Heliyon 10 (2024) e30002

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

# Analysis of variables to determine their influence on renewable energy forecasting using ensemble methods

Carlos M. Travieso-González [a,b], Sergio Celada-Bernal [a,*], Alejandro Lomoschitz [c], Fidel Cabrera-Quintero [b]

[a] Institute for Technological Development and Innovation in Communications, IDeTIC, University of Las Palmas de Gran Caanria, ULPGC, Las Palmas de Gran Canaria, E35017, Spain
[b] Signals and Communications Departament, University of Las Palmas de Gran Caanria, Las Palmas de Gran Canaria, ULPGC, E35017, Spain
[c] Instituto de Oceanografía y Cambio Global, IOCAG, Universidad de Las Palmas de Gran Canaria, ULPGC, E35017, Las Palmas de Gran Canaria, Spain

ARTICLE INFO

*Index Terms:*
Renewable energy
Ensemble methods
Neural networks
Machine learning
Solar energy

ABSTRACT

Forecasting is of great importance in the field of renewable energies because it allows us to know the quantity of energy that can be produced, and thus, to have an efficient management of energy sources. However, determining which prediction system is more adequate is very complex, as each energy infrastructure is different. This work studies the influence of some variables when making predictions using ensemble methods for different locations. In particular, the proposal analyzes the influence of the aspects: the variation of the sampling frequency of solar panel systems, the influence of the type of neural network architecture and the number of ensemble method blocks for each model. Following comprehensive experimentation across multiple locations, our study has identified the most effective solar energy prediction model tailored to the specific conditions of each energy infrastructure. The results offer a decisive framework for selecting the optimal system for accurate and efficient energy forecasting. The key point is the use of short time intervals, which is independent of type of prediction model and of their ensemble method.

## 1. Introduction

In recent times, there has been a lot of talk about climate change, and the imperative necessity to mitigate its noticeable effects and the increase in global warming, its main cause. Global warming is mainly caused by greenhouse effect gases (which principal components are carbon dioxide ($CO_2$), methane ($CH_4$) and ozone ($O_3$)). These gases generate a sort of envelope around the atmosphere, containing solar rays and therefore maintaining temperature at an adequate level to allow life on the planet. However, the Industrial Revolution saw the rise of the emission of these aforementioned gases to an unhealthy level, due to the use of fossil fuels, like coal or petrol, to generate energy, which in turn has caused temperatures to skyrocket globally [1]. Deforestation also played a big part in this increase in global warming, since a decrease in forest mass meant fewer green plants to convert carbon dioxide ($CO_2$) into breathable oxygen ($O_2$).

With these issues acknowledged and the ever-growing global population, which needs more of both energy and oxygen, countries

* Corresponding author.
*E-mail address:* sergio.celada101@alu.ulpgc.es (S. Celada-Bernal).

https://doi.org/10.1016/j.heliyon.2024.e30002
Received 14 December 2023; Received in revised form 4 April 2024; Accepted 18 April 2024
Available online 21 April 2024

are finally starting to act accordingly and in December 2015, in Paris, 195 of them signed the historic Paris Agreement [2]. One of the long-terms goals of this agreement was to reduce the emission of greenhouse gases by promoting the use of renewable energy, which does not generate greenhouse effect gases [3]. This sort of energy sources can nowadays be even cheaper than their fossil counterparts [4], but despite this, they still need careful planning and foresight of the zone they aimed to be installed in to estimate their effectiveness [5].

The aim of this work is to explore and analyze three distinct variables: the sampling frequency of solar panel systems, the type of neural network architecture used, and the ensemble method's configuration. This innovative study provides detailed insights for designing solar energy predictors, applied across various locations to evaluate their impact on predictive accuracy and select the most efficient settings for different energy infrastructures. A pivotal aspect of our approach is the utilization of short time intervals in prediction models, which is independent of the prediction model type and ensemble method employed. The principal contributions of this manuscript are as follows.

1. Innovative Analysis of Key Variables: We delve into the influence of sampling frequency, neural network architecture type, and ensemble configuration on solar power prediction accuracy. This analysis is pivotal for understanding and improving forecasting performance across different energy infrastructures.
2. Optimization of Solar Energy Forecasting Models: Our research identifies optimal model configurations for various locations, offering a pathway to enhance solar energy prediction models tailored to specific environmental conditions.
3. Advancement in Short Interval Forecasting Techniques: Highlighting the importance of short time intervals in prediction models, we demonstrate their impact on improving forecasting precision, crucial for effective energy management and grid integration.
4. Comprehensive Evaluation of Ensemble Methods: We assess how ensemble methods can bolster prediction stability and accuracy, providing insights into their application in solar energy forecasting and setting a foundation for future research.

These contributions signify our manuscript's endeavor to refine and advance the methodologies for solar power generation forecasting, underlining the importance of precision and reliability in renewable energy systems. Our methodology is distinguished by its robust evaluation of three primary variables: sampling frequency of solar panel systems, the architecture of neural networks, and the configuration of ensemble methods. We developed a series of experiments to systematically assess the impact of each variable on the forecasting accuracy for solar energy production. The data used has been provided by different sources within the Canary Islands (Spain), which will be explained more in depth in following sections of this paper, that have measurements taken at different time intervals. Data taken from the islands of Gran Canaria and Tenerife will be used as reference, while data taken from the inverters in El Hierro will be used in the study integrating its values to match the references. After making the predictions, the results will be compared with those of different journals and papers to see the time interval that best matches acceptable standards.

The following process will be structured as follows: Section 2 will discuss previous related work, Section 3 will go over the starting data and the different installations where data was gathered, Section 4 will talk about the data processing procedure and the neural

**Table 1**
Overview of solar energy prediction models and outcomes.

| Model/Approach | Results/Main Findings | References |
|---|---|---|
| Predictive models for solar energy | Challenges due to the stochastic behavior of meteorological parameters. | Prakash & Singh [6] |
| Primitive models (clear sky model, regressive methods) | Demonstrated a foundational approach but overshadowed by the effectiveness of advanced AI methods. | Jain, Rathee, Kumar, Sambasivam, Boadh, Choudhary, Kumar & Kumar Singh [7]; Inman, Pedro, Coimbra[8] |
| Artificial Intelligence (AI) methods, especially Artificial Neural Networks (ANNs) | AI methods, notably ANNs, have shown superior effectiveness over primitive models. | Sfetsos & Coonick [9]; Nguyen, Pham, Duong & Vu [10]; Malik, Gehlot, Singh, Gupta & Thakur [11] |
| Multilayer Perceptron (MLP) and Feed-Forward Back-Propagation Networks | Used to predict solar irradiance on horizontal surfaces with a net error significantly lower than some linear methods. | Gardner & Dorling [12]; Paoli, Voyant, Muselli, & Nivet [13]; Achite, Banadkooki, Ehteram, Bouharira, Ahmed & Elshafie [14] |
| Precision evaluation using RMSE, nRMSE, and MAPE | Mixed results in predictions, hovering around 20–30 % accuracy in some cases. | Ben Ammar, Ben Ammar & A. Oualha [15]; Notton, Voyant, Fouilloy, Duchaud & Nivet [16]; Benali, Notton, Fouilloy, Voyant, Dizene [17]; Gao, Huang & Shi [18] |
| Random Forest in predicting global radiance | nRMSE error around 20–30 % for global radiance predictions. | Babar, Luppino, Boström & Anfinsen [19] |
| Importance of sampling intervals | The choice of sampling frequency significantly impacts the estimated amount of solar radiation. | Fathima, Nedumpozhimana, Lee, Winkler & Dev [20]; Kadirgama, Amirruddin & Bakar [21]; Notton, Voyant, Fouilloy, Duchaud & Nivet [16]; Schuss, Eichberger & Rahkonen [22]; Reikard [23] |
| Machine learning models for immediate rainfall forecasting | Proposed models for immediate probabilistic rainfall forecasting at 10-min intervals for short lead times. | Pirone, Cimorelli, Del Giudice & Pianés [24] |
| Integrating photovoltaic performance model with machine learning | Prediction of power output several hours in advance with a 5-min resolution. | Pombo, Bacher, Ziras, Bindner, Spataru, & Sørensen [25] |
| Ensemble models for solar power prediction | Ensemble models based on random forests and gradient boosting algorithms showed high performance in predicting power output. | Raj et al. [26]; Debani et al. [27]; Rahimi et al. [28] |

network system while Section 5 will go over the methodology employed to apply the neural network system. Finally, Section 6 will showcase the results of the experiment and spark the discussion that will eventually lead to the conclusions that will be showcased in Section 6.

## 2. Related work

In the quest to enhance the prediction accuracy of solar energy outputs, the scientific community has embarked on a journey from elementary forecasting models to the adoption of sophisticated artificial intelligence (AI) and machine learning (ML) algorithms. This transition reflects a broader trend towards utilizing computational intelligence for environmental and renewable energy applications.

Table 1 provides a curated overview of seminal and recent studies that have significantly contributed to this field. It highlights various modeling approaches, ranging from traditional statistical methods to advanced neural networks, and encapsulates their key findings and contributions to solar energy forecasting. By examining these works, we aim to showcase the evolution of predictive techniques and underline the pivotal role of AI and ML in driving forward the accuracy and reliability of solar irradiance predictions. This comparative analysis serves as a foundation for understanding current capabilities and identifying future research avenues in the pursuit of optimizing solar energy utilization. Our work builds on the fundamental ideas of the aforementioned studies by extending their methodologies. The proposed approach allows for a more detailed understanding of predictive dynamics and sets a new benchmark for future research in this field.

Following the synthesis of related works presented in the table, it is evident that the field of solar energy prediction is rapidly evolving, benefiting significantly from advancements in artificial intelligence and machine learning. The continuous refinement of predictive models, particularly through the application of Artificial Neural Networks (ANNs) and ensemble methods, demonstrates a promising trajectory towards achieving greater precision in solar irradiance forecasting. This evolution not only enhances the reliability of solar energy systems but also contributes to the optimization of renewable energy integration into the global energy mix. Future research directions are likely to focus on further improving the accuracy of predictions through the integration of multi-source data and the exploration of novel machine learning techniques. The collective efforts in this domain underscore the scientific community's commitment to addressing the complexities of renewable energy forecasting, ultimately paving the way for a more sustainable and resilient energy future.

In light of the evolving landscape of solar power generation forecasting, this study has meticulously selected methods that exhibit robustness in handling the inherent variabilities and nonlinearities of solar irradiance data. The deployment of Artificial Neural Networks (ANNs), including Multilayer Perceptron (MLP), and ensemble methods, is predicated on their proven efficacy in capturing complex patterns within time-series data, which is crucial for accurately predicting solar power output. These methods have been chosen due to their ability to process sequential data, learn from temporal dependencies, and generalize from historical data to unseen scenarios. Comparative evaluations with other deep learning techniques, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), were considered during the preliminary phase of this research. The selection process was guided by literature that demonstrates the particular effectiveness of ANNs and MLPs in solar energy forecasting, underscoring their capacity for high-dimensional data handling and adaptability to various meteorological conditions.

## 3. Materials

The starting data for our tests can be separated into two different categories: data provided by AEMET (Agencia Estatal de Meteorología) and data provided by a private person.

### 3.1. Data provided by AEMET

This set of data has been obtained from two different locations: Reina Sofía airport in Tenerife and Gran Canaria's airport.


**Fig. 1.** Location of reina sofía airport in tenerife.

The former is in the south of the island of Tenerife in the municipality of Granadilla de Abona (60 km away from the capital), near the localities of Atogo, Las Chafiras and San Isidro to the north and El Médano and Los Abrigos to the south. The climate near Reina Sofia Airport in Tenerife is also subtropical and markedly consistent, with a slight variation in temperature between seasons. This area benefits from a significant number of sunny days annually, which, coupled with moderate rainfall primarily in the winter months, provides a conducive environment for solar energy harvesting. The presence of trade winds helps mitigate extreme temperatures, ensuring stable conditions for solar panel efficiency. Fig. 1 shows this location.

The latter is in the south of the island of Gran Canaria, between the municipalities of Ingenio and Telde (17 km away from the capital) and between the localities of El Goro and Carrizal. Gran Canaria Airport is characterized by a subtropical climate that is mild and pleasant throughout the year. This region enjoys abundant sunshine with low precipitation, making it an ideal location for solar energy production. The consistent weather patterns, dominated by clear skies and minimal cloud cover, contribute to high solar irradiance levels, maximizing the potential for solar power generation. Again, Fig. 2 shows this location.

According to the documentation provided, at Reina Sofia Airport in Tenerife, samples are taken every 60 min. These measurements have been collected for each day of the years 2003, 2004, 2005, 2006 and 2007. The measurements indicate solar radiation in kJ/m2. According to the documentation provided, at Gran Canaria's airport, samples are taken every 60 min. These measurements have been collected for each day of the years 2004, 2005, 2006 and 2007. The measurements indicate solar radiation in kJ/m2.

For confidentiality and security reasons, AEMET does not specify the exact location of the photovoltaic panels or any specific details about the installation or the way in which the data is obtained.

### 3.2. Data provided by a private person

This set of data has been taken from a country house in the south of the small island of El Hierro. The southern part of El Hierro experiences a climate that varies more significantly with elevation but generally offers favorable conditions for solar energy due to its geographical location. The area is exposed to consistent solar radiation year-round, with higher altitudes receiving more precipitation but also benefiting from clear skies. The unique climatic conditions, including the occasional presence of horizontal rain in higher altitudes, highlight the island's potential for diverse renewable energy sources, including solar power. Fig. 3 shows the Location of the private building in El Hierro.

The installation is composed of three single phase inverter of 3.3 kW nominal power each from INGETEAM. All panels were bought from Spanish manufactures ISOFOTON. Each of them is connected to a set of panels as follows.

- Inverter 2 gives 3.6 kW peak and is connected to 24 panels of 150 W.
- Inverter 3 gives 3.3 kW peak and is connected to 22 panels of 150 W.

According to the provided documentation inverter 2 takes sample every 2 min while inverter 3 takes them every 15 min. In these cases the measurements are in terms of generated power (W).These measurements have been compiled into a spreadsheet for every day from 2016, 2017 and 2018.

## 4. Methods

Raw data from the sources can't be utilized until processed and treated, then and only then it can be implemented into our system.

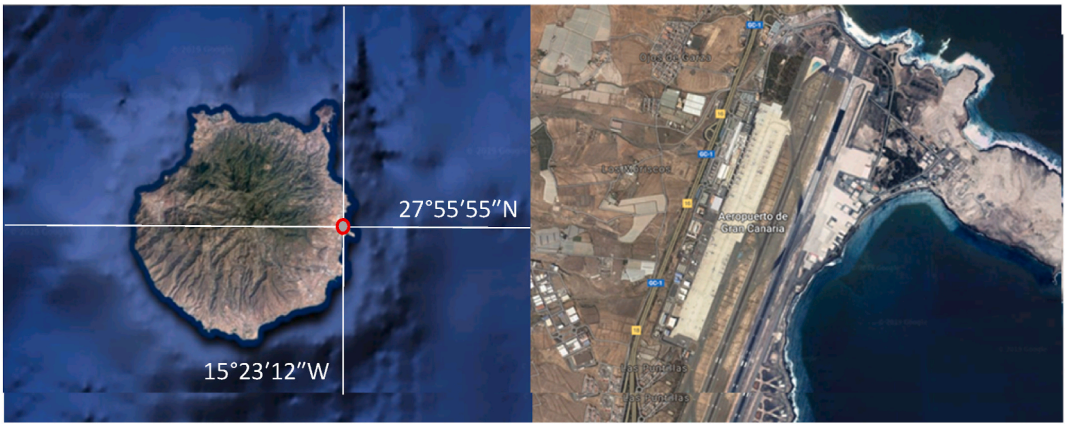


**Fig. 2.** Location of Gran Canaria's airport.

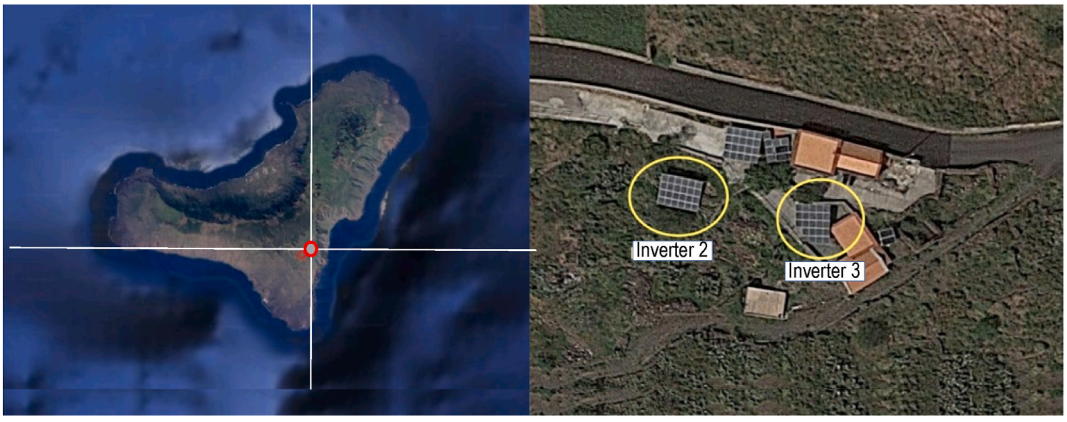


**Fig. 3.** Location of the private building in El Hierro.

### 4.1. Integrating the data

Before importing, an integration of each year's data will be carried out to compare the different time intervals between measurements and all data, as explained in Fig. 4.

This will require the loading of the appropriately arranged data and the execution of the following flux diagram in Fig. 5.

The function following this flux diagram must create a new matrix that takes data from the input matrix two at a time, adds them up and then divides them by two. This makes it so that if the input matrix had samples every 2 min, then the output matrix has them every 4 min. Repeating this process and changing the input will provide data for all desired time intervals shown in Fig. 4.

### 4.2. Neural network system

Artificial neural networks date back to the 1940s [29], however, they were not successful, as the computational resources required were simply non-existent back then. The idea of this ANNs was to imitate the way our neurons work, by creating a set of connections through experience and learning by working together that would lead to knowledge being retained [30]. More complex architectures of ANNs need additional intermediate perceptron to apply extra conditions and information to improve results. Various layers, all connected to all inputs, are therefore needed. Every layer aims to find a new characteristic to better classify the data. Fig. 6 shows the generic schematic of an ANN architecture.

There are multiple types of ANNs architectures [31]. To obtain the results of this paper, we have decided to select 3 of them. The first one is Multi-Layer Perceptron (MLP), the second one is Long/Short Term Memory (LSTM) and the third one is Gated Recurrent Units (GRU). The main differences are explained below.

#### 4.2.1. Multi-layer perceptron (MLP)

Multi-Layer Perceptron (MLP) is one of the most widely used types of ANNs. An input layer serves as a buffer that distributes input signals to the next layer, which is a hidden layer. Each neuron in the hidden layer communicates with all neurons in the next hidden

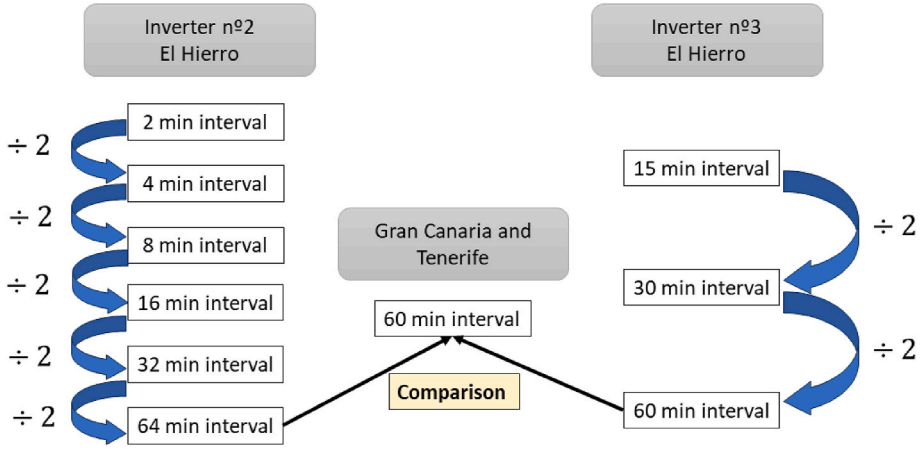


**Fig. 4.** Compression of the different data series.

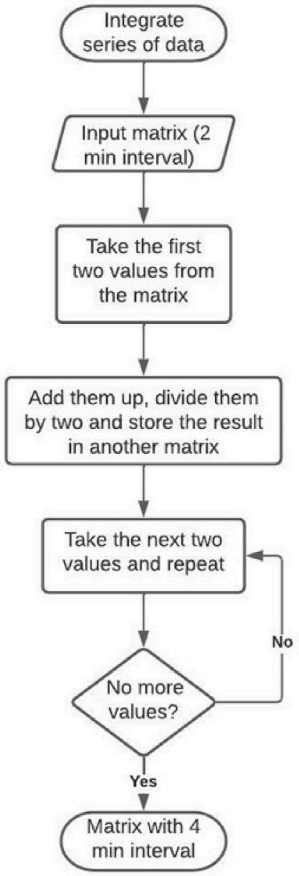


**Fig. 5.** Flux diagram explaining the integration of the data series.

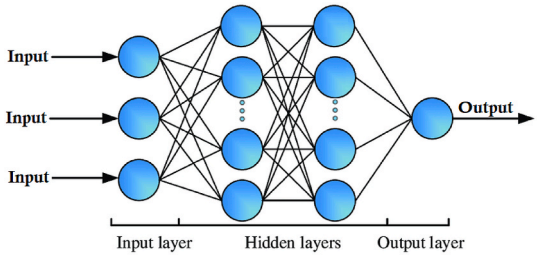


**Fig. 6.** Generic schematic of an ANN architecture.

layer having in each connection a typical weight factor. Then, each unitary artificial neuron in the hidden layer adds up its input, processes it with a transfer and distributes the result to the output layer [32].

#### *4.2.2. LSTM*

Long/Short Term Memory (LSTM) networks try to combat the vanishing/exploding gradient problem by introducing gates and an explicitly defined memory cell. These are inspired mostly by circuitry, not so much biology. Each neuron has a memory cell and three gates: input, output and forget. The function of these gates is to safeguard the information by stopping or allowing the flow of it. The input gate determines how much of the information from the previous layer gets stored in the cell. The output layer takes the job on the other end and determines how much of the next layer gets to know about the state of this cell [33].

#### *4.2.3. GRU*

Gated Recurrent Units (GRU) networks are variation of LSTM networks. They have one less gate and are wired slightly differently: instead of an input, an output and a forget gate, they have an update gate. This update gate determines how much information to keep

from the last state and how much information to let in from the previous layer. The reset gate works very similarly to the forget gate of an LSTM but is located slightly differently. They always send their full state, they do not have an exit gate [34].

#### 4.2.4. Ensemble method

When performing the different experiments with a model with a single neural network, the results may be very different from one experiment to the next. It is not recommended to use a model with a single neural network because the selection of a model with a single network ignores potentially useful information. In other words, not all a model with a single network store the same information, which means that the a model with a single network chosen as optimal may not contain all the information contained in the data on which it was trained.

To solve this variability in the results, one possible solution is to make an ensemble method. The operation is quite simple, consisting of applying multiple models with a single neural network independently, i.e. *n* models with a single neural network working in parallel, to subsequently perform the arithmetic mean of all the results. With this technique, a set is created by averaging the results provided by each model. In this way, greater accuracy will be obtained by discriminating models that do not fit the mean. Fig. 7 shows the diagram, of the ensemble method.

In our design we have opted for a design consisting of 3 ensemble methods. Each ensemble method in turn consists of 1000 models with a neutral network with a different architecture for each ensemble. For ensemble number 1 each of the models will have a neural network with an MLP architecture, for ensemble number 2 an LSTM architecture and for ensemble number 3 a GRU architecture. Fig. 8 shows the proposed design.

The proposed ensemble design offers a flexible and robust solution for enhancing the accuracy of solar power generation forecasts. By incorporating three distinct ensemble methods, each utilizing a different neural network architecture (MLP, LSTM, and GRU), our approach provides a comprehensive framework that can adapt to various data characteristics and forecasting scenarios. This diversity in model architecture allows for a nuanced analysis of temporal and spatial data patterns, ensuring that the ensemble method selected for a given situation is the most suited for maximizing predictive performance. Furthermore, the use of 1000 models within each ensemble method significantly increases the reliability of predictions, as it reduces the risk associated with overfitting and ensures that the forecasts are robust against variability in the input data. The flexibility of our design to accommodate any neural network architecture also future-proofs the approach, making it adaptable to advancements in neural network designs and architectures. Overall, this multi-ensemble strategy not only elevates the accuracy of solar power generation forecasts but also offers a scalable and adaptable framework that can be tailored to meet the specific needs of different energy infrastructures, making it a valuable tool for optimizing the integration of solar energy into power grids.

## 5. Experimental methodology

The proposed experimental methodology in this study aims to analyze the impact of various variables on solar energy prediction using 3 ensemble methods with neural networks across different locations. To conduct the experiment, multiple tests will be conducted at diverse sites, taking into account factors such as varying sampling frequencies of solar panel systems, the choice of neural network architecture, and the number of ensemble method blocks for each model. These experiments will facilitate the identification of the most appropriate solar energy predictor for each energy infrastructure, considering the unique conditions of each location. A crucial aspect of this methodology is the utilization of short time intervals, regardless of the prediction model and ensemble method employed. The outcomes obtained from this study will yield valuable data and statistics to determine the optimal system for solar energy

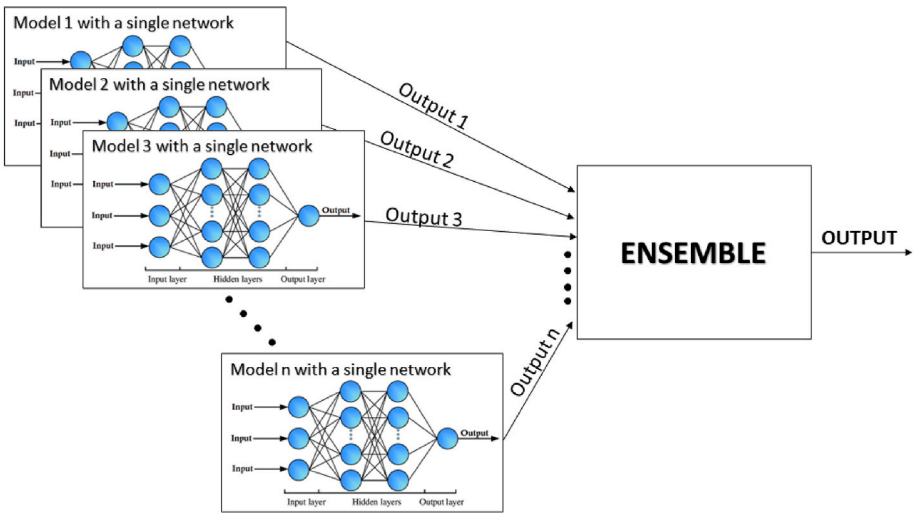


**Fig. 7.** Diagram of ensemble method and connections.

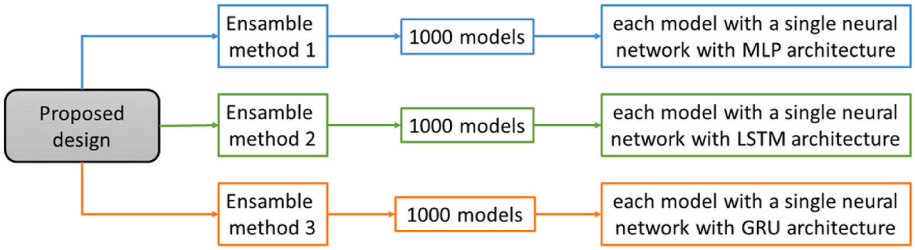


**Fig. 8.** Diagram of the proposed design.

prediction in various scenarios, thereby contributing to the more efficient management of renewable energy sources.

For our tests, a neural network will be used and generated using MATLAB similar to Ref. [35]. This network utilizes supervised error-correction learning, making the weights of the different neurons change during training depending on the error in the output layer [36]. This algorithm is composed of two different phases.

- Training phase. During the training phase, the network processes a set of input variables to forecast subsequent values. Following the prediction, it performs an error assessment by comparing the predicted values against the actual data. This step is critical for the network to adjust and optimize its parameters, enhancing its predictive accuracy over time.
- Test phase. In this stage several series are taken in order to make a prediction to compare with the original series. This is made with all the different series it is intended to work with. Finally, with each test different parameters will be calculated and stored in a matrix. This matrix's columns store the series of data utilized while the rows contain the different errors achieved.

Table 2 shows the data used for the proposed design according to location.

### 5.1. *Applying the training phase*

The first stage when applying this backpropagation network will be to create the necessary variables to work on the training phase. This training phase follows the following flow chart shown in Fig. 9 and will be explained throughout this section.

First, we set the initial variables, notably the size of the window, which is related to the number of neurons in the input layer, and the number of repetitions for the training phase, which in this paper is 500. The P and T matrices are then created and are traversed and filled. This process prepares a data set for training using a sliding window technique. For each data point, a time segment(tv) is created that serves as the input(P), while the next data point is used as the target value(T) to predict. This method captures the time dependence in the data by populating the P and T matrices with consecutive segments and their respective target values. Finally, the orientation of the P matrix is adjusted to meet the requirements of the model's input format.

The *P* matrix will be the one containing the predictions. The *T* matrix will be the one used to evaluate the accuracy of our predictions.

### 5.2. *Applying the test phase*

Next comes the test phase of our network, which consist of a test for every year we have data with the exception of the one we performed the training phase with. This particular phase starts similarly to the last one, by creating a set of empty matrices which will contain the future predictions and follows the diagram shown in Fig. 10.

First, the initial variables are created. Then the Tr matrices are created, which are the actual values of the data, and the *Y20XX* matrix, which corresponds to the predicted values.

For each neural network, matrices *Tr_1, Y20XX_1, Tr_2, Y20XX_2, ….,Tr_n,Y20XX_n*, are created and populated, *n* being the number of neural networks that form the ensemble.

Once we have all these matrices, we proceed to calculate the arithmetic mean and obtain a single value for each matrix (*Tr, Y20XX*). The following expressions are used to calculate the values *Y20XX* and *Tr*:

**Table 2**
Data used for the proposed design.

| Location | Data training phase | Data test phase |
|---|---|---|
| Gran Canaria | Data Year 2004 | Data years 2005,2006 and 2007 |
| Tenerife | Data year 2005 | Data years 2004, 2005, 2006 and 2007 |
| El Hierro inverter 2 | Data year 2016 | Data years 2017 and 2018 |
| El Hierro inverter 3 | Data year 2006 | Data yeard 2017 and 2018 |

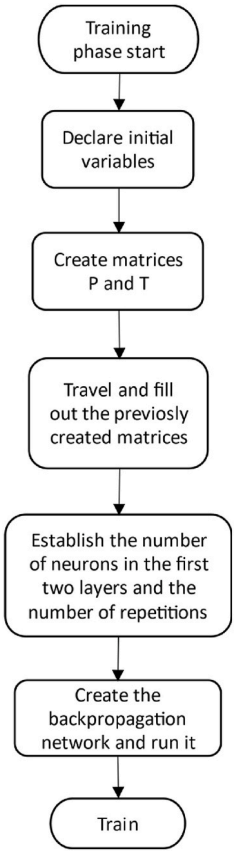


**Fig. 9.** Flux diagram explaining the training phase.

$$Y20XX = \frac{1}{n} \cdot \sum_{i=1}^{n} Y20XX\_n \tag{1}$$

$$Tr = \frac{1}{n} \cdot \sum_{i=1}^{n} Tr\_n \tag{2}$$

In our case $n = 1000$. That is to say, we are going to work with 1000 networks working in parallel so that we can see how the increase in the number of networks in the ensemble affects, in order to determine an optimal number.

Finally, the matrix with the final results of the errors obtained is created. In this matrix each column will refer to a year, and each row to a different error, according to the following list.

- First row will contain the minimal absolute error.
- Second row will contain the maximum absolute error.
- Third row will contain the mean absolute error.
- Fourth row the standard deviation.
- Fifth row will contain the *Root-Mean-Square-Error,* calculated as:

$$RMSE = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} (Tr - Y20XX)^2} \tag{3}$$

- Sixth row will contain the *normalized Root-Mean-Square-Error (nRMSE),* calculated as:

$$nRMSE = \frac{RMSE}{mean\ (Tr)} \bullet 100 \tag{4}$$

- Seventh row will contain the *Mean Absolute Percentage Error,* calculated as:

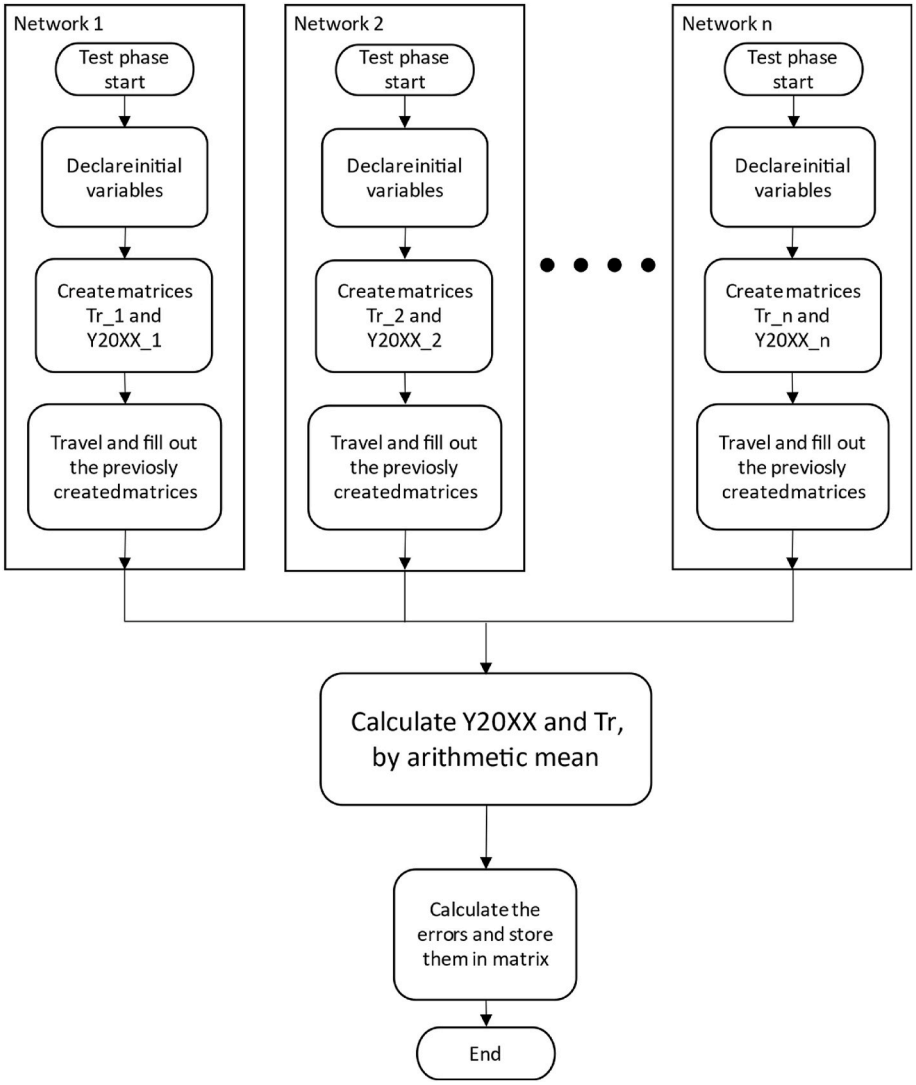


**Fig. 10.** Flux diagram of ensemble of networks explaining the test phase.

$$MAPE = \frac{1}{n} \cdot \sum_{i=1}^{n} \frac{Tr - Y20XX}{Tr} \tag{5}$$

## 6. Results and discussion

The results of our study demonstrate the superior effectiveness and reliability of our proposed forecasting model. By employing short sampling intervals and optimizing ensemble configurations, we achieved a significant reduction in forecasting error rates compared to traditional methods. This section will be organised as numerical results and graphical results. In the former, the errors stored in the matrix for each year and for each type of neural network architecture, will be shown in tabular form. In the second, the real values (*Tr*) and the values obtained in the prediction (*Y20XX*), for the year and for the type of architecture showing the best values from the previous tables, are shown in graphical form.

The first results shown will be the results from the island of Gran Canaria.

### *6.1. Results from Gran Canaria*

For the Gran Canaria data, the year 2004 will be used as the training year and for the test phase the results of the years 2005, 2006 and 2007 will be obtained.

#### 6.1.1. Numerical results

For this subsection each test is run once for each year with recorded data, changing the test year each time, and calculating the mean value of each error for each type of neural network architecture. Table 3 shows these results.

#### 6.1.2. Graphical results

Analysing the results of the previous table, when in the test phase the 2004 data is used and a GRU typology is applied in the neural network, the best values are obtained. Fig. 11 shows the comparison of actual and predicted values, refers to the solar radiation measurements (kJ/m2), in the first 24 h (1 day) and in the first 168 h (1 week) of the year 2006.

### 6.2. Results from tenerife

Numerical results of Tenerife (Table 4) follow the same reasoning explained in the Gran Canaria part of this section. For the Tenerife data, the year 2003 will be used as the training year and for the test phase the results of the years 2004, 2005, 2006 and 2007 will be obtained.

#### 6.2.1. Numerical results

Numerical results of Tenerife (Table 4) follow the same reasoning explained in the Gran Canaria part of this section.

#### 6.2.2. Graphical results

Just like with the data from Gran Canaria, the first year was used for the training phase for the graphical results displayed in the following subsection. Analysing the results of the previous table, when in the test phase the 2007 data is used and a GRU typology is applied in the neural network, the best values are obtained. Fig. 12 shows the comparison of real and predicted values, refers to the solar radiation measurements (kJ/m2), in the first 24 samples/hours (1 day) and in the first 168 samples/hours (1 week) of the year 2007.

### 6.3. Results from El Hierro

Data from El Hierro can be divided based on two things: the inverter it comes from, and the time interval between samples.

#### 6.3.1. Inverter 2

Intervals for this series of data will range from 2 min to up to 64 min as mentioned previously.

*6.3.1.1. Numerical results.* Having data from the years 2016,2017 and 2018, the first year will be used as the training year and data from 2016 will be showcased here for different time intervals between samples. The following tables show the results obtained for sample intervals between 2 min and 64 min for the years 2017 (Table 5) and 2018 (Table 6).

*6.3.1.2. Graphical results.* The year 2016 was used for the training phase for the graphical results displayed in the following subsection. Analysing the results of the previous table, when in the test phase the 2018 data, the interval between samples is 2 min is used

**Table 3**
Results obtained from the island of Gran Canaria.

| | | 2005 | 2006 | 2007 |
|---|---|---|---|---|
| MLP | Min. Error (kJ/m$^2$) | $1{,}21 \cdot 10^{-2}$ | $7{,}49 \cdot 10^{-3}$ | $2{,}70 \cdot 10^{-3}$ |
| | Max. Error (kJ/m$^2$) | 270,03 | 310,09 | 260,21 |
| | Mean Error (kJ/m$^2$) | 15,80 | 13,76 | 16,97 |
| | T.Deviation (kJ/m$^2$) | 24,28 | 21,42 | 24,96 |
| | RMSE (kJ/m$^2$) | 28,98 | 25,48 | 30,18 |
| | nRMSE (%) | 42,86 | 35,95 | 37,23 |
| | MAPE (%) | $7{,}64 \cdot 10^{-3}$ | $4{,}10 \cdot 10^{-3}$ | $2{,}97 \cdot 10^{-3}$ |
| LSTM | Min. Error (kJ/m$^2$) | $1{,}34 \cdot 10^{-2}$ | $1{,}71 \cdot 10^{-3}$ | $3{,}96 \cdot 10^{-5}$ |
| | Max. Error (kJ/m$^2$) | 213,03 | 264,68 | 238,42 |
| | Mean Error (kJ/m$^2$) | 12,86 | 10,51 | 12,55 |
| | T.Deviation (kJ/m$^2$) | 22,16 | 18,95 | 22,56 |
| | RMSE (kJ/m$^2$) | 25,62 | 21,67 | 25,82 |
| | nRMSE (%) | 39,24 | 32,84 | 34,00 |
| | MAPE (%) | $2{,}47 \cdot 10^{-3}$ | $1{,}14 \cdot 10^{-3}$ | $8{,}54 \cdot 10^{-4}$ |
| GRU | Min. Error (kJ/m$^2$) | $8{,}84 \cdot 10^{-3}$ | $8{,}35 \cdot 10^{-4}$ | $1{,}17 \cdot 10^{-4}$ |
| | Max. Error (kJ/m$^2$) | 203,64 | 262,37 | 235,09 |
| | Mean Error (kJ/m$^2$) | 13,00 | 10,55 | 12,74 |
| | T.Deviation (kJ/m$^2$) | 21,89 | 18,65 | 22,52 |
| | RMSE (kJ/m$^2$) | 25,46 | 21,42 | 25,87 |
| | nRMSE (%) | 38,82 | 32,39 | 34,10 |
| | MAPE (%) | $2{,}84 \cdot 10^{-3}$ | $1{,}35 \cdot 10^{-3}$ | $9{,}75 \cdot 10^{-4}$ |

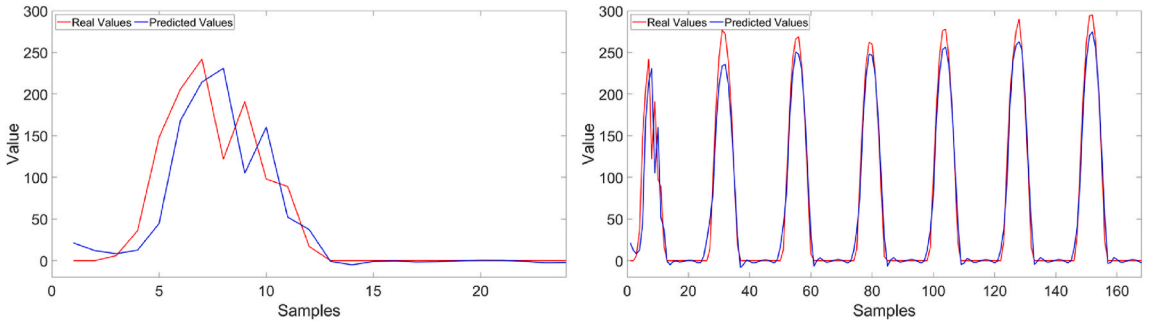


**Fig. 11.** Graphical results of Gran Canaria.

**Table 4**
Results obtained from the island of Tenerife.

| | | 2004 | 2005 | 2006 | 2007 |
|---|---|---|---|---|---|
| MLP | Min. Error (kJ/m$^2$) | $2{,}06{\cdot}10^{-4}$ | $4{,}14{\cdot}10^{-3}$ | $9{,}36{\cdot}10^{-4}$ | $8{,}47{\cdot}10^{-3}$ |
| | Max. Error (kJ/m$^2$) | 241,08 | 222,27 | 261,90 | 245,97 |
| | Mean Error (kJ/m$^2$) | 19,32 | 18,13 | 22,15 | 22,63 |
| | T.Deviation (kJ/m$^2$) | 27,10 | 25,50 | 30,04 | 30,31 |
| | RMSE (kJ/m$^2$) | 33,28 | 31,29 | 37,32 | 37,83 |
| | nRMSE (%) | 42,86 | 39,83 | 40,36 | 39,09 |
| | MAPE (%) | $3{,}15{\cdot}10^{-3}$ | $2{,}98{\cdot}10^{-3}$ | $2{,}17{\cdot}10^{-3}$ | $2{,}21{\cdot}10^{-3}$ |
| LSTM | Min. Error (kJ/m$^2$) | $1{,}51{\cdot}10^{-4}$ | $9{,}99{\cdot}10^{-5}$ | $1{,}76{\cdot}10^{-4}$ | $4{,}64{\cdot}10^{-4}$ |
| | Max. Error (kJ/m$^2$) | 226,14 | 189,43 | 230,37 | 207,35 |
| | Mean Error (kJ/m$^2$) | 14,16 | 13,60 | 16,04 | 16,15 |
| | T.Deviation (kJ/m$^2$) | 23,33 | 21,35 | 23,88 | 24,12 |
| | RMSE (kJ/m$^2$) | 27,29 | 25,32 | 28,76 | 29,02 |
| | nRMSE (%) | 38,32 | 35,15 | 35,11 | 34,08 |
| | MAPE (%) | $2{,}03{\cdot}10^{-3}$ | $1{,}90{\cdot}10^{-3}$ | $1{,}27{\cdot}10^{-3}$ | $1{,}30{\cdot}10^{-3}$ |
| GRU | Min. Error (kJ/m$^2$) | $2{,}91{\cdot}10^{-5}$ | $2{,}17{\cdot}10^{-3}$ | $4{,}31{\cdot}10^{-4}$ | $4{,}62{\cdot}10^{-4}$ |
| | Max. Error (kJ/m$^2$) | 220,73 | 198,07 | 241,68 | 204,42 |
| | Mean Error (kJ/m$^2$) | 13,27 | 12,27 | 14,47 | 14,56 |
| | T.Deviation (kJ/m$^2$) | 22,22 | 19,63 | 22,29 | 22,43 |
| | RMSE (kJ/m$^2$) | 25,88 | 23,15 | 26,57 | 26,74 |
| | nRMSE (%) | 38,38 | 34,98 | 35,03 | 34,01 |
| | MAPE (%) | $1{,}46{\cdot}10^{-3}$ | $1{,}30{\cdot}10^{-3}$ | $8{,}70{\cdot}10^{-4}$ | $8{,}83{\cdot}10^{-4}$ |

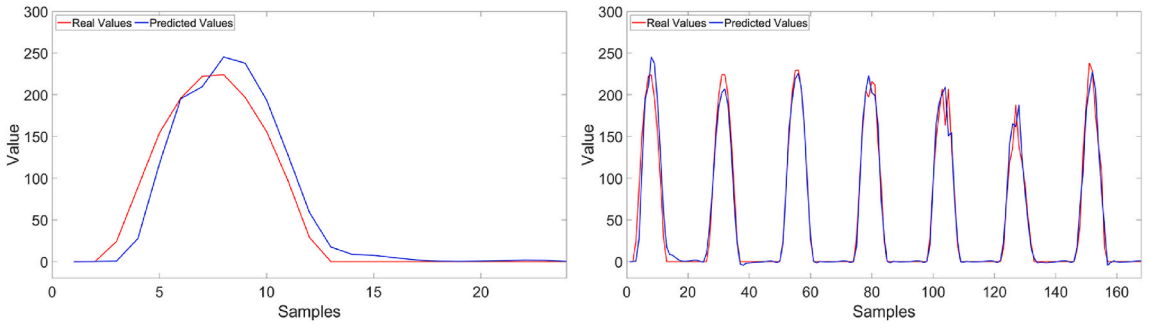


**Fig. 12.** Graphical results of Tenerife.

and a GRU typology is applied in the neural network, the best values are obtained. Fig. 13 shows the comparison of real and predicted values, measurements in terms of generated power (W), in the interval 4330 to 5040 samples (seven day) and in the first 4970 samples (1 week) of the year 2018.

### *6.3.2. Inverter 3*

Intervals for this series of data will range from 15 min to up to 60 min as mentioned previously. The year 2016 was used for the training phase.

**Table 5**
Results obtained from the island of El Hierro from the inverter 2, year 2017.

| | | 2 min. | 4 min | 8 min. | 16 min. | 32 min. | 64 min. |
|---|---|---|---|---|---|---|---|
| MLP | Min. Error (W) | $5{,}21{\cdot}10^{-4}$ | $8{,}69{\cdot}10^{-5}$ | $5{,}25{\cdot}10^{-3}$ | $1{,}99{\cdot}10^{-3}$ | $1{,}23{\cdot}10^{-3}$ | $2{,}96{\cdot}10^{-3}$ |
| | Max. Error (W) | 2746,33 | 2620,60 | 2380,17 | 2180,32 | 2261,30 | 2343,04 |
| | Mean Error (W) | 57,78 | 65,78 | 73,00 | 80,94 | 98,27 | 135,82 |
| | T.Deviation (W) | 157,51 | 161,56 | 162,60 | 170,23 | 188,05 | 196,13 |
| | RMSE (W) | 167,77 | 174,44 | 178,24 | 188,49 | 212,17 | 238,56 |
| | nRMSE (%) | 26,73 | 27,79 | 28,39 | 30,02 | 33,79 | 37,98 |
| | MAPE (%) | $8{,}83{\cdot}10^{-6}$ | $2{,}03{\cdot}10^{-5}$ | $4{,}32{\cdot}10^{-5}$ | $9{,}93{\cdot}10^{-5}$ | $2{,}42{\cdot}10^{-4}$ | $7{,}25{\cdot}10^{-4}$ |
| LSTM | Min. Error (W) | $2{,}52{\cdot}10^{-4}$ | $5{,}71{\cdot}10^{-5}$ | $5{,}24{\cdot}10^{-4}$ | $1{,}42{\cdot}10^{-3}$ | $4{,}04{\cdot}10^{-4}$ | $4{,}34{\cdot}10^{-3}$ |
| | Max. Error (W) | 2722,92 | 2573,37 | 2398,33 | 2120,81 | 1886,15 | 1999,33 |
| | Mean Error (W) | 60,26 | 62,24 | 73,73 | 77,78 | 92,83 | 105,66 |
| | T.Deviation (W) | 155,55 | 159,80 | 157,82 | 162,46 | 168,48 | 182,23 |
| | RMSE (W) | 166,81 | 171,49 | 174,19 | 180,12 | 192,36 | 210,63 |
| | nRMSE (%) | 26,57 | 27,32 | 27,75 | 28,69 | 30,63 | 33,53 |
| | MAPE (%) | $8{,}59{\cdot}10^{-6}$ | $1{,}93{\cdot}10^{-5}$ | $3{,}98{\cdot}10^{-5}$ | $8{,}83{\cdot}10^{-5}$ | $2{,}44{\cdot}10^{-4}$ | $5{,}53{\cdot}10^{-4}$ |
| GRU | Min. Error (W) | $4{,}93{\cdot}10^{-4}$ | $3{,}32{\cdot}10^{-4}$ | $3{,}08{\cdot}10^{-5}$ | $2{,}05{\cdot}10^{-3}$ | $4{,}83{\cdot}10^{-4}$ | 5,24E-04 |
| | Max. Error (W) | 2810,09 | 2575,19 | 2422,09 | 2156,39 | 1854,23 | 2024,47 |
| | Mean Error (W) | 57,71 | 68,92 | 69,07 | 79,52 | 92,44 | 108,86 |
| | T.Deviation (W) | 155,84 | 158,35 | 159,84 | 164,78 | 169,62 | 183,28 |
| | RMSE (W) | 166,18 | 172,70 | 174,13 | 182,96 | 193,17 | 213,17 |
| | nRMSE (%) | 26,47 | 27,51 | 27,74 | 29,14 | 30,76 | 33,93 |
| | MAPE (%) | $7{,}71{\cdot}10^{-6}$ | $1{,}26{\cdot}10^{-5}$ | $3{,}56{\cdot}10^{-5}$ | $7{,}45{\cdot}10^{-5}$ | $2{,}17{\cdot}10^{-4}$ | $3{,}50{\cdot}10^{-4}$ |

**Table 6**
Results obtained from the island of El Hierro from the inverter 2, year 2018.

| | | 2 min. | 4 min | 8 min. | 16 min. | 32 min. | 64 min. |
|---|---|---|---|---|---|---|---|
| MLP | Min. Error (W) | $8{,}16{\cdot}10^{-4}$ | $8{,}09{\cdot}10^{-4}$ | $9{,}13{\cdot}10^{-4}$ | $5{,}62{\cdot}10^{-3}$ | $3{,}07{\cdot}10^{-3}$ | $6{,}01{\cdot}10^{-3}$ |
| | Max. Error (W) | 2533,07 | 2528,08 | 2223,09 | 2042,98 | 1982,90 | 2201,44 |
| | Mean Error (W) | 57,49 | 64,28 | 70,79 | 79,07 | 94,55 | 134,80 |
| | T.Deviation (W) | 156,28 | 156,20 | 156,24 | 163,43 | 178,81 | 197,44 |
| | RMSE (W) | 166,51 | 168,91 | 171,53 | 181,55 | 202,27 | 239,06 |
| | nRMSE (%) | 26,14 | 26,51 | 26,92 | 28,49 | 31,74 | 37,50 |
| | MAPE (%) | $8{,}79{\cdot}10^{-6}$ | $1{,}97{\cdot}10^{-5}$ | $4{,}12{\cdot}10^{-5}$ | $9{,}45{\cdot}10^{-5}$ | $2{,}29{\cdot}10^{-4}$ | $7{,}24{\cdot}10^{-4}$ |
| LSTM | Min. Error (W) | $2{,}72{\cdot}10^{-4}$ | $2{,}19{\cdot}10^{-5}$ | $3{,}53{\cdot}10^{-3}$ | $6{,}87{\cdot}10^{-4}$ | $4{,}16{\cdot}10^{-3}$ | $2{,}47{\cdot}10^{-3}$ |
| | Max. Error (W) | 2536,60 | 2533,55 | 2079,17 | 2131,62 | 1629,15 | 1707,04 |
| | Mean Error (W) | 59,88 | 60,10 | 71,31 | 74,95 | 89,06 | 102,80 |
| | T.Deviation (W) | 154,58 | 154,53 | 151,13 | 155,71 | 160,13 | 179,53 |
| | RMSE (W) | 165,77 | 165,80 | 167,11 | 172,80 | 183,22 | 206,87 |
| | nRMSE (%) | 26,02 | 26,02 | 26,23 | 27,12 | 28,75 | 32,45 |
| | MAPE (%) | $8{,}45{\cdot}10^{-6}$ | $1{,}83{\cdot}10^{-5}$ | $3{,}78{\cdot}10^{-5}$ | $8{,}26{\cdot}10^{-5}$ | $2{,}32{\cdot}10^{-4}$ | $5{,}38{\cdot}10^{-4}$ |
| GRU | Min. Error (W) | $8{,}53{\cdot}10^{-5}$ | $4{,}84{\cdot}10^{-4}$ | $1{,}23{\cdot}10^{-4}$ | $1{,}60{\cdot}10^{-4}$ | $3{,}88{\cdot}10^{-5}$ | $3{,}41{\cdot}10^{-3}$ |
| | Max. Error (W) | 2536,84 | 2548,37 | 2106,26 | 2458,89 | 1719,80 | 1658,33 |
| | Mean Error (W) | 57,19 | 66,69 | 66,45 | 77,16 | 88,39 | 105,99 |
| | T.Deviation (W) | 154,92 | 153,16 | 153,29 | 158,00 | 161,09 | 180,87 |
| | RMSE (W) | 165,14 | 167,05 | 167,07 | 175,83 | 183,74 | 209,63 |
| | nRMSE (%) | 25,92 | 26,22 | 26,22 | 27,59 | 28,83 | 32,88 |
| | MAPE (%) | $7{,}52{\cdot}10^{-6}$ | $1{,}13{\cdot}10^{-5}$ | $3{,}31{\cdot}10^{-5}$ | $7{,}06{\cdot}10^{-5}$ | $2{,}01{\cdot}10^{-4}$ | $3{,}29{\cdot}10^{-4}$ |

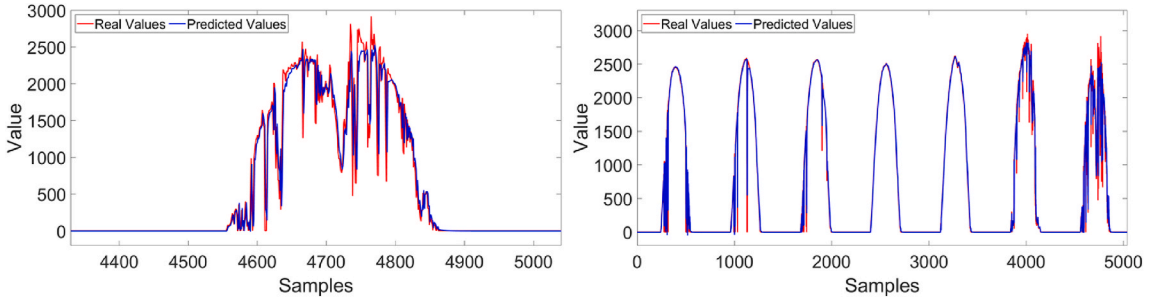


**Fig. 13.** Graphical results of inverter 2 of El Hierro.

*6.3.2.1. Numerical results.* Having data from the years 2016,2017 and 2018, the first year will be used as the training year and data from 2016 will be showcased here for different time intervals between samples. The following tables show the results obtained for the years 2017 (Table 7) and 2018 (Table 8).

*6.3.2.2. Graphical results.* Just like inverter 2, the year 2016 was used for the training phase for the graphical results displayed in the following subsection. Analysing the results of the previous table, when in the test phase the 2018 data, the interval between samples is 15 min is used and a LSTM typology is applied in the neural network, the best values are obtained. Fig. 14 shows the comparison of real and predicted values, measurements in terms of generated power (W), in the first 96 samples (1 day) and in the first 672 samples (1 week) of the year 2018.

### *6.4. Discussion*

#### *6.4.1. Influence of the type of architecture and sampling time*

As mentioned above, the parameter used to evaluate the results is the normalized root mean square error or nRMSE. To facilitate the analysis of this parameter, the results for each location when a single neural network is used for the tests are plotted, thus complementing the results of the previous tables which were based on tests with ensemble neural networks.

The following tables show the values for Gran Canaria (Table 9) and Tenerife (Table 10).

As can be seen in the tables, with the data from Gran Canaria and Tenerife (samples every 60 min), the type of architecture has a considerable influence. Using LSTM or GRU type architectures the nRMSE value is reduced up to 11 points with respect to MLP with the Tenerife data and 7 points with the Gran Canaria data.

Table 11 shows the data for inverter 2. As discussed throughout the paper, the sampling data varies from 2 to 64 min.

In order to visualise the above table more intuitively, Fig. 15 Comparison nRMSE inverter 2 El Hierro. graphs are shown.

Analyzing the data presented in the table and the corresponding graphs, it becomes evident that the choice of neural network architecture becomes increasingly influential as the time interval between data sampling extends. This suggests that the interval duration plays a significant role in the predictive performance of different architectures. When samples are taken every 2, 4, and 8 min, the nRMSE results are very similar. However, as the time interval between samples increases, GRU and LSTM give better results than MLP.

As with inverter 2, Table 12 and Fig. 16 of the nRMSE results varying the sampling time are shown for inverter 3.

In this case, as the first reference interval is one data every 15 min, it can already be seen that there is a small difference depending on the type of architecture. As was the case for inverter 2, as the time interval between samples increases, GRU and LSTM give better results than MLP.

#### *6.4.2. Sensitivity of the proposed design*

The sensitivity of the proposed design is shown in Fig. 17. The figure shows this analysis for the 2006 Tenerife data (measured in kJ/m2) using an ensemble method with 1000 models with a GRU architecture. Also, the year 2018 data from inverter 3 (measurements in W) an ensemble method with 1000 models with a GRU architecture is used.

The sensitivity analysis plots reveal the high accuracy shown by the predictive proposed design used to estimate solar panel power values. The points of both plots align closely with the diagonal line, signifying a consistent proximity between the predictions generated by the model and the actual observed values. This result underscores the effectiveness of the proposed design in capturing relevant data patterns and generating reliable predictions. In conclusion, our study demonstrates that the neural network model employed here serves as a robust and accurate tool for solar energy prediction based on input variables.

#### *6.4.3. Uncertainty analysis of the proposed design*

In addition, it is necessary to perform a proposed design uncertainty analysis. This analysis helps to determine how reliable the proposed design predictions are. An analysis of proposed design uncertainty and the distribution of prediction errors is shown in Fig. 18 of the year 2006 data from Tenerife (measurements in kJ/m2) using an ensemble method with 1000 models with a GRU architecture. Fig. 19 shows the proposed design uncertainty and prediction error distribution using 2018 data from inverter 3 (W measurements) and an ensemble method with 1000 models with a GRU architecture.

Figures reveal that most predictions fall within one standard deviation, indicating a high degree of consistency and precision in the proposed design. 85 % of the data points fall within one standard deviation. This result signifies a pronounced level of consistency and precision in the predictions of our proposed design. Visually, it is evident as a tightly clustered distribution of data points around the central line on a scatterplot.

This high proportion of points within the range of one standard deviation implies that the proposed design's predictions tend to closely align with the actual values in the majority of cases. The resulting figure illustrates a distinct concentration of data around the reference line, portraying the coherence in the model's estimates.

#### *6.4.4. Ensemble benefits*

The benefits of the proposed ensemble method are shown in Fig. 20. Fig. 20 shows the results of the nRMSE value of Tenerife Island as a function of the number of networks in the ensemble. Li et al. [37], Lee, Kim, and Pedrycz [38], Peng and Zhu [39], and Afan et al. [40] also demonstrated that neural network ensembles minimize generalization errors, provide higher accuracy, and improve

**Table 7**
Results obtained from the island of El Hierro from the inverter 3, year 2017.

| | | 15 min. | 30 min | 60 min. |
|---|---|---|---|---|
| MLP | Min. Error (W) | $4{,}38{\cdot}10^{-3}$ | $1{,}44{\cdot}10^{-3}$ | $2{,}52{\cdot}10^{-2}$ |
| | Max. Error (W) | 1819,16 | 2280,91 | 1383,57 |
| | Mean Error (W) | 72,26 | 83,78 | 107,16 |
| | T.Deviation (W) | 140,55 | 152,18 | 164,56 |
| | RMSE (W) | 158,03 | 173,72 | 196,37 |
| | nRMSE (%) | 29,47 | 32,39 | 36,60 |
| | MAPE (%) | $7{,}77{\cdot}10^{-5}$ | $1{,}90{\cdot}10^{-4}$ | $4{,}26{\cdot}10^{-4}$ |
| LSTM | Min. Error (W) | $2{,}20{\cdot}10^{-4}$ | $9{,}77{\cdot}10^{-4}$ | $1{,}53{\cdot}10^{-2}$ |
| | Max. Error (W) | 1921,71 | 1688,50 | 1229,22 |
| | Mean Error (W) | 64,95 | 81,42 | 90,40 |
| | T.Deviation (W) | 134,33 | 138,59 | 150,80 |
| | RMSE (W) | 149,21 | 160,73 | 175,81 |
| | nRMSE (%) | 27,83 | 29,97 | 32,77 |
| | MAPE (%) | $6{,}71{\cdot}10^{-5}$ | $3{,}24{\cdot}10^{-4}$ | $5{,}90{\cdot}10^{-4}$ |
| GRU | Min. Error (W) | $6{,}86{\cdot}10^{-4}$ | $3{,}38{\cdot}10^{-3}$ | $1{,}58{\cdot}10^{-2}$ |
| | Max. Error (W) | 1906,79 | 1690,91 | 1278,37 |
| | Mean Error (W) | 69,09 | 75,91 | 89,07 |
| | T.Deviation (W) | 134,32 | 139,27 | 150,69 |
| | RMSE (W) | 151,04 | 158,61 | 175,04 |
| | nRMSE (%) | 28,17 | 29,58 | 32,63 |
| | MAPE (%) | $6{,}91{\cdot}10^{-5}$ | $1{,}34{\cdot}10^{-4}$ | $3{,}57{\cdot}10^{-4}$ |

**Table 8**
Results obtained from the island of El Hierro from the inverter 3, year 2018.

| | | 15 min. | 30 min | 60 min. |
|---|---|---|---|---|
| MLP | Min. Error (W) | $3{,}01{\cdot}10^{-3}$ | $5{,}33{\cdot}10^{-3}$ | $4{,}48{\cdot}10^{-3}$ |
| | Max. Error (W) | 1533,00 | 1418,77 | 1558,55 |
| | Mean Error (W) | 73,29 | 84,55 | 107,04 |
| | T.Deviation (W) | 141,99 | 152,65 | 168,31 |
| | RMSE (W) | 159,79 | 174,50 | 199,46 |
| | nRMSE (%) | 29,44 | 32,14 | 36,73 |
| | MAPE (%) | $8{,}49{\cdot}10^{-5}$ | $2{,}09{\cdot}10^{-4}$ | $4{,}86{\cdot}10^{-4}$ |
| LSTM | Min. Error (W) | $1{,}73{\cdot}10^{-4}$ | $3{,}75{\cdot}10^{-4}$ | $1{,}17{\cdot}10^{-4}$ |
| | Max. Error (W) | 1536,63 | 1345,67 | 1448,51 |
| | Mean Error (W) | 65,13 | 81,61 | 90,38 |
| | T.Deviation (W) | 135,36 | 140,59 | 152,58 |
| | RMSE (W) | 150,21 | 162,56 | 177,33 |
| | nRMSE (%) | 27,67 | 29,94 | 32,65 |
| | MAPE (%) | $6{,}90{\cdot}10^{-5}$ | $3{,}50{\cdot}10^{-4}$ | $6{,}25{\cdot}10^{-4}$ |
| GRU | Min. Error (W) | $3{,}60{\cdot}10^{-4}$ | $4{,}67{\cdot}10^{-4}$ | $5{,}94{\cdot}10^{-3}$ |
| | Max. Error (W) | 1533,15 | 1349,02 | 1402,03 |
| | Mean Error (W) | 69,46 | 75,74 | 89,21 |
| | T.Deviation (W) | 135,81 | 141,14 | 152,71 |
| | RMSE (W) | 152,54 | 160,17 | 176,85 |
| | nRMSE (%) | 28,10 | 29,50 | 32,57 |
| | MAPE (%) | $7{,}65{\cdot}10^{-5}$ | $1{,}43{\cdot}10^{-4}$ | $3{,}64{\cdot}10^{-4}$ |

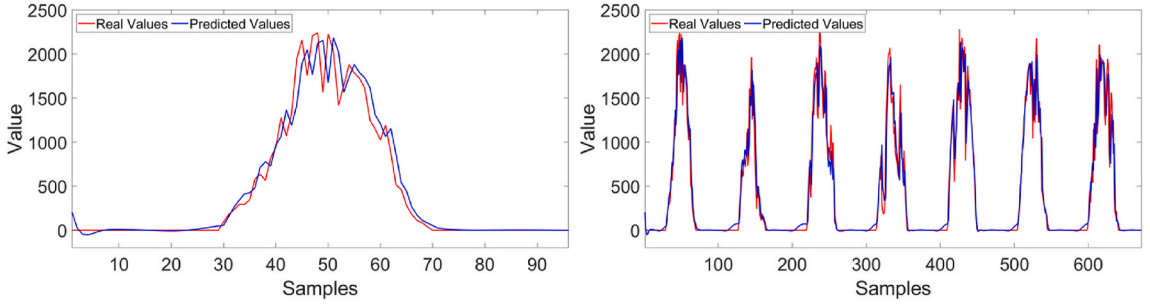


**Fig. 14.** Graphical results of inverter 3 of El Hierro.

**Table 9**
Values of nRMSE for Gran Canaria.

| | 2005 | 2006 | 2007 |
|---|---|---|---|
| MLP | 42,86 | 35,95 | 37,23 |
| LSTM | 39,24 | 32,84 | 34,00 |
| GRU | 38,82 | 32,39 | 34,10 |

**Table 10**
Values of nRMSE for Tenerife.

| | 2004 | 2005 | 2006 | 2007 |
|---|---|---|---|---|
| MLP | 42,86 | 39,83 | 40,36 | 39,09 |
| LSTM | 38,32 | 35,15 | 35,11 | 34,08 |
| GRU | 38,38 | 34,98 | 35,03 | 34,01 |

**Table 11**
nRMSE value based on the time interval and year for inverter 2.

| | | 2017 | 2018 |
|---|---|---|---|
| MLP | 2 min | 26,73 | 26,14 |
| | 4 min | 27,79 | 26,51 |
| | 8 min | 28,39 | 26,92 |
| | 16 min | 30,02 | 28,49 |
| | 32 min | 33,79 | 31,74 |
| | 64 min | 37,98 | 37,50 |
| LSTM | 2 min | 26,57 | 26,02 |
| | 4 min | 27,32 | 26,02 |
| | 8 min | 27,75 | 26,23 |
| | 16 min | 28,69 | 27,12 |
| | 32 min | 30,63 | 28,75 |
| | 64 min | 33,53 | 32,45 |
| GRU | 2 min | 26,47 | 25,92 |
| | 4 min | 27,51 | 26,22 |
| | 8 min | 27,74 | 26,22 |
| | 16 min | 29,14 | 27,59 |
| | 32 min | 30,76 | 28,83 |
| | 64 min | 33,93 | 32,88 |

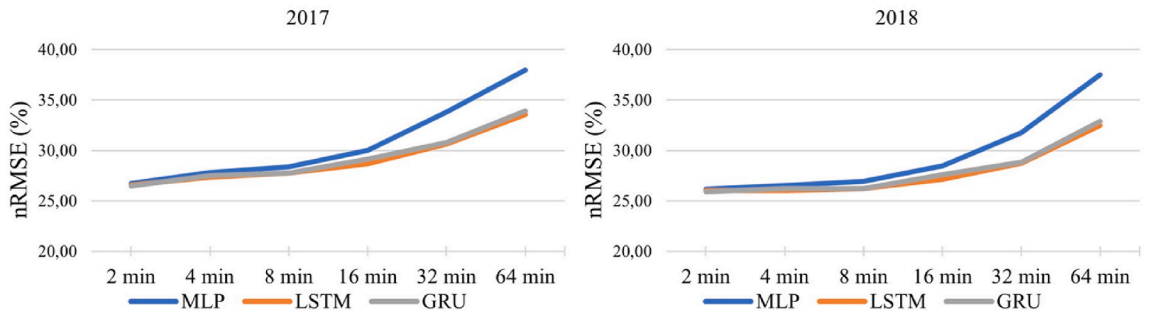


**Fig. 15.** Comparison nRMSE inverter 2 El Hierro.

learning.

*6.4.5. Novelty, limitations, and future directions*

Previous studies, such as the work of Chakraborty et al. [41], have investigated the impact of certain variables, like weather conditions, on solar power predictions. The novelty of our research, however, is twofold: firstly, it comprehensively analyzes how these variables affect solar power predictions when applying ensemble methods, a technique not extensively explored before in this context. Secondly, by conducting experiments across diverse locations, our study is uniquely positioned to identify the most effective predictor of solar energy for each specific energy infrastructure. We provide empirical data to determine the most suitable prediction system, taking into account the unique environmental and operational conditions present at each site. Additionally, the key point of this paper

**Table 12**
nRMSE value based on the time interval and year for inverter 3.

| | | 2017 | 2018 |
|---|---|---|---|
| MLP | 15 min | 29,47 | 29,44 |
| | 30 min | 32,39 | 32,14 |
| | 60 min | 36,60 | 36,73 |
| LSTM | 15 min | 27,83 | 27,67 |
| | 30 min | 29,97 | 29,94 |
| | 60 min | 32,77 | 32,65 |
| GRU | 15 min | 28,17 | 28,10 |
| | 30 min | 29,58 | 29,50 |
| | 60 min | 32,63 | 32,57 |

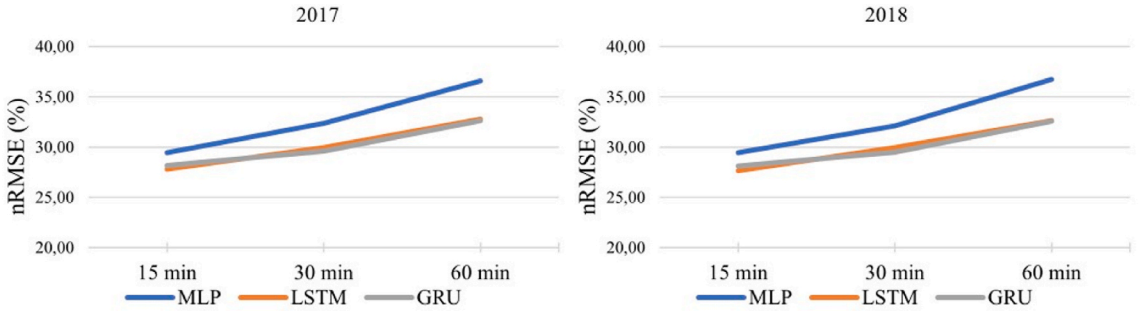


**Fig. 16.** Comparison nRMSE inverter 3 El Hierro.

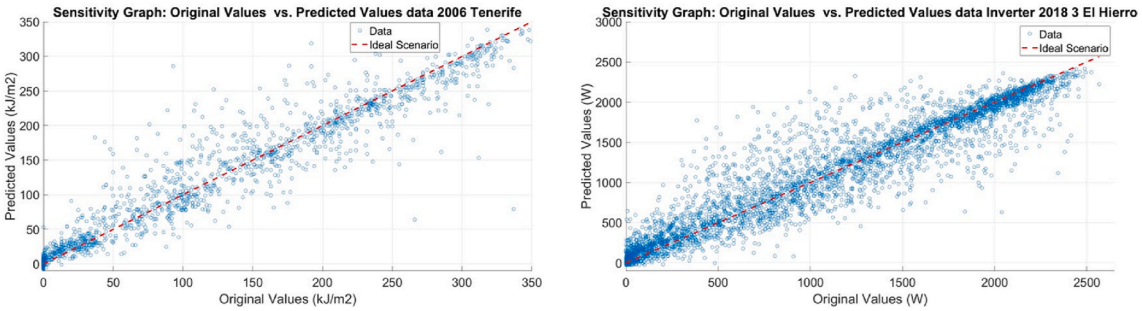


**Fig. 17.** Model sensitivity.

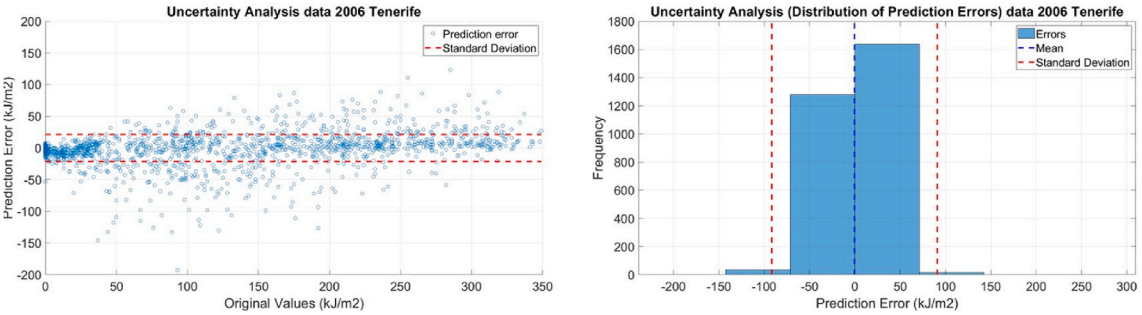


**Fig. 18.** Model uncertainty analysis data from Tenerife.

is the use of short time intervals, which is independent of the type of prediction model and its ensemble method. This approach can provide a more accurate and efficient management of energy sources for renewable energy infrastructures. An uncertainty analysis of the model was performed. When comparing the original values with those obtained in the model prediction, it was found that 85 % of the values are within the standard deviation. Capturing 85 % of the variations in data is considered quite solid and useful for

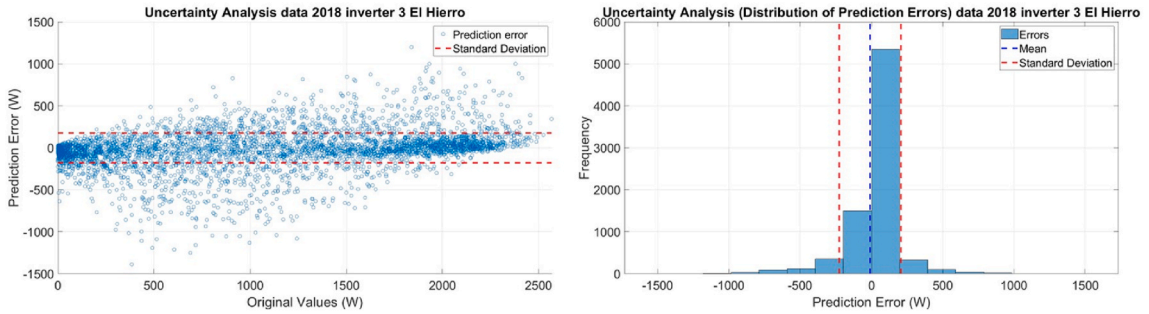


**Fig. 19.** Model uncertainty analysis data from inverter 3 El Hierro.

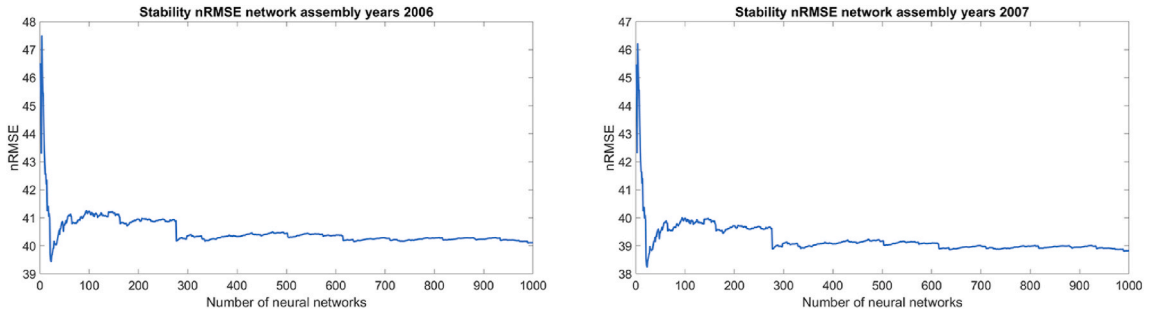


**Fig. 20.** Stability of the nRMSE as a function of the number of networks on the island of Tenerife.

decision-making in solar energy management [42,43]. Therefore, reasonably accurate estimates of future values are provided, which can be useful for planning and optimizing solar energy production.

It is important to keep in mind that new deep learning methods are always being developed and need to be compared with existing methods to determine their effectiveness and novelty in time series forecasting. One potential avenue for future research is the exploration of advanced neural network architectures such as Transformer Neural Networks (TNNs), networks utilizing attention mechanisms, and Dense Stream Neural Networks (DSNNs). These novel approaches offer promising enhancements for time series prediction by leveraging complex patterns in data more effectively.

While our study presents several innovative approaches to forecasting solar power generation, we recognize certain limitations that warrant mention. Firstly, the specificity of data sources primarily from the Canary Islands may limit the generalizability of our findings to other geographical locations with different meteorological conditions and solar irradiance patterns. Additionally, while our analysis incorporates advanced ensemble methods and neural network architectures, the rapid evolution of machine learning techniques suggests that newer models could offer further improvements in prediction accuracy. Another limitation lies in the scope of variables analyzed; considering additional factors such as atmospheric conditions, panel degradation over time, and the impact of shadowing could refine the predictive models further. Finally, the computational complexity and resource requirements of the deployed models may pose challenges for real-time application in energy management systems. Future research will aim to address these limitations, exploring broader datasets, emerging predictive models, and strategies to enhance computational efficiency.

## 7. Conclusion

The aim of this paper was to study and analyze different variables of the forecasting process, as the optimal time interval between samples needed to make an acceptable prediction and the use of different architectures with their ensemble. When looking at AEMET data, which has all been taken with a 1-h interval between samples, it can clearly be seen that the error is on average higher than 30 %. Looking at the El Hierro inverter data, we can see that using an inter-sampling time interval of 15–16 min or less results in an nRMSE value of around 30 % and lower depending on the architecture.

The type of architecture has more influence when the time interval between samples increases. If a sampling time of 2, 4 or 8 min is used, the results of the 3 types of architectures (GRU, LSTM or MLP) are very similar. However, as the time between samples increases it can be determined that it is a better solution to use GRU or LSTM architectures for predictions than the MLP architecture. Regarding the use of network assemblies, one point in favour is that stability is achieved in the results. The LSTM and GRU networks are quite stable with a low number of blocks works in the ensemble, however if networks with the MLP architecture are used, the variation is more noticeable, and it is necessary to use more networks in the ensemble. It could be determined that an ensemble with 100 neural networks is an optimal solution as stability is achieved for any type of architecture and the processing cost is not high. In conclusion, in order to forecast solar energy production accurately short time interval between samples (2, 4 or 8 min), any type of neural network

architecture proves to be the optimal choice and provides a fair compromise between reliability and costs, therefore it is independent of prediction system, due to the number of samples and their variation of the meteorological parameters in short intervals. But if conditions are different for the time interval, increasing its value, the type of neural network architecture and its ensemble has the principal role. Because, for bigger intervals, the pattern is clearer, and the prediction systems can model it, where the deep learning approaches and their ensemble learning can reduce the error measure.

## Data availability

Sharing research data helps other researchers evaluate your findings, build on your work and to increase trust in your article. We encourage all our authors to make as much of their data publicly available as reasonably possible. Please note that your response to the following questions regarding the public data availability and the reasons for potentially not making data available will be available alongside your article upon publication.

Has data associated with your study been deposited into a publicly available repository?

**Response**: No.

Question:

Please select why. Please note that this statement will be available alongside your article upon publication.

as follow-up to "Data Availability.

Sharing research data helps other researchers evaluate your findings, build on your work and to increase trust in your article. We encourage all our authors to make as much of their data publicly available as reasonably possible. Please note that your response to the following questions regarding the public data availability and the reasons for potentially not making data available will be available alongside your article upon publication.

Has data associated with your study been deposited into a publicly available repository?

**Response**: The data that has been used is confidential.

## CRediT authorship contribution statement

**Carlos M. Travieso-González:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Sergio Celada-Bernal:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Alejandro Lomoschitz:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Fidel Cabrera-Quintero:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

## Declaration of competing interest

The authors declare that they have no competing interests that could influence the objectivity or integrity of the research presented in this manuscript. The disclosure of interests aims to provide transparency and allow readers to form their own judgments regarding potential bias.

We assure the editors and readers of Applied Energy that the research presented in this manuscript has been conducted with the utmost integrity and objectivity.

## Acknowledgement

This work is supported under Grant CEI2021-06, from direct agreement SD-21/08 by Consejería de Economía, Industria, comercio y conocimiento from Gobierno de Canarias to ULPGC.

## References

[1] B. Eliasson, M. Robertson, The global environmental issues and the power industry, in: 21st International Telecommunications Energy Conference. INTELEC '99 (Cat. No.99CH37007), 1999, p. 204, https://doi.org/10.1109/INTLEC.1999.794057. Copenhagen, Denmark.

[2] A. Robbins, How to understand the results of the climate change summit: Conference of Parties21 (COP21) Paris 2015, J. Publ. Health Pol. 37 (2) (2016) 129–132, https://doi.org/10.1057/JPHP.2015.47.

[3] S. Paul, T. Dey, P. Sha, S. Dey, R. Sen, Review on the development scenario of renewable energy in different country, Innovations in Energy Management and Renewable Resources 52042 (2021) 1–2, https://doi.org/10.1109/IEMRE52042.2021.9386748, 2021.

[4] B. Sørensen, A history of renewable energy technology, Energy Pol. 19 (1) (1991) 8–12, https://doi.org/10.1016/0301-4215(91)90072-V.

[5] I. Guita-Pradas, I. Marques-Perez, B. Segura, A. Gallego, Criteria for identifying more favourable areas for photovoltaic installations: case of East Spain, 6th International Renewable and Sustainable Energy Conference (IRSEC) 2018 (2018) 1–5, https://doi.org/10.1109/IRSEC.2018.8702891.

[6] A. Prakash, S.K. Singh, Towards an efficient regression model for solar energy prediction, in: 2014 Innovative Applications of Computational Intelligence on Power, Energy and Controls with Their Impact on Humanity (CIPECH), 2014, pp. 18–23, https://doi.org/10.1109/CIPECH.2014.7019040. Ghaziabad, India.

[7] S. Jain, S. Rathee, A. Kumar, A. Sambasivam, R. Boadh, T. Choudhary, P. Kumar, P. Kumar Singh, Prediction of temperature for various pressure levels using ann and multiple linear regression techniques: a case study, in: Materials Today: Proceedings, vol. 56, 2022, pp. 194–199, https://doi.org/10.1016/j.matpr.2022.01.067, international Conference on Materials, Machines and Information Technology-2022.

[8] R.H. Inman, H.T. Pedro, C.F. Coimbra, Solar forecasting methods for renewable energy integration, Prog. Energy Combust. Sci. 39 (2013) 535–576, https://doi.org/10.1016/j.pecs.2013.06.002.

[9] A. Sfetsos, Coonick, A. Univariate and multivariate forecasting of hourly solar radiation with artificial intelligence techniques, Sol. Energy 68 (2000) 169–178, https://doi.org/10.1016/S0038-092X(99)00064-X.

[10] T.-A. Nguyen, M.-H. Pham, T.-K. Duong, M.-P. Vu, A recent invasion wave of deep learning in solar power forecasting techniques using ann, in: IEEE International Future Energy Electronics Conference (IFEEC), 2021, pp. 1–6, https://doi.org/10.1109/IFEEC53238.2021.9661747. Taipei, Taiwan.

[11] P. Malik, A. Gehlot, R. Singh, L.R. Gupta, A.K. Thakur, A review on ANN based model for solar radiation and wind speed prediction with real-time data, Arch. Comput. Methods Eng. (2022) 1–19, https://doi.org/10.1007/s11831-021-09687-3.

[12] M.W. Gardner, S.R. Dorling, Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences, Atmospheric environment 32 (14–15) (1998) 2627–2636, https://doi.org/10.1016/S1352-2310(97)00447-0.

[13] C. Paoli, C. Voyant, M. Muselli, M.L. Nivet, Forecasting of preprocessed daily solar radiation time series using neural networks, Sol. Energy 84 (12) (2010) 2146–2160, https://doi.org/10.1016/j.solener.2010.08.011.

[14] M. Achite, F.B. Banadkooki, M. Ehteram, A. Bouharira, A.N. Ahmed, A. Elshafie, Exploring Bayesian model averaging with multiple ANNs for meteorological drought forecasts, Stoch. Environ. Res. Risk Assess. (2022) 1–26, https://doi.org/10.1007/s00477-021-02150-6.

[15] R. Ben Ammar, M. Ben Ammar, A. Oualha, Photovoltaic power forecasting through temperature and solar radiation estimation, in: 16th International Multi-Conference on Systems, Signals Devices (SSD), 2019, pp. 691–699, https://doi.org/10.1109/SSD.2019.8893174. Istanbul, Turkey.

[16] G. Notton, C. Voyant, A. Fouilloy, J.L. Duchaud, M.L. Nivet, Some applications of ANN to solar radiation estimation and forecasting for energy applications, Appl. Sci. 9 (1) (2019) 209, https://doi.org/10.3390/app9010209.

[17] L. Benali, G. Notton, A. Fouillov, C. Voyant, R. Dizene, Solar radiation forecasting using artificial neural network and random forest methods: application to normal beam, horizontal diffuse and global components, Renew. Energy 132 (2019) 871–884, https://doi.org/10.1016/j.renene.2018.08.044.

[18] B. Gao, X. Huang, J. Shi, Research on 24-h forecasting of solar irradiance based on multilayer perceptron model, in: Optical Metrology and Inspection for Industrial Applications VI, vol. 11189, SPIE, 2019, pp. 384–390, https://doi.org/10.1117/12.2538768, 2019.

[19] B. Babar, L.T. Luppino, T. Boström, S.N. Anfinsen, Random forest regression for improved mapping of solar irradiance at high latitudes, Sol. Energy 198 (2020) 81–92, https://doi.org/10.1016/j.solener.2020.01.034.

[20] T.A. Fathima, V. Nedumpozhimana, Y.H. Lee, S. Winkler, S. Dev, Predicting solar irradiance in Singapore, Photonics Electromagnetics Research Symposium - Fall (PIERS - Fall), Xiamen, China (2019) 3164–3167, https://doi.org/10.1109/PIERS-Fall48861.2019.9021313.

[21] K. Kadirgama, A.K. Amirruddin, R.A. Bakar, Estimation of solar radiation by artificial networks: east coast Malaysia, Energy Proc. 52 (2014) 383–388, https://doi.org/10.1016/j.egypro.2014.07.090.

[22] Schuss, C., Eichberger, B., Rahkonen, T. Impact of sampling interval on the accuracy of estimating the amount of solar energy, IEEE International Instrumentation and Measurement Technology Conference Proceedings, Taipei, Taiwan, pp. 1–6. doi:10.1109/I2MTC.2016.7520566.

[23] G. Reikard, Predicting solar radiation at high resolutions: a comparison of time series forecasts, Sol. Energy 83 (3) (2009) 342–349, https://doi.org/10.1016/j.solener.2008.08.007.

[24] D. Pirone, L. Cimorelli, G. Del Giudice, D. Pianés, Short-term precipitation forecast using cumulative precipitation fields from station data: a probabilistic machine learning approach, J. Hydrol. 617 (2023) 128949, https://doi.org/10.1016/j.jhydrol.2022.128949.

[25] D.V. Pombo, P. Bacher, C. Ziras, H.W. Bindner, S.V. Spataru, P.E. Sørensen, Benchmarking physics-informed machine learning-based short term PV-power forecasting tools, Energy Rep. 8 (2022) 6512–6520, https://doi.org/10.1016/j.egyr.2022.05.006.

[26] V. Raj, S.-Q. Dotse, M. Sathyajith, M.I. Petra, H. Yassin, Ensemble machine learning for predicting the power output from different solar photovoltaic systems, Energies 16 (2023) 671, https://doi.org/10.3390/en16020671.

[27] P.M. Debani, S. Jena, R. Senapati, A. Panigrahi, R.S. Surender, Global solar radiation forecast using an ensemble learning approach, Int. J. Power Electron. Drive Syst. 14 (1) (2023) 496–505, https://doi.org/10.11591/ijpeds.v14.i1.pp496-505.

[28] N. Rahimi, S. Park, W. Choi, B. Oh, S. Kim, Y. Cho, S. Ahn, C. Chong, D. Kim, C. Jin, D. Lee, A comprehensive review on ensemble solar power forecasting algorithms, Journal of Electrical Engineering & Technology 18 (2023) 719–733.

[29] M. Chow, S.O. Yee, Methodology for on-line incipient fault detection in single-phase squirrel-cage induction motors using artificial neural networks, IEEE Trans. Energy Convers. 6 (3) (1991) 536–545, https://doi.org/10.1109/60.84332.

[30] S. Walczak, N. Cerpa, Artificial neural networks, in: third ed.), third ed., in: R.A. Meyers (Ed.), Encyclopedia of Physical Science and Technology, vol. 2003, Academic Press, New York, 2003, pp. 631–645, https://doi.org/10.1016/B0-12-227410-5/00837-1.

[31] S. Aslam, H. Herodotou, N. Ayub, S.M. Mohsin, Deep learning based techniques to enhance the performance of microgrids: a review, in: International Conference on Frontiers of Information Technology (FIT), 2019, pp. 116–1165, https://doi.org/10.1109/FIT47737.2019.00031. Islamabad, Pakistan.

[32] F. Nawab, A.S. Abd Hamid, A. Alwaeli, M. Arif, M.F. Fauzan, A. Ibrahim, Evaluation of artificial neural networks with satellite data inputs for daily, monthly, and yearly solar irradiation prediction for Pakistan, Sustainability 14 (13) (2022) 7945, https://doi.org/10.3390/su14137945.

[33] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780, https://doi.org/10.1162/neco.1997.9.8.1735.

[34] Z. Liang, X. Wang, Z. Chen, X. Luo, Body temperature prediction with recurrent neural network and its variants, in: 11th International Conference on Intelligent Control and Information Processing (ICICIP), 2021, pp. 149–153, https://doi.org/10.1109/ICICIP53388.2021.9642166. Dali, China.

[35] G. Perveen, P. Anand, A. Kumar, Short-term power prediction using ann, in: IEEE International Conference on Signal and Image Processing Applications (ICSIPA), 2021, pp. 233–237, https://doi.org/10.1109/ICSIPA52582.2021.9576813. Kuala Terengganu, Malaysia.

[36] R. Hecht-Nielsen, Iii.3 - theory of the backpropagation neural network**based on “nonindent” by robert hecht-nielsen, which appeared in proceedings of the international joint conference on neural networks 1, 593–611, june 1989. Ieee, in: H. Wechsler (Ed.), Neural Networks for Perception, vol. 1992, Academic Press, 1989, pp. 65–93, https://doi.org/10.1016/B978-0-12-741252-8.50010-8.

[37] K. Li, H. Huang, X. Ye, L. Cui, A selective approach to neural network ensemble based on clustering technology, in: Proceedings of the International Conference on Machine Learning and Cybernetics, vols. 1–7, 2004, pp. 3229–3233. Shanghai, China, 26–29 August 2004.

[38] H. Lee, E. Kim, W. Pedrycz, A new selective neural network ensemble with negative correlation, Appl. Intell. 37 (2012) 488–498.

[39] S. Peng, S. Zhu, Application of neural network ensemble in nonlinear time-series forecasts, in: Proceedings of the ICICTA: 2009 Second International Conference on Intelligent Computation Technology and Automation, Vol I, Proceedings, Changsha, China, 10–11 October 2009; IEEE Intelligent Computat Soc; IEEE Comp Soc; Res Assoc Intelligent Computat Technol & Automat; Changsha Univ Sci & Technol; Hunan Univ Sci & Technol, vol. 2009, IEEE Computer Society, Washington, DC, USA, 2009, pp. 45–47.

[40] H.A. Afan, A.I.A. Osman, Y. Essam, A.N. Ahmed, Y.F. Huang, O. Kisi, M. Sherif, A. Sefelnasr, K. wing Chau, A. El-Shafie, Modeling the fluctuations of groundwater level by employing ensemble deep learning techniques, Engineering Applications of Computational Fluid Mechanics 15 (2021) 1420–1439.

[41] D. Chakraborty, J. Mondal, H.B. Barua, A. Bhattacharjee, Computational solar energy – ensemble learning methods for prediction of solar power generation based on meteorological parameters in Eastern India, Renewable Energy Focus 44 (2023) 277–294, https://doi.org/10.1016/j.ref.2023.01.006.

[42] P. Aupke, A. Kassler, A. Theocharis, M. Nilsson, M. Uelschen, Quantifying uncertainty for predicting renewable energy time series data using machine learning, Engineering Proceedings 5 (1) (2021) 50, https://doi.org/10.3390/engproc2021005050.

[43] I. Jamil, J. Zhao, L. Zhang, S.F. Rafique, R. Jamil, Uncertainty analysis of energy production for a 3× 50 MW AC photovoltaic project based on solar resources, Int. J. Photoenergy 2019 (2019).