CONESCAPANHONDURAS2025paper82.pdf

Institute of Electrical and Electronics Engineers (IEEE)

Document Details

Submission ID

trn:oid:::14348:477770002

Submission Date

Jul 31, 2025, 11:04 PM CST

Download Date

Aug 12, 2025, 2:48 PM CST

File Name

CONESCAPANHONDURAS2025paper82.pdf

File Size

595.4 KB

4 Pages

2,887 Words

15,704 Characters

14% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Match Groups

- **19** Not Cited or Quoted 11%
  Matches with neither in-text citation nor quotation marks
- **0** Missing Quotations 0%
  Matches that are still very similar to source material
- **4** Missing Citation 3%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

13%	Internet sources
9%	Publications
0%	Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

- **19** Not Cited or Quoted 11%
  Matches with neither in-text citation nor quotation marks
- **0** Missing Quotations 0%
  Matches that are still very similar to source material
- **4** Missing Citation 3%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

13% Internet sources
9% Publications
0% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Internet	cbr.robocup.org.br	2%
2	Internet	pubs2.ascee.org	1%
3	Internet	revistas.ups.edu.ec	1%
4	Publication	Irene Monasterolo, Antoine Mandel, Stefano Battiston, Andrea Mazzocchetti et al...	<1%
5	Publication	Vahid Nourani, Nazanin Behfar, Martijn J. Booij, Anne Ng, Chunwei Zhang, Sepide...	<1%
6	Internet	www.unido.org	<1%
7	Internet	ptx-hub.org	<1%
8	Internet	www.diva-portal.org	<1%
9	Internet	www.mdpi.com	<1%
10	Internet	repositorio.unal.edu.co	<1%

11 Internet

dokumen.pub <1%

12 Internet

openarchive.usn.no <1%

13 Publication

Eneko Artetxe, Oscar Barambones, Imanol Martín Toral, Jokin Uralde, Isidro Calv... <1%

14 Internet

ams.confex.com <1%

15 Internet

www.intechopen.com <1%

16 Publication

Antonello Rosato, Rodolfo Araneo, Amedeo Andreotti, Massimo Panella. "2-D Con... <1%

17 Publication

Carlos Meza. "A review on the Central America electrical energy scenario", Renew... <1%

18 Internet

vdocuments.mx <1%

19 Publication

Utpal Kumar Das, Kok Soon Tey, Mehdi Seyedmahmoudian, Saad Mekhilef et al. "... <1%

20 Publication

Jun Abraham. "Investment prospects for geothermal power in El Salvador's electr... <1%

Evaluation of Prediction Models for Optimizing the Integration of Photovoltaic Systems in El Salvador

1st Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

2nd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

3rd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

***Abstract*—This study explores the integration of machine learning and neural networks for photovoltaic energy prediction on a weekly basis.**

Data were obtained from sensors on photovoltaic panel installed in urban buildings. Weather data were obtained through third party sites and later combined with the information from panels. Four models were trained on this data and afterwards compared their RMSE, MSE and the graphs between predicted and real values.

For new installations without years of data, machine learning ensemble models like Random Forest Regressor were found to be better performant that neural networks and other machine learning algorithms. The results obtained allow the development and integration of tools in new installations without the need for years of data to create accurate models.

***Index Terms*—Solar Energy Prediction Neural Networks, Machine Learning, SDG 7 – Affordable and Clean Energy**

I. Introduction

Energetic transition into renewable energy and its inclusion into electrical supply is a topic that has been gaining momentum in the last decades, due to a plurality of reasons like population growth, changing lifestyles and increase in energy consumption [8] [10]. This increase of energetic demand implied a greater consumption of resources used to generate electricity like carbon, natural gas and other fossil fuels, which negatively impact climate and the environment. In order to reduce its impact, renewable energy is looked upon to cover this need, where solar and wind energy are first in line thanks to its availability and, in the case of the former, the ease of access [9] [10].

El Salvador has experienced remarkable progress in the adoption of photovoltaic energy; the country has promoted its energy matrix diversification with large-scale solar power plants installation and photovoltaic systems incorporation on commercial and business sectors [3] [4]. In the recent years, its installed capacity of solar energy generation has greatly expanded with multiple photovoltaic plants, being the main ones are Albireo I y II, Sonsonate Solar, Ecosolar y Provincia Solar, representing a vital part in the country and region energetic supply [6].

This growth has allowed a constant increase in total solar energy installed capacity, going from 474.4 MW in 2020 to 676.01 MW in the first semester of 2024, being a significant increase in the last years [5]. Fig. 1 shows the country's evolution of photovoltaic energy installed capacity.

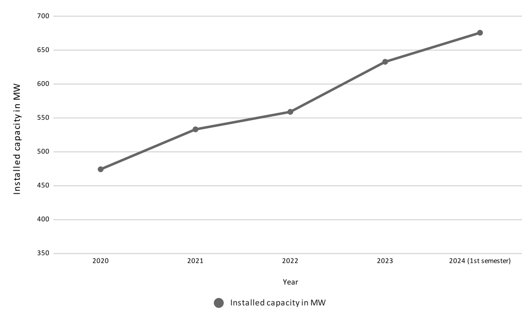

Fig. 1. Growth of installed capacity (MW).

Photovoltaic energy has had a high share of the electricity matrix, injecting 1,121.44 GWH into the Salvadorean market in 2022, representing 15.07% of the total net injection of electrical energy. In 2023, photovoltaic energy injection was 1,186.71 GWh (15.29%), while in 2024, it increased slightly to 1,260.03 GWh (15.31%), indicating a stable trend that supports the country's energy demand [5].

Although its accessibility, photovoltaic energy generation presents several challenges and difficulties to overcome: uncertainty and instability of climatic factors, variability in energy production relying on sunlight availability, storage, capacity to meet demand, and a certain degree of social rejection [7].

Efforts to overcome these challenges have been made through democratization and transparency in projects like the ones done by Universidad José Simeón Cañas (UCA) and its pilot program that integrates photovoltaic panels and public access to its energy production data.

Due to this situation, there is a need to reduce uncertainty as soon as possible. Methods such as the creation of tools that will allow for greater certainty when making decisions or determining the availability of renewable energy over a specific period, can be of great help in this labor. These tools should also be adjusted for the weather conditions where they will be used, since not every location is exactly the same weather-wise. Based on this, this article explores the possibility of creating this type of tool based on solar panel

18 installations with limited data adjusted for urban areas in San
14 Salvador.

II. Methodology

A. *Recollection, exploratory analysis and union of datasets.*

UCA has placed five photovoltaic systems around the university campus aiming to reduce energy consumption from the national electrical grid. These count with a meter for each system that registers historic measures per minute. These records are open to the public through UCA Departamento de Electrónica e Informática (DEI) website: https://dei.uca.edu.sv/energy/. Thanks to access facilities and the presence of multiple photovoltaic variables, it was decided to use website data.

In the analysis of UCA photovoltaic data, the energy variables available in the system were sometimes divided by L1, L2, and L3, which refers to a three-phase model, that handles three single-phase currents with similar frequency and amplitude, but separated by a phase difference of 120 electrical degrees. Among the variables available on the data, Active Power Output was chosen as the target to predict, since this represents the real and useful value of the energy produced by the solar panels.

The data generated was downloaded in CSV files for data preprocessing. Afterwards, an exploratory analysis was performed for each sensor to understand the structure and quality of the dataset. Three columns of interest were identified (Measurement, Timestamp, and Value) for subsequent manipulation. A temporal continuity analysis was also performed, where duplicate records belonging to the same minute were eliminated, grouped into time intervals, and finally averaged to a single hour value.

TABLE I
Temporal Continuity Analysis.

Interval	Theoretical Lecture	Complete Data	Missing or Incomplete Data
Minuto	172,801	172,378	423
10 minutos	17,281	17,274	7
Hora	2,881	2,880	1

The previous process was performed for the five sensors available, with the values in Table 1 corresponding to the one with the best results.

The decision to group the records and average them into hourly measures is influenced by the weather information. Since no up-to-date weather data was found that could be used for predictions in public institutions nor provided by the sensors in the solar installation, a third-party site, OpenMeteo (https://open-meteo.com/), was consulted to get the required data for the climatological variables. The information obtained from the site was provided in records grouped hourly, therefore the decision to group the power output derives from this.

Once all the data was gathered, both collections were fused in a single dataset for ease of use and manipulation. A single dataset with the following columns was obtained Potencia activa L1, Potencia activa L2, Potencia activa L3, temperature_2m, relative humidity_2m, wind_speed_10m, wind_gusts_10m, surface_pressure, shortwave_radiation, dew_point_2m, wind_direction_10m, et0_fao_evapotranspiration, rain, cloud_cover, direct_radiation y direct_normal_irradiance.

The resulting dataset was analyzed for outliers, Interquartile range and correlation between target and predictor variables. From this the need for normalizing the data previous training was found due to the presence of outliers and some variation in magnitude between variables.

TABLE II
Variables with outliers using the IQR method.

Variable	IQR	Lower Limit	Upper Limit	Outlier Count
wind_speed_10m (km/h)	6.20	-4.10	20.70	25
wind_gusts_10m (km/h)	16.60	-9.10	57.30	14
surface_pressure (hPa)	2.40	921.30	930.90	3
dew_point_2m (°C)	4.60	9.40	23.80	50
rain (mm)	0.00	0.00	0.00	410

For the correlation analysis, a minimum value for r was established as a threshold in order to consider or drop a column, with this value being $r > |0.4|$ in relation to the target column. Based on this, the following columns were dropped: wind_speed_10m ($r = 0.32$), rain ($r = -0.01$), cloud_cover ($r = -0.01$). Additionally, the three Power Output columns were combined into a single one that averaged the total in order to simplify the prediction.

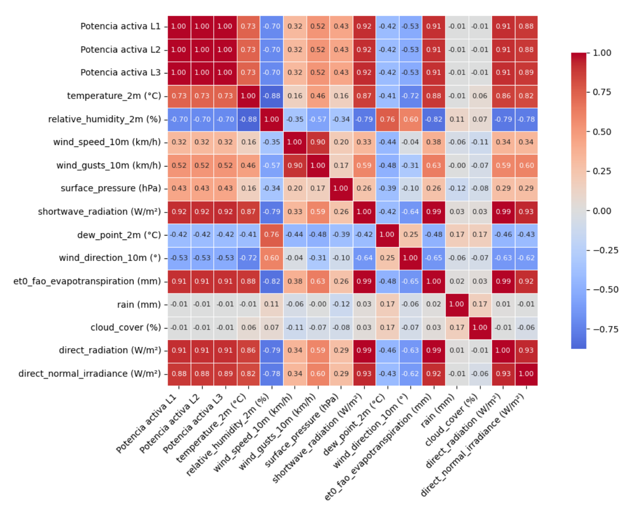

Fig. 2. Correlation between photovoltaic and climatic variables.

B. *Data preprocessing and separation for training and testing.*

Once the dataset was constructed only by relevant variables, a transformation process for the target variable was performed in order to predict a whole week instead of a single hour. For this, the records were moved from one to 168 hours to represent a full week of power output information and related

11 weather variables. Following this, the dataset was divided in an 80/20 split, where 80% of the information was used for training and 20% was set aside for testing purposes on unseen data.

As previously stated, a need for normalization was identified due to the varying magnitudes in the data, but it was also needed for a better learning process for the models. Hence, a normalization process was conducted using scikit-learn's
12 MinMaxScaler. In this process, two scalers were created: one for the target variable and the other for independent weather variables. These scalers were fitted using the training data, which was later transformed, while the testing data was only transformed (but not fitted), in order to avoid leaking information to the models during training time.

Finally, the datasets were duplicated for reorganization in columns, without affecting the inner data, since two of the models required a specific format (records, timesteps, features) that the other two did not. This process only affected the matrix shapes but did not operate inside its data.

C. Model implementation.

1) Baseline Model (M1).: A baseline model was developed to establish a minimum performance that the other algorithms had to meet in order to be considered. This model assumed that the following week's prediction would be the same as the previous week's

2) Support Vector Regression with Multi-Output (M2).: Vector regression machines are an adaptation of the Support Vector Machine algorithm used for classification and used in time series forecasting problems.

This nonlinear algorithm transforms input data, for which it is very difficult to apply linear regression in the original plane, to a higher plane, and then applies linear regression in this new plane or dimension. Our implementation was extended using multi-output to allow for full-range predictions, as the default algorithm cannot perform this.

9 *3) Random Forest Regressor (M3).:* This is an ensemble algorithm based on decision trees. This method for regression problems trains multiple trees using random sampling with repetition. Once the trees have completed training, the algorithm averages the results and returns a tree based on them.

16 *4) Long Short Term Memory (M4).:* The third model is a variant of recurrent neural networks that avoids the gradient explosion and vanishing problem, allowing for selective management of information to be remembered and forgotten.

15 Long Short Term Memory (LSTM) networks are recurrent networks that recognize patterns over long periods of time by using cells containing gates to decide which information to preserve or forget. Gated Recurrent Unit

9 *5) Gated Recurrent Unit (GRU) (M5).:* Another type of recurrent neural network, which, unlike LSTM, only works with two gates: reset and update. The reset gate determines how much of the previous information is forgotten, and the update gate determines how much information is incorporated into the new secret state. The advantages of this model over LSTM are that it is much easier to train and therefore requires fewer resources.

III. Results

Once the models were implemented, they were trained with the same dataset to ensure a level playing field when comparing results. After training, predictions were obtained for the validation dataset, which consisted of 509 records.

To evaluate performance, four metrics were used: the Mean Square Error (MSE) and Root Mean Square Error (RMSE) values, as well as a graph comparing the actual values and the predictions made by each model, taking four samples of the predictions, along with the distribution of errors produced by each model.

After the predictions were made, the following results were obtained:

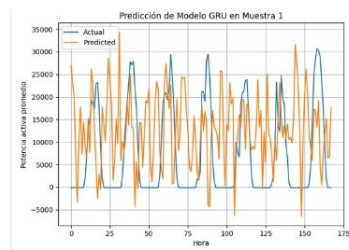

Fig. 3. Modelo GRU.

The base model obtained a threshold RMSE of 12,254.14, and the visual comparison shows that the model shows an advantage in its predictions compared to the actual values.

From this, M5 was the model with the worst performance in the evaluation, being inferior to the base model. Obtaining the highest values in the MSE and RMSE error measurements, which indicates that its predictions are far from the actual values, and comparing the graph of predictions versus actual values, it is observed that the model is unable to fit the series trend. This led to the model being discarded.

The remaining models, M2, M3, and M4, on the other hand, performed better than M1. Of these three models, M3 fits the data better and its variations, reflecting factors such as the lowest RMSE and MSE values, with 3144.43 and 9887,469.68, respectively. A graphical comparison shows that the model generates values very close to the real ones and adapts well to the valleys within the data, areas where M2 and M4 struggle.

Likewise, M3 is more consistent in modeling peaks or maximum values. Although it is more conservative than M2, most likely due to its use of averages, the latter tends to predict values that exceed the maximums, and it does not adapt well to atypical days. A graphical comparison of these methods shows what was previously indicated, where M2 and M4 have difficulty adjusting to the valleys in the data, and M4 also has some degrees of delay.

From these comparisons, it can be concluded that the model that best fits is M3; however, the limitation of the amount of data, Opens the door for future research to compare these models to see if this trend is maintained, especially in the case

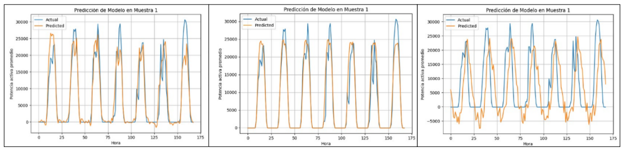

Fig. 4. Model Comparison.

of M4, which, being a model belonging to neural networks, usually requires a greater amount of data.

IV. Conclusions.

This paper presents one of the firsts artificial intelligence based models suitable for El Salvador climate conditions, applied to residential and commercial urban areas, which represents not only a progress in the incorporation of this kind of tools inside the electric grid, but also the support of renewable energy inclusion with a higher confidence level alongside the fact of being a not so explored area inside the country.

Besides, it opens the door to the development of an energy administration system, through the supplement with consumption predictions by buildings, which will remain pending for future research. This will allow better energy administration and better renewable energy incorporation inside a growing market.

It was observed that in recent installments with data quantity no bigger than months, reliable enough tools for energy prediction can be created through machine learning with algorithms like Random Forest. This allows to reduce insertion and development costs inside the market by decreasing the quantity of necessary data for tool adjustment.

Furthermore, it is observed that obtaining climate data by third parties does not affect the models projection in the context of local information lacking provided by the system itself or sensors corresponding to the area in question.

V. Acknowledgments

We appreciate Universidad José Simeón Cañas for providing access to data from the solar panels located at its facilities. We thank OpenMeteo for providing access to the climate data used in this research. Last but not least, we would like to thank PhD Erick Varela for his support and guidance throughout the development of this article.

References

[1] Open-Meteo, "Open-Meteo: Weather API," 2025. [Online]. Available: https://open-meteo.com/

[2] Universidad Centroamericana José Simeón Cañas, "Energy DEI UCA," 2025. [Online]. Available: https://dei.uca.edu.sv/energy/

[3] Consejo Nacional de Energía (CNE), "Política Energética Nacional 2010–2024," 2010. [Online]. Available: https://estadisticas.dgehm.gob.sv/wp-content/uploads/2016/09/Politica_Energetica.pdf

[4] Consejo Nacional de Energía (CNE), "Plan Indicativo de la Generación de la Expansión 2021–2031," 2021. [Online]. Available: https://estadisticas.dgehm.gob.sv/wp-content/uploads/2021/06/Plan-indicativo-de-la-generaci%C3%B3n-de-la-expansi%C3%B3n-2021-2031.pdf

[5] Superintendencia General de Electricidad y Telecomunicaciones (SIGET), "Visualizador dinámico de Estadísticas Eléctricas," 2024. [Online]. Available: https://www.siget.gob.sv/gerencias/electricidad/informe-de-mercado-y-estadisticas-electricas/estadisticas-electricas-bi/

[6] Superintendencia General de Electricidad y Telecomunicaciones (SIGET), "Mercado Eléctrico de El Salvador 2020," 2021. [Online]. Available: https://www.siget.gob.sv/wp-content/uploads/2021/05/Mercado-Electrico-de-El-Salvador-2020-2.pdf

[7] C. Roldán-Blay, M. F. Abad-Rodríguez, V. Abad-Giner, and X. Serrano-Guerrero, "Interval-based solar photovoltaic energy predictions: A single-parameter approach with direct radiation focus," Renewable Energy, vol. 230, pp. 120821, 2024. [Online]. Available: https://dx.doi.org/10.1016/j.renene.2024.120821

[8] A. Shahsavari and M. Akbari, "Potential of solar energy in developing countries for reducing energy-related emissions," Renewable Sustainable Energy Reviews, vol. 90, pp. 275–291, 2018. [Online]. Available: https://dx.doi.org/10.1016/j.rser.2018.03.065

[9] U. K. Das et al., "Forecasting of photovoltaic power generation and model optimization: A review," Renewable Sustainable Energy Reviews, vol. 81, pp. 912–928, 2018. [Online]. Available: https://dx.doi.org/10.1016/j.rser.2017.08.017

[10] World Bank Group, "Climate and development: An agenda for action - emerging insights from World Bank Group 2021-22 country climate and development reports," Washington, DC: World Bank, 2022. [Online]. Available: https://hdl.handle.net/10986/38220