value  $y_i$  and the predicted value  $\hat{y}_i$ .

#### 5.1. Scenario 1

The results of the three proposed models (ST\_CNN\_v1, ST\_CNN\_v2, and ST\_Dilated\_CNN) for the various meteorological stations in Scenario 1 and comparisons with the baseline models are presented in Table 1 (by means of the FS, MAE, and RMSE metrics achieved by the trained models).

**Table 1.** Results of the baseline and proposed models for Scenario 1 and the selected stations.

Metric Station	Linear Regression			Random Forest			ST_CNN_v1			ST_CNN_v2			ST_Dilated_CNN		
	FS (%)	MAE (W/m <sup>2</sup> )	RMSE (W/m <sup>2</sup> )	FS (%)	MAE (W/m <sup>2</sup> )	RMSE (W/m <sup>2</sup> )	FS (%)	MAE (W/m <sup>2</sup> )	RMSE (W/m <sup>2</sup> )	FS (%)	MAE (W/m <sup>2</sup> )	RMSE (W/m <sup>2</sup> )	FS (%)	MAE (W/m <sup>2</sup> )	RMSE (W/m <sup>2</sup> )
BU04	26.17	42.33	68.76	26.02	41.12	68.90	26.43	39.76	68.52	28.28	37.72	66.79	<b>28.71</b>	40.54	66.39
LE05	34.37	34.90	56.51	35.63	33.25	55.43	34.16	32.43	56.69	35.97	32.27	55.13	<b>36.30</b>	33.44	54.85
P03	25.54	43.38	69.78	24.89	42.03	70.39	25.90	39.98	69.45	26.83	41.45	68.58	<b>26.97</b>	41.63	68.44
P07	31.89	43.71	69.44	30.01	42.83	71.736	30.10	42.38	71.27	30.99	41.01	70.36	<b>32.31</b>	40.93	69.02
SG01	27.05	40.67	70.03	25.28	44.07	71.73	27.36	39.62	69.74	28.49	38.48	68.65	<b>28.74</b>	38.89	68.42
SO01	26.51	40.66	61.68	23.65	41.13	64.08	19.93	41.00	67.21	26.97	38.47	61.30	<b>28.65</b>	38.35	59.89
ZA02	34.70	38.56	57.90	34.87	38.50	57.74	36.11	34.56	56.65	<b>37.23</b>	34.48	55.66	37.22	34.08	55.67

Note: Entries in bold represent the best results.

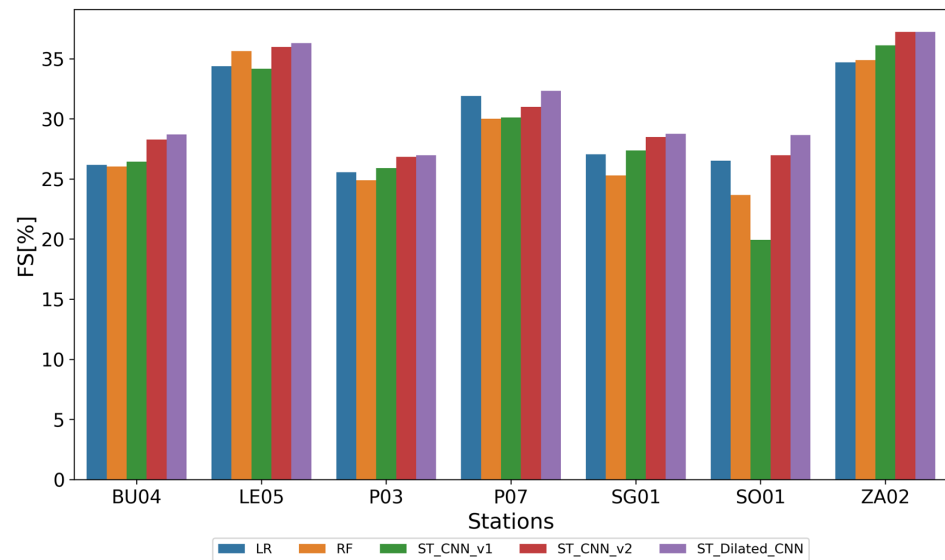
When analyzed in detail, the metrics reveal that the proposed ST\_Dilated\_CNN model obtains higher values for the FS percentages systematically, indicating its ability to better capture the variance within the dataset. The high performance of this model is especially noticeable at stations ZA02 and LE05, where it outperforms other models by a difference of one to two percentage points.

The ST\_CNN\_v2 model has the lowest mean absolute error (MAE) values for most stations, suggesting that its predictions are closer to the observed data. As an illustration of its high performance, the MAE of ST\_CNN\_v2 for station LE05 is significantly lower than that of the linear regression model. The ST\_CNN\_v2 model outperforms the ST\_Dilated\_CNN model in terms of accuracy in forecasting both mean and extreme values, despite the latter’s general superiority in capturing the variability of the data. For instance, ST\_Dilated\_CNN outperforms both ST\_CNN\_v1 and ST\_CNN\_v2 at station BU04 with an FS of 28.71%, an MAE of 40.54 W/m<sup>2</sup>, and an RMSE of 66.39 W/m<sup>2</sup>. Similar trends are shown at the LE05, P03, P07, SG01, SO01, and ZA02 stations, where the prediction accuracy of ST\_Dilated\_CNN is consistently high.

In the context of the table provided, the linear regression and Random Forest models serve as benchmark comparisons to evaluate the performance of more complex models, such as those proposed in this work (ST\_CNN\_v1, ST\_CNN\_v2, and ST\_Dilated\_CNN). The linear regression model, often used as a starting point in predictive analysis, shows moderate performance in all seasons. However, the ST\_CNN\_v2 model consistently outperforms it. In general, the Random Forest model achieves higher FS percentages and lower RMSE values, suggesting greater efficiency in handling variance and outliers. The ST\_CNN\_v1 model has mixed results; in stations like P03, BU04, ZA02, it outperforms the baseline models, but it obtains the worst results for station SO01.

Despite this, the ST\_CNN\_v2 and ST\_Dilated\_CNN models demonstrate higher FS. When analyzed especially in terms of absolute error, this highlights the advantages of employing advanced deep learning techniques for this type of solar data analysis.

Figure 7 presents a plot of the FS values achieved by the trained models for each of the stations studied. The models are shown in the following order: LR, RF, ST\_CNN\_v1, ST\_CNN\_v2, ST\_Dilated\_CNN.



**Figure 7.** Comparative FS performance of three distinct models—ST\_CNN\_v1, ST\_CNN\_v2, and ST\_Dilated\_CNN—across various meteorological stations in Scenario 1.

It is shown that, for station ZA02, the ST\_CNN\_v1 model improves on the baseline models, although it has a discrete result compared to those obtained by the other two deep learning models. This model presents less homogeneous results since, for stations SG01, P03, and BU04, it improves on the baselines, while for stations P05 and SO01, it delivers worse results. The ST\_Dilated\_CNN and ST\_CNN\_v2 models present the highest FS value for stations ZA02 and LE05. Although the ST\_CNN\_v2 model can equal this at station ZA02, the same does not occur for the rest of the stations, where the ST\_Dilated\_CNN model outperforms the rest of the models.

For stations LEO5 and ZA02, the RF model outperforms the LR model, with the opposite being observed for the rest of the stations. ST\_CNN\_v2 is also capable of outperforming the rest of the baseline results, although the performance is always below that of the ST\_Dilated\_CNN model, except for the abovementioned case (ZA02 station), where it equals it.

### 5.2. Scenario 2

Table 2 summarizes the FS, MAE, and RMSE metrics of the two baseline models and the three proposed models supplemented with astronomical input variables. The analysis is performed for several meteorological stations to better understand the nuanced impacts on the accuracy of solar radiation forecasting.

**Table 2.** Results of the baseline and proposed models for the selected stations in Scenario 2.

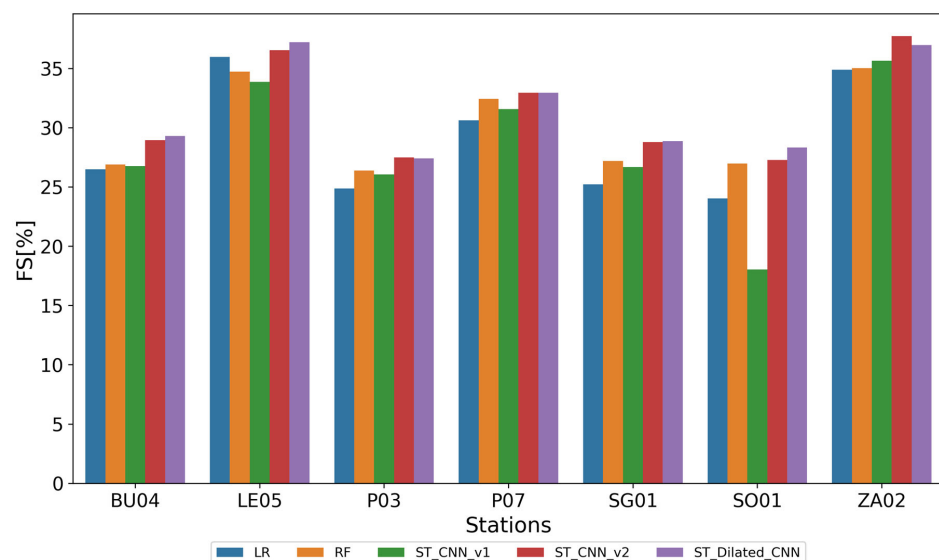
Metric Station	Linear Regression			Random Forest			ST_CNN_v1			ST_CNN_v2			ST_Dilated_CNN		
	FS (%)	MAE (W/m <sup>2</sup> )	RMSE (W/m <sup>2</sup> )	FS (%)	MAE (W/m <sup>2</sup> )	RMSE (W/m <sup>2</sup> )	FS (%)	MAE (W/m <sup>2</sup> )	RMSE (W/m <sup>2</sup> )	FS (%)	MAE (W/m <sup>2</sup> )	RMSE (W/m <sup>2</sup> )	FS (%)	MAE (W/m <sup>2</sup> )	RMSE (W/m <sup>2</sup> )
BU04	26.89	41.83	69.69	26.49	40.72	68.46	26.74	38.75	68.23	28.93	38.16	66.19	<b>29.28</b>	38.53	65.86
LE05	34.72	34.50	66.01	35.96	33.00	55.14	33.85	32.78	56.96	36.53	31.03	54.66	<b>37.21</b>	31.12	54.07
P03	26.37	42.33	71.51	24.85	42.17	70.43	26.05	40.40	69.31	<b>27.47</b>	41.06	67.97	27.39	40.81	68.05
P07	32.41	43.67	69.78	30.61	42.64	70.75	31.57	40.81	69.77	32.93	40.21	68.38	<b>32.93</b>	43.99	68.38
SG01	27.19	43.60	72.97	25.21	44.51	71.81	26.66	40.05	70.41	28.79	38.66	68.37	<b>28.86</b>	39.57	68.30
SO01	26.97	40.81	66.79	24.02	40.84	63.78	18.03	41.18	68.81	<b>27.27</b>	38.15	61.05	<b>28.32</b>	40.46	60.17
ZA02	35.01	38.52	68.53	34.87	38.46	57.76	35.65	35.14	57.06	<b>37.72</b>	34.09	55.22	36.96	34.57	55.90

Note: Entries in bold represent the best results.

In contrast to Scenario 1, where the ST\_Dilated\_CNN model achieves the best results for all stations over the rest of the models, in Scenario 2, the greatest FS percentages have values between 27.39 and 37.21 (except for stations P03 and ZA02, where the ST\_CNN\_v2 model delivers a better FS with 27.47 and 37.72, respectively). The LR and RF models achieve higher RMSE values than the proposed models. Although the FS, MAE, and RMSE values obtained by the ST\_CNN\_v1 model demonstrate an imbalanced performance, the ST\_CNN\_v2 and ST\_Dilated\_CNN models stand out for delivering the best metrics. Nonetheless, the ST\_Dilated\_CNN model appears to be capable of capturing data variability.

Each station presents unique conditions that test the limits of the models. For example, station LE05 shows the strengths of the ST\_CNN\_v2 model with the lowest MAE values, indicating its strong predictive accuracy. The two baseline models have a similar behavior, with a difference in some cases of one percentage point. Meanwhile, the ST\_CNN\_v1 model has a difference of six in terms of FS for the SO01 station, which shows that the shallow depth of the network prevents it from capturing patterns that are more complex.

Figure 8 shows how the ST\_CNN\_v2 and the ST\_Dilated\_CNN models stand out with a better forecast skill than the rest of the models for all seasons. These models show a more stable behavior, unlike the baselines that seem to have behaviors linked to the data of each station. The RF model for stations BU04, P03, P07, SG01, and SO01 improves on the other baseline model, while the LR model improves on the other baseline model for station LE05. The ST\_CNN\_v1 model only improves the two baseline models for the ZA02 station. In the P03 and P07 stations, it outperforms LR, but not the RF model.



**Figure 8.** Comparative FS performance of three distinct models proposed—ST\_CNN\_v1, ST\_CNN\_v2, and ST\_Dilated\_CNN—and baseline models across various meteorological stations in Scenario 2.

In this scenario, it can be observed that the ST\_CNN\_v2 model improves the FS results of the ST\_Dilated\_CNN model for station ZA02, while the ST\_Dilated\_CNN model outperforms it for the rest of the stations. In Figure 8, it can also be appreciated that stations LE05 and ZA02 are the stations with the highest FS for all models, with a difference of up to 5 points with respect to the rest of the stations. Stations BU04, P03, P07, and SG01 have similar results, with the ST\_CNN\_v2 and the ST\_Dilated\_CNN models always exceeding the baseline. The atypical case is station SO01, with the lowest FS values for all models.

### 5.3. Scenario 3

This section explores the results related to Scenario 3, which are shown in Table 3. The table shows the FS, MAE, and RMSE metrics for the two baseline models and the three proposed models.

**Table 3.** Results of the baseline and proposed models for the selected stations in Scenario 3.

Metric Station	Linear Regression			Random Forest			ST_CNN_v1			ST_CNN_v2			ST_Dilated_CNN		
	FS (%)	MAE (W/m <sup>2</sup> )	RMSE (W/m <sup>2</sup> )	FS (%)	MAE (W/m <sup>2</sup> )	RMSE (W/m <sup>2</sup> )	FS (%)	MAE (W/m <sup>2</sup> )	RMSE (W/m <sup>2</sup> )	FS (%)	MAE (W/m <sup>2</sup> )	RMSE (W/m <sup>2</sup> )	FS (%)	MAE (W/m <sup>2</sup> )	RMSE (W/m <sup>2</sup> )
BU04	25.59	43.76	69.30	26.26	40.90	68.67	24.99	43.00	69.86	28.50	38.95	66.58	<b>28.90</b>	39.10	66.21
LE05	33.97	35.93	56.86	36.00	32.96	55.11	34.27	34.11	56.60	36.56	30.92	54.63	<b>36.62</b>	31.93	54.57
P03	25.51	43.65	69.81	24.77	42.27	70.51	24.96	41.24	70.33	<b>27.26</b>	39.62	68.17	27.18	40.71	68.25
P07	30.26	46.76	71.10	30.51	42.71	70.85	31.24	43.49	70.11	32.03	41.89	69.30	<b>33.22</b>	42.47	68.09
SG01	26.64	44.74	70.43	25.22	44.53	71.79	27.69	41.89	69.43	28.78	39.74	68.38	<b>28.97</b>	38.65	68.19
SO01	25.67	42.76	62.39	23.27	41.25	64.41	19.34	45.02	67.70	<b>27.72</b>	38.11	60.67	27.70	39.77	60.69
ZA02	35.77	38.74	56.96	35.92	37.80	56.82	35.19	37.69	57.46	<b>38.91</b>	33.19	54.16	38.83	34.76	54.24

Note: Entries in bold represent the best results.

With these results, we can analyze how adding extra weather details affects the performance of the ST\_CNN\_v1, ST\_CNN\_v2, and ST\_Dilated\_CNN models in predicting GHI at different weather stations.

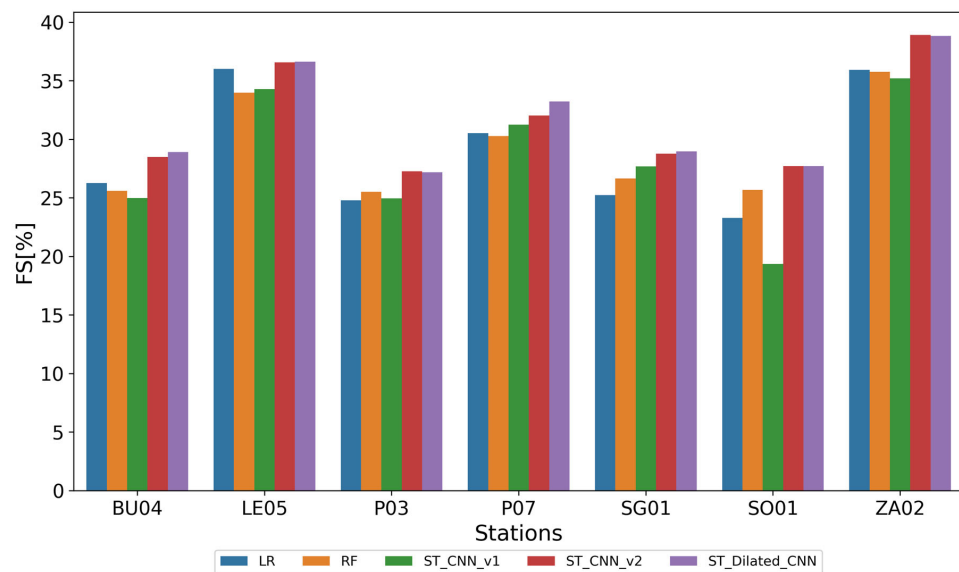
It can also be observed that the LR and RF models have similar results in model forecast skill for stations ZA02 and P07. However, for stations LE05 and B04, RF shows higher values, while LR improves on RF for stations P03, SG01, and SO01. These results show that these models are more closely tied to the variability present at each of the stations evaluated. Nevertheless, it can be concluded that both models have similar predictive capacity for this study.

The ST\_CNN\_v1 model improves the FS of the LR model at stations LE05, P07, and SG01, and the RF model at stations P03, P07, and SG01. The ST\_CNN\_v2 and ST\_Dilated\_CNN models improve on the FS of all baseline models for all seasons. The ST\_Dilated\_CNN model improves on the ST\_CNN\_v1 and ST\_CNN\_v2 models for almost all stations, except for P03 and SO01, on which the ST\_CNN\_v2 model outperforms the others with an FS of 27.76 and 27.72, respectively.

It should be noted that CNN-based models generally perform better than LR and RF models, although there may be changes depending on the specific station. In particular, the ST\_CNN\_v2 and ST\_Dilated\_CNN models show their ability to learn the variability present by improving the FS for all stations.

Figure 9 shows the FS of the models for all the stations studied, and evidences the ability of the ST\_CNN\_v1, ST\_CNN\_v2, and ST\_Dilated\_CNN models to improve the quality of the forecast for all stations, with a significant difference with respect to the baseline models.

In Figure 9, it can be observed that at stations such as ZA02 and LE05, the FS values are generally higher, which could indicate that their weather conditions are more predictable and therefore the models are able to better fit the data measured at the site. At the same time, there are stations such as BU04, P03, and SO01 that generally show lower FS values. These changes may indicate that these sites are subject to greater climatic variability and therefore it may be necessary to better adapt the model parameters to achieve a better result.



**Figure 9.** Comparative FS performance of three distinct models—ST\_CNN\_v1, ST\_CNN\_v2, and ST\_Dilated\_CNN—and baseline models across various meteorological stations in Scenario 3.

The LR, RF, and ST\_CNN\_v1 models show variable behavior dependent on particular stations. The LR model for stations LE05 and BU04 improves on the RF and ST\_CNN\_v1 models, while RF improves on the LR and ST\_CNN\_v1 models for stations P03 and SO01. The ST\_CNN\_v1 model improves on the baselines for stations P07 and SG01.

In this scenario, the ST\_Dilated\_CNN model improves the FS of the ST\_CNN\_v2 model for stations BU04, LE05, P07, and SG01, while the ST\_CNN\_v2 model improves it for the rest of the stations, although the margin of improvement is below one percentage point. Both models improve the FS compared to all other models.

#### 5.4. Comparison between Scenarios

Comparing the results of the three scenarios (each applying a different approach to data input) provides a better understanding of the impact of including variables in the solar forecast. The three scenarios involve the use of different input features: the first scenario exploits only the global horizontal irradiance of the target station and its close neighbors with 48 lags; the second scenario introduces astronomical variables in addition to the initial features; and the third scenario additionally incorporates astronomical and meteorological variables.

In Scenario 1, the modelling of global horizontal irradiance only is simple and computationally more efficient, since it focuses only on historical irradiance data, but there is a limitation of potentially missing certain spatio-temporal nuances and astronomical patterns that contribute to solar irradiance dynamics.

In Scenario 2, when astronomical variables are added, the predictive capability of the models is improved, as demonstrated by higher FS values and lower MAE and RMSE values at several stations. Stations such as SO01, P07, and SG01 benefit notably from the addition of astronomical variables, showing improved accuracy and better model performance.

In Scenario 3, when the modelling combines astronomical and meteorological variables, further improvements in the performance of the models can be found, resulting in reduced MAE and RMSE values, particularly at stations P07 and SO01.

Upon closer examination of the results in the results tables for the three scenarios, it can be observed that the incorporation of astronomical variables produces noticeable impacts on the prediction of solar radiation at different meteorological stations. In particular, the ST\_Dilated\_CNN model shows better performance in FS, MAE, and RMSE in a systematic way when compared to the ST\_CNN\_v1 and ST\_CNN\_v2 models, with different degrees of improvement observed at different stations.

The station that shows a distinct behavior is station SO01, which does not clearly benefit from the inclusion of the astronomical and meteorological variables. The ST\_Dilated\_CNN model obtains the best FS results (28.32%) for Scenario 1 and reduced MAE and RMSE values, which underlines its ability to capture the solar radiation dynamics at this location.

The P07 and SG01 stations also experience notable benefits from the integration of astronomical variables, as evidenced by consistently elevated FS values and decreased MAE and RMSE metrics. The ST\_Dilated\_CNN's performance surpasses its counterparts, accentuating its proficiency in leveraging astronomical features for improved predictions.

Stations with marginal benefit are BU04, LE05, P03, and ZA02; while still benefiting from the inclusion of astronomical variables, these stations exhibit more modest improvements. ST\_Dilated\_CNN consistently outperforms ST\_CNN\_v1 and ST\_CNN\_v2, but the differences in FS, MAE, and RMSE are less pronounced compared to the aforementioned stations.

When analyzing the ST\_Dilated\_CNN model for the three scenarios, it is found that the FS values of Scenarios 2 and 3 improve with respect to Scenario 1, with differences of one to two percentage points (Appendices A and B can be consulted for a visual representation of the delta FS values). The ST\_CNN\_v2 model is the only one that benefits in all stations for all scenarios. The ST\_Dilated\_CNN model, although it also benefits for stations SO01 and ZA02, shows different behavior in Scenario 2, and for station SO01 in Scenario 3. It is estimated that this different behavior is due to the fact that these two stations have the most spatially distant neighbors of all the stations studied, as can be seen in Figure 1. Therefore, the input of data from neighbors can create noise, reducing the network's ability to extract valid information. On the other hand, ST\_CNN\_v1 frequently appears as the worst model in terms of highest MAE and RMSE. This suggests that, for the scenarios evaluated, ST\_CNN\_v1 may not be as effective compared to the other models evaluated. By comparing Scenarios 2 and 3, it can be found that, for four of the seven stations studied, higher results are achieved in Scenario 2, indicating a significant influence of astronomical variables in the prediction of the proposed models.

##### 5.5. On the Variability of Model Performance when Grouped by the Considered Scenarios

The results of the evaluation using the subsampling method proposed in Section 4.5 are presented here. Tables 4–6 show the means and standard deviations of the metrics obtained for each of the stations, across all scenarios and models. In Scenario 1 (Table 4), the ST\_CNN\_v2 model performs consistently better for most stations, showing low MAE and RMSE values and high R2 score values. The ST\_CNN\_v1 model generally performs slightly worse, with higher MAE and RMSE values and lower R2 score values compared to the other models. Although the ST\_Dilated\_CNN model differs between 0.01 and 0.04 in mean R2 scores with respect to ST\_CNN\_v2, it maintains a consistent performance.

**Table 4.** Mean and standard deviation values of the performance metrics obtained in Scenario 1.

Station	Model	MAE		RMSE		R2 Score	
		Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation
BU04	ST_CNN_v1	73.95	11.69	106.98	19.43	0.8	0.03
	ST_CNN_v2	41.03	7.42	66.54	13.19	0.92	0.02
	ST_Dilated_CNN	62.11	8.9	82.07	14.82	0.88	0.02
LE05	ST_CNN_v1	62.68	1.87	89.1	2.38	0.87	0.04
	ST_CNN_v2	38.26	1.52	60.65	4.0	0.94	0.01
	ST_Dilated_CNN	48.42	6.65	65.58	7.7	0.93	0
P03	ST_CNN_v1	61.96	7.37	93.68	14.24	0.86	0.02
	ST_CNN_v2	47.6	5.75	71.46	10.15	0.92	0.01
	ST_Dilated_CNN	74.54	8.77	95.01	13.83	0.85	0.01

Table 4. Cont.

Station	Model	MAE		RMSE		R2 Score	
		Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation
P07	ST_CNN_v1	80.77	11.52	109.86	17.75	0.77	0.02
	ST_CNN_v2	55.31	10.83	80.5	18.29	0.88	0.02
	ST_Dilated_CNN	62.95	14.72	88.86	20.95	0.85	0.03
SG01	ST_CNN_v1	66.99	9.43	98.06	15.55	0.86	0.02
	ST_CNN_v2	49.12	7.7	73.68	12.32	0.92	0.01
	ST_Dilated_CNN	56.24	6.61	76.46	11.19	0.91	0.01
SO01	ST_CNN_v1	128.94	17.55	168.98	26.34	0.53	0.13
	ST_CNN_v2	104.02	18.54	138.56	22.69	0.69	0.07
	ST_Dilated_CNN	66.79	4.58	82.3	7.57	0.89	0.01
ZA02	ST_CNN_v1	64.78	1.76	89.41	4.46	0.88	0.02
	ST_CNN_v2	45.58	4.31	61.60	7.09	0.94	0
	ST_Dilated_CNN	49.57	7.53	68.71	11.08	0.93	0.01

In Scenario 2 (Table 5), the response of the models remains the same, with ST\_CNN\_V2 and ST\_Dilated\_CNN having the best metrics. For example, at station BU04, ST\_CNN\_v2 has an MAE of  $50.28 \pm 7.04$ , an RMSE of  $75.91 \pm 11.97$ , and an R2 score of  $0.90 \pm 0.02$ , outperforming ST\_CNN\_v1 and ST\_Dilated\_CNN on all metrics. These results are consistent with those presented in Section 5.2.

Table 5. Mean and standard deviation values of the performance metrics obtained in Scenario 2.

Station	Model	MAE		RMSE		R2 Score	
		Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation
BU04	ST_CNN_v1	82.8	6.61	111.09	14.38	0.78	0.04
	ST_CNN_v2	50.28	7.04	75.91	11.97	0.9	0.02
	ST_Dilated_CNN	69.98	7.3	87.57	9.8	0.86	0.02
LE05	ST_CNN_v1	60.06	2.78	87.39	1.65	0.87	0.05
	ST_CNN_v2	44.22	1.95	68.4	3.12	0.92	0.02
	ST_Dilated_CNN	49.63	6.98	67.85	7.88	0.93	0
P03	ST_CNN_v1	58.49	7.07	89.56	13.16	0.87	0.02
	ST_CNN_v2	46.71	7.24	72.12	12.06	0.92	0.01
	ST_Dilated_CNN	73.1	11.05	93.03	14.39	0.86	0.01
P07	ST_CNN_v1	83.86	11.25	113.29	15.07	0.76	0.03
	ST_CNN_v2	60.02	13.27	86.8	20.44	0.86	0.02
	ST_Dilated_CNN	86.33	17.37	115.91	25.94	0.75	0.04
SG01	ST_CNN_v1	75.44	5.68	104.37	13.34	0.84	0.02
	ST_CNN_v2	48.14	5.02	73.57	9.46	0.92	0.01
	ST_Dilated_CNN	72.44	12.53	96.35	18.53	0.86	0.02
SO01	ST_CNN_v1	115.87	19.15	154.87	28.11	0.6	0.16
	ST_CNN_v2	88.24	25.99	134.97	42	0.69	0.16
	ST_Dilated_CNN	56.39	9.33	77.56	10.91	0.9	0.01
ZA02	ST_CNN_v1	89.93	3.17	115.49	4.42	0.81	0.04
	ST_CNN_v2	71.25	3.32	93.49	8.60	0.86	0.01
	ST_Dilated_CNN	58.47	6.42	74.22	8.81	0.92	0.03

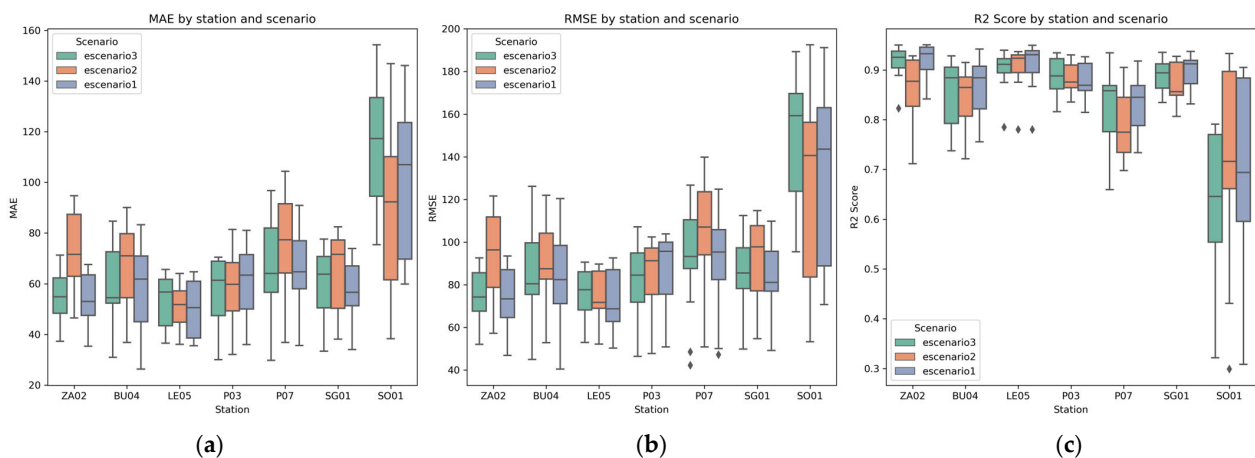
For Scenario 3, as shown in Table 6, the ST\_CNN\_v2 model stands out as the best in most of the stations evaluated. For stations such as LE05 and P03, ST\_CNN\_v2 consistently exhibits the lowest MAE and RMSE, as well as the highest R2 score compared to the other models, indicating a better fit and predictive performance. In contrast, ST\_CNN\_v1 tends

to show the worst performance at several stations, such as P07 and SO01, where it has the highest MAE and RMSE values, and the lowest R2 score values.

**Table 6.** Mean and standard deviation values of the performance metrics obtained in Scenario 3.

Station	Model	MAE		RMSE		R2 Score	
		Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation
BU04	ST_CNN_v1	77.39	8.82	111.99	18.56	0.78	0.02
	ST_CNN_v2	53.07	5.82	78.64	11.76	0.89	0.02
	ST_Dilated_CNN	47.48	8.51	71.25	13.49	0.91	0.02
LE05	ST_CNN_v1	63.5	1.82	88.03	1.88	0.87	0.04
	ST_CNN_v2	42.86	2.21	67.09	3.47	0.93	0.02
	ST_Dilated_CNN	54.12	9.14	74.06	10.72	0.91	0.0
P03	ST_CNN_v1	66.63	5.4	97.15	12.37	0.85	0.02
	ST_CNN_v2	44.37	7.08	67.69	10.5	0.93	0.01
	ST_Dilated_CNN	61.15	11.22	82.74	14.37	0.89	0.01
P07	ST_CNN_v1	88.54	8.41	115.57	11.88	0.74	0.05
	ST_CNN_v2	52.87	12.87	83.17	22.72	0.87	0.03
	ST_Dilated_CNN	56.63	12.2	82.03	18.16	0.87	0.02
SG01	ST_CNN_v1	71.42	8.92	100.68	15.63	0.85	0.01
	ST_CNN_v2	47.51	7.01	74.9	12.6	0.92	0.01
	ST_Dilated_CNN	63.62	7.97	83.89	13.08	0.9	0.01
SO01	ST_CNN_v1	119.83	20.57	162.58	32.13	0.57	0.15
	ST_CNN_v2	133.11	15.09	165.49	12.85	0.54	0.14
	ST_Dilated_CNN	89.27	9.33	117.33	13.61	0.78	0.01
ZA02	ST_CNN_v1	65.55	3.78	88.21	3.07	0.89	0.03
	ST_CNN_v2	54.52	2.39	73.37	4.07	0.92	0.01
	ST_Dilated_CNN	44.53	4.44	64.14	6.55	0.94	0.01

In summary, these new experiments have shown that the ST\_CNN\_v2 and ST\_Dilated\_CNN models are consistent models, with the ST\_CNN\_v2 model being the best of the two, although the difference between them is not significant. As for the variability in the stations, the SO01 station is the one with the greatest variability, which is attributed to the fact that its neighbors are the furthest away, so the weight of the information they provide should be less. It is recommended to carry out more experimental repetitions to further validate these findings. The performance results analyzed in this section are also illustrated in the form of boxplots in Figure 10.



**Figure 10.** Boxplots of the performance metrics achieved by the proposed models when grouped by the scenarios considered in terms of (a) MAE, (b) RMSE, and (c) R2 score values computed.

### 5.6. On the Statistical Validation of the Models

The evaluation results of the three scenarios using the MAE, RMSE, and R2 score metrics, accompanied by their respective F-statistic and  $p$ -value values, are shown in Table 7. In Scenario 1, it is observed that the F-statistic for MAE is 11.43 with a  $p$ -value of  $3.23 \times 10^{-5}$ , indicating a significant difference in the mean absolute errors between the models evaluated. Similarly, for RMSE and R2 score, the F-statistic values are 14.50 and 14.13 with  $p$ -values of  $2.75 \times 10^{-6}$  and  $3.67 \times 10^{-6}$ , respectively, suggesting that the differences in the mean square errors and coefficients of determination between the models are highly significant.

**Table 7.** F-statistics and  $p$ -values computed by applying the ANOVA test to the considered performance metrics, grouped by scenario.

Scenario	Performance Metric	F-Statistic	$p$ -Value
Scenario 1	MAE	11.43	$3.23 \times 10^{-5}$
	RMSE	14.50	$2.75 \times 10^{-6}$
	R2 score	14.13	$3.67 \times 10^{-6}$
Scenario 2	MAE	13.18	$7.81 \times 10^{-6}$
	RMSE	9.02	0.0002
	R2 score	8.15	0.0005
Scenario 3	MAE	7.03	0.0014
	RMSE	8.87	0.0003
	R2 score	7.42	0.0010

In Scenario 2, the results show an F-statistic of 13.18 for MAE. The values for RMSE and R2 score are 9.02 and 8.15, with  $p$ -values less than 0.05 for all. These results indicate significant differences in the error metrics and explanatory power of the models evaluated.

Finally, in Scenario 3, the F-statistic values for MAE, RMSE, and R2 score are 7.03, 8.87, and 7.42, with  $p$ -values also below 0.05. Although these values are lower compared to the first two scenarios, they are still statistically significant, implying that the observed differences in the error metrics and the coefficients of determination between the models are not attributable to chance.

In summary, the F-statistic values and  $p$ -values in all scenarios and for all metrics indicate that the observed differences between the models are highly statistically significant. This suggests that the evaluated models perform significantly differently in terms of MAE, RMSE, and R2 score, which is crucial for selecting the most appropriate model for each specific scenario. The  $p$ -value also indicates that the improvements in the performance achieved by the proposed models are highly statistically significant.

## 6. Conclusions

In this study, three new deep learning architectures that are applicable to the field of solar forecast have been presented, namely the spatio-temporal dilated convolutional neural network (ST\_Dilated\_CNN) and spatio-temporal convolutional neural network versions 1 and 2 (ST\_CNN\_v1 and ST\_CNN\_v2). These artificial neural networks were evaluated against three scenarios with different sets of input features, including irradiance data, astronomical variables, and a combination of astronomical and meteorological features.

For each of the scenarios evaluated, the proposed networks were proven to be superior to the baseline models, outperforming them by up to 3 percentage points in some cases. A dataset available to the scientific community was used for the study. It was a dataset from the Spanish region of Castile and Leon that has records from 37 meteorological stations. It has a temporal resolution of 30 min and covers a region of up to 1000 km<sup>2</sup>.

The results show that the best combinations of input characteristics for the models included astronomical variables. Scenarios 2 and 3 benefited the most when assessing the forecast skill of the models, achieving significantly better FS values than Scenario 1, where only irradiance data were modelled.

The  $p$ -values computed with the ANOVA test were highly statistically significant (lower than 0.001) for all metrics and all the evaluated scenarios, suggesting that there was a significant difference in performance. This reinforces the validity of the improvements in performance achieved by the proposed models, and suggests that the models evaluated performed significantly differently in terms of MAE, RMSE, and R2 score.

In terms of future work, the evaluation of the models with public datasets from other regions is proposed to study if changes in location (and, therefore, atmospheric conditions) affect the prediction quality of the models. Additionally, variables of physical models, such as atmospheric visibility, 200 hpa geopotential, cloud liquid condensation (CLC), and cloud ice condensation (CIC), could also be added to the training and evaluation of the proposed models. Subsequent works could also explore the application of feature selection algorithms to reduce the dimensionality of the input and ultimately minimize the computational cost of training the models. Another line of research may be associated with finding the distance at which neighbor data are no longer correlated, and therefore create noise instead of providing relevant information.

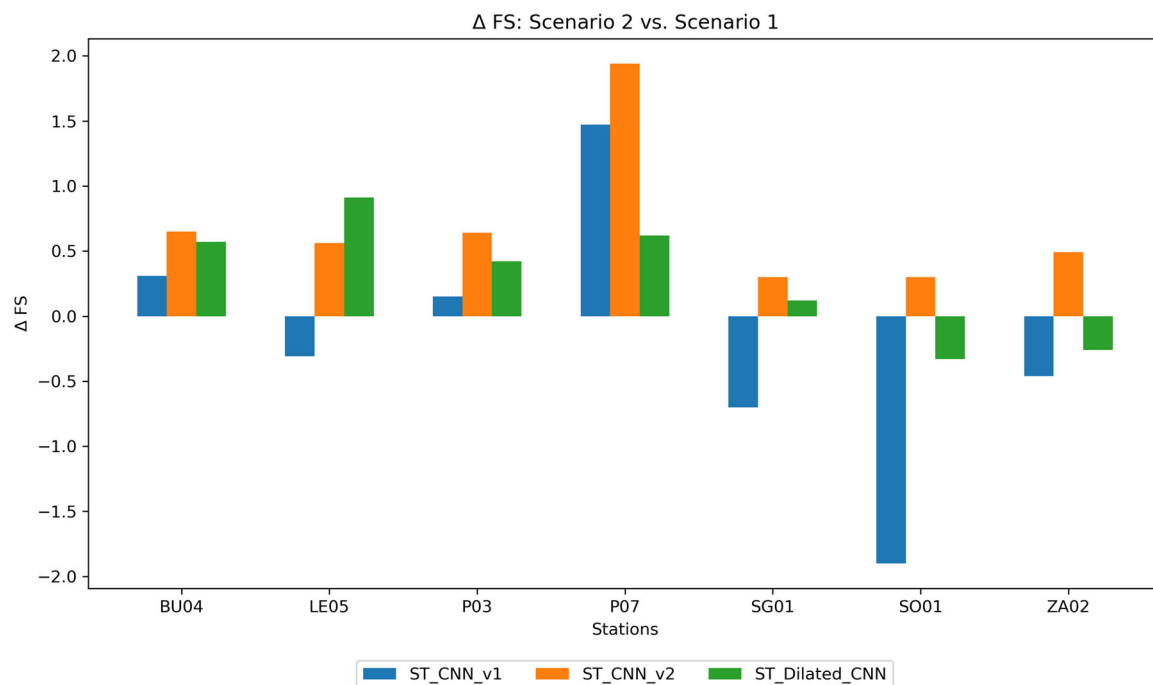
**Author Contributions:** Conceptualization, L.B.C., C.-I.C. and M.-Á.M.-C.; methodology, L.B.C. and M.-Á.M.-C.; software, L.B.C.; validation, L.B.C., M.-Á.M.-C. and C.-I.C.; investigation, L.B.C.; resources, M.-Á.M.-C. and C.-I.C.; data curation, L.B.C.; writing—original draft preparation, L.B.C.; writing—review and editing, L.B.C., M.-Á.M.-C. and C.-I.C.; visualization, L.B.C.; supervision, M.-Á.M.-C. and C.-I.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** The CyL-GHI dataset used in this study is openly available in the Zenodo repository (<https://doi.org/10.5281/zenodo.7404167>).

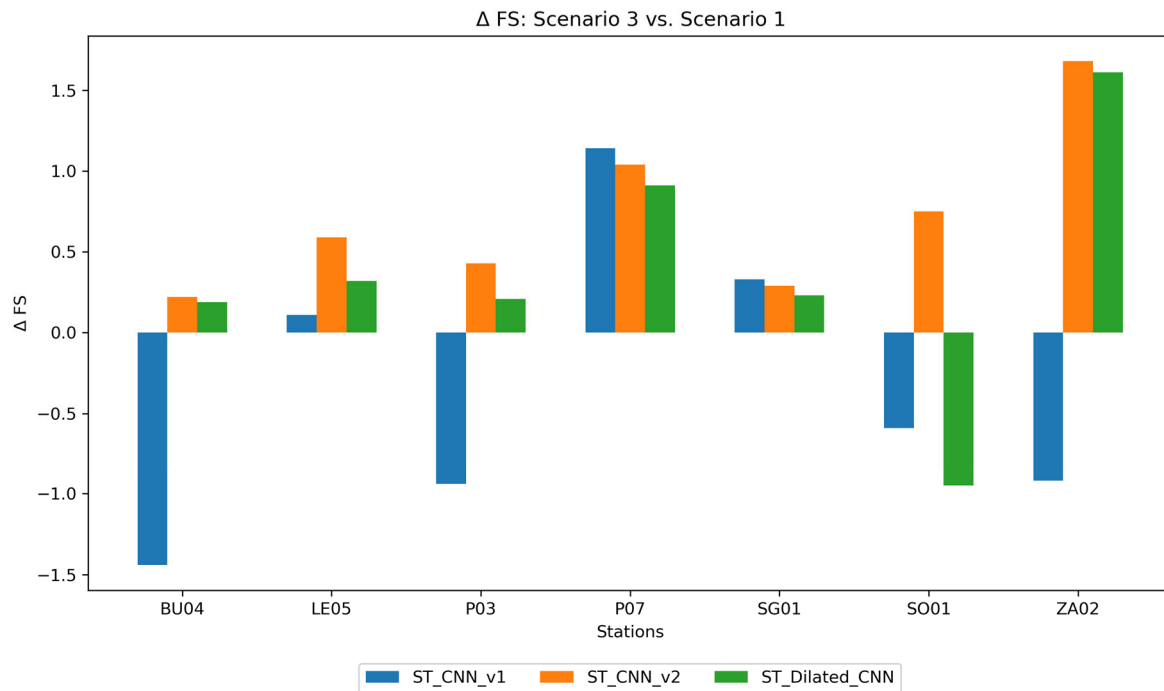
**Conflicts of Interest:** The authors declare no conflicts of interest.

## Appendix A



**Figure A1.** Impact on forecast skill: A comparison between Scenario 2 and Scenario 1.

## Appendix B



**Figure A2.** Impact on forecast skill: A comparison between Scenario 3 and Scenario 1.

## References

- BOE. Resolución de 30 de Diciembre de 2020, de la Dirección General de Calidad y Evaluación Ambiental, por la Que se Formula la Declaración Ambiental Estratégica del Plan Nacional Integrado de Energía y Clima 2021–2030. Madrid, 2021. Available online: [https://www.boe.es/diario\\_boe/txt.php?id=BOE-A-2021-421](https://www.boe.es/diario_boe/txt.php?id=BOE-A-2021-421) (accessed on 16 June 2024).
- Gandhi, O.; Zhang, W.; Kumar, D.S.; Rodríguez-Gallegos, C.D.; Yagli, G.M.; Yang, D.; Reindl, T.; Srinivasan, D. The value of solar forecasts and the cost of their errors: A review. *Renew. Sustain. Energy Rev.* **2024**, *189*, 113915. [\[CrossRef\]](#)
- Yang, D.; Wang, W.; Gueymard, C.A.; Hong, T.; Kleissl, J.; Huang, J.; Perez, M.J.; Perez, R.; Bright, J.M.; Xia, X.; et al. A review of solar forecasting, its dependence on atmospheric sciences and implications for grid integration: Towards carbon neutrality. *Renew. Sustain. Energy Rev.* **2022**, *161*, 112348. [\[CrossRef\]](#)
- Cesar, L.B.; e Silva, R.A.; Callejo, M.M.; Cira, C.-I. Review on Spatio-Temporal Solar Forecasting Methods Driven by In Situ Measurements or Their Combination with Satellite and Numerical Weather Prediction (NWP) Estimates. *Energies* **2022**, *15*, 4341. [\[CrossRef\]](#)
- Zhou, Y.; Liu, Y.; Wang, D.; Liu, X.; Wang, Y. A review on global solar radiation prediction with machine learning models in a comprehensive perspective. *Energy Convers. Manag.* **2021**, *235*, 113960. [\[CrossRef\]](#)
- Feng, C.; Zhang, J. SolarNet: A deep convolutional neural network for solar forecasting via sky images. In Proceedings of the 2020 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), Washington, DC, USA, 17–20 February 2020.
- Le Guen, V.; Thome, N. A Deep Physical Model for Solar Irradiance Forecasting with Fisheye Images. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, 14–19 June 2020; pp. 2685–2688.
- Simeunovic, J.; Schubnel, B.; Alet, P.-J.; Carrillo, R.E. Spatio-Temporal Graph Neural Networks for Multi-Site PV Power Forecasting. *IEEE Trans. Sustain. Energy* **2021**, *13*, 1210–1220. [\[CrossRef\]](#)
- Sun, Y.C.; Venugopal, V.; Brandt, A.R. Short-term solar power forecast with deep learning: Exploring optimal input and output configuration. *Sol. Energy* **2019**, *188*, 730–741. [\[CrossRef\]](#)
- Ziyabari, S.; Du, L.; Biswas, S. A Spatio-temporal Hybrid Deep Learning Architecture for Short-term Solar Irradiance Forecasting. In Proceedings of the 2020 47th IEEE Photovoltaic Specialists Conference (PVSC), Calgary, AB, Canada, 15 June–21 August 2020; pp. 0833–0838.
- Suresh, V.; Janik, P.; Rezmer, J.; Leonowicz, Z. Forecasting Solar PV Output Using Convolutional Neural Networks with a Sliding Window Algorithm. *Energies* **2020**, *13*, 723. [\[CrossRef\]](#)
- Ruan, Z.; Sun, W.; Yuan, Y.; Tan, H. Accurately forecasting solar radiation distribution at both spatial and temporal dimensions simultaneously with fully-convolutional deep neural network model. *Renew. Sustain. Energy Rev.* **2023**, *184*, 113528. [\[CrossRef\]](#)
- Ziyabari, S.; Du, L.; Biswas, S.K. Multibranch Attentive Gated ResNet for Short-Term Spatio-Temporal Solar Irradiance Forecasting. *IEEE Trans. Ind. Appl.* **2021**, *58*, 28–38. [\[CrossRef\]](#)
- Yu, F.; Koltun, V. Multi-Scale Context Aggregation by Dilated Convolutions. *arXiv* **2015**, arXiv:1511.07122.

15. Mishra, K.; Basu, S.; Maulik, U. A Dilated Convolutional Based Model for Time Series Forecasting. *SN Comput. Sci.* **2021**, *2*, 1–11. [CrossRef]
16. Ma, Y.-J.; Shuai, H.-H.; Cheng, W.-H. Spatiotemporal Dilated Convolution With Uncertain Matching for Video-Based Crowd Estimation. *IEEE Trans. Multimed.* **2021**, *24*, 261–273. [CrossRef]
17. Pham, N.T.; Dang, D.N.M.; Nguyen, N.D.; Nguyen, T.T.; Nguyen, H.; Manavalan, B.; Lim, C.P.; Nguyen, S.D. Hybrid data augmentation and deep attention-based dilated convolutional-recurrent neural networks for speech emotion recognition. *Expert Syst. Appl.* **2023**, *230*, 120608. [CrossRef]
18. Salehi, A.; Balasubramanian, M. DDCNet: Deep dilated convolutional neural network for dense prediction. *Neurocomputing* **2023**, *523*, 116–129. [CrossRef] [PubMed]
19. Contreras, J.; Ceberio, M.; Kreinovich, V. Why Dilated Convolutional Neural Networks: A Proof of Their Optimality. *Entropy* **2021**, *23*, 767. [CrossRef] [PubMed]
20. Chen, Y.; Xie, Z. Multi-channel fusion graph neural network for multivariate time series forecasting. *J. Comput. Sci.* **2022**, *64*, 101862. [CrossRef]
21. Liang, J.; Tang, W. Ultra-Short-Term Spatiotemporal Forecasting of Renewable Resources: An Attention Temporal Convolutional Network-Based Approach. *IEEE Trans. Smart Grid* **2022**, *13*, 3798–3812. [CrossRef]
22. Fan, T.; Sun, T.; Liu, H.; Xie, X.; Na, Z. Spatial-Temporal Genetic-Based Attention Networks for Short-Term Photovoltaic Power Forecasting. *IEEE Access* **2021**, *9*, 138762–138774. [CrossRef]
23. Rajagukguk, R.A.; Ramadhan, R.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]
24. Paletta, Q.; Terrén-Serrano, G.; Nie, Y.; Li, B.; Bieker, J.; Zhang, W.; Dubus, L.; Dev, S.; Feng, C. Advances in solar forecasting: Computer vision with deep learning. *Adv. Appl. Energy* **2023**, *11*, 100150. [CrossRef]
25. Cesar, L.B.; Callejo, M.M.; Cira, C.-I.; Alcarria, R. CyL-GHI: Global Horizontal Irradiance Dataset Containing 18 Years of Refined Data at 30-Min Granularity from 37 Stations Located in Castile and León (Spain). *Data* **2023**, *8*, 65. [CrossRef]
26. Cesar, L.B.; Callejo, M.Á.M.; Cira, C.-I.; Garrido, R.P.A. CyL\_GHI. Zenodo, 2022. Available online: <https://zenodo.org/doi/10.5281/zenodo.7404166> (accessed on 6 June 2023).
27. Eschenbach, A.; Yepes, G.; Tenllado, C.; Gomez-Perez, J.I.; Pinuel, L.; Zarzalejo, L.F.; Wilbert, S. Spatio-Temporal Resolution of Irradiance Samples in Machine Learning Approaches for Irradiance Forecasting. *IEEE Access* **2020**, *8*, 51518–51531. [CrossRef]
28. Gutierrez-Corea, F.-V.; Manso-Callejo, M.-A.; Moreno-Regidor, M.-P.; Manrique-Sancho, M.-T. Forecasting short-term solar irradiance based on artificial neural networks and data from neighboring meteorological stations. *Sol. Energy* **2016**, *134*, 119–131. [CrossRef]
29. Borovykh, A.; Bohte, S.; Oosterlee, C.W. Dilated convolutional neural networks for time series forecasting. *J. Comput. Financ.* **2018**, *22*, 73–101. [CrossRef]
30. He, Z.; Zhao, C.; Huang, Y. Multivariate Time Series Deep Spatiotemporal Forecasting with Graph Neural Network. *Appl. Sci.* **2022**, *12*, 5731. [CrossRef]
31. Bai, S.; Kolter, J.Z.; Koltun, V. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. *arXiv* **2018**, arXiv:1803.01271.
32. Sutskever, I.; Vinyals, O.; Le, Q.V. Sequence to Sequence Learning with Neural Networks. *arXiv* **2014**, arXiv:1409.3215.
33. Nauta, M.; Bucur, D.; Seifert, C. Causal Discovery with Attention-Based Convolutional Neural Networks. *Mach. Learn. Knowl. Extr.* **2019**, *1*, 312–340. [CrossRef]
34. Chollet, F. Keras. GitHub. 2015. Available online: <https://github.com/fchollet/keras> (accessed on 6 June 2023).
35. Nair, V.; Hinton, G.E. Rectified Linear Units Improve Restricted Boltzmann Machines. In Proceedings of the 27th International Conference on Machine Learning (ICML-10), Haifa, Israel, 21–24 June 2010; pp. 807–814.
36. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization, CoRR. *arXiv* **2014**, arXiv:1412.6980v9.
37. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
38. Silva, R.A.E.; da Silva, L.C.C.T.; Brito, M.C. Support vector regression for spatio-temporal PV forecasting PV variability The need for PV forecasting. In Proceedings of the 35th EUPVSEC 2018, Brussels, Belgium, 24–28 September 2018. [CrossRef]
39. Dambreville, R.; Blanc, P.; Chanussot, J.; Boldo, D. Very short term forecasting of the Global Horizontal Irradiance using a spatio-temporal autoregressive model. *Renew. Energy* **2014**, *72*, 291–300. [CrossRef]
40. Boland, J. Spatial-temporal forecasting of solar radiation. *Renew. Energy* **2015**, *75*, 607–616. [CrossRef]
41. 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]
42. Ghimire, S.; Deo, R.C.; Raj, N.; Mi, J. Deep Learning Neural Networks Trained with MODIS Satellite-Derived Predictors for Long-Term Global Solar Radiation Prediction. *Energies* **2019**, *12*, 2407. [CrossRef]
43. Park, J.; Moon, J.; Jung, S.; Hwang, E. Multistep-Ahead Solar Radiation Forecasting Scheme Based on the Light Gradient Boosting Machine: A Case Study of Jeju Island. *Remote Sens.* **2020**, *12*, 2271. [CrossRef]

- 
44. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Duchesnay, É. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830. [[CrossRef](#)]
  45. Härdle, W.; Horowitz, J.; Kreiss, J. Bootstrap Methods for Time Series. *Int. Stat. Rev.* **2003**, *71*, 435–459. [[CrossRef](#)]

**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.