

Article

Application of Temporal Fusion Transformer for Day-Ahead PV Power Forecasting

Miguel López Santos, Xela García-Santiago , Fernando Echevarría Camarero, Gonzalo Blázquez Gil and Pablo Carrasco Ortega *

Galicia Institute of Technology (ITG), 15003 A Coruña, Spain; mlopez@itg.es (M.L.S.); xgarcia@itg.es (X.G.-S.); fechevarria@itg.es (F.E.C.); gblazquez@itg.es (G.B.G.)

* Correspondence: pcarrasco@itg.es; Tel.: +34-981-173-206

Abstract: The energy generated by a solar photovoltaic (PV) system depends on uncontrollable factors, including weather conditions and solar irradiation, which leads to uncertainty in the power output. Forecast PV power generation is vital to improve grid stability and balance the energy supply and demand. This study aims to predict hourly day-ahead PV power generation by applying Temporal Fusion Transformer (TFT), a new attention-based architecture that incorporates an interpretable explanation of temporal dynamics and high-performance forecasting over multiple horizons. The proposed forecasting model has been trained and tested using data from six different facilities located in Germany and Australia. The results have been compared with other algorithms like Auto Regressive Integrated Moving Average (ARIMA), Long Short-Term Memory (LSTM), Multi-Layer Perceptron (MLP), and Extreme Gradient Boosting (XGBoost), using statistical error indicators. The use of TFT has been shown to be more accurate than the rest of the algorithms to forecast PV generation in the aforementioned facilities.

Keywords: photovoltaic power forecast; solar energy; Temporal Fusion Transformer; deep learning; artificial intelligence



Citation: López Santos, M.; García-Santiago, X.; Echevarría Camarero, F.; Blázquez Gil, G.; Carrasco Ortega, P. Application of Temporal Fusion Transformer for Day-Ahead PV Power Forecasting. *Energies* **2022**, *15*, 5232. <https://doi.org/10.3390/en15145232>

Academic Editors: Fouzi Harrou, Ying Sun, Bilal Taghezouit and Dairi Abdelkader

Received: 15 June 2022

Accepted: 16 July 2022

Published: 19 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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

Renewable energy is rapidly increasing worldwide since it became an economical alternative to conventional energy sources such as fossils fuels, which are responsible for greenhouse emissions (GHG), have a limited supply, and their prices are becoming increasingly unpredictable. Furthermore, the European Commission targets aim to reduce emissions by 50% to 55% by 2030 in comparison with 1990 and to become climate neutral in 2050 [1].

Electricity generated by solar photovoltaic (PV) systems has increased by 23% in 2020 to reach 821 TWh worldwide [2]. Projections indicate that the global installed capacity of this technology could rise by more than three times, reaching 2.840 GW in 2030, and 8.519 GW in 2050 [3]. The reason is solar energy presents many advantages: it is an environmentally friendly renewable source, is abundant, has a long service life [4], and as shown in Figure 1, it is becoming one of the most competitive power generation technologies after more than ten years of cost declines with a promising future ahead.

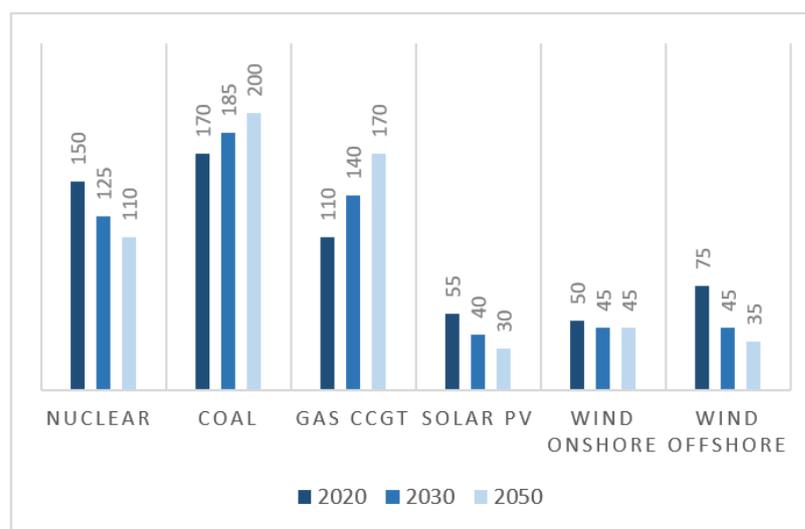


Figure 1. Levelized cost of electricity (USD/kWh) in European Union in the Stated Policies Scenario [5].

However, variations in solar irradiance and meteorological conditions cause fluctuation in solar power generation and lead to uncertainty in power output. This results in a power imbalance between the demand and the supply side of the grid [6]. Also, the unpredictable output significantly impacts the economic dispatch and the scheduling, stability, and reliability of the power system operation [7]. Precise forecasting of the energy produced is essential for grid operators, considering that deviations need to be compensated by the remaining generation technologies. Moreover, it is not only beneficial for system operators as PV plant managers can prevent potential penalties arising from differences between produced and predicted energy [8,9]. Therefore, PV power forecasting contributes to stabilize and optimize the operation and performance of the grid and renewable energy microgrids [10–13] by means of the comparative analysis between required and predicted energy. It also positively affects the customer by reducing the uncertainty and the generated energy cost [8,14].

Solar power forecasting ranges from ultra-short term to long-term forecasts. Long-term forecasts are especially relevant for planning the long duration power system. However, industry and researchers currently focus primarily on short-term or day-ahead predictions as these can account for cloud cover variability and provide more accurate results [15]. There are several approaches for solar power forecasting: physical models [16], persistence forecast [17], statistical models [18], artificial intelligence [19], including machine learning (ML) and deep learning (DL) [20,21], ensemble and hybrid-based models [14,22–26]. Common ML methods applied to PV power forecasting in literature are artificial neural networks (ANN) [20,27,28], support vector machine (SVM) [29,30] and extreme learning machine (ELM) [31,32]. Among all PV forecasting approaches, ANN have shown superior accuracy, being successfully applied in short-term forecast horizons [33–36]. In addition, machine learning methods, like ANN, can handle large amounts of data and produce accurate predictions for short-term periods, without the complexity of mathematical and physical relationships.

Some methods used for PV power forecasting present certain limitations. Persistence and statistical methods cannot handle nonlinear data. Physical methods depend on local meteorological data that can have limited access. Numerical weather prediction (NWP) models are widely used to provide forecasts of weather conditions, but they often lack sufficient spatial and temporal resolution. ANN are involved with problems of local minima, overfitting, and complex structure. DL methods have been developed to solve these limitations [37,38]. DL is an advanced branch of machine learning with the ability to process non-linear and complex relationships between various inputs and the forecasted output.

Examples of DL are Convolutional Neural Networks (CNN) [39–41], Recurrent Neural Networks (RNN) [42,43], Long Short-Term Memory (LSTM) [21,44] or Convolutional Self-Attention LSTM [45]. These methods have proven to achieve better prediction results when applied to solar power forecasting compared to other machine learning methods, especially those based on LSTM networks [46–49].

Time series can have seasonal patterns or increasing and decreasing trends. LSTM can learn to remember useful information from several time steps back that is relevant for producing the predicted output. Traditional recurrent networks do not have the ability to decide which information to remember. Nevertheless, LSTM has difficulty in determining long-term dependencies.

More complex deep learning architectures are constantly created that outperform previous ones. For example, Temporal Fusion Transformer (TFT) [50] is an Attention-Based Deep Neural Network for multi-horizon time series forecasting that integrates LSTM in its architecture. TFT also uses the attention mechanism to learn long-term dependencies in time series data, which means that it can directly learn patterns during training. This makes the TFT more advantageous than other time series methods in terms of interpretability. Moreover, TFT minimizes a quantile loss function, which enables it to generate a probabilistic forecast with a confidence interval.

This work aims to predict hourly day-ahead PV power generation by applying TFT as the forecasting method. TFT incorporates an interpretable explanation of temporal dynamics and high-performance forecasting over multiple horizons. It uses specialized components to choose important attributes and a series of gating layers to remove non-essential elements, resulting in high performance in a wide range of applications.

This paper contributes to extending the application of innovative DL methods to PV power forecasting. The novel aspect of this work is the use of TFT for PV power generation forecasting as there is no previous literature found that applies and compares this method for predicting PV output performance. TFT potentially enhances the accuracy of forecasts compared to other methods by learning short and long-term temporal relationships. This not only benefits the management of PV production systems but can also influence the stability and operation of the power system.

The predicting method is evaluated with common datasets used in literature. They include real data from six photovoltaic systems located in Germany and Australia. The results of the proposed DL method are compared with the following methods: Auto Regressive Integrated Moving Average (ARIMA), Multi-Layer Perceptron (MLP), LSTM, and Extreme Gradient Boosting (XGBoost).

2. Materials and Methods

The methodology applied in this study includes the following steps. (1) Data gathering, where data from different sources is collected to define the final dataset. The electrical energy produced by a photovoltaic system depends on variables such as horizontal irradiation, temperature, humidity, and solar zenith and azimuth, among others. (2) Data pre-processing, to transform raw data into a form more suitable for modeling, including data cleaning, feature selection, and data transformation (3) Model selection. (4) Training and tuning, where the selected model is trained for different combinations of hyperparameters, selecting the model with the best performance. (5) Results evaluation.

Figure 2 summarizes the process carried out.

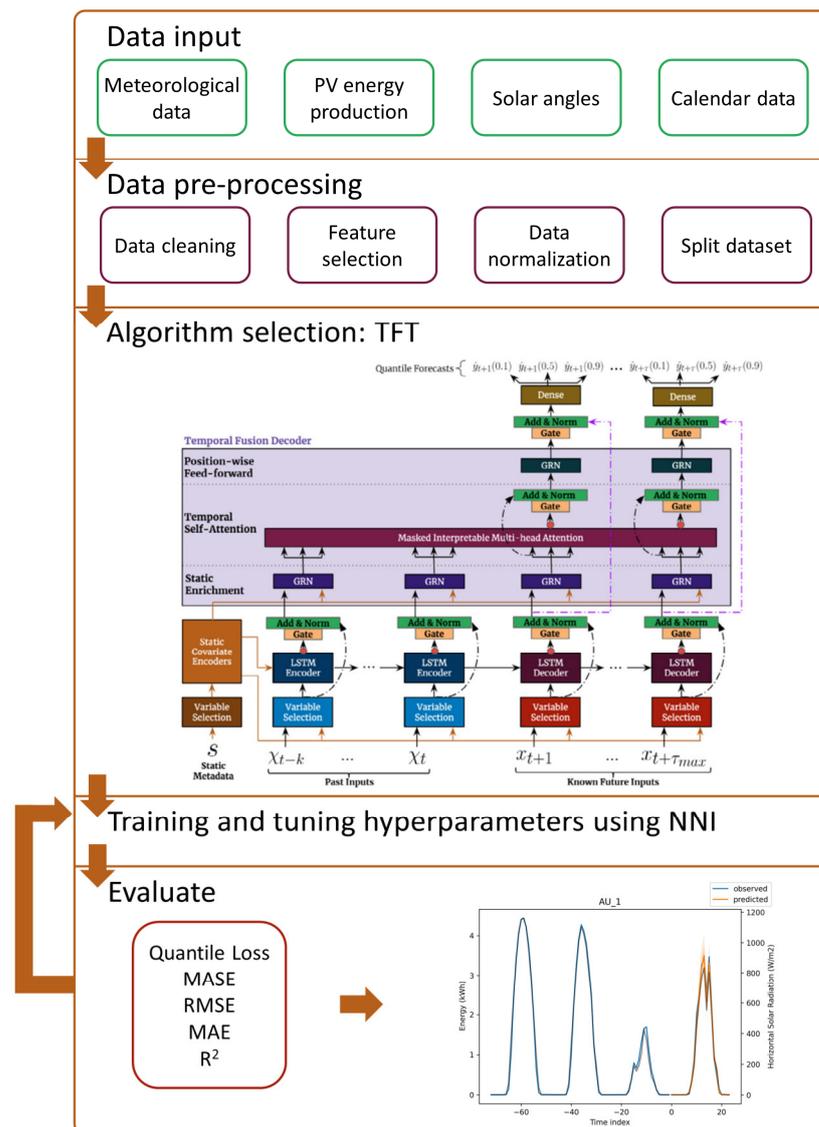


Figure 2. Structure of the proposed model. TFT architecture was adapted from [50].

2.1. Type of Data

The dataset used in this study includes historical power generation data from six different facilities as the dependent variables, and meteorological data, solar angles, and calendar data as independent variables.

The six facilities are located in Germany and Australia. Those located in Germany consist of three roof mounted systems situated on industrial and residential buildings in the city of Konstanz [51]. In this city, the average global horizontal irradiation is 1212 kWh/m² per year, the hottest month is July, with an average high of 25 °C and low of 15 °C, and the coldest month is January with an average low of −2 °C and high of 4 °C. The data is provided with a 5-min resolution.

The facilities in Australia are located at Desert Knowledge Australia Solar Centre (DKASC), in Alice Springs, which is a real-life demonstration of solar technologies in an area of high solar resources [52]. In this area, the average global horizontal irradiation is 2271 kWh/m² per year. The hottest month of the year in Alice Springs is January, with an average high of 35 °C and a low of 22 °C. The coldest month of the year in Alice Springs is July, with an average low of 5 °C and a high of 20 °C. The data is provided with 5-min resolution. Table 1 shows the specific information of each PV system and its data.

Table 1. Main characteristics of the PV facilities and datasets.

Facility	Array Rating (kW)	Characteristics	Start Day	Dataset Length
GE_1	17.0	Fixed, Roof mounted	25 October 2015	498 days
GE_2	5.0	Fixed, Roof mounted	29 February 2016	940 days
GE_3	10.0	Fixed, Roof mounted	11 October 2015	776 days
AU_1	4.9	Fixed, Roof mounted	13 October 2018	667 days
AU_2	6.0	Fixed, Ground mounted	13 October 2018	667 days
AU_3	5.8	Fixed, Ground mounted	13 October 2018	667 days

The meteorological variables used include global horizontal irradiance, wind speed, precipitation, temperature, and humidity. These factors affect to a greater or lesser extent the PV energy output of the system. The DKASC dataset used includes measurements of these variables, while for the systems of Konstanz this information is collected from a nearby weather station [53].

The solar angles, zenith and azimuth, were calculated based on the timestamp and the location of each PV system for each record of the dataset. The solar zenith angle is the angle between the sun's rays and the vertical direction, defining the sun's apparent altitude. The azimuth represents the sun's relative direction along the local horizon.

Figures 3 and 4 show the aforementioned variables during a five-day period in one of the facilities from Germany and one from Australia.

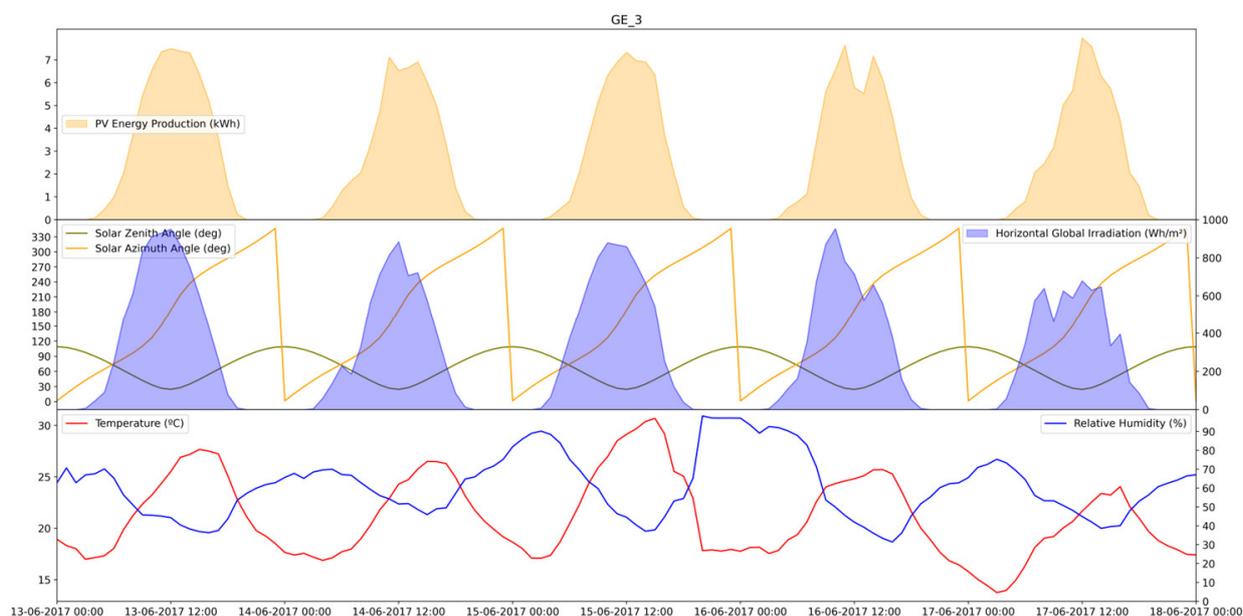


Figure 3. Photovoltaic (PV) energy production (kWh), horizontal global irradiation (Wh/m^2), solar zenith ($^\circ$), solar azimuth ($^\circ$), temperature ($^\circ\text{C}$), and relative humidity (%) from one of the PV facilities in Germany (GE_3) during five days.

Regarding the calendar data, cyclical calendar variables need to be transformed to represent the data sequentially. For example, December and January are 1 month apart, although those months will appear to be separated by 11 months. To avoid this, 2 new features were created, deriving a sine and a cosine transform of the original feature.

Finally, the data was converted to hourly data and merged to create the final dataset, consisting of eleven independent variables, five meteorological variables, four calendar factors, and two variables to represent solar angles. Table 2 shows the variables initially considered.

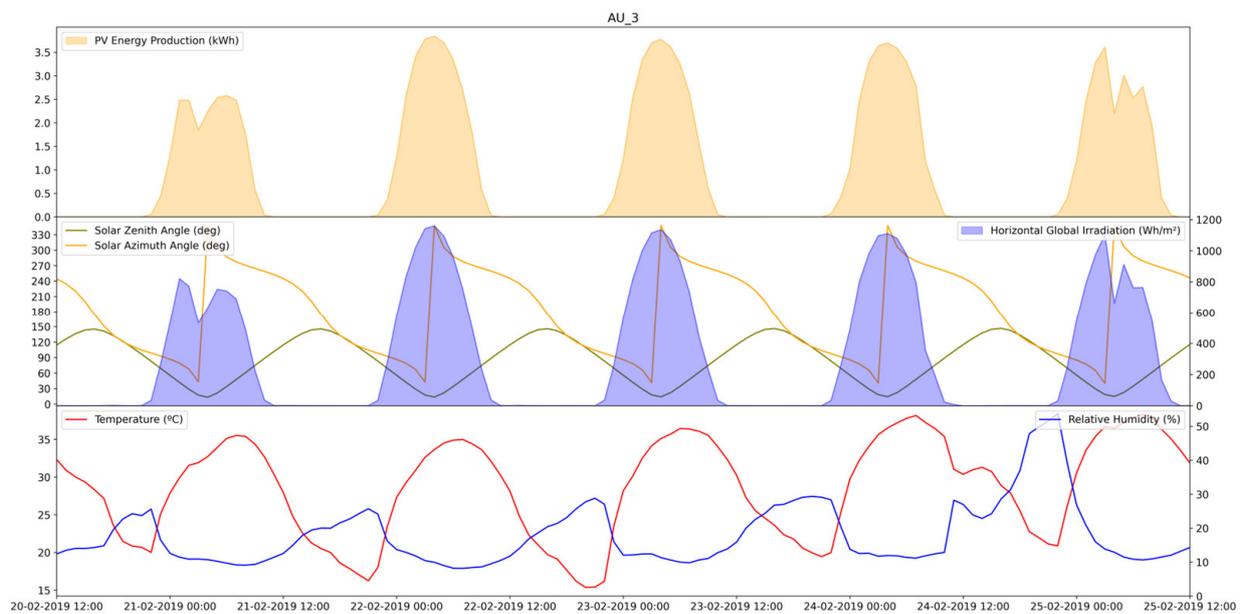


Figure 4. PV energy production (kWh), horizontal global irradiation (Wh/m^2), solar zenith ($^\circ$), solar azimuth ($^\circ$), temperature ($^\circ\text{C}$), and relative humidity (%) from one of the PV facilities in Australia (AU_3) during five days.

Table 2. Dependent and independent variables initially considered.

Type of Variable	Variable Name	Source *
Dependent Variable	PV energy Production	PP
	Horizontal global irradiation	WS
	Solar zenith	DV
	Solar azimuth	DV
	Temperature	WS
	Relativity humidity	WS
Independent Variable	Wind speed	WS
	Precipitation	WS
	Hour sin	DV
	Hour cosine	DV
	Month sin	DV
	Month cosine	DV

* PP: Power Plant, WS: Weather Station, DV: Derived Variable.

2.2. Data Pre-Processing

Data pre-processing is defined as the transformation of raw data into a form more suitable for modeling. In this study, the following pre-processing tasks are included: data cleaning, feature selection, and data transformations.

Data cleaning:

The raw dataset has to be cleaned to identify and correct errors that may negatively impact the performance of the predictive model. For that, one important task is the detection of outliers taking into account the physical behavior of the system. DBSCAN (Density-Based Spatial Clustering of Application with Noise) was used to identify outliers in the dataset. This algorithm creates groups depending on a distance measurement between points, usually using Euclidean distance. A point is included in a cluster if the distance between this point and another point of the cluster is less than a parameter ϵ , taking into account a minimum number of points for each cluster. As a result, it identifies outliers as the points that do not belong to a cluster, which are those that are in low-density regions [54]. Considering that the PV energy production should fall within certain bounds for a given

global horizontal irradiation, zenith angle and azimuth angle, these four features were considered to apply this algorithm.

Next, the missing values and the values of solar energy production identified as outliers were replaced taking into account the averaged coefficient between the PV generation and the horizontal global irradiation at the same hour on previous days. The new value for solar energy production is obtained by multiplying this coefficient by the horizontal global irradiation at the considered index.

Feature selection:

Feature selection is applied using wrapper methods. These methods train models using a subset of the input variables, where the objective is to find the subset that builds the model that yields the best results. As this process is computationally very expensive, we use TFT, a model that already has interpretable feature selection within its architecture to guide it.

The wrapper method used is Backward Elimination [55]. Initially, a TFT model is trained with all the variables, then, at each iteration, the least significant variable in the TFT model is removed, which should improve the performance of the model. Thus, this process is stopped when no improvement is observed after discarding another feature.

Data transforms:

Continuous input variables have different scales, which may result in a slow or unstable learning process. For that, they need to be normalized to make every datapoint have the same importance. In this study, standardization (of z-score normalization) was used to rescale the data. This is a technique where the final values have a zero mean and a unit standard deviation. The formula of the z-score normalization is:

$$x_{stand} = \frac{x_i - \mu}{\sigma} \quad (1)$$

where x_i is the input data, μ is the average value of the feature, and σ is the standard deviation.

For the target variable, different transformations are applied. First the natural logarithm of the values, then the values are scaled by dividing each value by the median in its series.

2.3. Data Splitting

The dataset is divided into three sets for training purposes:

Training set: The sample of the data used to fit the model.

Validation set: The sample of data used to provide an unbiased evaluation of a fitted model while tuning its hyperparameters.

Test set: Part of the data used to fairly evaluate the final model fit on the training set.

This division is not done evenly; each photovoltaic installation in the dataset is divided so that 70% of its randomly selected samples go into the training set, 20% into the validation set, and 10% into the test set.

2.4. Forecasting Models

The solar energy production was predicted using TFT, although its performance was compared with other algorithms like ARIMA, MLP, LSTM, and XGBoost. MLP, LSTM, and TFT models were trained using the implementation in Pytorch Forecasting [56], while ARIMA and XGBoost used their own packages for Python.

2.4.1. TFT

TFT is an attention-based neural network architecture specially designed to combine multi-horizon forecasting with interpretable insights into temporal dynamics [50]. TFT utilizes specialized building blocks to select relevant features and leave out unused components, enabling high performance over a broad range of tasks. Their main components are:

- Gating mechanisms, named Gated Residual Network (GRN). They allow us to filter out any unnecessary elements of the architecture and avoid non-linear processing if it is not needed.
- Variable selection networks to choose important input features at each time step.
- Static covariate encoders that incorporate static variables into the network to condition temporal dynamics.
- Temporal processing to learn both long and short-term temporal patterns from both known and observed time-varying inputs. A temporal self-attention decoder, based on the Masked Multi-Head Attention layer, is employed to capture long-term dependencies, while a sequence-to-sequence layer is used to process local information.
- Allows us to calculate prediction intervals via quantile forecast to determine the likelihood of each target value.

TFT improves the interpretability of time series forecasting through the identification of the globally-important variables, the persistent temporal patterns, and the significant events for the prediction problem. It explains how and why the results are generated by the model, in contrast with the concept of a “black box” where it is difficult to explain how models arrive at their predictions. This makes the model’s output more trustworthy and easier to work with, which is the objective of Explainable AI (XAI) [57].

2.4.2. ARIMA

ARIMA is a statistical model that predicts future values based on past time series data [58]. This algorithm consists of 3 components:

- Autoregression (AR), takes into account the dependence between an observation and certain past values.
- Integrated (I), uses the differencing of raw observations necessary to make the series stationary.
- Moving Average (MA), considers the dependent relationship between an observation and a residual error of a moving average model applied to certain past observations.

2.4.3. MLP

A MLP is a fully connected class of feedforward ANN [59]. It is made up of many interconnected computational units, called neurons. An artificial neuron receives inputs from other neurons, which are weighted and summed, establishing the impact of the information going through them. This weighted sum is then transformed by an activation function to generate the output of the neuron. These activation functions are needed to learn complex patterns and non-linear curves.

In a neural network, the neurons are arranged into multiple layers: an input layer, one or more hidden layers, and an output layer. The number of neurons of each layer and the number of hidden layers are the main parameters to optimize.

2.4.4. LSTM

LSTM networks are a special class of RNN, capable of learning long-term dependencies [60]. They can retain previous information and learn temporal correlations between consecutive data points. The memory block, its building structure, is composed of three interacting gates.

In LSTM, the core variable is the cell state, which allows for information to be carried from previous steps throughout the memory block. Its interacting gates control the addition or removal of information from the cell state. In particular, the first gate is called forget gate, and determines which information from previous steps to throw away from the cell state. The second one, the “input gate”, controls the updating of the cell state. Finally, the last interacting gate, called the “output gate”, provides the final output, based on a filtered version of the updated cell state.

2.4.5. XGBoost

XGBoost is a supervised predictive algorithm that uses the boosting principle [61]. The idea behind boosting is to generate multiple “weak” prediction models sequentially, each of these takes the results of the previous model to generate a “stronger” model. It uses different types of weak decision trees as its models. To get a stronger model from these weak models, an optimization algorithm is used, in this case, Gradient Descent.

2.5. Hyperparameter Tuning

The different hyperparameters of the model are predefined with default values, although better performance is achieved if they are optimized for the problem under study. Neural Network Intelligence (NNI) is used to tune these hyperparameters [62]. This tool allows us to define a search space, generating different configurations of the parameters. Next, NNI performs training for each combination, in order to find the best outcome. NNI trains the model by generating its own random training, validation, and test sets from the dataset, and training it with a combination of the defined hyperparameters.

2.6. Evaluation Metrics

To evaluate and compare the forecasting performances, several statistical error metrics were employed: root mean square error (*RMSE*), mean absolute error (*MAE*), mean absolute scaled error (*MASE*), coefficient of determination (R^2), and quantile loss. These evaluation metrics are defined as:

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

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (3)$$

$$MASE = \frac{MAE}{MAE_{naive}} \quad (4)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - y_{avg})^2} \quad (5)$$

$$QuantileLoss = \sum_{q \in Q} \max((q(y_i - \hat{y}_i), (q - 1)(y_i - \hat{y}_i)) \quad (6)$$

where y_i and \hat{y}_i represent the real and forecasted values, respectively, y_{avg} shows the mean of real values, MAE_{naive} is the *MAE* of a naïve model that predicts the values by shifting actual values into the future, and Q is the set of quantile values to be fitted in our study $Q = \{0.02, 0.1, 0.25, 0.5, 0.75, 0.9, 0.98\}$.

3. Results and Discussion

3.1. Outliers Detection

The percentage of missing values and outliers detected in each dataset are shown in Table 3. The number of outliers identified is greater in the facilities from Germany than in those from Australia. The reason for this behavior could be the distance between the PV facilities and the weather station in Konstanz. Although the tower is located in the same city not far from the facilities, the irradiance recorded at the same moment can vary significantly on partially cloudy days. The analysis of outliers is, therefore, of great importance for those facilities that do not have their own weather station.

Table 3. Missing values and outliers detected in each dataset.

Facility	Missing Values (%)	Outliers (%)	Facility	Missing Values (%)	Outliers (%)
GE_1	0.60	0.49	AU_1	1.83	0.00
GE_2	1.13	1.85	AU_2	1.83	0.00
GE_3	0.68	1.05	AU_3	1.83	0.00

3.2. Data Analysis

The dataset was analyzed once it was cleaned. First, Figure 5 shows the probability density function (PDF) of the PV energy production data, taking only into account values between sunrise and sunset. These datasets present a bimodal distribution for the Australian facilities, with a peak close to zero and another at high PV energy production. On the other hand, the distributions are left-skewed for the installations located in Germany, with its highest point at low PV energy production. In both cases, the first peak is derived from the hours near sunset and sunrise, where the global horizontal radiation is low.

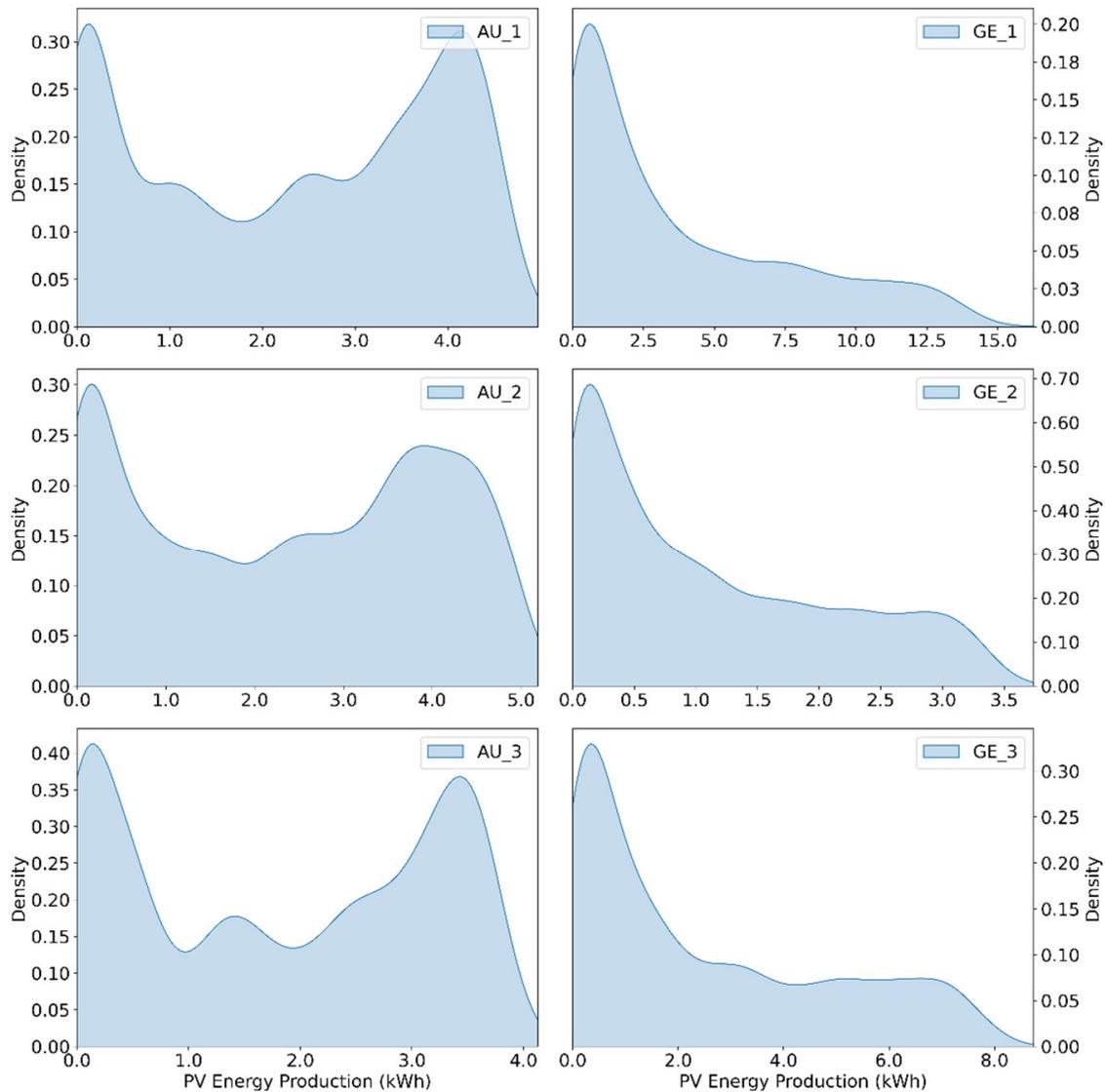


Figure 5. Probability density function (PDF) of the PV energy production for the different facilities.

Figure 6 shows the autocorrelation function (ACF) of PV energy production for two of the time series considered, one for each location. This ACF represents the similarity between two observations depending on the lag time among them. In this case, the results indicate that the PV energy production presents a clear periodic pattern with an interval of 24h. These results are equivalent for the rest of the facilities.

The Pearson correlation coefficient was also calculated to measure the correlation between the PV energy production and both the meteorological variables and the solar angles (Table 4).

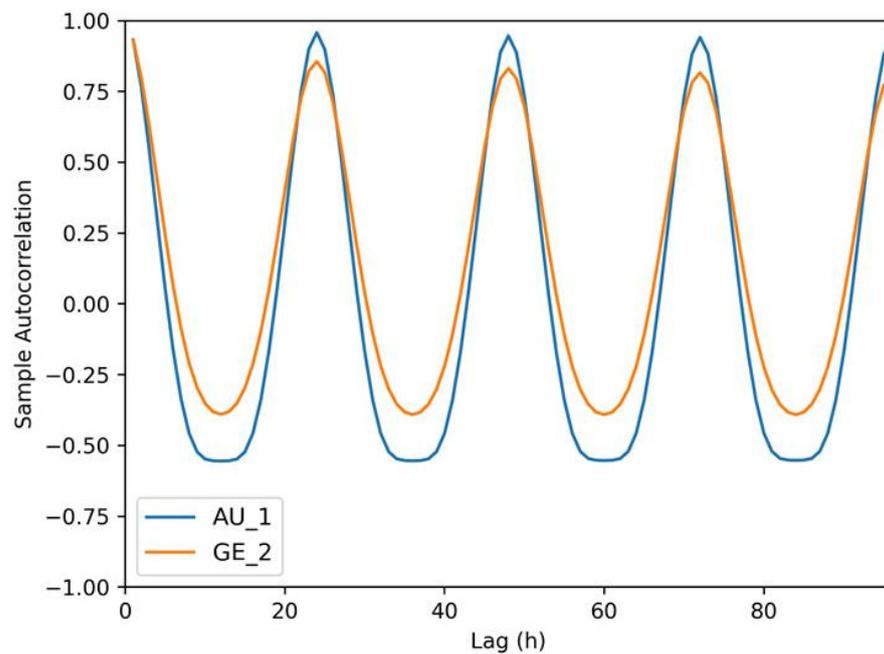


Figure 6. Autocorrelation function (ACF) of PV energy production time series data.

Table 4. Pearson correlation coefficient (PCC) between target value (PV energy production) and the meteorological variables and solar angles.

PCC	GE_1	GE_2	GE_3	AU_1	AU_2	AU_3
Horizontal global irradiation	0.91	0.97	0.89	0.98	0.99	0.98
Temperature	0.56	0.57	0.54	0.40	0.44	0.40
Relative humidity	−0.74	−0.69	−0.74	−0.37	−0.39	−0.36
Wind speed	0.08	0.07	0.10	0.09	0.10	0.09
Precipitation	−0.13	−0.11	−0.12	−0.03	−0.03	−0.03
Solar zenith	−0.77	−0.83	−0.77	−0.84	−0.85	−0.83
Solar azimuth	0.21	0.11	0.25	0.03	0.04	0.05

The correlation between PV energy production and the different meteorological variables varies between the location, showing a different influence of these features depending on the climatic conditions. With respect to the correlation with the horizontal global irradiation, it is stronger in the case of the facilities from Australia. Among the facilities in Germany, the highest correlation was found in the GE_2 facility. The differences in the correlations obtained may be due to several factors. First, the sky conditions in Konstanz are much more diverse than the sky conditions in Alice Springs, which has an impact in the relation between horizontal irradiation and plane-of-array (POA) irradiation [63], and therefore, in energy production. Secondly, the GE_1 and GE_3 installations are not oriented towards the south, but slightly towards the southwest, which produces a certain gap between the maximum generation and the maximum irradiation points, as shown in Figure 7. Thirdly, the tilt of the different systems is not the same: the lower the tilt, the greater the correlation. Finally, the facilities in Germany are located on roofs of urbanized areas, which results in a more obstructed horizon that also decreases the correlation.

Finally, Figure 8 represents the relation between the PV energy production and the horizontal global irradiation for the different facilities, highlighting the differences between the months of the year.

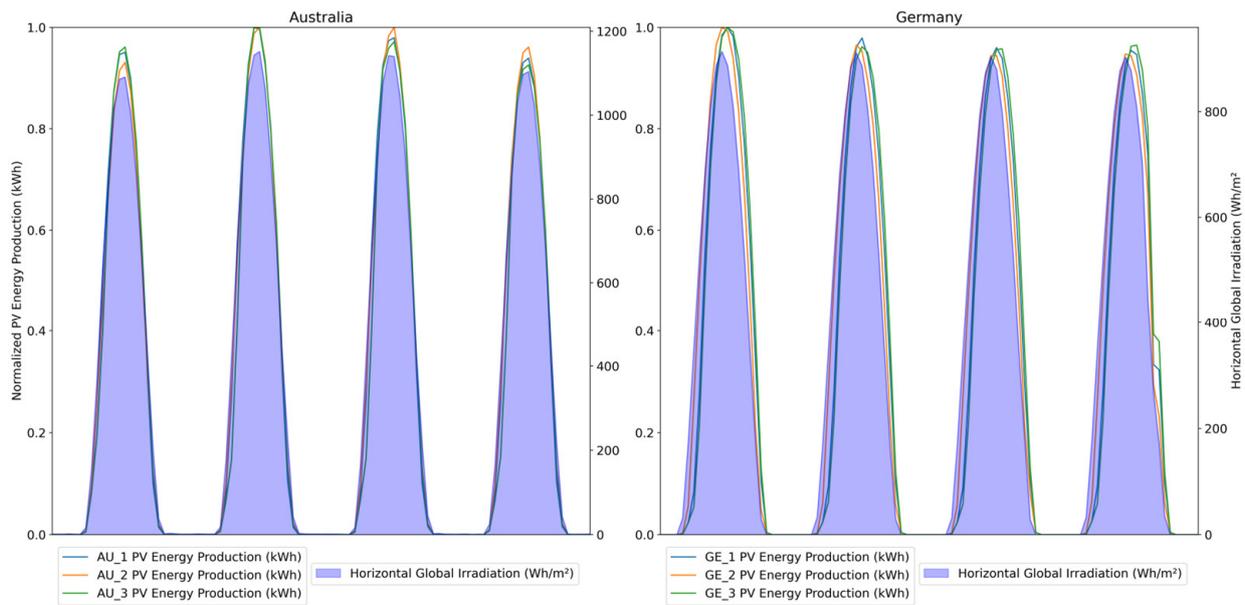


Figure 7. Normalized PV energy production and horizontal solar irradiation for the facilities from Germany and Australia during a four-day period.

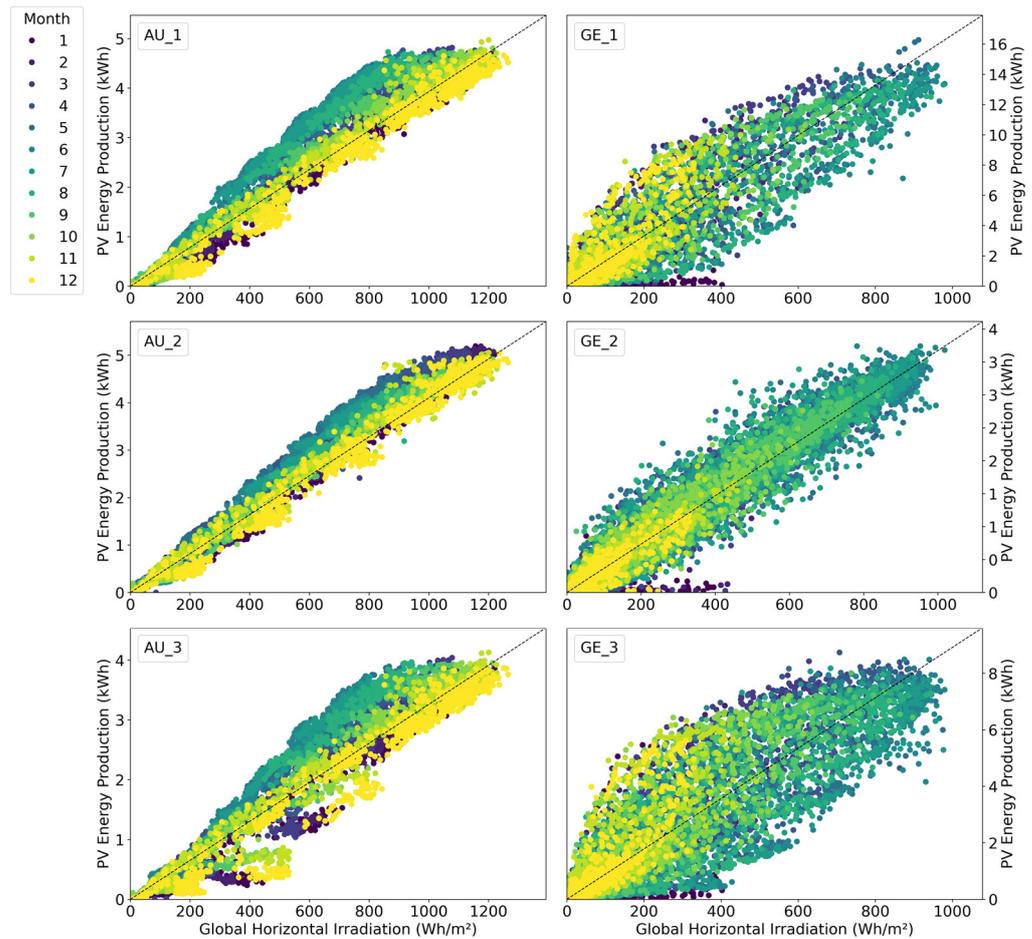


Figure 8. Relation between PV energy production (kWh) and global horizontal irradiation (Wh/m^2) for the different locations.

3.3. Feature Selection

From the dataset, solar horizontal irradiation, temperature, humidity, zenith angle, azimuth angle, and the sine and cosine transformation of the month were found to be valuable parameters for the model, whereas the rest of the variables considered were discarded. Although different feature sets were used, temperature and solar angles were also selected variables in [64], while precipitation was also discarded.

3.4. Forecasting Results

Using the variables selected, the hourly day-ahead PV power generation was predicted using different models for the two locations. The hyperparameters of MLP, LSTM, and XGBoost were established based on the best models in [28,48,65]. For TFT, the hyperparameters were tuned with NNI and the best performance was obtained for the values in Table 5.

Table 5. Selected hyperparameters for TFT model.

Hyperparameter	Value
LSTM layers	2
Hidden size	60
Hidden continuous size	30
Attention head size	2
Learning rate	0.001
Encoder length	72
Dropout	0.7
Batch size	1024

Figure 9 shows the results for different combinations of hyperparameters.

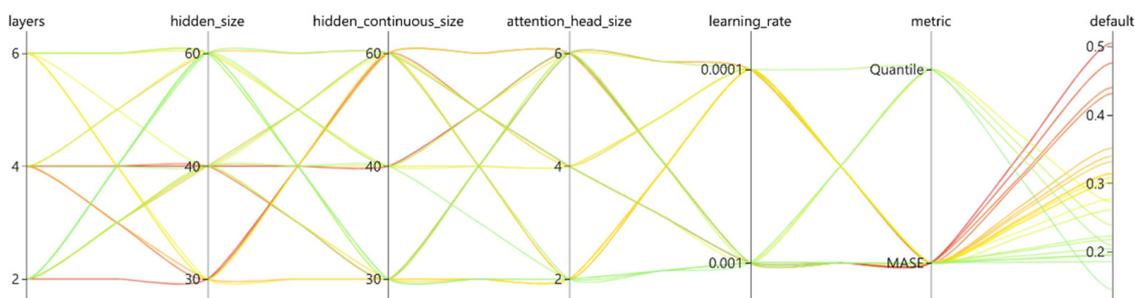


Figure 9. Results during the tuning of the hyperparameters for TFT with NNI.

The variables related to the complexity of the neural network based models are shown in Table 6. TFT is the most complex model considering the three variables: number of parameters, model size, and training time per epoch showing an 82.6% and 86.6% increase in size and in time complexity, respectively, compared to LSTM.

Table 6. Variables related to neural networks complexity.

	Total Parameters (K)	Estimated Parameter Size (MB)	Training Time per Epoch (s)
MLP	8.7	0.035	15
LSTM	48.1	0.192	16
TFT	361	1.44	92

The forecasting errors using these different models are shown in Table 7. As can be seen, TFT outperforms the other models with lowest values for all the indicators in both locations. Comparing TFT with LSTM, the second-best model, the indicators of RMSE, MAE, and MASE, for Australia are 46%, 48%, 48% lower, respectively. The same behavior

can be observed for Germany, with a reduction of 20%, 26%, and 54% in RMSE, MAE, and MASE from TFT with respect to LSTM. These results may be explained by the different components integrated in TFT. This algorithm incorporates temporal self-attention decoder to capture long-term dependencies, allowing to learn relationships at different scales. It also supports to efficiently address features for each input type: static or time-invariant features, past-observed time varying input and future-known time varying input. It was also found to exceed other predictions models in other areas like the forecasting of wind speed or freeway traffic speed [66,67].

Table 7. Forecasting errors for the different models in both locations: Germany and Australia.

	RMSE		MAE		MASE		R ²	
	GE	AU	GE	AU	GE	AU	GE	AU
ARIMA	1.441	0.219	1.106	0.187	-	-	0.640	0.978
MLP	0.786	0.593	0.370	0.331	1.458	0.961	0.816	0.853
LSTM	0.343	0.118	0.162	0.064	0.846	0.191	0.975	0.994
XGBoost	1.124	0.336	0.497	0.137	1.105	1.242	0.577	0.945
TFT	0.276	0.064	0.120	0.033	0.390	0.100	0.983	0.998

LSTM showed good accuracy and better performance than XGBoost, MLP, and ARIMA. This behavior was also found in [47,48]. Unlike XGBoost and MLP, both LSTM and TFT can retain previous information and learn temporal correlations between consecutive data points, giving these models a better performance. Besides, while the ARIMA model was only developed taking into account a linear relationship between the exogenous variables and the target variable, both LSTM and TFT can include non-linear approximations. Therefore, these results indicate that both mechanisms are important for the forecasting of day-ahead solar production.

The performance of TFT on each facility was compared using MASE and R², the two metrics that are scale invariant. The results for Germany are always worse than those for Australia, with a significant reduction in accuracy for both metrics. The forecasting errors with TFT for the different indicators and the different series considered are shown in Table 8.

Table 8. TFT performance for the 6 locations.

	RMSE	MAE	MASE	R ²	Quantile Loss
GE_1	0.432	0.187	0.346	0.983	0.053
GE_2	0.105	0.050	0.513	0.987	0.014
GE_3	0.291	0.123	0.312	0.980	0.035
AU_1	0.062	0.032	0.091	0.999	0.009
AU_2	0.070	0.035	0.097	0.998	0.010
AU_3	0.061	0.032	0.112	0.998	0.009

The highest accuracy in forecasting the day-ahead PV power generation is always achieved in AU_1, although the performance of the TFT is very similar for the three Australian PV plants. In this case, the coefficient of variation, which establish the extent of the variability in relation to the mean, is 3.73% and 0.05% for MASE and R², respectively. In the case of Germany, the coefficients of variation are 27.62% and 0.29% for MASE and R², showing a great variability between the performance of the model in German facilities.

The forecast accuracy in the German facilities is still not as favorable as the one in the facilities in Australia, this may be due to the greater variability in the weather conditions experienced by the solar facilities in Konstanz.

However, the forecast in the facilities in Germany has improved drastically, especially in GE_1 and GE_3. As mentioned, the other factors which could be damaging the Pearson correlation between horizontal irradiation and energy production (the different orientation

and tilt of the panels and the more blocked horizon) could have been corrected with the implementation in the model of the solar angles (zenith and azimuth), the meteorological variables (temperature and humidity) and the calendar data. The solar angles help to predict the relationship between the horizontal irradiation and the irradiation in the plane of the panels [68], which is reinforced with the meteorological variables [63] that also influence other factors, such as the efficiency of the panels [69].

Figures 10 and 11 show the results of PV power generation forecasting based on TFT and the real solar energy production for representative rainy and sunny days in both locations. These graphs represent the global horizontal irradiation and the observed values from the past 72 h to the prediction length. The predicted value is represented with the prediction intervals, showing the uncertainty of the forecasted values. This information is important when managing the operation of the facilities.

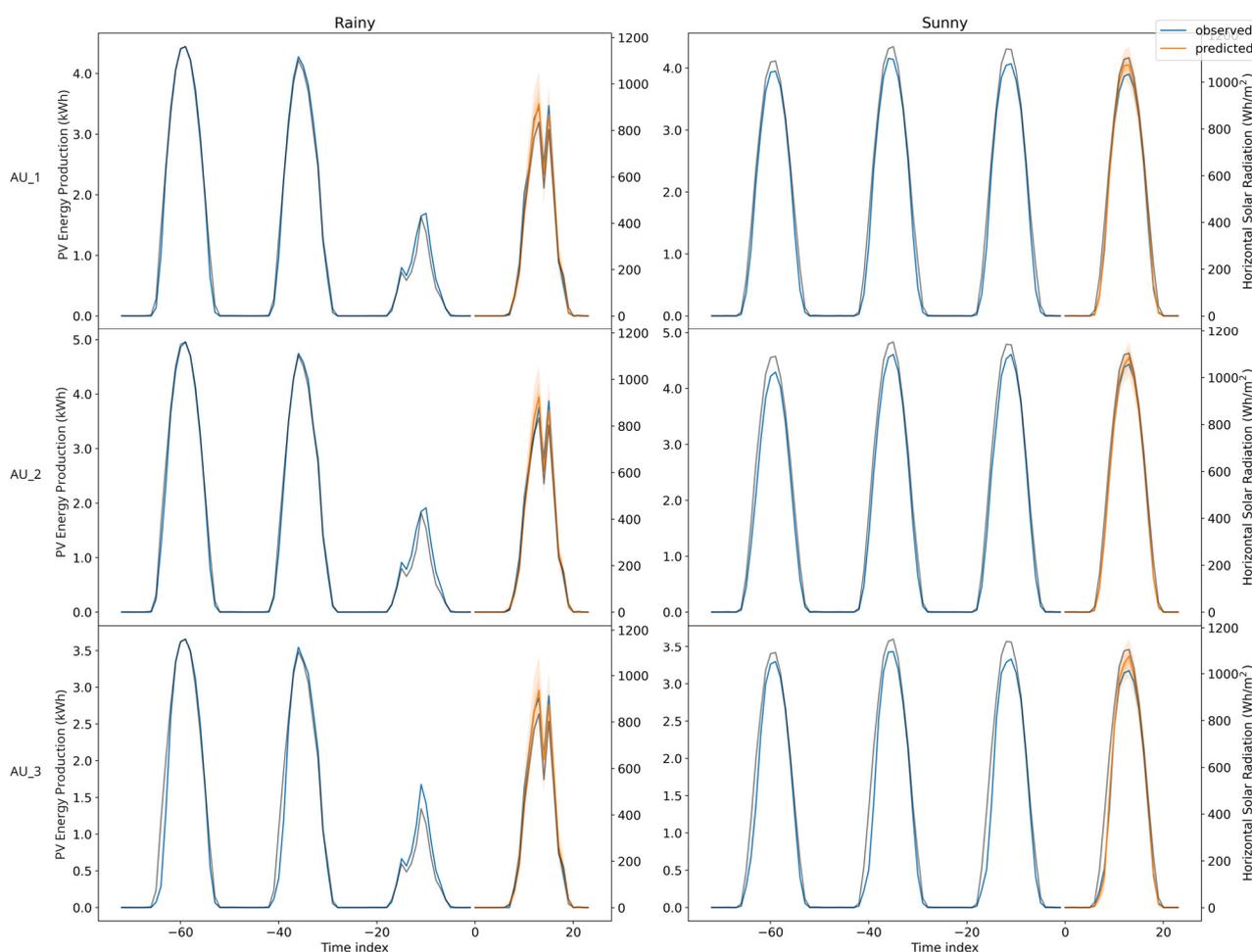


Figure 10. Observed and predicted PV energy production forecasting for a sunny and a rainy day in the facilities from Australia. Global horizontal radiation is represented in grey.

For AU_1 on a rainy day, the maximum solar energy production predicted is 3.50 kWh, where 3.22 kWh and 3.76 kWh are the 0.1 and 0.9 percentiles, respectively. The interval between the tenth and the ninetieth percentile is 0.54 kWh. Considering the hours with energy production, the mean and standard deviation for half of this interval are 0.19 kWh and 0.07 kWh, respectively, with a maximum value of 0.27 kWh and a minimum of 0.07 kWh. For the example of a sunny day, the value predicted for the hour with maximum solar output is 4.05 kWh. In this case, the mean and standard deviation of the mid-range between the considered percentiles during the day is 0.12 kWh and 0.03 kWh,

respectively, indicating lower variation in these values. Analyzing these results for this case, it seems that the uncertainty is higher and more varied for rainy days.

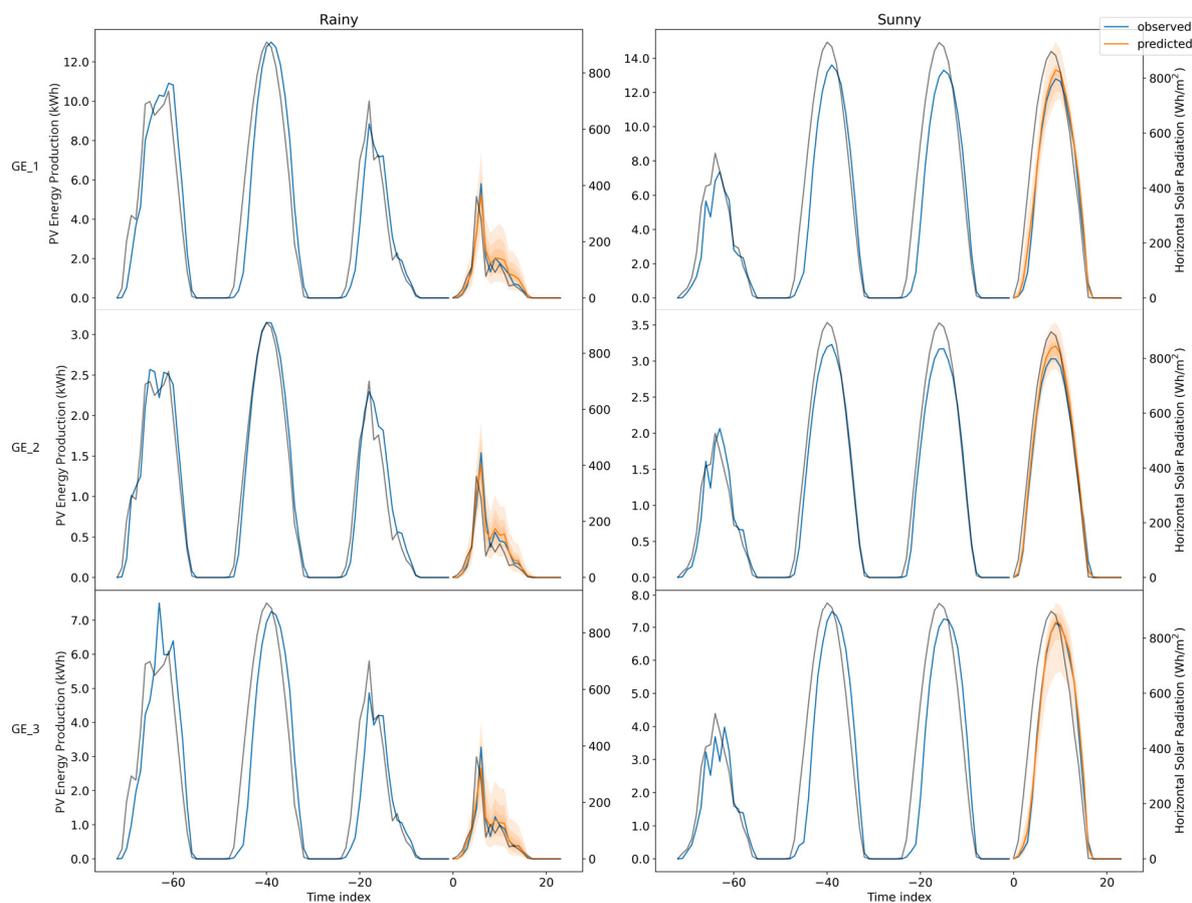


Figure 11. Observed and predicted PV energy production forecasting for a sunny and a rainy day in the facilities from Germany. Global horizontal irradiation is represented in grey.

For the case of GE_2, the maximum production on the example for a rainy day is 1.39 kWh, and the mean and standard deviation of the half of the interval between 0.1 and 0.9 percentiles along the day is 0.20 kWh and 0.07 kWh, respectively. The maximum is 0.35 kWh and the minimum is 0.08 kWh. In the case of a sunny day, the values are 0.17 kWh and 0.02 kWh for the mean and the standard deviation, respectively, with a solar energy production of 3.2 kWh. In this case, the uncertainty is higher than in the case of Australia, although it is constant throughout the day. Comparing the examples of a rainy and a sunny day, the mean values are similar, although the rainy day is over a peak of solar energy production of 1.39 kWh, whereas on a sunny day, the maximum output is 3.2 kWh. This indicates that the uncertainty is also higher on rainy days as in the case of Germany.

One possible explanation for the lower accuracy during cloudy or rainy days could be the higher variability of the irradiance levels recorded during such days. This implies that the training of the network, and therefore, the prediction of PV energy production, becomes less reliable under these circumstances.

3.5. TFT Interpretability

TFT improves the interpretability of time series forecasting through the calculation of the variable importance scores for the different types of features, and also through the representation of the attention weight patterns. These results can be seen in Figure 12, which shows the importance of the past-observed time varying inputs, Figure 13, representing the

importance of the future-known time varying features, and Figure 14 showing the attention weight patterns for one-step-ahead forecasts.

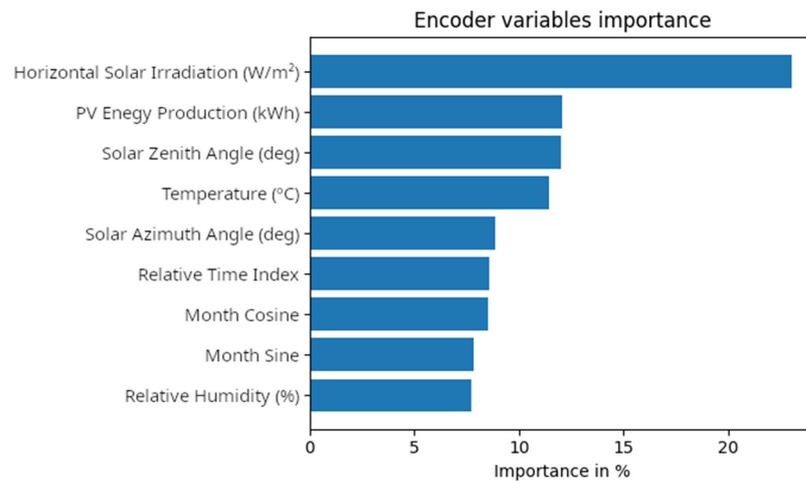


Figure 12. Importance of encoder variables.

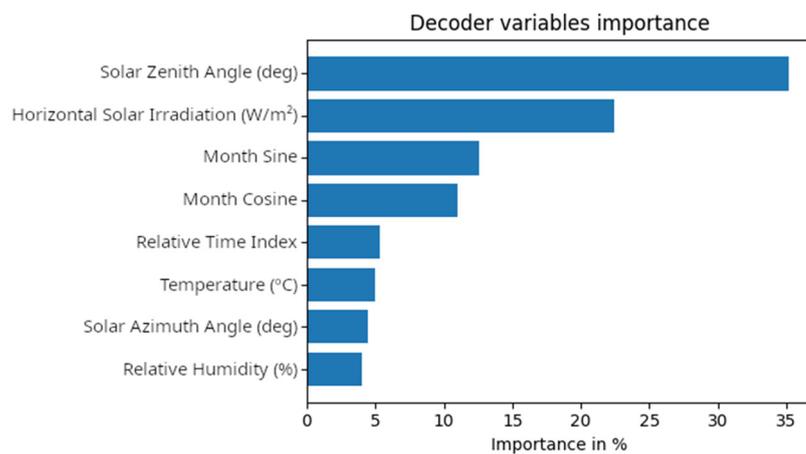


Figure 13. Importance of decoder variables.

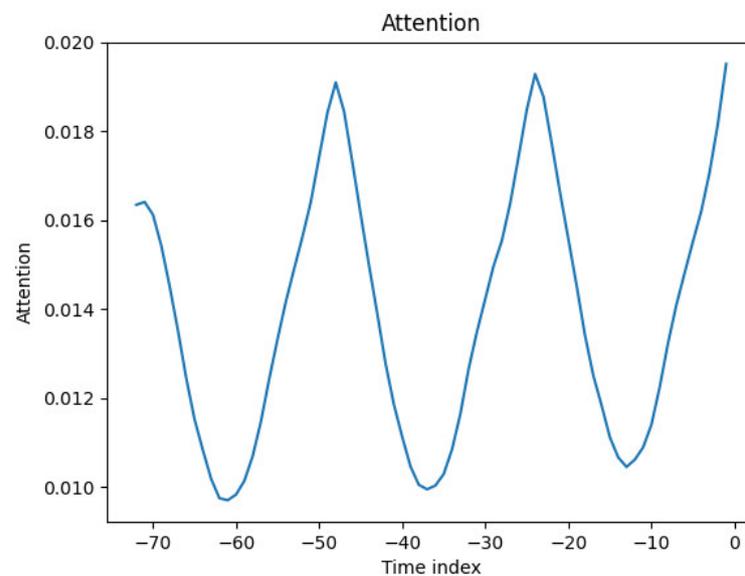


Figure 14. Attention weight patterns.

These datasets also have three static inputs: target center, target scale, and the identification of the facility, aiming at providing additional information and context to the model. This last variable is needed to identify each series, since TFT allows to train all of them together, learning also from general patterns amongst series. The first two are related to the standardization of the target variable, and included as static variables in the model. For these series, target center has a value of 0 and target scale is the median of the time series. The importance of these variables is as follows: target center > series id > target scale.

The encoder variables contain the features for which past values are known at prediction time, and consist of the features selected previously in addition to the index indicating the relative time. In this case, the relative time takes values between -72 (input length) and 0. The importance of each of these variables can be seen in Figure 12. The horizontal solar irradiation received most of the attention (almost 25%), followed by the target variable and the solar zenith (around 12.5% each). Relative humidity and the sine and cosine transformation of the month are variables with a limited weight on the encoder side.

Decoder variables are those features for which future values are known at prediction time. For the decoder, the relative time index takes values between -72 and 24 (prediction length). The importance of each of these variables for the prediction model can be observed in Figure 13. Solar zenith and horizontal solar irradiation play a significant role, summing almost 60% of the weighted importance.

The results from both the encoder and decoder weighted importance highlight the necessity of having a good representation of solar zenith and horizontal solar irradiation to achieve high prediction performance.

Finally, Figure 14 represents the attention weight patterns for one-step-ahead forecasts, which can be used to understand the most important past time steps that the TFT model focused on. It shows that the attention displays a cyclic pattern, with clear attention peaks at daily intervals. Thus, the attention is focused on the closest values, and in values at the same hour of previous days. This behavior is understandable since the PV energy production follows the same periodic pattern, as can be seen in Figure 6.

4. Conclusions

PV energy production varies significantly due to many factors, including weather conditions, time of day, season, solar irradiance, panel degradation, shading, and other uncontrollable factors. These variations create challenges for power systems because they cause grid instability and imbalances in the supply and demand of electricity amongst other complications. An accurate forecast of the PV power generation helps to guarantee the operation of the power system.

In this study, the TFT, a novel deep learning algorithm, is tested to address the hourly day-ahead PV energy production on datasets from different facilities. This algorithm is able to learn long-term and short-term temporal relationships respectively, and also helps to efficiently build feature representations of static variables, observed and known time-varying inputs. These different components allow the model to be successful in producing accurate forecasts, outperforming other advanced methods like LSTM.

TFT outperformed the other models for predicting PV energy production in the six facilities analyzed, providing better values for all the accuracy indicators. Comparing TFT with LSTM, the second-best model, the indicators of RMSE, MAE, and MASE, for the facilities in Australia are 46%, 48%, 48% lower, and 20%, 26%, and 54% lower for the facilities in Germany.

The TFT model has also performed well in facilities which did not have an on-site weather station where the meteorological variables had to be collected from the nearest meteorological station. This is of great importance since a large number of small and medium solar PV facilities do not have a pyranometer on site. The study shows the importance of detection and replacement of outliers in these types of facilities.

The importance of the decoder and encoder variables has been also calculated, revealing that solar horizontal irradiation and the zenith angle are the key variables for the model.

This model can also calculate the prediction intervals, estimating the uncertainty of the forecast which facilitates the operation and management of the facilities and for economic dispatch optimization.

In relation to the limitations and complexity of the model used, it should be noted that TFT needs a larger amount of data to achieve good prediction results. Besides, it is also a complex model, requiring higher maintenance due to higher computational and storage resources. When this is not an issue, the outcomes show that this research can be of great benefit to grid operators, plant managers and customers. The use of more complex models, such as TFT, provides more accurate PV production predictions, which contributes to improve the management of the power system.

Author Contributions: Conceptualization, M.L.S., X.G.-S., F.E.C., P.C.O. and G.B.G.; methodology, M.L.S.; software, M.L.S.; validation, M.L.S., X.G.-S.; formal analysis, M.L.S.; investigation, M.L.S., X.G.-S. and F.E.C.; data curation, M.L.S.; writing—original draft preparation, M.L.S., X.G.-S. and F.E.C.; writing—review and editing, M.L.S., X.G.-S., G.B.G. and F.E.C.; visualization, M.L.S., X.G.-S., G.B.G.; supervision, G.B.G., F.E.C. and P.C.O.; project administration, F.E.C. and P.C.O.; funding acquisition, F.E.C. and P.C.O. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by CERVERA Research Program of CDTI, the Industrial and Technological Development Centre of Spain, through the Research Projects HySGrid+, grant number CER-20191019, and CEL.IA, grant number CER-20211022. The APC was funded by CERVERA Research Program of CDTI, through the Research Project HySGrid+, grant number CER-20191019.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. European Commission. *The European Green Deal*; European Commission: Brussels, Belgium, 2019.
2. IEA. *Solar PV*; International Energy Agency: Paris, France, 2021.
3. International Renewable Energy Agency. *Future of Solar Photovoltaic Deployment, Investment, Technology, Grid Integration and Socio-Economic Aspects A Global Energy Transformation Paper About IRENA*; International Renewable Energy Agency: Masdar City, Abu Dhabi, 2019; ISBN 978-92-9260-156-0.
4. Raza, M.Q.; Nadarajah, M.; Ekanayake, C. On Recent Advances in PV Output Power Forecast. *Sol. Energy* **2016**, *136*, 125–144. [[CrossRef](#)]
5. International Energy Agency. *World Energy Outlook 2021*; International Energy Agency: Paris, France, 2021.
6. Akhter, M.N.; Mekhilef, S.; Mokhlis, H.; Ali, R.; Usama, M.; Muhammad, M.A.; Khairuddin, A.S.M. A Hybrid Deep Learning Method for an Hour Ahead Power Output Forecasting of Three Different Photovoltaic Systems. *Appl. Energy* **2022**, *307*, 118185. [[CrossRef](#)]
7. Shivashankar, S.; Mekhilef, S.; Mokhlis, H.; Karimi, M. Mitigating Methods of Power Fluctuation of Photovoltaic (PV) Sources—A Review. *Renew. Sustain. Energy Rev.* **2016**, *59*, 1170–1184. [[CrossRef](#)]
8. 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](#)]
9. Pierro, M.; Moser, D.; Perez, R.; Cornaro, C. The Value of PV Power Forecast and the Paradox of the “Single Pricing” Scheme: The Italian Case Study. *Energies* **2020**, *13*, 3945. [[CrossRef](#)]
10. Zahraoui, Y.; Alhamrouni, I.; Mekhilef, S.; Basir Khan, M.R. Chapter One-Machine Learning Algorithms Used for Short-Term PV Solar Irradiation and Temperature Forecasting at Microgrid. In *Applications of AI and IOT in Renewable Energy*; Academic Press: Cambridge, MA, USA, 2022; pp. 1–17. ISBN 978-0-323-91699-8.

11. López, E.; Monteiro, J.; Carrasco, P.; Sáenz, J.; Pinto, N.; Blázquez, G. Development, Implementation and Evaluation of a Wireless Sensor Network and a Web-Based Platform for the Monitoring and Management of a Microgrid with Renewable Energy Sources. In Proceedings of the 2019 International Conference on Smart Energy Systems and Technologies (SEST), Porto, Portugal, 9–11 September 2019; pp. 1–6.
12. Hao, Y.; Dong, L.; Liang, J.; Liao, X.; Wang, L.; Shi, L. Power Forecasting-Based Coordination Dispatch of PV Power Generation and Electric Vehicles Charging in Microgrid. *Renew. Energy* **2020**, *155*, 1191–1210. [[CrossRef](#)]
13. Aslam, M.; Lee, J.M.; Kim, H.S.; Lee, S.J.; Hong, S. Deep Learning Models for Long-Term Solar Radiation Forecasting Considering Microgrid Installation: A Comparative Study. *Energies* **2019**, *13*, 147. [[CrossRef](#)]
14. Ramsami, P.; Oree, V. A Hybrid Method for Forecasting the Energy Output of Photovoltaic Systems. *Energy Convers. Manag.* **2015**, *95*, 406–413. [[CrossRef](#)]
15. Ahmed, R.; Sreeram, V.; Mishra, Y.; Arif, M.D. A Review and Evaluation of the State-of-the-Art in PV Solar Power Forecasting: Techniques and Optimization. *Renew. Sustain. Energy Rev.* **2020**, *124*, 109792. [[CrossRef](#)]
16. Dolara, A.; Leva, S.; Manzolini, G. Comparison of Different Physical Models for PV Power Output Prediction. *Sol. Energy* **2015**, *119*, 83–99. [[CrossRef](#)]
17. Dutta, S.; Li, Y.; Venkataraman, A.; Costa, L.M.; Jiang, T.; Plana, R.; Tordjman, P.; Choo, F.H.; Foo, C.F.; Puttgen, H.B. Load and Renewable Energy Forecasting for a Microgrid Using Persistence Technique. *Energy Procedia* **2017**, *143*, 617–622. [[CrossRef](#)]
18. Shireen, T.; Shao, C.; Wang, H.; Li, J.; Zhang, X.; Li, M. Iterative Multi-Task Learning for Time-Series Modeling of Solar Panel PV Outputs. *Appl. Energy* **2018**, *212*, 654–662. [[CrossRef](#)]
19. Ardila, S.; Maciel, V.M.; Ledesma, J.N.; Gaspar, D.; Dinho Da Silva, P.; Pires, L.C.; María, V.; Nunes Maciel, J.; Javier, J.; Ledesma, G.; et al. Fuzzy Time Series Methods Applied to (In)Direct Short-Term Photovoltaic Power Forecasting. *Energies* **2022**, *15*, 845. [[CrossRef](#)]
20. Almonacid, F.; Pérez-Higueras, P.J.; Fernández, E.F.; Hontoria, L. A Methodology Based on Dynamic Artificial Neural Network for Short-Term Forecasting of the Power Output of a PV Generator. *Energy Convers. Manag.* **2014**, *85*, 389–398. [[CrossRef](#)]
21. Wang, F.; Xuan, Z.; Zhen, Z.; Li, K.; Wang, T.; Shi, M. A Day-Ahead PV Power Forecasting Method Based on LSTM-RNN Model and Time Correlation Modification under Partial Daily Pattern Prediction Framework. *Energy Convers. Manag.* **2020**, *212*, 112766. [[CrossRef](#)]
22. Massucco, S.; Mosaico, G.; Saviozzi, M.; Silvestro, F. A Hybrid Technique for Day-Ahead PV Generation Forecasting Using Clear-Sky Models or Ensemble of Artificial Neural Networks According to a Decision Tree Approach. *Energies* **2019**, *12*, 1298. [[CrossRef](#)]
23. Li, P.; Zhou, K.; Lu, X.; Yang, S. A Hybrid Deep Learning Model for Short-Term PV Power Forecasting. *Appl. Energy* **2020**, *259*, 114216. [[CrossRef](#)]
24. Raza, M.Q.; Mithulananthan, N.; Summerfield, A. Solar Output Power Forecast Using an Ensemble Framework with Neural Predictors and Bayesian Adaptive Combination. *Sol. Energy* **2018**, *166*, 226–241. [[CrossRef](#)]
25. Dolara, A.; Grimaccia, F.; Leva, S.; Mussetta, M.; Ogliari, E. A Physical Hybrid Artificial Neural Network for Short Term Forecasting of PV Plant Power Output. *Energies* **2015**, *8*, 1138–1153. [[CrossRef](#)]
26. Lotfi, M.; Javadi, M.; Osório, G.J.; Monteiro, C.; Catalão, J.P.S. A Novel Ensemble Algorithm for Solar Power Forecasting Based on Kernel Density Estimation. *Energies* **2020**, *13*, 216. [[CrossRef](#)]
27. Radicioni, M.; Lucaferri, V.; de Lia, F.; Laudani, A.; Presti, R.L.; Lozito, G.M.; Fulginei, F.R.; Schioppo, R.; Tucci, M. Power Forecasting of a Photovoltaic Plant Located in ENEA Casaccia Research Center. *Energies* **2021**, *14*, 707. [[CrossRef](#)]
28. Cervone, G.; Clemente-Harding, L.; Alessandrini, S.; Delle Monache, L. Short-Term Photovoltaic Power Forecasting Using Artificial Neural Networks and an Analog Ensemble. *Renew. Energy* **2017**, *108*, 274–286. [[CrossRef](#)]
29. Pan, M.; Li, C.; Gao, R.; Huang, Y.; You, H.; Gu, T.; Qin, F. Photovoltaic Power Forecasting Based on a Support Vector Machine with Improved Ant Colony Optimization. *J. Clean. Prod.* **2020**, *277*, 123948. [[CrossRef](#)]
30. Zendejboudi, A.; Baseer, M.A.; Saidur, R. Application of Support Vector Machine Models for Forecasting Solar and Wind Energy Resources: A Review. *J. Clean. Prod.* **2018**, *199*, 272–285. [[CrossRef](#)]
31. Tang, P.; Chen, D.; Hou, Y. Entropy Method Combined with Extreme Learning Machine Method for the Short-Term Photovoltaic Power Generation Forecasting. *Chaos Solitons Fractals* **2016**, *89*, 243–248. [[CrossRef](#)]
32. Hossain, M.; Mekhilef, S.; Danesh, M.; Olatomiwa, L.; Shamshirband, S. Application of Extreme Learning Machine for Short Term Output Power Forecasting of Three Grid-Connected PV Systems. *J. Clean. Prod.* **2017**, *167*, 395–405. [[CrossRef](#)]
33. Almonacid, F.; Rus, C.; Hontoria, L.; Muñoz, F.J. Characterisation of PV CIS Module by Artificial Neural Networks. A Comparative Study with Other Methods. *Renew. Energy* **2010**, *35*, 973–980. [[CrossRef](#)]
34. Oudjana, S.H.; Hellal, A.; Mahammed, I.H. Power Forecasting of Photovoltaic Generation. *Int. J. Electr. Comput. Eng.* **2013**, *7*, 627–631.
35. Monteiro, R.V.A.; Guimarães, G.C.; Moura, F.A.M.; Albertini, M.R.M.C.; Albertini, M.K. Estimating Photovoltaic Power Generation: Performance Analysis of Artificial Neural Networks, Support Vector Machine and Kalman Filter. *Electr. Power Syst. Res.* **2017**, *143*, 643–656. [[CrossRef](#)]
36. Matteri, A.; Ogliari, E.; Nespoli, A.; Rojas, F.; Herrera, L.J.; Pomare, H. Enhanced Day-Ahead PV Power Forecast: Dataset Clustering for an Effective Artificial Neural Network Training. *Eng. Proc.* **2021**, *5*, 16. [[CrossRef](#)]

37. Mellit, A.; Pavan, A.M.; Lughi, V. Deep Learning Neural Networks for Short-Term Photovoltaic Power Forecasting. *Renew. Energy* **2021**, *172*, 276–288. [[CrossRef](#)]
38. Mishra, M.; Byomakesha Dash, P.; Nayak, J.; Naik, B.; Kumar Swain, S. Deep Learning and Wavelet Transform Integrated Approach for Short-Term Solar PV Power Prediction. *Measurement* **2020**, *166*, 108250. [[CrossRef](#)]
39. Zang, H.; Cheng, L.; Ding, T.; Cheung, K.; Liang, Z.; Wei, Z.; Sun, G. A Hybrid Method for Short-Term Photovoltaic Power Forecasting Based on Deep Convolutional Neural Network. *IET Gener. Transm. Distrib.* **2018**, *12*, 4557–4567. [[CrossRef](#)]
40. 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](#)]
41. Yu, D.; Lee, S.; Lee, S.; Choi, W.; Liu, L. Forecasting Photovoltaic Power Generation Using Satellite Images. *Energies* **2020**, *13*, 6603. [[CrossRef](#)]
42. Yona, A.; Senjyu, T.; Funabashi, T.; Kim, C. Determination Method of Insolation Prediction with Fuzzy and Applying Neural Network for Long-Term Ahead PV Power Output Correction. *IEEE Trans. Sustain. Energy* **2013**, *4*, 527–533. [[CrossRef](#)]
43. Li, G.; Wang, H.; Zhang, S.; Xin, J.; Liu, H. Recurrent Neural Networks Based Photovoltaic Power Forecasting Approach. *Energies* **2019**, *12*, 2538. [[CrossRef](#)]
44. Maitanova, N.; Telle, J.S.; Hanke, B.; Grottko, M.; Schmidt, T.; von Maydell, K.; Agert, C. A Machine Learning Approach to Low-Cost Photovoltaic Power Prediction Based on Publicly Available Weather Reports. *Energies* **2020**, *13*, 735. [[CrossRef](#)]
45. Yu, D.; Choi, W.; Kim, M.; Liu, L. Forecasting Day-Ahead Hourly Photovoltaic Power Generation Using Convolutional Self-Attention Based Long Short-Term Memory. *Energies* **2020**, *13*, 4017. [[CrossRef](#)]
46. Abdel-Nasser, M.; Mahmoud, K. Accurate Photovoltaic Power Forecasting Models Using Deep LSTM-RNN. *Neural Comput. Appl* **2019**, *31*, 2727–2740. [[CrossRef](#)]
47. Zheng, J.; Zhang, H.; Dai, Y.; Wang, B.; Zheng, T.; Liao, Q.; Liang, Y.; Zhang, F.; Song, X. Time Series Prediction for Output of Multi-Region Solar Power Plants. *Appl. Energy* **2020**, *257*, 114001. [[CrossRef](#)]
48. Luo, X.; Zhang, D.; Zhu, X. Deep Learning Based Forecasting of Photovoltaic Power Generation by Incorporating Domain Knowledge. *Energy* **2021**, *225*, 120240. [[CrossRef](#)]
49. Kim, B.; Suh, D.; Otto, M.O.; Huh, J.S. A Novel Hybrid Spatio-Temporal Forecasting of Multisite Solar Photovoltaic Generation. *Remote Sens.* **2021**, *13*, 2605. [[CrossRef](#)]
50. Lim, B.; Arik, S.; Loeff, N.; Pfister, T. Temporal Fusion Transformers for Interpretable Multi-Horizon Time Series Forecasting. *Int. J. Forecast.* **2021**, *37*, 1748–1764. [[CrossRef](#)]
51. Data Platform—Open Power System Data. Available online: https://data.open-power-system-data.org/household_data/ (accessed on 24 March 2022).
52. DKASC, Alice Springs DKA Solar Centre. Available online: <http://dkasolarcentre.com.au/locations/alice-springs> (accessed on 24 March 2022).
53. Wetter Und Klima-Deutscher Wetterdienst—Our Services—Open Data Server. Available online: <https://www.dwd.de/EN/ourservices/opendata/opendata.html> (accessed on 24 March 2022).
54. Ram, A.; Kumar, M. A Density Based Algorithm for Discovering Density Varied Clusters in Large Spatial Databases Sunita Jalal. *Int. J. Comput. Appl.* **2010**, *3*, 975–8887.
55. Witten, I.H.; Frank, E.; Hall, M.A. Data Transformations. In *Data Mining: Practical Machine Learning Tools and Techniques*; Elsevier: Amsterdam, The Netherlands, 2011; pp. 305–349. [[CrossRef](#)]
56. PyTorch Forecasting Documentation—Pytorch-Forecasting Documentation. Available online: <https://pytorch-forecasting.readthedocs.io/en/stable/#> (accessed on 2 June 2022).
57. Gunning, D.; Stefik, M.; Choi, J.; Miller, T.; Stumpf, S.; Yang, G.Z. XAI—Explainable Artificial Intelligence. *Sci. Robot.* **2019**, *4*, eaay7120. [[CrossRef](#)]
58. Asteriou, D.; Hall, S.G. ARIMA Models and the Box-Jenkins Methodology. *Appl. Econom.* **2016**, *2*, 275–296. [[CrossRef](#)]
59. Haykin, S.S. *Neural Networks: A Comprehensive Foundation*; Prentice Hall PTR: Hoboken, NJ, USA, 1994.
60. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)]
61. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016. [[CrossRef](#)]
62. NNI Documentation—Neural Network Intelligence. Available online: <https://nni.readthedocs.io/en/stable/index.html> (accessed on 2 June 2022).
63. Cheng, H.Y.; Yu, C.C.; Hsu, K.C.; Chan, C.C.; Tseng, M.H.; Lin, C.L. Estimating Solar Irradiance on Tilted Surface with Arbitrary Orientations and Tilt Angles. *Energy* **2019**, *12*, 1427. [[CrossRef](#)]
64. Kraemer, F.A.; Palma, D.; Braten, A.E.; Ammar, D. Operationalizing Solar Energy Predictions for Sustainable, Autonomous IoT Device Management. *IEEE Internet Things J.* **2020**, *7*, 11803–11814. [[CrossRef](#)]
65. Wang, J.; Li, P.; Ran, R.; Che, Y.; Zhou, Y. A Short-Term Photovoltaic Power Prediction Model Based on the Gradient Boost Decision Tree. *Appl. Sci.* **2018**, *8*, 689. [[CrossRef](#)]
66. Wu, B.; Wang, L.; Zeng, Y.-R. Interpretable Wind Speed Prediction with Multivariate Time Series and Temporal Fusion Transformers. *Energy* **2022**, *252*, 123990. [[CrossRef](#)]
67. Zhang, H.; Zou, Y.; Yang, X.; Yang, H. A Temporal Fusion Transformer for Short-Term Freeway Traffic Speed Multistep Prediction. *Neurocomputing* **2022**, *500*, 329–340. [[CrossRef](#)]

-
68. Olmo, F.J.; Vida, J.; Foyo, I.; Castro-Diez, Y.; Alados-Arboledas, L. Prediction of Global Irradiance on Inclined Surfaces from Horizontal Global Irradiance. *Energy* **1999**, *24*, 689–704. [[CrossRef](#)]
 69. Said, S.A.M.; Hassan, G.; Walwil, H.M.; Al-Aqeeli, N. The Effect of Environmental Factors and Dust Accumulation on Photovoltaic Modules and Dust-Accumulation Mitigation Strategies. *Renew. Sustain. Energy Rev.* **2018**, *82*, 743–760. [[CrossRef](#)]