

## USO DE REDES NEURONALES ARTIFICIALES PARA LA ESTIMACIÓN DE LA RADIACIÓN SOLAR HORARIA BAJO DIFERENTES CONDICIONES DE CIELO

**Lucas Olivera<sup>1,2</sup>, Julissa Atia<sup>1,2</sup>, Leonardo Amet<sup>1</sup>, Jorge Osio<sup>1,3</sup>, Martín Morales<sup>1,4</sup>, Marcelo Cappelletti<sup>1,3</sup>**

<sup>1</sup> Programa TICAPPS, Universidad Nacional Arturo Jauretche, Av. Calchaquí 6200, Florencio Varela (1888), Argentina – e-mail: [lolivera.unaj@gmail.com](mailto:lolivera.unaj@gmail.com)

<sup>2</sup> Comisión de Investigaciones Científicas, Provincia de Buenos Aires (CICPBA), 526 E/10 y 11, La Plata (1900), Argentina

<sup>3</sup> Grupo de Control Aplicado (GCA), Area CeTAD, Instituto LEICI (UNLP-CONICET), Facultad de Ingeniería, Universidad Nacional de La Plata, 48 y 116, La Plata (1900), Argentina

<sup>4</sup> Centro UTN CODAPLI-FRLP, Av. 60 esq. 124, La Plata (1900), Argentina

**RESUMEN:** La radiación solar es un factor clave en numerosas aplicaciones, tales como sistemas fotovoltaicos o térmicos, en la arquitectura y en la agricultura. Sin embargo, no siempre es posible contar con datos experimentales de radiación solar en los lugares de interés. Por esta razón, anteriormente, una amplia variedad de modelos teóricos han sido desarrollados con el fin de estimar este parámetro. Este trabajo presenta un análisis comparativo de modelos de redes neuronales artificiales para la estimación de la radiación solar global horaria en la localidad de Florencio Varela, provincia de Buenos Aires, a partir de variables meteorológicas de sencilla obtención (temperatura y humedad relativa). Los resultados obtenidos muestran un pobre desempeño de los modelos cuando son entrenados con diferentes condiciones de cielo. Esto se debe fundamentalmente al conjunto limitado de datos utilizado y a la gran dispersión de valores de radiación solar medidos. Por el contrario, cuando se utiliza el índice de claridad  $K_t$ , y los modelos son entrenados con datos correspondientes a la condición de cielo despejado ( $K_t > 0.6$ ), los errores de estimación se reducen significativamente. Estos modelos podrían aplicarse en lugares donde no se dispone de valores de radiación solar medidos.

**Palabras clave:** Radiación solar. Redes neuronales artificiales. Nubosidad.

### INTRODUCCIÓN

El incremento continuo de la demanda energética a nivel mundial y la respectiva disminución de reservas de combustibles fósiles y el impacto ambiental asociado a su combustión, ha empujado el desarrollo hacia recursos energéticos alternativos. En este sentido, el estudio de fuentes de energía renovables es una temática de interés global y local. Ante el contexto planteado, surgen iniciativas tecnológicas para explorar y abordar esta problemática. Al respecto, la energía solar constituye una fuente limpia y renovable con bajo impacto ambiental, disponibilidad en el sitio de consumo y sostenibilidad. Esta energía se presenta entonces como uno de los recursos más importantes tendiente a sustituir las fuentes de energía no renovables (carbón, petróleo, gas natural), reducir las emisiones de dióxido de carbono a la atmósfera y proteger nuestro medio ambiente.

El conocimiento preciso de la radiación solar en un determinado instante y lugar es un factor clave para el dimensionamiento óptimo de sistemas con diferentes principios de operación, como los sistemas fotovoltaicos para la generación de electricidad y los colectores solares para el

calentamiento de fluidos. A su vez, la radiación solar es un recurso fundamental en el diseño arquitectónico y en la planificación agrícola. Por ejemplo, para este último caso, la distribución de radiación solar determina el rendimiento y desarrollo de los cultivos, y la baja disponibilidad o el exceso de ella puede producir daños irreparables. Y en el caso específico de los sistemas fotovoltaicos, la cantidad total de radiación solar que llega a los mismos tiene un rol determinante en la evaluación de la producción potencial de energía eléctrica que podría producirse en ese lugar.

El problema que se presenta es que no siempre es posible contar con datos experimentales de la radiación solar en los lugares de interés (Raichijk et al., 2005). Además, la radiación solar disponible en un punto determinado de la superficie terrestre, es fuertemente dependiente de la localización geográfica (latitud, longitud, altura sobre el nivel del mar) y de los microclimas locales del sitio (temperatura, humedad, entre otras variables meteorológicas), con lo cual se trata de una magnitud variable con el tiempo de manera instantánea, horaria, diaria y estacional.

La radiación solar global incidente sobre una superficie horizontal puede ser medida por piranómetros, los cuales son dispositivos costosos que requieren ser calibrados a intervalos regulares y con un apropiado mantenimiento para asegurar la calidad de los datos (Pandey et al., 2013). Previamente, han sido reportados en la literatura diferentes técnicas para modelar la radiación solar que permita evaluar el futuro desempeño de una planta fotovoltaica. Entre ellas se encuentra la ecuación de Ångström–Prescott y sus modificaciones, modelos lineales, polinómicos, exponenciales y logarítmicos (Assi et al., 2012), las técnicas geoestadísticas (Leal et al., 2009), las técnicas estocásticas (Kaplanis et al., 2010) y las técnicas basadas en imágenes satelitales (Polo Martínez, 2010). Sin embargo, todos los modelos mencionados pueden sufrir problemas de precisión debido a la incertidumbre de las condiciones iniciales, como consecuencia de la cobertura de nubes y su efecto en la atenuación de la radiación solar en una localización específica. Por esta razón, este campo de estudio se encuentra abierto a nuevas propuestas que presenten alto rendimiento en la precisión de los resultados, con un bajo costo de cómputo, facilidad de implementación y buena velocidad de convergencia. Al respecto, gracias al crecimiento de las capacidades de cálculo y al mejoramiento de los algoritmos implementados, en los últimos años se han comenzado a utilizar técnicas basadas en redes neuronales artificiales (RNA), como una buena alternativa para la estimación de la radiación solar con el mínimo error (Pang et al., 2020; Çoban et al., 2020; Neelamegam et al., 2016; Khatib et al., 2015; Sayago et al., 2011; Raichijk et al., 2008).

Usualmente, debido a la mayor disponibilidad de datos de variables meteorológicas mensuales o diarias, la mayoría de las técnicas mencionadas son utilizadas para determinar la radiación solar mensual o diaria. Sin embargo, para un diseño óptimo de sistemas fotovoltaicos y térmicos, es más apropiado el conocimiento de la radiación solar horaria. La confiabilidad de los sistemas fotovoltaicos o térmicos diseñados en base a datos de radiación solar horaria es mayor que los sistemas diseñados en base a perfiles de radiación solar diarios o mensuales (Erdinc et al., 2012). Trabajos previos han estimado la radiación solar horaria utilizando RNA a partir de variables meteorológicas horarias (Sayago et al., 2011; Pang et al., 2020), como por ejemplo temperatura, humedad relativa, velocidad y dirección del viento, cuyos datos son provistos por las mayorías de las estaciones meteorológicas automáticas.

Las RNA están inspiradas en el funcionamiento del sistema nervioso de los seres vivos. Consisten en una arquitectura conformada por un conjunto de neuronas artificiales las cuales se conectan entre sí intercambiando información o estímulos al mismo tiempo que están organizadas en capas. Las RNA son una de las técnicas claves del aprendizaje automático (*Machine Learning*), cuyo objetivo principal es lograr que las computadoras sean capaces de “aprender” a tomar decisiones mediante modelos generales entrenados a partir de datos de ejemplos particulares o experiencias pasadas (Jordan et al., 2015; Harrington, 2012). El “aprendizaje” de las computadoras se refiere a la capacidad para identificar patrones en grandes conjuntos de datos y a través de ellos tomar decisiones, o hacer una predicción acerca de comportamientos futuros de los sistemas bajo estudio. Las características principales de los

algoritmos de aprendizaje automático es que son capaces de resolver problemas no lineales; aprender de ejemplos; encontrar correlaciones entre datos diversos; manejar diferentes tipos de datos (numéricos, textuales, imágenes, etc.); tratar con grandes conjuntos de datos y/o conjuntos de datos de alta dimensión; ser tolerantes a fallos, es decir sobrellevar el ruido y datos incompletos o atípicos; realizar predicciones y generalizaciones a altas velocidades; y realizar procesamiento en tiempo real. Si bien las primeras teorías sobre el aprendizaje automático se desarrollaron en la década del 80, existen dos razones principales por las que solo han empezado a resultar útiles recientemente: requieren de grandes cantidades de datos y de una potencia de cálculo significativa.

En este trabajo se presenta el desarrollo y evaluación del desempeño de diferentes modelos de redes neuronales para la estimación de la radiación solar global horaria en la ciudad de Florencio Varela, provincia de Buenos Aires, a partir de variables meteorológicas de sencilla adquisición. Los valores estimados podrían ser utilizados por ejemplo para la evaluación y dimensionamiento de nuevos emprendimientos fotovoltaicos; o para restaurar un conjunto de datos de radiación solar en caso de tener datos faltantes debido a interrupciones del sistema de monitoreo; o bien, para utilizar su valor en un sistema de seguimiento del punto de máxima potencia de sistemas fotovoltaicos.

## **METODOLOGÍA**

La metodología general llevada a cabo en este trabajo comprende las siguientes etapas:

- ***Recolección de datos***

Los datos meteorológicos utilizados en esta investigación son datos obtenidos de una estación meteorológica automática Davis Instruments Vantage Pro2, ubicada en Florencio Varela, provincia de Buenos Aires, más precisamente en la Universidad Nacional Arturo Jauretche (34.77° S; 58.26° W; 25 m.s.n.m.). Desde su puesta en funcionamiento en el mes de diciembre de 2017, dicha estación recopila datos meteorológicos tales como temperatura exterior media, mínima y máxima, humedad relativa, temperatura del punto de rocío, velocidad y dirección del viento, precipitación, presión atmosférica y radiación solar global, entre otros.

Los parámetros seleccionados para el análisis (y su correspondiente unidad) fueron: temperatura media (°C), humedad relativa (%), velocidad de viento media (km/h), presión atmosférica (hPa) y radiación solar global (W/m<sup>2</sup>). La frecuencia de muestreo de los datos recopilados por la estación meteorológica es de 5 minutos. Sin embargo, para este estudio, fueron considerados los valores promedios tomados cada 30 minutos.

Para los resultados presentados en este trabajo, se utilizaron los datos recopilados del mes de enero de 2019, para el entrenamiento y validación de los modelos formulados. Si bien es sabido que las RNA producen modelos más potentes y precisos a medida que se introducen más datos en la red, un trabajo reciente (Pang et al., 2020) utiliza datos de solo una semana para el entrenamiento de los modelos, obteniendo resultados auspiciosos.

- ***Tratamiento y procesado de datos***

La recopilación de datos, conjuntamente con su procesado, constituyen quizás las fases más importantes del análisis. Como parte de la preparación de los datos, se seleccionaron solamente aquellos datos comprendidos entre las 6:00 y las 21:00 horas de los 31 días del mes de enero 2019, con el fin de considerar únicamente valores de radiación solar en el rango horario de interés donde se aplicarán los modelos de estimación. Con un almacenamiento de datos cada 30 minutos, se tienen 30 vectores de datos por día y 930 en total.

- **Entrenamiento y Testeo**

En base a la información presentada en la etapa previa, se desarrollaron modelos de redes neuronales para estimar la radiación solar global horaria. La arquitectura de la RNA utilizada consta de una capa de entrada que recibe los estímulos externos (datos de entrenamiento), de una capa de salida que ofrece la respuesta de la red (en este caso, radiación solar global horaria estimada), y de una o más capas intermedias denominadas capas ocultas. La Figura 1 muestra un esquema de uno de los modelos analizados en este trabajo, en el cual la capa de entrada recibe los valores de hora del día, temperatura media y humedad relativa. Cada neurona recibe estímulos como entrada, donde cada conexión posee un peso que determina la fuerza del estímulo recibido. La salida de una neurona está dada por la información pesada que recibe y una función de transferencia que determina el valor de salida de dicha neurona. El entrenamiento de una RNA consiste en un proceso iterativo donde mediante alguna técnica preestablecida, los pesos de las conexiones de las neuronas se adaptan de tal forma que la red comienza a producir en la salida la información esperada ante una entrada específica. Este tipo de red permite crear modelos con una gran capacidad para generalizar y aprender de patrones de entrada produciendo valores de salida ante la recepción de estímulos similares.

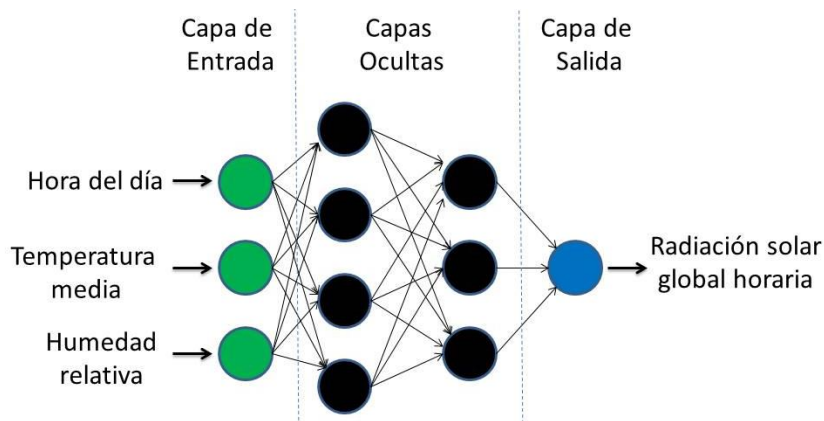


Figura 1: Arquitectura de una red neuronal artificial como la utilizada en este trabajo.

Los desempeños de cada uno de los modelos formulados fueron evaluados y comparados sobre la base de parámetros estadísticos, los cuales permiten elegir el mejor modelo con el mínimo error posible. Las métricas que normalmente se utilizan son: el coeficiente de determinación ( $R^2$ ), el error absoluto medio (*Mean Absolute Error, MAE*), la raíz cuadrática del error cuadrático medio (*Root Mean Square Error, RMSE*) y el coeficiente de variación del *RMSE (CV(RMSE))*. En las ecuaciones (1)-(4) se puede ver como estos valores son calculados, donde las variables  $y$ ,  $\hat{y}$ ,  $\bar{y}$  y  $n$ , que aparecen en las ecuaciones, representan los datos de entrenamiento, los datos estimados, el promedio de los datos de entrenamiento y el tamaño de los datos de entrenamiento, respectivamente. El coeficiente  $R^2$  determina el porcentaje de variabilidad de los datos que es explicada por el modelo. El parámetro *MAE* es definido como una cantidad utilizada para medir cuán próximos se encuentran los valores calculados respecto de los valores medidos. Por su parte, *RMSE* indica el nivel de dispersión que produce el modelo bajo análisis.

$$R^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (1)$$

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

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (3)$$

$$CV(RMSE) = \frac{RMSE}{\bar{y}} \cdot 100\% \quad (4)$$

## RESULTADOS

En la Figura 2 se puede observar el diagrama de dispersión entre la radiación solar global horaria y la temperatura (a), la humedad relativa (b), la velocidad del viento (c) y la presión atmosférica (d), para los 930 vectores de datos seleccionados para este estudio.

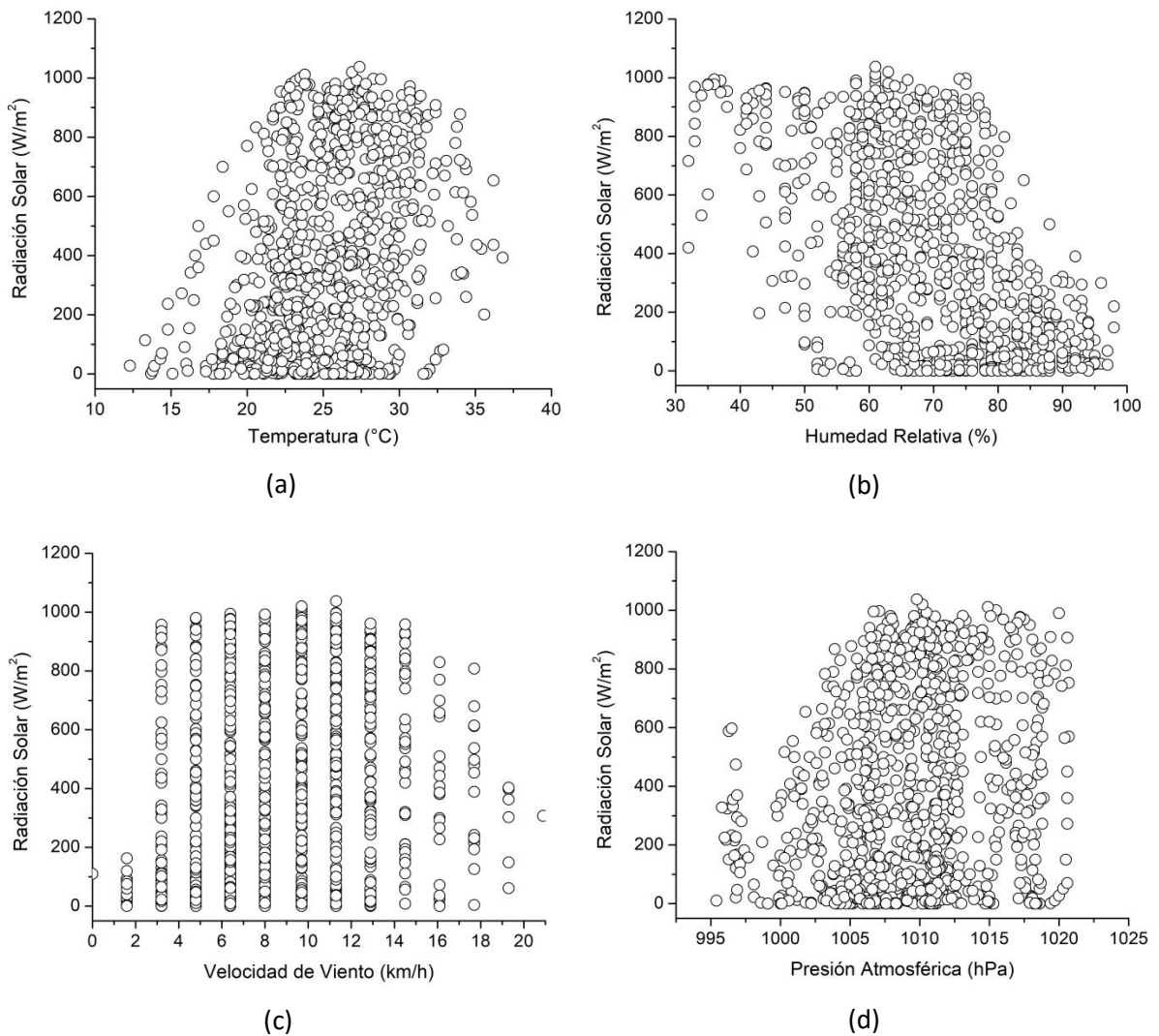


Figura 2: Diagrama de dispersión de la radiación solar global versus (a) Temperatura media; (b) Humedad relativa; (c) Velocidad de viento; y (d) Presión atmosférica.

Con el propósito de evaluar la relación lineal entre la radiación solar con las restantes magnitudes meteorológicas consideradas, se calculó el coeficiente de correlación de Pearson, obteniendo correlación positiva entre la radiación solar y las variables temperatura ( $R = 0.39$ ), velocidad del viento ( $R = 0.15$ ) y presión atmosférica ( $R = 0.02$ ), y correlación negativa con la humedad ( $R = -0.56$ ). Se puede observar que la temperatura y la humedad relativa, tienen la relación más fuerte con la radiación solar en comparación con las otras dos variables, lo cual sugiere que, para el conjunto de datos utilizados en este trabajo, la velocidad del viento y la presión atmosférica podrían excluirse de la entrada de los

modelos de RNA sin afectar de manera significativa los resultados obtenidos. En base a este análisis, solo la temperatura media y la humedad relativa fueron elegidas como las únicas variables meteorológicas de entrada a las RNA.

Por su parte, la Figura 3 presenta los valores correspondientes a la radiación solar global en función del horario de registro de la misma, para el mes completo de enero 2019. Puede observarse en esta figura una gran dispersión de valores de radiación solar a lo largo del mes, como consecuencia de la cobertura de nubes que atenúan la radiación solar global incidente en las diferentes horas del día.

A continuación, se realizó un análisis considerando dos modelos de RNA para la estimación de la radiación solar global cada 30 minutos. El modelo M1, que tiene como variables de entrada a la hora del día y a la temperatura media, y el modelo M2, el cual es similar al modelo M1 pero con el agregado de la humedad relativa como variable de entrada. También se realizaron pruebas con modelos de estimación incluyendo como variables de entrada a la velocidad del viento y a la presión atmosférica, no obteniendo mejoras en los resultados, lo cual confirma la conclusión del análisis de correlación de Pearson, que la temperatura y la humedad relativa son los parámetros de entrada más relevantes entre todas las variables meteorológicas tenidas en cuenta en este estudio. Cabe señalar que esta conclusión solo es cierta para este conjunto de datos. Las variables ambientales excluidas en este estudio podrían tener algún impacto en la precisión de los modelos en otras ubicaciones geográficas (Pang et al., 2020).

Un conjunto de 840 vectores de datos (28 días) fueron considerados para la fase de entrenamiento (90% del total), mientras que 90 vectores de datos, correspondientes a los días 7, 29 y 31 de enero 2019, fueron utilizados para la fase de validación de los modelos. La Figura 4 muestra la variación de la radiación solar para los tres días elegidos para la fase de validación, los cuales poseen diferentes condiciones de nubosidad: cielo despejado (círculos negros), parcialmente nublado (círculos azules) y mayormente nublado (círculos rojos).

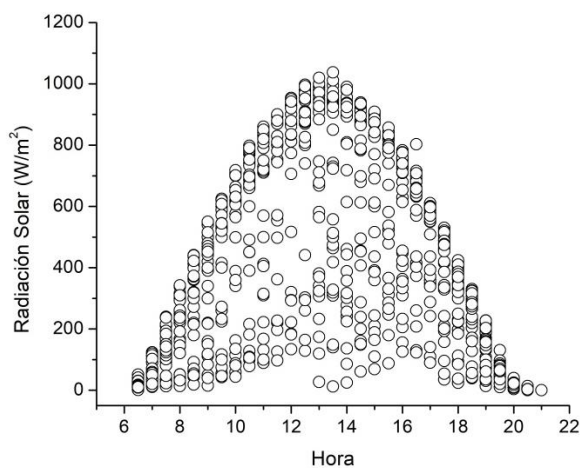


Figura 3. Valores medidos de radiación solar global horaria durante el mes de enero 2019 en Florencio Varela.

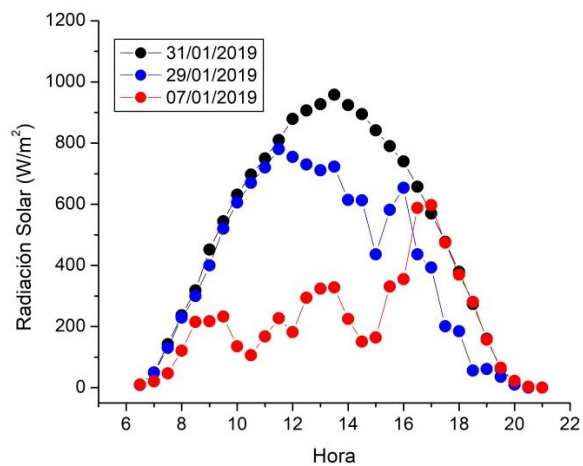


Figura 4. Radiación solar global horaria de los tres días del mes de enero 2019 utilizados para la fase de validación de los modelos de RNA.

Para el entrenamiento de los modelos se generó una estructura de red neuronal densa *feedforward*, donde todas las neuronas de una capa están conectadas a todas las neuronas de la siguiente, y no existe ninguna retroalimentación. Las principales características son:

- Algoritmo de aprendizaje: *RMSProp*, el cual es un algoritmo estocástico adaptativo para el entrenamiento de las redes neuronales profundas.
- Función de activación: *ReLU* (unidades lineales rectificadas).

- Capas ocultas: tres capas densas.
- Neuronas en las capas ocultas: 64, 24 y 16 en la primera, segunda y tercer capa, respectivamente.

Las Figuras 5 y 6 muestran los resultados obtenidos de la estimación de la radiación solar global horaria utilizando los modelos M1 (círculos rojos) y M2 (círculos azules), respectivamente. En estas figuras, pueden observarse además, a modo de comparación, los valores de radiación medidos por el piranómetro (círculos negros). Mientras que en las Figuras 7 y 8 se muestran los diagramas de dispersión correspondientes, conjuntamente con la recta identidad, para los modelos M1 y M2, respectivamente.

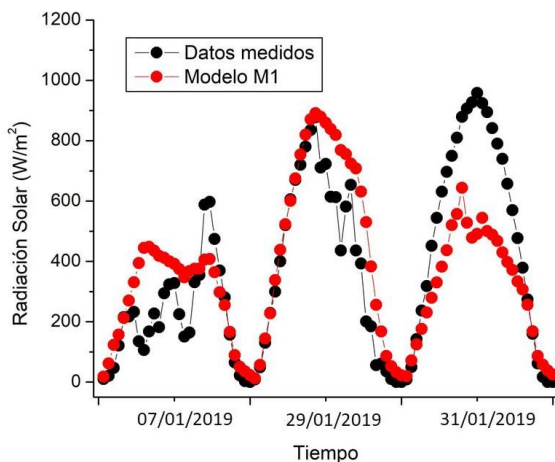


Figura 5: Estimación de la radiación solar utilizando el modelo M1 y la comparación con los datos medidos, para tres días diferentes.

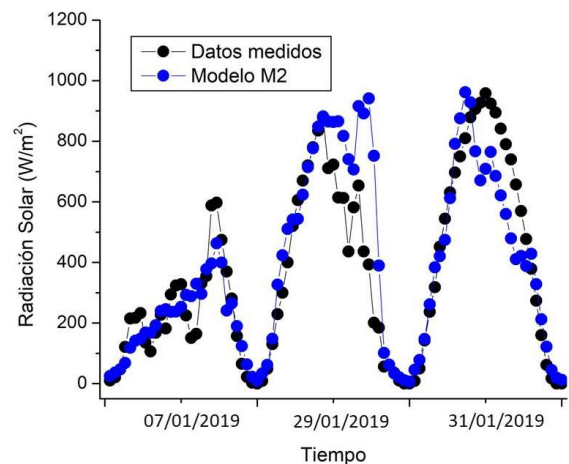


Figura 6: Estimación de la radiación solar utilizando el modelo M2 y la comparación con los datos medidos, para tres días diferentes.

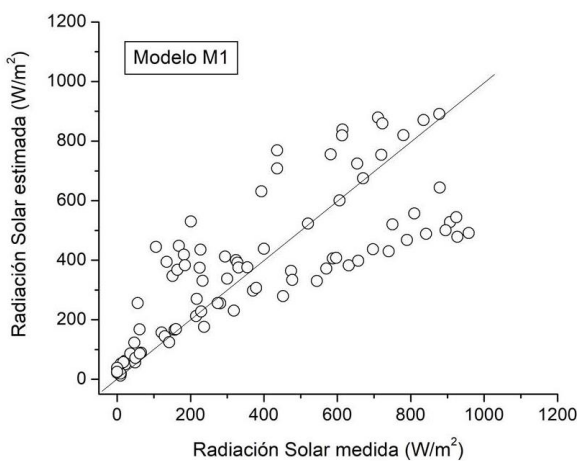


Figura 7: Diagrama de dispersión entre la radiación solar medida y la estimada por el modelo M1.

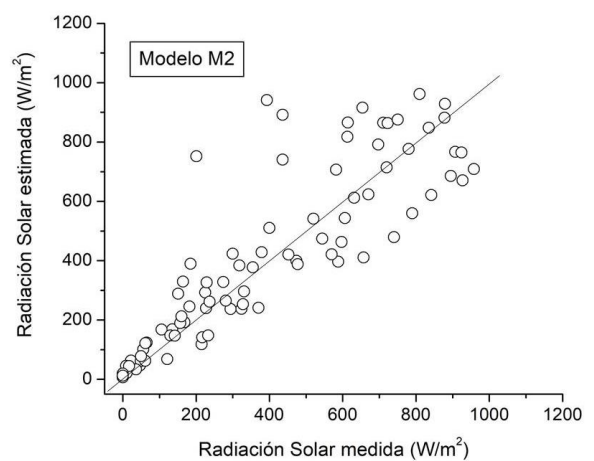


Figura 8: Diagrama de dispersión entre la radiación solar medida y la estimada por el modelo M2.

Para visualizar de manera más clara la diferencia obtenida entre la radiación solar global medida y la estimada por los modelos M1 y M2, la Figura 9 presenta el error de estimación para estos modelos, donde puede observarse que los errores son significativos, siendo el peor caso para el modelo M1 y para el día 31 de enero 2019, considerado como de cielo despejado. La Tabla 2 presenta los valores de las métricas obtenidas para cada uno de los modelos. El pobre desempeño de estos modelos puede deberse fundamentalmente al conjunto limitado de datos utilizado para el entrenamiento (28 días) y a la gran dispersión de valores de radiación solar, mostrada en la Figura 3.

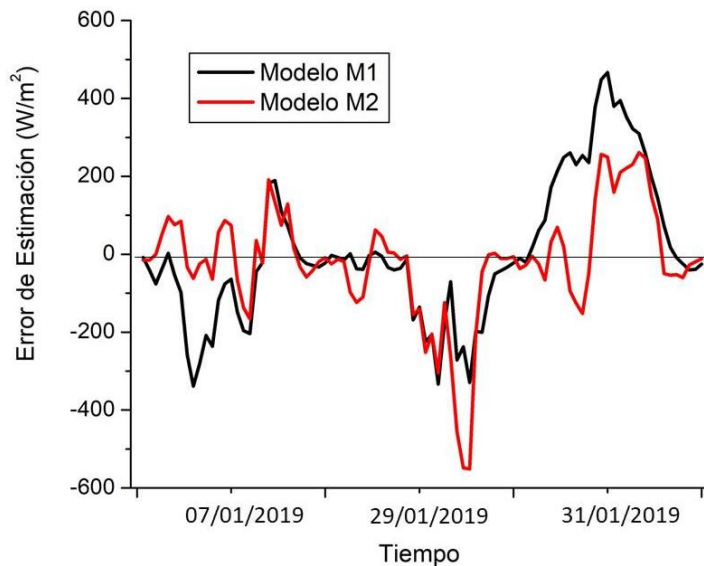


Figura 9: Comparación de errores de estimación para los modelos M1 y M2.

Modelo	R <sup>2</sup>	MAE	RMSE	CV(RMSE) (%)
M1	0.62	132.7	180.8	48.8
M2	0.75	98.9	147.9	39.9

Tabla 2: Valores de las métricas obtenidas, utilizadas para evaluar el desempeño de los modelos propuestos.

El análisis llevado a cabo hasta aquí, utiliza los modelos generales M1 y M2, los cuales han sido entrenados con todo tipo de nubosidad, y sus desempeños han sido evaluados para diferentes condiciones de cielo. Con el propósito de mejorar los resultados anteriores, se utilizó el índice de claridad  $K_t$  para los 31 días del mes de enero 2019, entre las 6:00 y las 21:00 horas, con un intervalo de 30 minutos. Este parámetro, definido como la razón entre la radiación solar global sensada en la superficie terrestre y la radiación solar extraterrestre, indica que tan clara es la atmósfera en ese instante, siendo sensible a la presencia de nubosidad y polución atmosférica. Para calcular  $K_t$ , se calculó previamente la radiación solar extraterrestre sobre un plano horizontal, a partir de ecuaciones que requieren como parámetros de entrada al número de día del año, la latitud, la longitud y la hora oficial (Grossi Gallegos, 2004).

Los valores de  $K_t$ , que varían entre 0 y 1, indican que fracción de la radiación solar extraterrestre es la que llega a la estación meteorológica a los diferentes instantes en los que se tomaron los valores de radiación solar global. La Tabla 3 presenta a modo de ejemplo los valores medidos de la radiación solar global ( $R_{Med.}$ ) y los valores calculados de la radiación solar extraterrestre ( $R_{Ext}$ ) y del índice de claridad  $K_t$ , para los días 7, 29 y 31 de enero 2019.

El índice de claridad es una herramienta muy valiosa y ampliamente utilizada para caracterizar las diferentes condiciones de cielo en un lugar específico (Okogbue et al., 2009). Además, posibilita la comparación entre diferentes épocas del año, e incluso entre localidades con diferentes alturas sobre el nivel del mar (Salazar et al., 2007). En este trabajo se adoptó la siguiente clasificación (Okogbue et al., 2009; Yousif et al., 2013):

- Cielo Nublado (CN):  $K_t < 0.2$
- Cielo Parcialmente Nublado (CPN):  $0.2 < K_t < 0.6$
- Cielo Despejado (CD):  $K_t > 0.6$



Hora oficial	07/01/2019			29/01/2019			31/01/2019		
	R_Med. (W/m <sup>2</sup> )	R_Ext. (W/m <sup>2</sup> )	Kt	R_Med. (W/m <sup>2</sup> )	R_Ext. (W/m <sup>2</sup> )	Kt	R_Med. (W/m <sup>2</sup> )	R_Ext. (W/m <sup>2</sup> )	Kt
6:30	10	63.94	0.16	9	81.67	0.11	9	68.64	0.13
7:00	21	198.81	0.11	49	224.30	0.22	50	204.04	0.25
7:30	47	334.23	0.14	130	367.47	0.35	142	339.97	0.42
8:00	121	467.87	0.26	229	508.71	0.45	237	474.11	0.50
8:30	215	597.46	0.36	300	645.61	0.46	318	604.15	0.53
9:00	217	720.77	0.3	400	775.83	0.52	452	727.89	0.62
9:30	233	835.69	0.28	520	897.14	0.58	544	843.19	0.65
10:00	135	940.26	0.14	606	1007.47	0.60	631	948.09	0.67
10:30	106	1032.69	0.1	670	1104.92	0.61	697	1040.79	0.67
11:00	168	1111.4	0.15	720	1187.84	0.61	750	1119.70	0.67
11:30	227	1175.03	0.19	780	1254.79	0.62	810	1183.48	0.68
12:00	182	1222.51	0.15	835	1304.64	0.64	879	1231.04	0.71
12:30	294	1253.01	0.23	878	1336.53	0.66	907	1261.55	0.72
13:00	324	1266.03	0.26	711	1349.92	0.53	927	1274.51	0.73
13:30	328	1261.33	0.26	723	1344.57	0.54	958	1269.67	0.75
14:00	225	1238.99	0.18	614	1320.59	0.46	924	1247.14	0.74
14:30	151	1199.4	0.13	613	1278.37	0.48	895	1207.30	0.74
15:00	164	1143.23	0.14	436	1218.65	0.36	842	1150.81	0.73
15:30	331	1071.45	0.31	582	1142.44	0.51	790	1078.67	0.73
16:00	355	985.28	0.36	654	1051.05	0.62	740	992.08	0.75
16:30	588	886.2	0.66	436	946.04	0.46	657	892.55	0.74
17:00	597	775.91	0.77	393	829.20	0.47	570	781.77	0.73
17:30	474	656.28	0.72	201	702.55	0.29	477	661.64	0.72
18:00	370	529.37	0.7	185	568.23	0.33	379	534.21	0.71
18:30	281	397.35	0.71	56	428.57	0.13	274	401.66	0.68
19:00	157	262.48	0.6	61	285.93	0.21	160	266.26	0.60
19:30	65	127.06	0.51	36	142.76	0.25	62	130.33	0.48
20:00	22	68	0.32	0	1.52	0.00	17	71	0.24
20:30	0	0	0	0	0	0.00	0	0	0.00
21:00	0	0	0	0	0	0.00	0	0	0.00

Tabla 3: Valores de radiación solar medida y extraterrestre para tres días diferentes de enero 2019.

Para los 930 valores de Kt calculados, se obtuvieron 210, 302 y 418 valores para CN, CPN y CD, respectivamente, lo cual implica que sólo el 45% del total de los datos corresponden a instantes considerados como Cielo Despejado. Para que el sistema propuesto pueda clasificar los datos en las tres categorías mencionadas: CN, CPN y CD, será necesario una variable de entrada adicional al sistema que advierta o informe sobre la nubosidad y se actualice cada 30 minutos, por ejemplo, utilizando imágenes satelitales, de cielo o bien el pronóstico del tiempo.

Posteriormente, con los 418 vectores de datos correspondientes a la categoría CD ( $Kt > 0.6$ ), se desarrollaron dos modelos de RNA, denominados M1\_CD y M2\_CD, utilizando el mismo criterio de los modelos M1 y M2 previamente definidos, respectivamente. En este caso, fueron elegidos para la fase de validación 21 valores correspondientes al día 31 de enero de 2019 (los comprendidos entre las 9:00 y las 19:00 horas), mientras que los restantes 397 fueron utilizados para la fase de entrenamiento de los modelos.

La Figura 10 muestra los resultados obtenidos de la estimación de la radiación solar global horaria utilizando los modelos M1\_CD (círculos rojos) y M2\_CD (círculos azules), los cuales son comparados con los valores de radiación medidos por el piranómetro (círculos negros). Finalmente, la Figura 11 presenta el error de estimación para estos modelos, y su comparación con los valores correspondientes a los modelos M1 y M2 (mostrados en la Figura 9) para el día 31 de enero de 2019. Se puede observar, a pesar de la poca cantidad de datos utilizados, una reducción significativa de los errores al entrenar los modelos únicamente con valores de cielo despejado. La Tabla 4 presenta los valores de las métricas obtenidas para cada uno de los modelos.

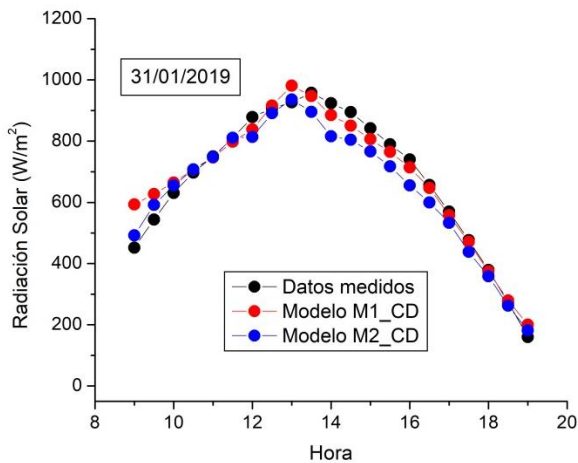


Figura 10: Estimación de la radiación solar utilizando los modelos M1\_CD y M2\_CD y su comparación con los datos medidos.

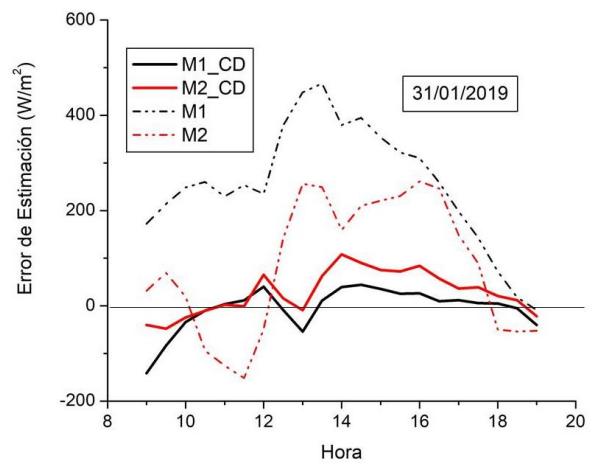


Figura 11: Errores de estimación de los modelos M1\_CD y M2\_CD entrenados con datos correspondientes a  $K_t > 0.6$ . Comparación con los errores de estimación de los modelos M1 y M2 de la Figura 9.

Modelo	R <sup>2</sup>	MAE	RMSE	CV(RMSE) (%)
M1_CD	0.96	30.8	44.3	6.5
M2_CD	0.94	42.6	52.5	7.7

Tabla 4: Valores de las métricas obtenidas, utilizadas para evaluar el desempeño de los modelos propuestos para la condición de cielo despejado.

Es importante tener en cuenta que los modelos de RNA desarrollados son entrenados con valores medidos con instrumentos que poseen errores de medición propios. Para una correcta evaluación del desempeño de los modelos, se debería analizar la sensibilidad de los mismos frente a estos errores instrumentales.

Un análisis similar al presentado para los datos categorizados como CD, fue llevado a cabo también para los días CPN y CN, aunque los errores de estimación obtenidos son comparables a los de los modelos M1 y M2, lo que implica que en estos casos la poca cantidad de datos sí es un factor clave para la obtención de mejores modelos.

## CONCLUSIONES

En este trabajo se desarrollaron modelos de redes neuronales para estimar la radiación solar global horaria en la localidad de Florencio Varela. Los resultados obtenidos muestran que cuando los modelos son entrenados con todo tipo de nubosidad, los errores obtenidos son apreciables. Esto se debe a la poca cantidad de datos utilizados y a la gran dispersión de valores de radiación solar global medidos. Mientras que cuando los modelos son entrenados únicamente con datos clasificados, de acuerdo al índice de claridad, como cielo despejado, el error de estimación se reduce considerablemente, a pesar del reducido número de datos utilizado. Por lo tanto, estos modelos podrían aplicarse en lugares donde

no se dispone de valores de radiación solar medidos, o bien se cuente con un número limitado de datos previos. Se puede advertir también de los resultados obtenidos, que los modelos M1\_CD y M2\_CD presentan un comportamiento muy aproximado entre sí, por lo que para simplificar la implementación del sistema, se podría considerar la exclusión de la variable humedad relativa, es decir, seleccionar el modelo M1\_CD, donde el único parámetro de entrada a medir es la temperatura.

## REFERENCIAS

- Assi A, Jama M, Al-Shamisi M. (2012). Prediction of Global Solar Radiation in Abu Dhabi. *Int. Scholarly Research Network ISRN Renewable Energy*. Vol. 2012, Article ID 328237, 10 pages.
- Çoban V, Onar S. (2020). In: Kahraman C. et al., *Intelligent and Fuzzy Techniques in Big Data Analytics and Decision Making*. INFUS 2019. *Advances in Intelligent Systems and Computing*, vol 1029. Springer.
- Erdinc O, Uzunoglu M. (2012). Optimum design of hybrid renewable energy systems: overview of different approaches. *Renewable and Sustainable Energy Reviews*. Vol. 16, no. 3. pp. 1412–1425.
- Grossi Gallegos H. (2004). *Notas sobre Radiación Solar*. Universidad Nacional de Lujan – Departamentos de Ciencias Básicas.
- Harrington P. (2012). *Machine Learning in Action*. Manning Publications Co.
- Jordan M, Mitchell T. (2015). Machine learning: Trends, perspectives, and prospects. *Science*. Vol. 349, Issue 6245, pp. 255-260.
- Kaplanis S, Kaplani E. (2010). Stochastic prediction of hourly global solar radiation for Patra, Greece. *Applied Energy*. Vol. 87, pp. 3748–3758.
- Khatib T, Elmenreich W. (2015). A Model for Hourly Solar Radiation Data Generation from Daily Solar Radiation Data Using a Generalized Regression Artificial Neural Network. *International Journal of Photoenergy*. Vol. 2015 Article ID 968024, 13 pages.
- Leal S, Tiba C, Piacentini R. (2009). Modelos estadísticos para determinação da irradiação solar UV diária no estado de Pernambuco. *Avances en Energías Renovables y Medio Ambiente* 13, pp. 11.41–11.47.
- Neelamegam P, Amirtham V. (2016). Prediction of solar radiation for solar systems by using ANN models with different back propagation algorithms. *Journal of Applied Research and Technology* Vol. 14. 206–214.
- Okogbue E, Adedokun J, Holmgren B. (2009). Hourly and daily clearness index and diffuse fraction at a tropical station, Ile-Ife, Nigeria. *International Journal of Climatology*. Vol. 29. 1035–1047.
- Pandey C, Katiyar K. (2013). Solar Radiation: Models and Measurement Techniques. *Journal of Energy*. Vol. 2013, Article ID 305207, 8 pages.
- Pang Z, Niu F, O’Neill Z. (2020). Solar radiation prediction using recurrent neural network and artificial neural network: A case study with comparisons. *Renewable Energy*. Vol. 156. 279–289.
- Polo Martínez J. (2010). *Optimización de modelos de estimación de la radiación solar a partir de imágenes de satélite*. Tesis Doctoral. Universidad Complutense de Madrid.
- Raichijk C, Grossi Gallegos H, Righini R. (2005). Evaluación de un método alternativo para la estimación de valores medios mensuales de irradiación global en Argentina. *Avances en Energías Renovables y Medio Ambiente* Vol. 9. pp. 11.05–11.08.
- Raichijk C. (2008). Estimación de la irradiación solar global en Argentina mediante el uso de redes neuronales artificiales. *Energías Renovables y Medio Ambiente* Vol. 22. pp. 1–6
- Salazar G, Saravia L, Romero G. (2007). Comparación de datos esporádicos de irradiancia medidos simultáneamente en dos localidades de la provincia de Salta. *Avances en Energías Renovables y Medio Ambiente*. Vol. 11. pp. 73–79.
- Sayago S, Bocco M, Ovando G, Willington E. (2011). Radiación solar horaria: modelos de estimación a partir de variables meteorológicas básicas. *Avances en Energías Renovables y Medio Ambiente* Vol. 15. pp. 11.51–11.57.
- Yousif C, Quecedo G, Santos J. (2013). Comparison of solar radiation in Marsaxlokk, Malta and Valladolid, Spain. *Renewable Energy*. Vol. 49. 203–206.

## **ABSTRACT**

Solar radiation is a key factor in many applications, such as photovoltaic or thermal systems, architecture and agriculture. However, experimental data on solar radiation may not be available in all geographical areas. For this reason, in the past, a wide variety of theoretical models have been developed in order to estimate this parameter. This paper presents a comparative analysis of artificial neural network models for estimating the hourly global solar radiation in Florencio Varela, province of Buenos Aires, from easily obtained meteorological data (temperature and relative humidity). The results obtained show a poor performance of the models when they are trained with different sky conditions. This is mainly due to the limited data set used and the large dispersion of measured solar radiation values. On the contrary, when the clarity index  $K_t$  is used, and the models are trained with data corresponding to the clear sky condition ( $K_t > 0.6$ ), the estimation errors are significantly reduced. These models could be applied at sites where measured solar radiation values are unavailable.

**KEYWORDS:** Solar radiation. Artificial neural network. Cloudiness.