



Universidad Autónoma de Querétaro

Facultad de Informática

Modelo de predicción de redes neuronales para radiación solar
global optimizado por colonia artificial de abejas

Tesis

Que como parte de los requisitos
para obtener el Grado de

Maestro en Ciencias de la Computación

Presenta

ISC Sergio Alejandro Pérez Rodríguez

Dirigido por:

Dr. José Manuel Álvarez Alvarado

Querétaro, Qro. a 08 de agosto de 2024

La presente obra está bajo la licencia:
<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es>



CC BY-NC-ND 4.0 DEED

Atribución-NoComercial-SinDerivadas 4.0 Internacional

Usted es libre de:

Compartir — copiar y redistribuir el material en cualquier medio o formato

La licenciante no puede revocar estas libertades en tanto usted siga los términos de la licencia

Bajo los siguientes términos:



Atribución — Usted debe dar [crédito de manera adecuada](#), brindar un enlace a la licencia, e [indicar si se han realizado cambios](#). Puede hacerlo en cualquier forma razonable, pero no de forma tal que sugiera que usted o su uso tienen el apoyo de la licenciante.



NoComercial — Usted no puede hacer uso del material con [propósitos comerciales](#).



SinDerivadas — Si [remezcla, transforma o crea a partir](#) del material, no podrá distribuir el material modificado.

No hay restricciones adicionales — No puede aplicar términos legales ni [medidas tecnológicas](#) que restrinjan legalmente a otras a hacer cualquier uso permitido por la licencia.

Avisos:

No tiene que cumplir con la licencia para elementos del material en el dominio público o cuando su uso esté permitido por una [excepción o limitación](#) aplicable.

No se dan garantías. La licencia podría no darle todos los permisos que necesita para el uso que tenga previsto. Por ejemplo, otros derechos como [publicidad, privacidad, o derechos morales](#) pueden limitar la forma en que utilice el material.



Universidad Autónoma de Querétaro
Facultad de Informática
Maestría en Ciencias de la Computación

Modelo de predicción de redes neuronales para radiación solar global
optimizado por colonia artificial de abejas

Tesis

Que como parte de los requisitos para obtener el Grado
Maestro en Ciencias de la Computación

Presenta

ISC Sergio Alejandro Pérez Rodríguez

Dirigido por:

Dr. José Manuel Álvarez Alvarado

Dr. José Manuel Álvarez Alvarado
Presidente

Dra. Diana Margarita Córdova Esparza
Secretario

Dr. Julio Alejandro Romero González
Vocal

Dr. Mauricio Arturo Ibarra Corona
Suplente

Dr. Juvenal Rodríguez Reséndiz
Suplente

Centro Universitario, Querétaro, Qro.
México
Agosto 2024

A mi familia, cuyo apoyo constante, atención y confianza en mí han sido invaluable. Su amor y aliento me han dado la fuerza necesaria para llegar hasta aquí, y siempre estaré agradecido por su presencia en mi vida.

A mis profesores, quienes en cada etapa de mi educación vieron en mí un potencial especial. Su motivación y enseñanza me hicieron creer en mis capacidades y aspirar a grandes cosas. Su influencia ha sido fundamental en mi desarrollo académico.

A mis amigos y compañeros, quienes con su apoyo, compañía y palabras de aliento hicieron este camino más fácil y gratificante. Juntos hemos creado lazos de confianza y amistad que llevaré conmigo siempre.

A mi pareja, cuya importancia en mi vida no tiene comparación. Su apoyo incondicional en todos los aspectos y la motivación constante que me ha brindado han sido claves para alcanzar mis objetivos. Sin su amor y confianza, este logro no habría sido posible.

Finalmente, a todos aquellos que de alguna manera contribuyeron a la realización de esta tesis, les expreso mi más sincero agradecimiento. Cada uno de ustedes ha dejado una huella significativa en este viaje, y siempre les estaré agradecido.

Agradecimientos

Se extienden agradecimientos al Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCYT) por financiar el proyecto de investigación.

Resumen

En la presente tesis se propone un modelo híbrido para la predicción de la radiación solar global (GSR) utilizando el algoritmo de optimización bioinspirado Colonia Artificial de Abejas (ABC) combinado con un Perceptrón Multicapa (MLP). La motivación principal es abordar la intermitencia y volatilidad de la GSR, que impactan significativamente en la seguridad del suministro energético, especialmente en el contexto del aumento del uso de paneles fotovoltaicos. La metodología implementada incluye la optimización de los hiperparámetros del modelo MLP mediante el algoritmo ABC, lo que permite una selección óptima del número de neuronas en las capas ocultas, mejorando así la precisión de las predicciones. La validez del modelo propuesto se verificó utilizando datos de estaciones meteorológicas automáticas en distintas ubicaciones de México. En la estación UAQ Aeropuerto, el modelo ABC-MLP mostró un RMSE de 48.4032 W/m^2 , un MAE de 19.5054 W/m^2 y R^2 de 0.9812 . Adicionalmente, se realizó una selección de características críticas como la temperatura del aire, humedad relativa, radiación solar global promedio, velocidad y dirección del viento, las cuales contribuyen a la precisión del modelo. La revisión sistemática de la literatura incluyó un análisis exhaustivo de las tendencias actuales y futuras direcciones de investigación en la predicción de GSR utilizando técnicas de machine learning, deep learning y algoritmos de optimización metaheurísticos. Los resultados obtenidos demuestran la robustez y generalización del modelo propuesto Selector ABC-MLP en diferentes contextos geográficos, destacando su superioridad en la predicción de la GSR. Este trabajo contribuye al desarrollo de sistemas de energía solar más eficientes y a la gestión optimizada de recursos energéticos, proporcionando una herramienta valiosa para la toma de decisiones en el sector energético.

Abstract

In this thesis, a hybrid model is proposed for the prediction of global solar radiation (GSR) using the bio-inspired optimization algorithm Artificial Bee Colony (ABC) combined with a Multilayer Perceptron (MLP). The main motivation is to address the intermittency and volatility of GSR, which significantly impacts the security of energy supply, especially in the context of the increased use of photovoltaic panels.

The implemented methodology includes the optimization of the MLP model's hyperparameters through the ABC algorithm, which allows for an optimal selection of the number of neurons in the hidden layers, thereby improving prediction accuracy. The validity of the proposed model was verified using data from automatic weather stations in various locations in Mexico. At the UAQ Aeropuerto station, the ABC-MLP model showed an RMSE of 48.4032 W/m², an MAE of 19.5054 W/m², and an R² of 0.9812.

Additionally, a selection of critical features such as air temperature, relative humidity, average global solar radiation, wind speed, and direction was performed, contributing to the model's accuracy. The systematic literature review included an exhaustive analysis of current trends and future research directions in GSR prediction using machine learning, deep learning, and metaheuristic optimization algorithms.

The obtained results demonstrate the robustness and generalization of the proposed ABC-MLP model in different geographical contexts, highlighting its superiority in GSR prediction. This work contributes to the development of more efficient solar energy systems and optimized energy resource management, providing a valuable tool for decision-making in the energy sector.

Índice general

Agradecimientos

Resumen

Abstract

1. Introducción	1
1.1. Antecedentes	4
1.1.1. MLP	6
1.1.2. CNN	8
1.1.3. Redes Neuronales Recurrentes	10
1.1.4. Otros Modelos de ML	11
1.2. Descripción del problema	17
1.3. Justificación	18
2. Fundamentación teórica	19
2.1. Redes neuronales	19
2.2. Técnicas de preprocesamiento de datos	19
2.2.1. Limpieza de datos	20
2.2.2. Integración de datos	20
2.2.3. Reducción de datos	20
2.2.4. Transformación de los datos	21
2.3. Pronóstico de series temporales	21
2.4. Optimización	22
2.4.1. Algoritmos de optimización bioinspirados	23
2.4.2. Optimización de hiperparámetros	23
2.5. Métricas de rendimiento del modelo	24
3. Hipótesis y Objetivos	27
3.1. Hipótesis	27
3.2. Objetivos	27

4. Metodología	28
4.1. Adquisición de datos	29
4.2. Análisis de datos	29
4.3. Construcción del modelo	32
4.4. Optimización de las capas del modelo MLP	33
4.5. Entrenamiento del modelo con base de datos preprocesada	38
4.6. Evaluación del Modelo	38
5. Resultados y discusión	40
5.1. Resultados de la revisión sistemática	40
5.2. Resultados de la comparativa de modelos	43
5.3. Validación con otras bases de datos	48
6. Conclusiones	51
6.1. Trabajos a futuro	52
7. Anexos	53
Referencias	66
Apéndice	67

Introducción

La energía ha jugado un papel crucial en una sociedad en crecimiento, y la energía solar ha surgido como una alternativa prometedoras y sostenible (Sharma¹ y Khurana, 2018). La creciente dependencia de la tecnología y la escasez de recursos energéticos tradicionales han impulsado la búsqueda de soluciones, subrayando la importancia de una predicción precisa de la radiación solar global (GSR) para una captura eficiente de la energía solar (J. Fan et al., 2018). En muchos países, especialmente aquellos en desarrollo, el alto costo de los instrumentos de medición de GSR ha dificultado la obtención de datos precisos. Para abordar este desafío, se han desarrollado varios modelos de predicción de SR empleando una variedad de técnicas, como el Aprendizaje Automático (ML), el Aprendizaje Profundo (DL) y algoritmos de optimización, utilizando datos meteorológicos e imágenes satelitales. Si bien los modelos de ML han demostrado un rendimiento notable, es crucial destacar que la búsqueda de parámetros se ha vuelto esencial en la construcción de la estructura subyacente de cualquier modelo de ML. Esto se ha logrado incorporando métodos de optimización que han permitido afinar y refinar los parámetros del modelo, jugando un papel crucial en su efectividad y precisión (Kumar et al., 2023a).

Se han identificado un total de 232 trabajos mediante la búsqueda en la base de datos de Scopus utilizando términos clave como 'global solar radiation', 'prediction', 'forecast', 'machine learning' y 'deep learning' entre los años 2000 y 2024, restringido a trabajos en inglés y exclusivamente a artículos. El interés en predecir la radiación solar global ha mostrado un crecimiento constante a lo largo de los años, como se evidencia en la Fig. 1.1. En los últimos 4 años, se han publicado un total de 74 artículos, lo que sugiere una oportunidad de investigación significativa en esta área.

Los algoritmos metaheurísticos han contribuido significativamente a mejorar varios modelos de aprendizaje automático, como el bosque aleatorio (RF) (Prasad et al., 2019) y la regresión de vectores de soporte (SVR) (Ghazvinian et al., 2019), facilitando la obtención de parámetros óptimos para cada modelo. De manera similar, han impactado positivamente en los modelos de aprendizaje profundo, como las redes neuronales convolucionales (CNN) (Ghimire, Deo, Casillas-Pérez, y Salcedo-Sanz, 2022) o el perceptrón multicapa (MLP) (Djeldjli et al., 2023), mejorando sustancialmente su eficacia en la predicción de SR.

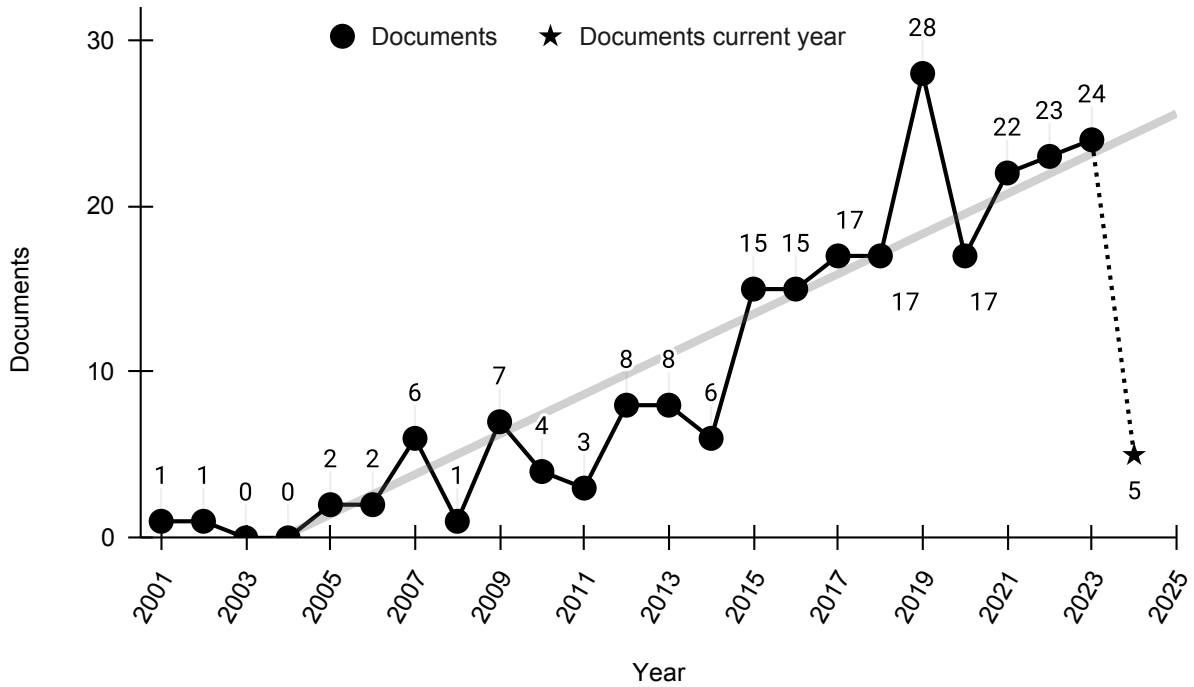


Figura 1.1. Número de trabajos que han utilizado diferentes algoritmos de IA para la predicción de la radiación solar global en Scopus.

El objetivo de este proyecto es realizar una comparativa de los métodos más recientes utilizados en la investigación y desarrollo en el campo de la predicción de SR, especialmente mediante la implementación de modelos híbridos basados en ML, DL y algoritmos metaheurísticos. Estos enfoques han sido diseñados para mitigar la carga computacional requerida, constituyendo una contribución clave en términos de eficiencia y precisión (Ng et al., 2023). El propósito fundamental de realizar una revisión sistemática ha radicado en obtener una visión más amplia y detallada del estado actual de la optimización de modelos mediante el uso de algoritmos metaheurísticos para la predicción de SR. Este análisis ha abarcado tanto técnicas de ML como de DL, además de considerar las variables meteorológicas más comúnmente utilizadas en estos modelos. Además, ha examinado el horizonte de tiempo de predicción adoptado por tales modelos, así como las métricas de evaluación empleadas para demostrar su efectividad. Este enfoque ha permitido discernir tanto la efectividad de los modelos como algunas de las configuraciones clave utilizadas en su desarrollo. Adicionalmente, se ha buscado comprender el propósito subyacente de los algoritmos de optimización empleados en los diversos trabajos de investigación revisados.

En esta tesis, se ha desarrollado un modelo para la predicción de GSR mediante la integración del algoritmo de optimización Colonia Artificial de Abejas (ABC) con un Perceptrón Multicapa (MLP). Este trabajo no solo aborda la necesidad de mejorar la precisión en las

predicciones de GSR, sino que también ofrece un enfoque robusto y eficiente que puede ser aplicado en diversas ubicaciones geográficas. A continuación, se detallan las principales aportaciones de esta investigación:

- Propuesta de un modelo híbrido Selector ABC-MLP: Desarrollo de un modelo híbrido que combina el algoritmo de optimización bioinspirado Colonia Artificial de Abejas (ABC) con un Perceptrón Multicapa (MLP) para la predicción de GSR, mostrando una mejora significativa en la precisión de las predicciones.
- Optimización de hiperparámetros: Implementación del algoritmo ABC para la optimización de los hiperparámetros del modelo MLP, lo que incluye la selección óptima del número de neuronas en las capas ocultas, obteniendo mejores resultados en comparación con modelos que no emplean técnicas de optimización.
- Validación con datos reales: Realización de validaciones cruzadas utilizando datos de diferentes estaciones meteorológicas automáticas en México, demostrando la robustez y generalización del modelo propuesto en distintas ubicaciones geográficas, con resultados comparativos que muestran la superioridad del modelo propuesto Selector ABC-MLP en términos de RMSE, MAE y R^2 .
- Selector de características: Empleando el selector de características incorporado en el modelo ABC-MLP propuesto, se identifican variables cruciales como la temperatura del aire, la humedad relativa, la radiación solar global promedio, así como la velocidad y dirección del viento. Esto contribuye significativamente a mejorar el rendimiento del modelo predictivo.
- Revisión sistemática de la literatura: Se llevó a cabo una revisión exhaustiva de la literatura sobre la predicción de la radiación solar mediante técnicas de machine learning y deep learning, además de algoritmos de optimización metaheurísticos. Esta revisión permitió identificar las tendencias actuales y las futuras direcciones de investigación en el ámbito de la predicción de radiación solar.

La tesis está organizada de la siguiente manera: La Sección 1 proporciona una introducción al tema, destacando su importancia y antecedentes, al tiempo que se esbozan las contribuciones previstas de este trabajo. La Sección 2 ofrece una visión general concisa de las técnicas de preprocesamiento de datos, la predicción de series temporales, los algoritmos de optimización bioinspirados, las redes neuronales artificiales (ANN) y las métricas de rendimiento. En la Sección 3, se describe la hipótesis y objetivos, tanto el general como los objetivos específicos. En la Sección 4 se detalla el marco metodológico utilizado en el presente trabajo. La Sección 5 presenta un resumen de los artículos que se centran en la predicción de SR con ML o DL y técnicas de optimización metaheurísticas, además de mostrar los resultados de una comparativa entre modelos y el modelo propuesto, así como los resultados de su validación. La Sección 6 muestra las conclusiones extraídas de este trabajo, sugerencias para futuras investigaciones, anexos y, finalmente, una lista de referencias bibliográficas.

1.1. Antecedentes

En el mundo se han ido desarrollando nuevas formas de obtener energía, dada la necesidad de reducir la contaminación generada por los combustibles fósiles, tales como la energía solar, la cual es una fuente de energía renovable usada en una variedad de aplicaciones, incluyendo la generación de electricidad, bombeo de agua, calentamiento de agua o aire y desalinización del agua. La información de la radiación solar global es crítica para estas aplicaciones (Chaibi, Benghoulam, Tarik, Berrada, y El Hmaidi, 2022). La búsqueda de un modelo de predicción de radiación solar está directamente motivada por el área de energía solar, ya que es un parámetro decisivo y crítico para la generación y administración de la energía solar (Tao et al., 2021; Premalatha et al., 2018). Dado el interés internacional de fuentes de energía renovables, se ha incrementado las aplicaciones de sistemas de energía inteligentes, tales como pronósticos de carga, pronósticos de salida de energía renovable, precio de energía, detección de perturbaciones en la calidad de energía y detección de fallos en sistemas y equipo de energía (Alkhayat y Mehmood, 2021).

La radiación solar es afectada por las estaciones y la geografía, por lo que tiene discontinuidades e incertidumbres. Estas características son la razón del enfoque de predecir la radiación solar a predecir la salida del sistema de energía solar (Wang y Shi, 2021). La entrada de energía en los paneles fotovoltaicos está fuertemente correlacionada a la radiación solar global, que es influenciada por muchos otros factores, como la latitud, la temporalidad o las condiciones del cielo (Premalatha et al., 2018). La radiación solar global es muy intermitente y caótica, incluso las pequeñas fluctuaciones en la radiación del sol pueden tener un gran impacto en la seguridad de suministro de energía. Esto hace que se requieran modelos predictivos cada vez más exactos debido al aumento potencial del uso de energía solar suministrada en paneles fotovoltaicos (Rajagukguk et al., 2020). Es por eso por lo que el crear un modelo predictivo de la radiación solar global es muy importante para asegurar una práctica óptima de despacho y gestión de la energía (Ghimire, Deo, Casillas-Pérez, y Salcedo-Sanz, 2022). De acuerdo con el estado del arte, se han creado diferentes modelos que usan técnicas de Machine learning, tales como máquinas de soporte de vectores, bosques aleatorios, al igual que técnicas de Deep learning, como las redes neuronales, siendo estas las que han presentado gran relevancia en los últimos años (Gensler et al., 2016; Kumari y Toshniwal, 2021).

Los parámetros o variables que inciden en la predicción de la radiación solar global son clasificados en tres, los cuales son las astronómicas (Besharat et al., 2013), geográficas y de fecha (Bounoua et al., 2021). Las variables se pueden observar en la Fig. 1.2.

El proceso de búsqueda de literatura relevante, los criterios de inclusión y exclusión aplicados, así como las bases de datos y herramientas analíticas utilizadas se describen a continuación. La Fig. 1.3 muestra un análisis del número de documentos publicados anualmente desde 2018 hasta 2024. Este gráfico es esencial para entender la evolución y el creciente interés en la búsqueda de mejores modelos híbridos para predecir la radiación solar, destacando cómo la cantidad de investigaciones ha aumentado a lo largo de los años.

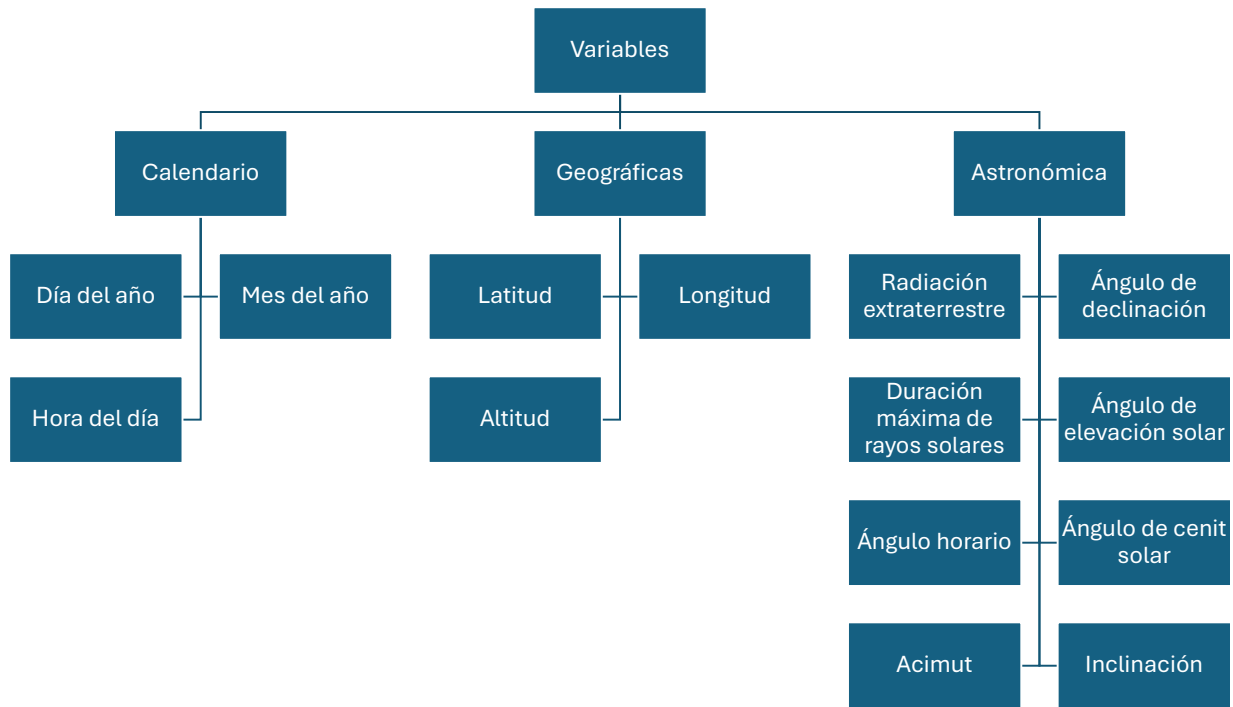


Figura 1.2. Variables que inciden en la predicción de la radiación solar global.

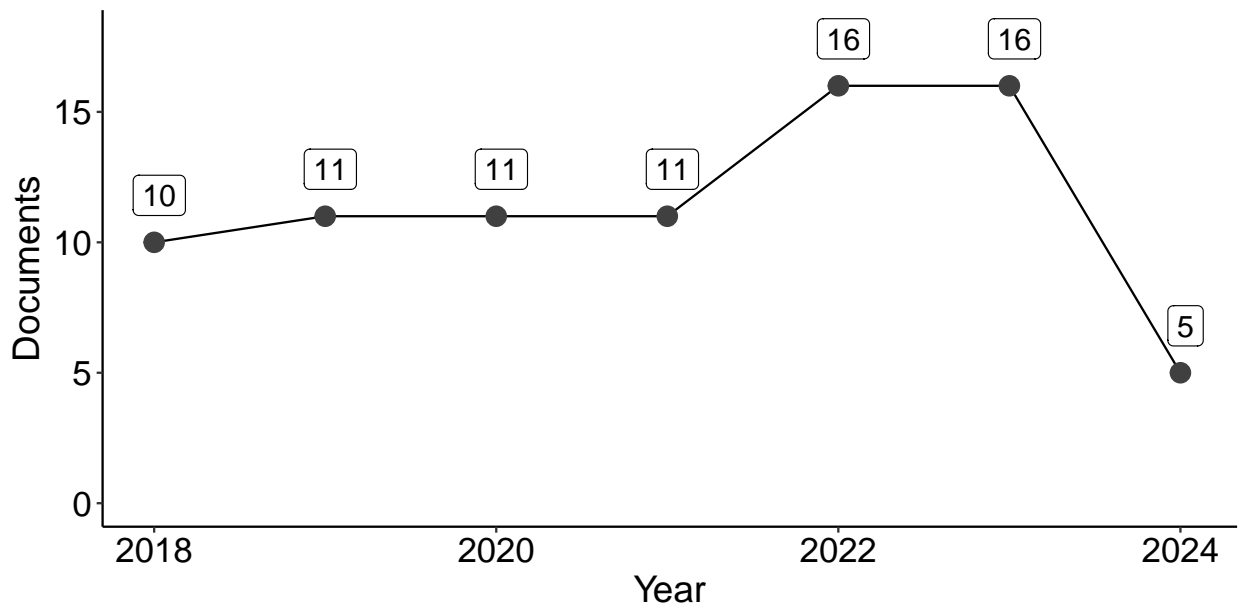


Figura 1.3. Informe de publicaciones desde 2018 hasta 2024, presentando los documentos publicados por año en Scopus.

1.1.1. MLP

Se han investigado múltiples estudios que han implementado MLP junto con varios algoritmos de optimización metaheurística para mejorar la precisión y eficiencia de las predicciones, como se ilustra en la Tabla 1.1. Cada artículo seleccionado para esta revisión ha aplicado una combinación de técnicas metaheurísticas con MLP, con el objetivo de optimizar desde la selección de características hasta el ajuste de parámetros del modelo. Se ha analizado la metodología, los resultados obtenidos y la relevancia de cada enfoque, proporcionando así una visión integral de cómo se pueden emplear estos métodos híbridos para lograr predicciones más robustas y confiables.

Tabla 1.1. Trabajos donde se utilizan MLP para la predicción de SR.

Autor	Algoritmo	Comparativa	Horizonte temporal	Optimización	Resultados
(Jadidi et al., 2018)	GA	-	Diario	Selección de características	RMSE 47.6955 Wh/m ² y MAE 24.7772 Wh/m ²
(Vaisakh y Jayabharathi, 2022)	GU-DHOA	CNN, RNN	Intra-horario	Neuronas ocultas	RMSE 212.27 Wh/m ² y MAE 136.51 Wh/m ²
(Y. Zhang et al., 2019)	PSO	-	Intra-horario	Pesos y sesgos	R ² 0.7649–0.9678
(Djeldjli et al., 2023)	FFA	SVM	Diario	Selección de características	R 0.9321 W/m ² , rRMSE % 11.96 y MAPE % 8.26
(Huang et al., 2022)	GOA	ANFIS	Mensual	No mencionado	nMAE 0.545
(Xu et al., 2024)	EO	-	Horario	Pesos y sesgos	RMSE 141.61 W/m ² , MAE 108.07 W/m ² y R ² 0.91

Jadidi et al. (2018) desarrollaron un modelo para predecir la irradiancia solar horizontal global (GHI) en Carolina del Norte, EE. UU. Utilizaron algoritmos genéticos (GA) y optimización por enjambre de partículas (PSO). Se observó que GA demostró un mejor rendimiento en la optimización de una red neuronal con 15 capas ocultas del tipo MLP, con una población inicial y un máximo de iteraciones de 50. La función de activación óptima identificada fue la tangente hiperbólica sigmooidal. El modelo logró un error cuadrático medio (RMSE) de 47.6955 Wh/m² y un error absoluto medio (MAE) de 24.7772 Wh/m². Además, se realizaron

pruebas de validación del modelo para períodos de 24 horas, 4 días y una semana.

Vaisakh y Jayabarathi (2022) compararon modelos MLP, CNN y RNN para la predicción de la irradiancia solar (SI), con el objetivo de minimizar el error entre los valores reales y los predichos. En este estudio, se empleó un enfoque híbrido que combina el Algoritmo de Optimización de Caza de Ciervos (DHOA) y la Optimización de Lobo Gris (GWO) para optimizar el número de neuronas ocultas en las redes neuronales. Los resultados mostraron un RMSE de 212.27 Wh/m^2 y un MAE de 136.51 Wh/m^2 .

X. Zhang y Wei (2019) realizaron un estudio para mejorar la precisión en la predicción diaria de GSR en el noroeste árido de China utilizando una Red Neuronal de Retropropagación (BP) y una BP optimizada por el algoritmo PSO (PSO-BP), con el fin de optimizar parámetros críticos de la red como pesos y sesgos. Se encontró que los modelos PSO-BP (con un coeficiente de determinación R^2 de $0.7649\text{--}0.9678$) eran más precisos que los modelos BP (R^2 de $0.7215\text{--}0.9632$) para la predicción diaria de GSR en las cuatro subzonas del noroeste árido de China.

Djeldjli et al. (2023) implementaron un MLP con 5 capas ocultas, utilizando la función de activación Tansig en las capas ocultas y Pureline en la capa de salida. Este modelo se utilizó para predecir el GSR diario en Argelia. Se empleó el algoritmo de luciérnagas con una población de 10 y un máximo de 50 generaciones para seleccionar las mejores características para el entrenamiento de la red neuronal. Los resultados obtenidos fueron $R = 0.9321 \text{ W/m}^2$, $\text{rRMSE} \% = 11.96$ y $\text{MAPE} \% = 8.26$.

Huang et al. (2022) desarrollaron un modelo para predecir la SR mensual en regiones semiáridas y húmedas de Irán. En su estudio, compararon modelos MLP y ANFIS, que fueron optimizados utilizando el algoritmo de saltamontes, con una población de 300 individuos. Este enfoque resultó en el mejor valor de la función objetivo de $3.12 \text{ (KWh/m}^2\text{/día)}$, con un nMAE de 0.54.

Xu et al. (2024) utiliza una Red Neuronal Artificial Feed-Forward (FFANN) que es un tipo de MLP, optimizada con varios algoritmos metaheurísticos, incluyendo la Optimización de Equilibrio (EO), la Optimización Inspirada en Óptica (OIO), la Optimización Impulsada por el Viento (WDO) y el Algoritmo de Araña Social (SOSA). La FFANN tiene una configuración de (8,6,1), lo que significa que tiene 8 neuronas en la capa de entrada, 6 en la capa oculta y 1 en la capa de salida, utilizando Tansig y Purelin como funciones de activación. Los resultados mostraron que el EO-FFANN superó a los otros modelos en términos de precisión, con un RMSE de 142.38 W/m^2 en el entrenamiento y 141.61 W/m^2 en la prueba, un MAE de 110.09 W/m^2 en el entrenamiento y 108.07 W/m^2 en la prueba, y un R^2 de 0.91. El EO optimiza los pesos y sesgos de la FFANN para minimizar el error de predicción.

En resumen, la implementación de modelos MLP combinados con varios algoritmos de optimización metaheurística ha mejorado significativamente la precisión y eficiencia de las predicciones de radiación solar. Los estudios revisados en esta sección demuestran el uso efectivo de algoritmos genéticos, optimización por enjambre de partículas, optimización de caza de ciervos, optimización de lobo gris, algoritmo de luciérnaga y algoritmo de saltamontes para optimizar diferentes aspectos de los modelos MLP, desde la selección de características hasta la afinación de parámetros. Los resultados destacan la robustez de estos métodos híbridos.

dos, con mejoras notables en métricas clave como RMSE, MAE y R^2 en diferentes regiones geográficas y horizontes temporales, reforzando el potencial de estas técnicas avanzadas para mejorar el rendimiento predictivo en la predicción de RS.

1.1.2. CNN

Los estudios se han centrado en integrar CNN con varios algoritmos de optimización metaheurística para refinar la selección de características y optimizar los parámetros del modelo, con el objetivo de mejorar la precisión y eficiencia de las predicciones, como se muestra en la Tabla 1.2. A través de un análisis detallado, se ha discutido cómo diferentes investigaciones han implementado estas técnicas para abordar desafíos específicos en la predicción solar, evaluando sus metodologías, resultados y contribuciones clave.

Ghimire, Deo, Casillas-Pérez, y Salcedo-Sanz (2022) desarrollaron un modelo híbrido avanzado para la predicción de GSR, integrando CNN y utilizando la Optimización de Halcones de Harris (HHO) para la selección efectiva de características, notable por su rendimiento óptimo con una población de 250, lo cual demostró ser lo más efectivo. Este modelo fue probado en varias granjas solares en Australia, demostrando una capacidad predictiva superior en comparación con métodos ML convencionales como ANN, ELM, DBN, DNN y MARS. En la granja solar de Dunblane, el modelo logró un R de 0.960, un RMSE de 2.118 MJ/m²/día y un MAE de 1.587 MJ/m²/día.

En el estudio de Ghimire, Deo, Casillas-Pérez, Salcedo-Sanz, Sharma, y Ali (2022), se desarrolló un modelo híbrido para la predicción de GSR combinando CNN, LSTM y MLP, donde se utilizó ReLU como función de activación, junto con la optimización del Algoritmo de Moho de Baba (SMA) para la selección de características. Este enfoque se aplicó a granjas solares en Queensland, Australia, demostrando un rendimiento superior en todas las granjas seleccionadas en comparación con ML, incluyendo ANN, RF y SADE-ELM. Con un valor de R entre 0.880 y 0.935, junto con un RMSE entre 2.207 y 2.792 MJ/m²/día en diferentes granjas.

En el estudio realizado por Ghimire, Bhandari, et al. (2022), se propuso un modelo híbrido para la predicción de GSR en Queensland, Australia, integrando una CNN con tres capas convolucionales, utilizando ReLU como función de activación. Para la selección óptima de características, se utilizó el algoritmo de Optimización de Búsqueda de Átomos (ASO), y los hiperparámetros se optimizaron utilizando el método HyperOpt. El rendimiento del modelo se evaluó en seis granjas solares, demostrando superioridad sobre varios modelos alternativos tanto en ML como en ML, como LSTM, RBF, DBN, M5TREE y MARS. Específicamente, el modelo logró un RMSE que varía entre 2.172 y 3.305 MJ/m²/día y un MAE que varía entre 1.624 y 2.370 MJ/m²/día.

Vaisakh y Jayabarathi (2022) desarrollaron un modelo híbrido para la predicción de la irradiación solar utilizando MLP, CNN y RNN, optimizando el número de neuronas ocultas mediante un algoritmo híbrido que combina DHOA y GWO, llamado GU-DHOA. El modelo GU-DHOA-CNN obtuvo una predicción de irradiación solar con un RMSE de 2.11×10^2 y MAE de 166.53.

El estudio de Irshad et al. (2023) introdujo un modelo de predicción de SR basado en

Tabla 1.2. Trabajos donde se utilizan CNN para la predicción de SR.

Autor	Algoritmo	Comparativa	Horizonte temporal	Optimización	Resultados
(Ghimire, Deo, Casillas-Pérez, y Salcedo-Sanz, 2022)	HHO	ML(ELM)	Diario	Selección de características	R 0.960, RMSE 2.118 MJ/m ² /día, MAE 1.587 MJ/m ² /día
(Ghimire, Deo, Casillas-Pérez, Salcedo-Sanz, Sharma, y Ali, 2022)	SMA	ANN, RF y SADE-ELM	Diario	Selección de características	R 0.880-0.935 y RMSE: 2.207-2.792 MJ/m ² /día
(Ghimire, Bhandari, et al., 2022)	ASO	LSTM, RBF, DBN, M5TREE y MARS	Intra-horario, diario	Selección de características	RMSE 2.172-3.305 MJ/m ² /día y MAE 1.624-2.370 MJ/m ² /día.
(Vaisakh y Jayabharathi, 2022)	GU-DHOA	MLP, RNN	Intra-horario	Neuronas ocultas	RMSE 2.11×10^2 y MAE 166.53
(Irshad et al., 2023)	AOA	-	Intra-horario	No especificado	MSE 0.18, RMSE 0.48 y MAE 0.32
(Guermoui et al., 2024)	SWD	-	Horario	Descomposición de series temporales, selección de características	NRMSE 0.75 % - 3.4 %

la Optimización Aritmética de Aprendizaje Profundo Híbrido (AOHDL-SRP), utilizando una CNN con memoria a largo plazo orientada a la atención (ALSTM). La optimización de hiperparámetros se realizó utilizando el Algoritmo de Optimización Aritmética (AOA), destacando su enfoque en tres etapas: preprocesamiento, predicción y optimización de hiperparámetros, con un R² del 100 % y valores de MSE, RMSE y MAE de 0.18, 0.43 y 0.32 respectivamente.

El artículo de Guermoui et al. (2024) analiza la mejora en la predicción de la Irradiancia Normal Directa (DNI) para sistemas de energía solar concentrada, a través de un modelo híbrido que combina la Descomposición del Modo Variacional (VMD), el Algoritmo de Des-

composición de Enjambre (SWD) utilizado para la descomposición de series temporales, la selección de características con Bosques Aleatorios (RF) y Redes Neuronales Convolucionales Profundas (DCNN). Este modelo fue probado en cuatro regiones con diferentes condiciones climáticas, demostrando una precisión sobresaliente con bajos errores de predicción en términos de nRMSE que varían entre 0.75 % y 3.4 % para pronósticos con una hora de antelación.

1.1.3. Redes Neuronales Recurrentes

La exploración se ha centrado en cómo las RNN, combinadas con varias técnicas de optimización metaheurística, se han aplicado para mejorar la precisión de las predicciones de SR, que se presentan en la Tabla 1.3. Cada estudio revisado contribuye a comprender la efectividad de las RNN en este campo, proporcionando información sobre las configuraciones de modelos y estrategias de optimización más efectivas que han producido los mejores resultados predictivos.

Tabla 1.3. Los siguientes artículos emplean algunos modelos optimizados de redes neuronales recurrentes.

Autor	Algoritmo	Comparativa	Horizonte temporal	Tipo	Optimización	Resultados
(Khafaga et al., 2023)	ADSSOA	-	Horario	LSTM	Selección de características	RMSE 0.000388
(Hasan et al., 2023)	SBO	-	Horario	GRU	Afinación de hiperparámetros	MSE 10.231, RMSE 3.199, MAE 1.748, R ² 0.974, MAPE 11.856 %.
(Vaisakh y Jayabarathi, 2022)	GU-DHOA	MLP, CNN	Intra-horario	RNN estándar	Número de neuronas ocultas	RMSE 3.15×10^2 y MAE 223.32
(Jayalakshmi et al., 2021)	CSO	-	Horario, Diario	LSTM	Afinación de hiperparámetros	MSE 0.2392 y MAE 0.3562
(Sansine et al., 2022)	PSO	-	Diario	LSTM	Selección de características	MAE 99.37 W/m ² , RMSE 154.84 W/m ² y R ² 0.82.

El estudio realizado por Khafaga et al. (2023) presentó un enfoque para mejorar la eficiencia de la predicción de SR utilizando un modelo híbrido que integra un Algoritmo de Optimización de Búsqueda Dinámica Adaptativa de Ardilla (ADSSOA) con LSTM. Este modelo híbrido, llamado ADSSOA-LSTM, fue diseñado para optimizar la predicción de SR utilizando factores climáticos, destacándose por su capacidad para manejar dependencias temporales complejas en los datos de SR. Específicamente, el modelo ADSSOA-LSTM logró un RMSE de 0.000388 en comparación con el RMSE de 0.001221 para el modelo LSTM estándar.

El artículo de Hasan et al. (2023) presenta un método avanzado para la predicción de SR utilizando la Optimización de Satin Bowerbird (SBO) con Aprendizaje Profundo Modificado, específicamente un modelo híbrido de Unidad Recurrente Gated Bidireccional Modificada (BGRU) y Máquina de Aprendizaje Extremo Secuencial en Línea (OSELM), llamado

OMBGRU-SRP. Este modelo incorpora mecanismos de atención y conexiones de salto para optimizar la predicción de SR, logrando un MSE de 10.231, RMSE de 3.199, MAE de 1.748 y un coeficiente de determinación R^2 de 0.974, junto con un Error Porcentual Absoluto Medio (MAPE) de 11.856

Vaisakh y Jayabarathi (2022) desarrollaron un modelo híbrido para la predicción de la irradiación solar utilizando MLP, CNN y RNN, optimizando el número de neuronas ocultas a través de un algoritmo híbrido que combina DHOA y GWO, llamado GU-DHOA. El modelo GU-DHOA-RNN obtuvo una predicción de irradiación solar con un RMSE de 3.15×10^2 y MAE de 223.32.

El estudio de Jayalakshmi et al. (2021) desarrolló un modelo de predicción de irradiación solar utilizando una Red Neuronal de Memoria a Largo Plazo (LSTM) optimizada por un algoritmo híbrido de Optimización de Enjambre de Pollo (CSO) y GWO, llamado LSTM-CSO-GWO. Este modelo se centra en predecir la irradiación solar en diferentes escalas de tiempo: horaria, diaria y semanal, empleando un enfoque de aprendizaje multitarea para compartir recursos de manera efectiva entre cada tarea. En términos de métricas estadísticas, para la escala de tiempo horaria, el modelo logró un MSE de 0.2392 y un MAE de 0.3562.

Sansine et al. (2022) analiza los pronósticos de irradiación solar utilizando modelos híbridos que combinan PSO con tres modelos independientes: XGboost (PSO-XGboost), PSO-LSTM y Algoritmo de Regresión de Potenciación de Gradiente (PSO-GBRT), aplicados a una microred combinada de generación de frío y electricidad en Tahití. Los modelos híbridos se compararon con otros modelos independientes como ANN, CNN, Bosques Aleatorios (RF), LSTM, GBRT y XGboost. Este modelo logró un MAE de 99.37 W/m², un RMSE de 154.84 W/m² y un R^2 de 0.82.

1.1.4. Otros Modelos de ML

Más allá de los modelos convencionales de redes neuronales, se han aplicado con éxito varios otros enfoques de ML en la predicción de SR. Varios estudios han examinado la utilización de diversos modelos, como Máquinas de Soporte Vectorial (SVM), Bosques Aleatorios (RF), entre otros, todos optimizados con técnicas metaheurísticas para mejorar su rendimiento predictivo, como se resume en la Tabla 1.4.

Cada uno de estos modelos ha ofrecido ventajas únicas en términos de manejo de diferentes tipos de datos y complejidades del problema. Se ha explorado cómo los investigadores implementaron estos modelos alternativos, integrando algoritmos de optimización para ajustar finamente sus estructuras y parámetros, con el objetivo de obtener predicciones de SR más precisas y efectivas. Además, se han discutido las configuraciones de modelos y estrategias de optimización que han demostrado ser más efectivas en el contexto específico de la predicción de SR.

El estudio de X. Zhang y Wei (2019) presenta un modelo híbrido para la predicción diaria de SR, combinando ELM optimizado por el Algoritmo de Murciélago (BA), con el uso de la Transformada Wavelet (WT) y el Análisis de Componentes Principales (PCA). Evaluado en varias ciudades como Beijing, Nueva York, Melbourne y Sao Paulo, destacó particularmente en Beijing por su precisión. Este modelo incluye una red con una capa oculta y 10 neuronas

Tabla 1.4. Los siguientes artículos emplean algunos modelos optimizados de ML.

Autor	Algoritmo	Comparativa	Horizonte temporal	Modelo	Optimización	Resultados
(X. Zhang y Wei, 2019)	BA	LSSVM, BPNN	Horario	ELM	Pesos y sesgos	RMSE 0.028, MAE 0.14, MAPE 2.2 y R ² 0.99
(Reddy et al., 2022)	EHO	SVR, RFR, GBR, XGB	Intra-horario	ELM	Pesos y sesgos	RMSE 16.30 W/m ² y MSE 265.8546 W/m ²
(El-Kenawy et al., 2021)	ASCA	PSO, WOA, GA	Intra-horario	KNN	Pesos y sesgos	RMSE 0.00175482, MAE 0.00161235 y MBE -0.00036521
(C. Zhang et al., 2022)	WT, IASO	-	Intra-horario	ELM	Pesos y umbrales	RMSE 26.1022 W/m ² , MAE 13.9023 W/m ² , R 0.9939 y MAPE 0.2209
(Ikram et al., 2022)	IMVO	GA, PSO	-	LSSVR	Afinación de hiperparámetros	RMSE 0.748, MAE 0.574, R ² 0.969, y NSE 0.964
(Prasad et al., 2019)	ACO	-	Mensual	RF	Selección de características	RMSE 0.89 MJ/m ² , MAE 0.69 MJ/m ² y R 0.984
(Dong y Jiang, 2019)	FA	SVM	Diario	RS	Pesos asignados a pronósticos individuales	RMSE 13.11-18.79 W/m ² y MAE 0.044-0.086
(Peng et al., 2023)	SSA	LSTM	Intra-horario	ELM	Hiperparámetros	RMSE 23.153 W/m ² , MAE 13.225 W/m ² y R 0.9923
(Ghazvinian et al., 2019)	IPSO	M5T, GP, MARS	Diario, Mensual	SVR	Selección de características	RMSE 9.01 y MAE 4.10
(Wu et al., 2023)	ACO, CS, GWO	DT, BP	Diario	SVM	Afinación de parámetros	RMSE 1.341 MJ/m ² /día, R ² 0.959, y NSE 0.956
(Xue y Zhou, 2019)	PSO	ELM, BPNN, BPNN	Diario	LSSVM	Afinación de parámetros	RMSE 0.7524, MAE 0.7257 y R 0.9773
(Ghimire et al., 2019)	PSO	-	Intra-horario, Horario	SVM	Afinación de hiperparámetros	R 0.982-0.995
(Sun et al., 2018)	GSA	-	Diario	LSSVR	Pesos y sesgos	NRMSE 2.96 %, MAPE 2.83 % y DS 88.24 %
(Guermoui et al., 2021)	ABC	LSSVM, CS	Horario	LSSVR	Afinación de parámetros	RMSE 116.22 Wh/m ² y R 94.3 %
(Khodayar et al., 2020)	GA	-	Diario	ELM	No especificado	Mejora de 3.64 % a 4.49 %
(Mirbolouki et al., 2022)	GA, PSO, GWO, HHO	-	Mensual	ANFIS	Selección de parámetros	RMSE 1.489 MJ/m ² , R ² 0.959 y MAE 1.248 MJ/m ²
(Kumar et al., 2023b)	MFO	-	Intra-horario	XGBoost	Refinamiento de hiperparámetros del modelo	RMSE 76.29 W/m ² y R ² 0.9337
(Bakali et al., 2023)	SSA	-	Diario	VMD	Optimización de parámetros	NRMSE 3.87 % y NMAPE 1.58 %
(Pani y Nayak, 2019)	GSA	-	Intra-horario, Horario	ELM	Optimización de variables del kernel	MAPE 0.71 %
(J. Fan et al., 2020)	BAT, PSO, WOA	MARS, XGBoost	Diario	SVM	Parámetros de entrada	RMSD 11.1 %, 10.0 % y 10.4 %
(Jia et al., 2023)	PSO	-	Diario, Mensual	GEM	Optimización de parámetros	RMSE 1.045-1.719 MJ/m ² /día, rRMSE 7.6-12.7 %, MAE 0.801-1.283 MJ/m ² /día, R ² 0.953-0.981
(Kumar et al., 2023b)	MFO	MLP, SVR	Diario	RFR	Afinación de hiperparámetros	RMSE 75.8613 W/m ² y R ² 0.9402
(Islam et al., 2021)	CA	PSO	Diario	RF	Afinación de hiperparámetros	RMSE 0.0366 y MAPE 2.98 %

ocultas, utilizando una función de activación sigmoideal. La optimización mediante BA se

configuró con una población inicial de 10 y hasta 500 iteraciones para ajustar el peso de entrada y el sesgo de la capa oculta. WT ayuda a descomponer y limpiar la serie temporal de SR, mientras que PCA reduce la dimensionalidad de los datos de entrada. El modelo arrojó resultados de MAE 0.14, MAPE 2.2 %, RMSE 0.028 y un coeficiente R^2 de 0.99.

El artículo de Reddy et al. (2022) introduce un modelo de Optimización de Rebaño de Elefantes con Máquina de Aprendizaje Extremo Profundo (EHO-DELM) para la predicción de SR utilizando pronósticos meteorológicos. La estructura de la red neuronal incluye una capa de entrada con 4 neuronas, 6 capas ocultas con 10 neuronas cada una, y una capa de salida con una neurona, utilizando la función de activación sigmoide logística. El algoritmo de Optimización de Rebaño de Elefantes (EHO) se emplea para ajustar óptimamente los pesos y sesgos del modelo DELM, mejorando la precisión predictiva, con un MSE de 265.8546 W/m^2 y un RMSE de 16.30 W/m^2 . La comparación con otros modelos destaca la efectividad de EHO-DELM, superando técnicas como SVR, RFR, GBR y XGB en la predicción de SR.

El estudio de El-Kenawy et al. (2021) propone un modelo de predicción de SR basado en el Algoritmo Seno Coseno (ASCA). Este modelo de conjunto optimizado consiste en fases de preprocesamiento y entrenamiento, utilizando la regresión de K-Nearest Neighbors (KNN) para el entrenamiento del modelo de conjunto. Se evalúa en comparación con varios algoritmos, incluyendo PSO, Whale Optimization Algorithm (WOA), GA, con un RMSE de 0.00175482, MAE de 0.00161235 y MBE de -0.00036521.

C. Zhang et al. (2022) desarrolló un modelo de predicción de SR utilizando WT, Descomposición Empírica Completa con Ruido Adaptativo (CEEMDAN), Optimización de Búsqueda de Átomos Mejorada (IASO) y Máquina de Aprendizaje Extremo Robusta ante Outliers (ORELM) con una función de activación sigmoide tangente hiperbólica. El modelo usa WT para la eliminación de ruido y CEEMDAN para descomponer los datos en modos intrínsecos. Luego, IASO optimiza los pesos y umbrales de ORELM, con una población de 50 y un máximo de 30 iteraciones. Este enfoque híbrido logró un RMSE de 26.1022 W/m^2 , MAE de 13.9023 W/m^2 , un R de 0.9939 y un MAPE de 0.2209.

El estudio presentado en el artículo de Ikram et al. (2022) se centra en la predicción de SR utilizando una versión mejorada del Algoritmo de Optimización de Multiverso (IMVO), integrado con la Máquina de Soporte Vectorial de Mínimos Cuadrados (LSSVM). Esta combinación se compara con otras configuraciones utilizando diferentes algoritmos de optimización con LSSVM, incluyendo GA, PSO y otros. La investigación se realizó utilizando datos de dos estaciones ubicadas en la región sureste de China. Las principales características de la red neuronal en este estudio incluyen el uso de LSSVM como modelo principal para la predicción, con IMVO utilizado para afinar los hiperparámetros del modelo LSSVM. En una de las estaciones y un escenario de división de datos de 75 % para entrenamiento y 25 % para prueba, el modelo LSSVM-IMVO logró un RMSE de 0.748, MAE de 0.574, R^2 de 0.969 y NSE de 0.964, destacándose como el enfoque más efectivo entre los analizados.

El artículo de Prasad et al. (2019) analiza la previsión de SR utilizando un modelo híbrido, integrando Descomposición Empírica Multivariante (MEMD), Optimización de Colonias de Hormigas (ACO) y Bosques Aleatorios (RF). Se seleccionaron características significativas utilizando ACO, empleando 1000 árboles y 10 predictores para la división en el modelo RF,

logrando un RMSE de 0.89 MJ/m², MAE de 0.69 MJ/m² y un R de aproximadamente 0.984 en el sitio de Springfield.

Para el artículo Dong y Jiang (2019), el enfoque principal es predecir con precisión el GSR horario utilizando conjuntos de datos de ocho ubicaciones en la provincia de Xinjiang, China. El método se basa en el aprendizaje conjunto con un método de Subespacio Aleatorio (RS) para dividir los datos originales en varios subespacios de covariables. Se introduce un nuevo método de selección de covariables llamado Desviación Absoluta Suavemente Recortada (SRSCAD), aplicado a cada subespacio para la extracción de covariables relevantes. Para combinar las predicciones obtenidas a través de RS y SRSCAD, se utiliza un Algoritmo de Luciérnagas (FA) para estimar los pesos asignados a las predicciones individuales, con un RMSE que varía entre 13.11 y 18.79 W/m² para la predicción con 24 horas de antelación y entre 17.79 y 21.12 W/m² para la predicción con 48 horas de antelación, y un MAE que varía de 0.044 a 0.086, superando a otros métodos comparados como las redes neuronales de Elman, SVM y los modelos empíricos de Angstrom-Prescott.

El estudio de Peng et al. (2023) propone un modelo híbrido para la predicción de SR, llamado OVMD-PACF-ISSA-DBN-OSELM, que incorpora Descomposición Modal Variacional Óptima (OVMD), Función de Autocorrelación Parcial (PACF) y un modelo híbrido de Red Neuronal de Creencias Profundas (DBN) con Máquina de Aprendizaje Extremo Secuencial en Línea (OSELM), optimizado por un Algoritmo de Búsqueda de Gorriones Mejorados (ISSA). El modelo logra un RMSE de 23.153 W/m², MAE de 13.225 W/m² y un R de 0.9923.

Ghazvinian et al. (2019) desarrollaron un modelo para predecir SR utilizando Regresión de Soporte Vectorial (SVR) con función de núcleo radial y un algoritmo de Optimización por Enjambre de Partículas Mejorada (IPSO) con una población de 40 partículas. Este estudio se centró en mejorar la capacidad de búsqueda global del algoritmo PSO e integrar SVR con IPSO para obtener valores óptimos de los parámetros de SVR. Se consideraron dos estaciones de SR en Turquía para evaluar el modelo, junto con varios modelos comparativos, incluyendo el modelo de árbol M5 (M5T), Programación Genética (GP), SVR integrado con otros algoritmos de optimización y Splines de Regresión Adaptativa Multivariante (MARS), con un RMSE de 9.01 y MAE de 4.10 para la estación de Antakya.

El estudio de Wu et al. (2023) optimizó modelos SVM para la predicción de GSR en diferentes zonas climáticas de China, utilizando los algoritmos ACO, CS y GWO para ajustar los parámetros internos de SVM. Entre los modelos independientes, SVM demostró la mayor precisión en todas las zonas, seguido del modelo de Árbol de Decisión (DT) y BP. Los modelos híbridos, especialmente GWO-SVM y CS-SVM, superaron al modelo SVM independiente en todas las zonas climáticas, donde GWO-SVM emergió como el modelo más efectivo, logrando un R² de hasta 0.959, RMSE de 1.341 MJ/m²/día y NSE de hasta 0.956 en la predicción diaria de GSR.

En el estudio de Xue y Zhou (2019), se desarrollaron cuatro tecnologías de IA para estimar GSR en Zhengzhou, China. Estas tecnologías incluyen ELM, un modelo híbrido que combina la Máquina de Soporte Vectorial de Mínimos Cuadrados con PSO (PSO-LSSVM), donde PSO se usa para optimizar los parámetros de LSSVM con función de núcleo radial, Red Neuronal de Retropropagación (BPNN) y Red Neuronal de Regresión Generalizada

(GRNN). Se seleccionaron seis variables meteorológicas como índices de evaluación, con GSR diario como salida. Para el modelo PSO-LSSVM, el R para los conjuntos de datos de entrenamiento, prueba y todas las muestras es 0.9812, 0.9773 y 0.9793, respectivamente, RMSE de 0.6324, 0.7524 y 0.7425, respectivamente, y para MAE son 0.5128, 0.7257 y 0.6228. Estos resultados demuestran que el modelo PSO-LSSVM tiene un rendimiento superior y se clasifica en orden de rendimiento como PSO-LSSVM $\hat{}$ ELM $\hat{}$ BPNN $\hat{}$ GRNN. Por lo tanto, el modelo híbrido PSO-LSSVM se puede utilizar para predecir GSR diario con alta precisión en Zhengzhou, China.

El estudio de Ghimire et al. (2019) empleó un enfoque que combina PSO para la selección de características, con un máximo de 500 iteraciones, un rango de población de 100 a 1000, y la Transformada Wavelet Discreta de Superposición Máxima (moDWT) para la descomposición de datos en tres ciudades solares australianas: Alice Springs, Coburg y Perth. Este método permitió ajustar los hiperparámetros del modelo SVR con Grid Search, logrando una notable correlación entre las predicciones y las observaciones para R de 0.982 a 0.995.

En este estudio, Sun et al. (2018) desarrollaron un enfoque de aprendizaje para la predicción de SR utilizando un método de Descomposición-Clustering-Ensemble (DCE) en Beijing, empleando la Descomposición Empírica en Modo Ensemble (EEMD) para descomponer los datos originales de SR. Para la predicción individual y el ensamblaje de componentes, se utilizó la Regresión de Soporte Vectorial de Mínimos Cuadrados (LSSVR), optimizada por el Algoritmo de Búsqueda Gravitacional (GSA). El clustering se realizó con el método Kmeans, aplicando un segundo método LSSVR-GSA para el ensamblaje final, basado en pesos de ensamblaje de clúster, logrando un Error Cuadrático Medio Normalizado (NRMSE) de 2.96%, MAPE de 2.83% y DS de 88.24% para la predicción con un día de antelación.

En el estudio realizado por Guermoui et al. (2021), se propuso un nuevo modelo híbrido llamado ABC-LS-SVM para la predicción de la radiación solar global (GHI) con varias horas de antelación, utilizando el Algoritmo de Colonia de Abejas Artificiales (ABC) para optimizar los parámetros libres del modelo LS-SVM. Este enfoque se validó con datos medidos durante cinco años, de 2013 a 2017, en la Unidad de Investigación Aplicada a las Energías Renovables (URAER) en Ghardaïa, en el sur de Argelia. Los resultados de la predicción de GHI con 12 horas de antelación utilizando el modelo ABC-LS-SVM llevaron a un RMSE de 116.22 Wh/m² y un R de 94.3%. En comparación, el modelo clásico de LS-SVM obtuvo un RMSE de 117.73 Wh/m² y un R de 92.42%, mientras que la combinación de CS con LS-SVM resultó en un RMSE de 116.89 Wh/m² y un R de 93.78%.

El estudio Khodayar et al. (2020) presenta un modelo generativo profundo, el Codificador Automático de Grafos Convolucionales (CGAE), para predecir la irradiancia solar probabilística espacio-temporal utilizando una red neuronal profunda. Captura características espacio-temporales de observaciones históricas de SR para generar muestras que estiman la irradiancia solar futura. El modelo CGAE demostró un rendimiento superior en términos de fiabilidad, nitidez y Puntuación de Rango Continuo (CRPS) en comparación con métodos de referencia tanto temporales como espacio-temporales. El modelo CGAE logró una mejora del 3.64% al 4.49% en fiabilidad sobre el mejor modelo temporal (ELM) para predicciones con una hora y seis horas de antelación, respectivamente.

El artículo Mirbolouki et al. (2022) desarrolla un enfoque para mejorar las predicciones de SR utilizando modelos climáticos globales y métodos avanzados de ML, específicamente ANFIS combinado con algoritmos metaheurísticos como GA, PSO, GWO y HHO. Se utilizan poblaciones iniciales de 500 y un máximo de 1000 iteraciones para ajustar eficientemente sus parámetros internos. Utilizando datos de temperatura como la única variable de entrada, se compararon varias configuraciones de división de datos para entrenamiento y prueba, encontrando que el ANFIS mejorado con algoritmos de optimización mostró una precisión superior en la predicción de SR. En particular, ANFIS combinado con optimización GWO (ANFIS-GWO) resultó ser el más efectivo, con un R^2 de 0.959 y los errores más bajos de RMSE y MAE de 1.489 MJ/m^2 y 1.248 MJ/m^2 , respectivamente, para la estación de Adana con una división de datos de entrenamiento-prueba del 80 %-20 %.

El artículo Kumar et al. (2023b) analiza un modelo de predicción de irradiancia solar utilizando la técnica de Boosting de Gradiente Extremo (XGB) con 100 a 1000 árboles y una profundidad máxima de 5, optimizado mediante algoritmos metaheurísticos como Optimización de la Llama de la Polilla (MFO) y GWO. Se experimentaron cinco tamaños de población (50, 100, 150, 200 y 500) y un máximo de 100 iteraciones para refinar los hiperparámetros del modelo XGB, con el objetivo de maximizar la precisión de la predicción de la irradiancia solar en una ubicación específica. Los resultados destacados incluyen un R^2 de 0.9337 para GHI, utilizando el modelo XGB-GWO, con un RMSE de 76.29 W/m^2 .

El artículo de Bakali et al. (2023) propone un modelo para predecir la irradiancia solar en un horizonte de 24 horas. Para esta predicción, se utilizan técnicas de Descomposición del Modo Variacional (VMD) y el algoritmo STACK con 200 neuronas en la primera capa oculta y 500 en la segunda, junto con la optimización de los parámetros de VMD utilizando el Algoritmo de Enjambre de Gorriones (SSA). Se enfatiza que la partición de datos estacionales y la selección adecuada de parámetros de VMD a través de SSA son cruciales para mejorar la precisión de la predicción. Los resultados muestran un Error Cuadrático Medio Normalizado (NRMSE) de 3.87 % y un Error Porcentual Absoluto Medio Normalizado (NMAPE) de 1.58 % para días nublados.

En el artículo de Pani y Nayak (2019), se propone un modelo de predicción basado en la Máquina de Aprendizaje Extremo con Núcleo (KELM) optimizada utilizando el Algoritmo de Búsqueda Gravitacional Caótica (CGSA) para predecir la irradiancia solar, crucial para la gestión de la energía fotovoltaica. El modelo optimizado, llamado OKELM, reduce significativamente el proceso de selección de parámetros al optimizar las variables del núcleo, logrando una predicción más precisa y eficiente. Se valida utilizando datos de una planta fotovoltaica de 1 MW en India, con resultados sobresalientes en la predicción de la irradiancia solar bajo diversas condiciones meteorológicas. El modelo KELM, antes de la optimización, no requiere definir el número de nodos ocultos, reduciendo la variabilidad de los resultados. La implementación de CGSA mejora la generalización y estabilidad del modelo, demostrado por métricas como MAPE, que en el caso optimizado alcanza un valor de 0.71 % para la temporada de verano.

El artículo J. Fan et al. (2020) analiza la predicción de la radiación solar difusa diaria (DSR) en regiones contaminadas utilizando máquinas de soporte vectorial híbridas con

algoritmos heurísticos: PSO, BA y WOA. Estos modelos se comparan con modelos SVM estándar, Splines de Regresión Adaptativa Multivariante (MARS) y Boosting de Gradiente Extremo (XGBoost). Los resultados indican que la incorporación de variables de contaminación del aire (PM2.5, PM10, O3) mejora significativamente la precisión del modelo en la predicción diaria de DSR, con disminuciones relativas promedio en la Desviación Cuadrática Media (RMSD) de 11.1 %, 10.0 % y 10.4 %. Además, el modelo SVM-BAT superó a los modelos SVM-PSO y SVM-WOA, así como a los modelos SVM estándar, MARS y XGBoost, demostrando ser el más efectivo para predecir la DSR diaria, especialmente en regiones con alta contaminación atmosférica.

El estudio de Jia et al. (2023) se centra en mejorar las predicciones diarias y mensuales de GSR en el noreste de China mediante la optimización de algoritmos PSO integrados con un modelo exponencial gaussiano (GEM). El modelo PSO-GEM se compara con otros cuatro modelos de ML y dos modelos empíricos utilizando datos meteorológicos diarios de 1997 a 2016 de cuatro estaciones en el noreste de China. Logró un RMSE entre 1.045 y 1.719 MJ/m²/día, RRMSE del 7.6 al 12.7 %, MAE de 0.801 a 1.283 MJ/m²/día y R² del 0.953 al 0.981 para estimaciones diarias. Para estimaciones mensuales, PSO-GEM también mostró la mayor precisión con un RMSE entre 0.197 y 0.575 MJ/m²/día, RRMSE del 1.5 al 7.0 %, MAE de 0.137 a 0.499 MJ/m²/día y R² del 0.999 al 1.

Kumar et al. (2023b) desarrollaron un modelo avanzado para la predicción de la irradiancia solar utilizando modelos de ML como SVR, MLP y RFR, optimizados con MFO, GWO y Optimización del Mejor de la Clase Evolutiva (ECTO). Para SVR, los parámetros óptimos incluyen una función de núcleo RBF y un parámetro de regularización C optimizado a 0.01. Para MLP, se identificaron 100 capas ocultas como óptimas con función de activación ReLu. RFR mostró un rendimiento superior con 100 árboles y una profundidad máxima de 5. Se observó que el modelo RFR-ECTO superó a los demás, ofreciendo los valores de R² más altos de 0.9402 con el valor de RMSE más bajo de 75.8613 W/m².

Islam et al. (2021) desarrollaron una técnica de predicción para SR y velocidad del viento utilizando un enfoque basado en bosques aleatorios de aprendizaje profundo, optimizado por un nuevo algoritmo Coot (CA). Este modelo superó al algoritmo PSO en términos de precisión. El estudio logró reducir el RMSE a 0.0366 para la irradiancia solar y 0.0602 para la velocidad del viento, mientras que MAPE se situó en 2.98 % y 4.78 % respectivamente.

1.2. Descripción del problema

La fluctuación de la radiación solar global afecta la estabilidad de los sistemas de energía fotovoltaicos y pone en peligro la seguridad de la red eléctrica (Zhu et al., 2021). Para monitorear esta variable se requiere invertir en estaciones meteorológicas en diferentes puntos. Sin embargo, hay una limitada cantidad de estaciones meteorológicas, las cuales monitorean la radiación solar global con instrumentos que son costosos, al igual que su mantenimiento (Q. Zhang et al., 2022). Bajo este contexto se requieren de algoritmos de predicción que sean precisos, pero que además el valor sea calculado lo más rápido posible ya que anteriormente los estudios se enfocaban en calcular los valores promedio diarios, mensuales y anuales (Li

et al., 2019). Los algoritmos deben ser capaces de calcular mediante datos históricos y modelos matemáticos, la variable de radiación solar global, esto permitirá a los operadores de la red eléctrica tomar decisiones informadas sobre la producción, generación y distribución de energía con mayor eficiencia (Hassan et al., 2022; Álvarez-Alvarado et al., 2021).

De acuerdo con el estado del arte, las técnicas convencionales presentan convergencia prematura dado que hacen una mala selección de los parámetros de control, principalmente en los métodos que usan algoritmos de optimización, por lo que cada vez se demanda en nuevas técnicas más exactas en la selección de parámetros de control (Langazane y Saha, 2022). Otro de los problemas que reportaron en la literatura fue el sobreajuste de los datos, lo que significa que los entrenamientos que realizaron en sus modelos de predicción fueron muy cercanos o exactamente el conjunto de datos que se estudiaba, lo que provoca que no se pueda ajustar a otros conjuntos de datos o pierda la capacidad de predicción ya que no se ajusta adecuadamente lo que puede volver demasiado específico el modelo de predicción y por lo tanto no generalizar bien ante nuevas entradas de datos (Malik et al., 2022; Zhou et al., 2021a). Lo anterior puede caer en la reducción de la confiabilidad del modelo (Mann et al., 2022).

1.3. Justificación

El pronóstico de la radiación solar es una solución rápida, económica y eficaz en la optimización de diferentes operaciones de microrredes, como el recorte de picos, la reducción del impacto de la incertidumbre, problema de despacho económico y en el sistema eléctrico (Rajagukguk et al., 2020). La creación de un modelo de predicción eficiente de la radiación solar global es importante para mejorar la obtención de energía eléctrica vía energía solar y observando los modelos existentes en el estado del arte, podemos visualizar que tienen problemas muy importantes como lo son la convergencia prematura y sobreajuste de los datos, en cuanto a los modelos que usan algoritmos de optimización y técnicas de probabilidad. Los modelos vistos que usan redes neuronales presentan buen rendimiento, sin embargo, es necesario el seguir contando con modelos que presenten aún un error más pequeño debido a la creciente infraestructura para la creación de base de datos y desarrollo de tecnologías que aprovechen aún mejor el recurso solar disponible, por lo que ahora es necesario optimizarlos, ya que se necesita obtener el valor de radiación solar global más confiable para poder administrar de mejor manera la energía solar obtenida.

Fundamentación teórica

2.1. Redes neuronales

Las redes neuronales son estructuras paralelas fuertemente interconectadas con neuronas (nodos) organizadas jerárquicamente, diseñadas para interactuar con objetos del mundo real de manera similar a como lo hacen los sistemas nerviosos biológicos (Kohonen, 2012).

Las redes neuronales son sistemas paralelos altamente interconectados con nodos (neuronas) organizados de manera jerárquica, concebidos para interactuar con el entorno de forma similar a los sistemas nerviosos biológicos (Kohonen, 2012). En ciencias de la computación, las redes neuronales se definen como modelos matemáticos con numerosos parámetros. Estos modelos comprenden varias funciones, como se ilustra en 2.1, que se anidan múltiples veces. Los algoritmos de aprendizaje para redes neuronales suelen estar respaldados por rigurosas pruebas matemáticas. Un ejemplo de red neuronal es el perceptrón, un clasificador binario que consiste en dos capas de neuronas. La primera capa recibe las señales externas y las envía a la segunda capa, que actúa como una unidad lógica de umbral, también conocida como neurona M-P (Zhou et al., 2021b).

$$y_j = f\left(\sum_i W_i X_i - \theta\right) \quad (2.1)$$

2.2. Técnicas de preprocesamiento de datos

Las bases de datos son altamente susceptibles a contener datos ruidosos, incompletos o inconsistentes debido al gran volumen y la variedad de fuentes de datos que se manejan actualmente. Existen varias técnicas de preprocesamiento de datos, tales como: la limpieza de datos, que se aplica para eliminar el ruido y corregir las inconsistencias; la integración de datos, que combina datos de múltiples fuentes en un lugar coherente, como un almacén de datos; la reducción de datos, que disminuye el volumen de datos mediante la agregación, eliminación de características redundantes o agrupamiento; y la transformación de datos, que incluye la normalización, donde los datos se escalan a un rango menor, como de 0.0 a

1.0 (C. Fan et al., 2021). Las técnicas de preprocesamiento de datos se muestran en la Fig. 2.1.

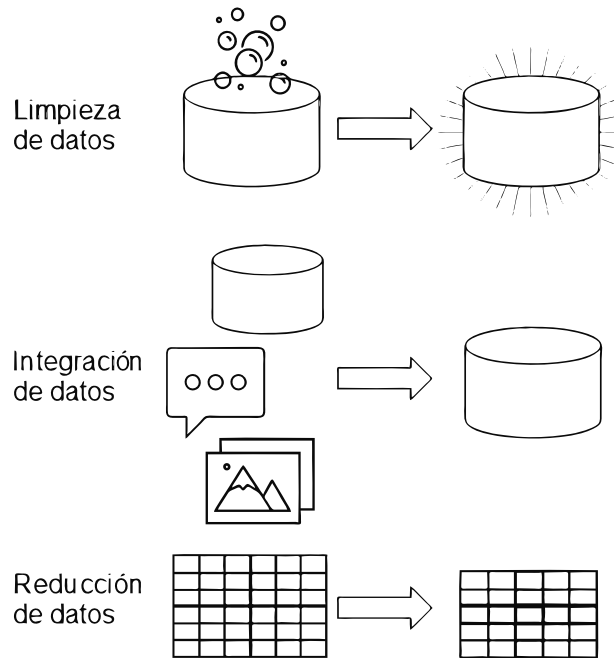


Figura 2.1. Técnicas de preprocesamiento de datos.

2.2.1. Limpieza de datos

El primer paso en el proceso de limpieza de datos es la detección de discrepancias, que pueden surgir debido a diversos factores como formularios mal diseñados, errores humanos o intencionales, y datos desactualizados.

2.2.2. Integración de datos

La combinación de múltiples fuentes de datos requiere una integración cuidadosa para reducir y evitar redundancias e inconsistencias en el conjunto de datos resultante. Esto mejora la precisión y la velocidad de los procesos de inteligencia artificial. La heterogeneidad en la semántica y la estructura de los datos presenta grandes desafíos en la integración de datos.

2.2.3. Reducción de datos

Las técnicas de reducción de datos se aplican para obtener una representación de menor volumen del conjunto de datos, preservando la integridad de los datos originales. Algunas

de estas técnicas incluyen la reducción de dimensiones, la reducción de numerosidad y la compresión de datos.

2.2.4. Transformación de los datos

Los datos se transforman para ser más eficientes y facilitar la identificación de patrones comprensibles. Algunas estrategias para la transformación de datos incluyen el suavizado, la construcción de atributos, la agregación, la normalización, la discretización y la creación de jerarquías de conceptos para datos nominales (C. Han et al., 2011).

2.3. Pronóstico de series temporales

Los métodos de series temporales, fundamentales en el análisis estadístico, son una de las principales técnicas utilizadas para estimar la radiación solar (SR). Los modelos de series temporales han demostrado una alta eficacia en la estimación de SR, aprovechando datos históricos para realizar predicciones (Gürel et al., 2023).

Una serie temporal se compone de observaciones registradas en momentos específicos. Las series temporales se caracterizan por componentes clave como la tendencia, que muestra la dirección general (ascendente o descendente), los ciclos, que reflejan patrones repetitivos sin un período fijo, y las variaciones estacionales, que indican fluctuaciones periódicas (Brockwell y Davis, 2002).

Una serie temporal es estacionaria cuando sus propiedades estadísticas, como la media y la desviación estándar, se mantienen constantes a lo largo del tiempo. La invariabilidad temporal es crucial, ya que implica una alta probabilidad de que la serie temporal repita su comportamiento en el futuro, lo que facilita significativamente su predicción (Ensaifi et al., 2022).

El pronóstico de series temporales abarca una amplia gama de metodologías. Estas incluyen desde la selección automatizada de modelos paramétricos hasta técnicas de aprendizaje automático (ML) convencionales, como la regresión de kernel y la regresión de vectores de soporte. Además, los procesos gaussianos son ampliamente utilizados en la predicción de series temporales. Los avances recientes incluyen el desarrollo de procesos gaussianos profundos y la integración de la paralelización en ML mediante procesos neuronales (Lim y Zohren, 2021).

Como se ilustra en la Fig. 2.2, la fiabilidad eléctrica del sistema depende de su capacidad para responder eficazmente a las fluctuaciones previstas e imprevistas en la producción y el consumo, así como a las interrupciones, manteniendo la calidad y la continuidad del servicio para los clientes. Por lo tanto, el proveedor de energía debe gestionar el sistema a través de diversos horizontes temporales (Voyant et al., 2017).

Además de la optimización metaheurística, las técnicas de IA se utilizan a menudo debido a la eficacia de los modelos de vanguardia, como las redes neuronales recurrentes (RNN), incluidas las redes de memoria a largo corto plazo (LSTM), que sobresalen en la captura de dependencias temporales intrincadas (Z. Han et al., 2021; Lara-Benítez et al., 2021; Lim

y Zohren, 2021; Hong y Martinez, 2024). Estas técnicas son valiosas para determinar los hiperparámetros adecuados y seleccionar los marcos más apropiados (Torres et al., 2021; Kolambe, 2024), en particular los Transformers de series temporales, para tareas como la previsión, la detección de anomalías y la clasificación (Zeng et al., 2023). Su efectividad se debe en gran medida a su capacidad para capturar dependencias e interacciones de largo alcance.

Predicción de Radiación Solar

La RS es altamente intermitente y caótica; incluso pequeñas fluctuaciones en la radiación solar pueden impactar significativamente la seguridad del suministro energético. Esto requiere modelos predictivos cada vez más precisos, especialmente dado el potencial aumento en el uso de energía solar suministrada por paneles fotovoltaicos (Rajagukguk et al., 2020). Por lo tanto, la creación de un modelo predictivo para la radiación solar global es crucial para garantizar prácticas óptimas de despacho y gestión de energía (Ghimire, Deo, Casillas-Pérez, y Salcedo-Sanz, 2022). Según el estado del arte, se han desarrollado varios modelos utilizando técnicas de aprendizaje automático como SVR (Dahmani et al., 2024; Yu, 2023; Ghimire et al., 2019) y RF (Chaibi, Benghoulam, Tarik, Berrada, y Hmaidi, 2022; Prasad et al., 2019), así como técnicas de aprendizaje profundo como RNN (Bamisile et al., 2022; Vaisakh y Jayabarathi, 2022), MLP (Jadidi et al., 2018; Ghimire, Deo, Casillas-Pérez, y Salcedo-Sanz, 2022; Mfetoum et al., 2024), que han ganado una relevancia significativa en los últimos años (Kumari y Toshniwal, 2021).

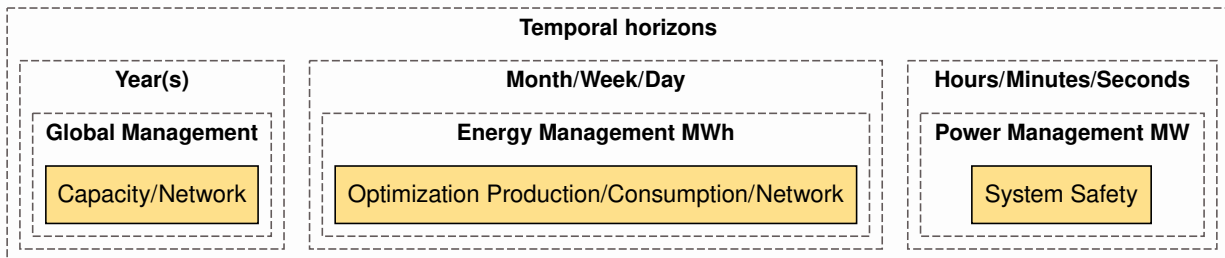


Figura 2.2. Horizontes de predicción para la gestión de energía solar en una red eléctrica.

2.4. Optimización

La optimización es obtener los valores óptimos para un conjunto de variables sobre cuál sea el objetivo de la función y que puede ser restringida o no restringida. Cualquier conjunto de valores asignados a las variables siempre produce una salida, pero el conjunto óptimo produce la salida óptima. Los objetivos del problema y las limitaciones pueden formularse en términos de funciones o ecuaciones matemáticas. La expresión matemática que representa el objetivo de la optimización se denomina función objetivo.

La función objetivo podría definirse para la maximización o minimización dependiendo del dominio del problema.

2.4.1. Algoritmos de optimización bioinspirados

Los algoritmos bioinspirados son algoritmos metaheurísticos que son desarrollados con los principios de la evolución biológica, el comportamiento de enjambre, procesos físicos y químicos. Son una técnica de inteligencia computacional bioinspirada desde que incorporan inteligencia en los algoritmos. Algunos de los campos donde se usan los algoritmos bioinspirados es en la ingeniería, especialmente ciencias de la computación, economía, diseño mecánico (Vasuki, 2020).

Colonia artificial de abejas (ABC)

El algoritmo ABC (Karaboga et al., 2005) se propuso para resolver problemas de optimización numérica unimodal y multimodal, pero con los años se ha convertido en una técnica prometedora para resolver problemas de optimización del mundo real.

El principio fundamental del algoritmo ABC es la comunicación, la cual es el intercambio de información entre las abejas. Es el paso más importante para la recolección de conocimiento acerca de las fuentes de alimento. Cuando la información ha sido compartida a otras abejas, la abeja observadora decide explotar las fuentes de alimento, dependiendo en el valor de probabilidad de las fuentes de alimento. El primer paso para encontrar fuentes de alimento en el algoritmo ABC es la inicialización de la población. Inicialmente, las primeras fuentes de alimento son encontradas aleatoriamente por una abeja exploradora usando la ecuación 2.2:

$$x_i^j = x_{\min}^j + \text{rand}(0, 1) (x_{\max}^j - x_{\min}^j) \quad (2.2)$$

en donde x_i^j es la fuente de alimento actual, x_{\min}^j y x_{\max}^j son el límite inferior y el límite superior de las fuentes de alimento, $\text{rand}(0,1)$ es un valor aleatorio entre 0 y 1 (Nadira y Rustam, 2018). En la Fig. 2.3, se puede observar el diagrama de flujo del algoritmo ABC.

2.4.2. Optimización de hiperparámetros

La optimización de hiperparámetros es un proceso crítico en ML y DL que implica seleccionar el mejor conjunto de hiperparámetros para un modelo dado (Bischl et al., 2023). Como se muestra en la Fig.??, se ilustra el proceso común de optimización de hiperparámetros utilizando algoritmos metaheurísticos. Los hiperparámetros son los parámetros ajustables que controlan el proceso de entrenamiento y la estructura del modelo, como la tasa de aprendizaje, el número de capas ocultas y el tamaño del lote para las redes neuronales artificiales, y otros hiperparámetros para modelos específicos de ML como C para SVR y el número máximo de niveles en cada árbol de decisión para RF (Yang y Shami, 2020).

A diferencia de los parámetros del modelo, que se aprenden durante el entrenamiento, los hiperparámetros deben establecerse antes de que comience el proceso de aprendizaje

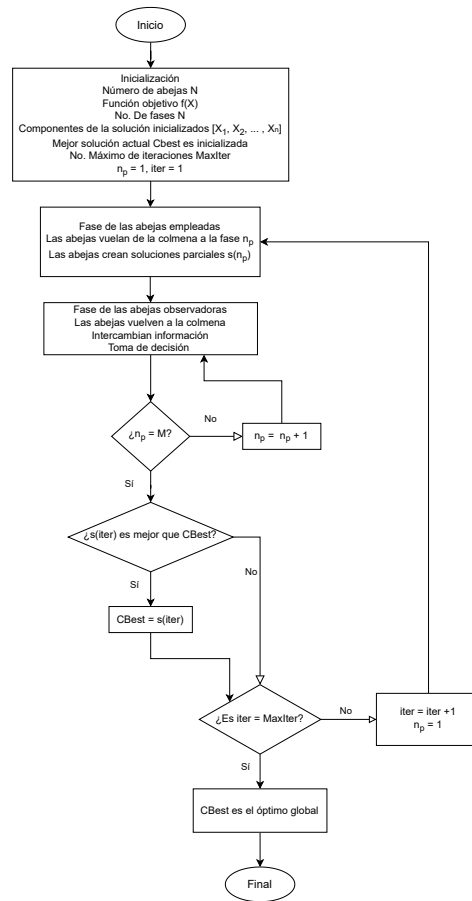
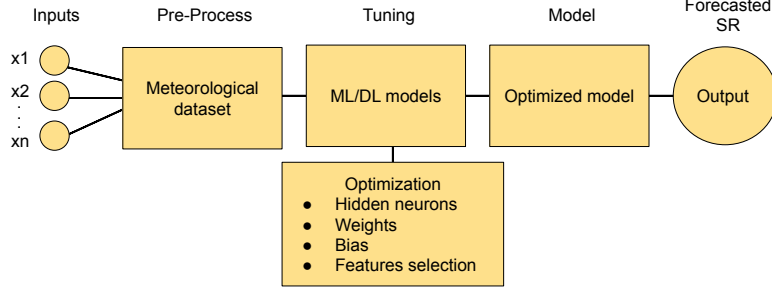


Figura 2.3. Diagrama de flujo del algoritmo de la colonia artificial de abejas (Vasuki, 2020).

(Elgeldawi et al., 2021). Una optimización efectiva de los hiperparámetros mejora el rendimiento del modelo al encontrar la configuración óptima que minimiza el error y mejora la precisión predictiva. Este proceso a menudo emplea varias técnicas, incluidas la búsqueda en cuadrícula, la búsqueda aleatoria y métodos avanzados como la optimización bayesiana y los algoritmos metaheurísticos, para navegar eficientemente el vasto espacio de hiperparámetros e identificar las configuraciones más efectivas (Abbas et al., 2023).

2.5. Métricas de rendimiento del modelo

Al crear un modelo de inteligencia artificial, se necesita evaluar la calidad del modelo, para evaluar el rendimiento de la etapa de entrenamiento con el cuál fue construida y qué



Metodologías comunes para la afinación de hiperparámetros de modelos ML/DL utilizando algoritmos metaheurísticos, donde x_1, x_2 a x_n son variables de entrada.

tan preciso es para predecir nuevas observaciones de pruebas.

Para esto, se necesitan diferentes métricas de evaluación del modelo, algunas de las cuales son las siguientes:

- R^2 es una medida que representa la proporción de variación en el resultado que puede atribuirse a las variables predictoras. En los modelos de regresión múltiple, R^2 indica la correlación cuadrada entre los valores de salida observados y los valores predictores utilizados en el modelo. Un valor más alto de R^2 sugiere un mejor rendimiento del modelo.
- El Error Cuadrático Medio (RMSE) evalúa el error típico del modelo en la predicción de los resultados de las observaciones. Se deriva matemáticamente como la raíz cuadrada del Error Cuadrático Medio (MSE), representando la varianza entre los valores observados reales y los valores predichos por el modelo. Un RMSE más bajo indica una mayor precisión del modelo.
- El Error Absoluto Medio (MAE) cuantifica la precisión de la predicción, similar al RMSE. Calcula la disparidad absoluta promedio entre los valores observados y predichos. Su versión normalizada (nMAE o NMAE) se utiliza comúnmente para facilitar la comparación entre diferentes escalas (Huang et al., 2022). El MAE muestra una menor sensibilidad a los valores atípicos en comparación con el RMSE.
- El Error Porcentual Absoluto Medio (MAPE) se obtiene mediante la ecuación 2.3, (Djeldjli et al., 2023).

$$\text{MAPE} = \frac{100}{N} \sum_{i=1}^N \left| \frac{G_{\text{Sim},i} - G_{\text{Act},i}}{X_{\text{Act},i}} \right| \quad (2.3)$$

- El Error Porcentual Absoluto Medio Normalizado (NMAPE) se calcula con la ecuación 2.4 (Bakali et al., 2023):

$$\text{NMAPE} = \frac{\frac{1}{N} \sum_{i=1}^N \left| \hat{X}_{\text{season},i} - X_{\text{season},i} \right|}{\frac{1}{N} \sum_{i=1}^N X_{\text{season},i}} \quad (2.4)$$

- El Error Cuadrático Medio Relativo (rRMSE) se obtiene mediante la ecuación 2.5 (Djeldjli et al., 2023):

$$rRMSE = \frac{1}{\overline{G}_{Act}} \sqrt{\frac{\sum_{i=1}^N (G_{Sim,i} - G_{Act,i})^2}{N}} \quad (2.5)$$

- El Error Medio de Sesgo (MBE) (El-Kenawy et al., 2021)
- La Eficiencia de Nash-Sutcliffe (NSE) se basa en la ecuación 2.6 (Ikram et al., 2022).

$$NSE = 1 - \frac{\sum_{i=1}^N [(SR_0)_i - (SR_c)_i]^2}{\sum_{i=1}^N [(SR_0)_i - \overline{SR_0}]^2}, \quad (2.6)$$

$$-\infty < NSE \leq 1 \quad (2.7)$$

- El Error Cuadrático Medio Normalizado (NRMSE) se obtiene mediante la ecuación 2.8

$$NRMSE = \frac{100}{\overline{x}} \sqrt{\frac{1}{T} \sum_{t=1}^T (x_t - \hat{x}_t)^2}, \quad (2.8)$$

- La Desviación Cuadrática Media (RMSD) se calcula mediante la ecuación 2.9, (J. Fan et al., 2020).

$$RMSD = \sqrt{\frac{1}{m} \sum_{i=1}^m (Y_{i,o} - Y_{i,p})^2} \quad (2.9)$$

Hipótesis y Objetivos

3.1. Hipótesis

El sobre entrenamiento de una red neuronal se verá disminuido a través de un modelo de optimización por la colonia artificial de abejas al explorar la relación entre la precisión y las técnicas de predicción considerando las variables en series de tiempo climáticas.

3.2. Objetivos

Diseñar un modelo de redes neuronales entrenado con datos meteorológicos y optimizado mediante un algoritmo de colonia artificial de abejas para la predicción de la radiación solar global.

Objetivos particulares

- Analizar la base de datos meteorológica con técnicas de preprocesamiento de datos para eficientizar los datos.
- Desarrollar modelo predictivo de redes neuronales para la predicción de la radiación solar global.
- Optimizar el modelo mediante el algoritmo de la colonia artificial de abejas para incrementar la precisión del modelo.
- Entrenar el modelo ya optimizado con la base de datos preprocesada para calcular valores de radiación solar global en tiempo real.
- Medir la precisión del modelo mediante indicadores estadísticos para validar la calidad del modelo desarrollado.

Metodología

El presente capítulo describe la metodología utilizada para predecir la radiación solar global en la región de Querétaro, México, utilizando modelos de aprendizaje automático, profundo, además de probar el modelo híbrido ABC-MLP. El proceso metodológico se divide en varias etapas clave, que incluyen la recopilación de datos, el análisis de datos, la construcción y el entrenamiento del modelo, y la evaluación del modelo. Cada una de estas etapas se detalla a continuación. La 4.1 ilustra el flujo de trabajo completo de la metodología aplicada.

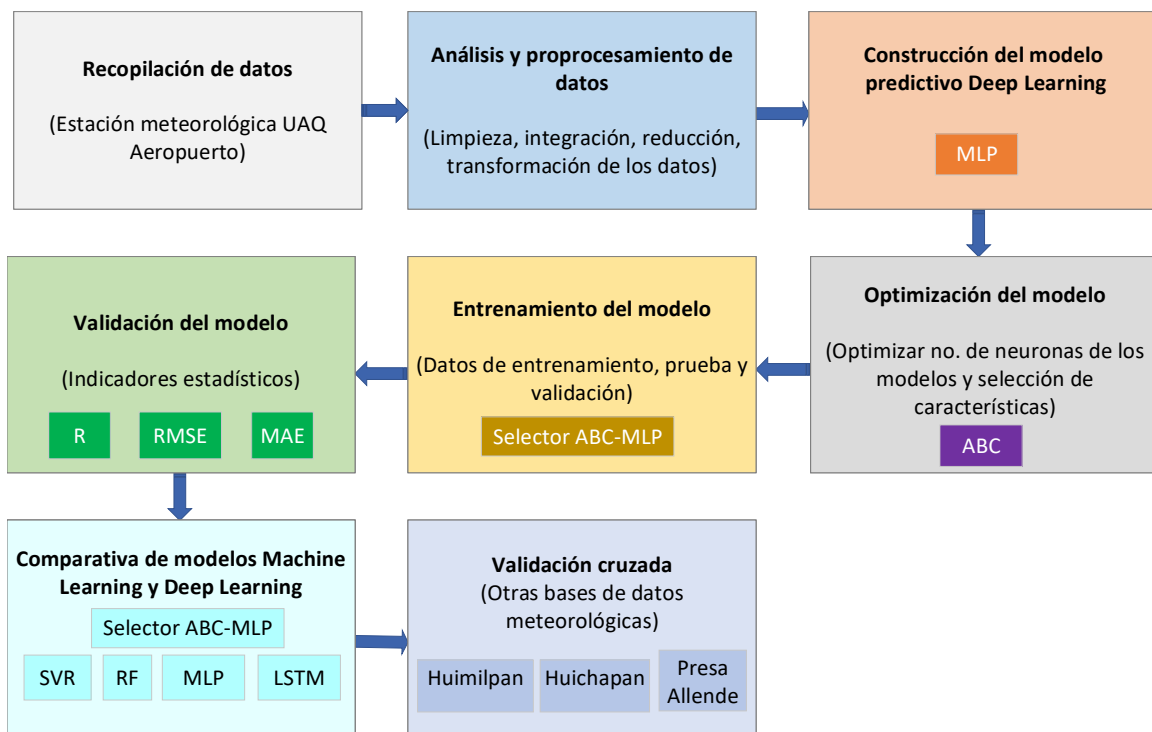


Figura 4.1. Diagrama de metodología propuesta.

En la 4.2 se puede observar el cronograma de actividades que se irán desarrollando a lo largo de la investigación propuesta.

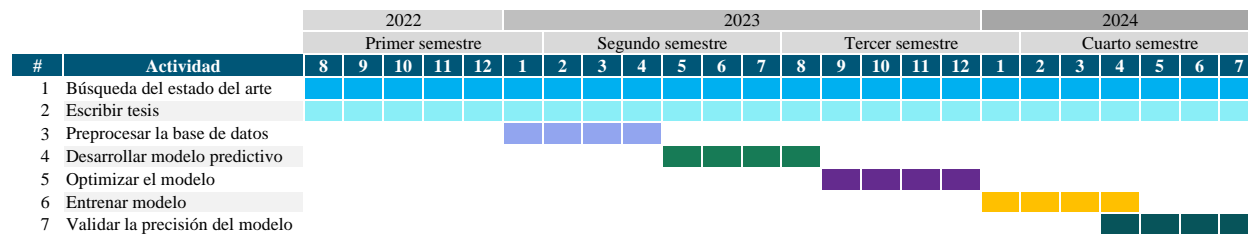


Figura 4.2. Cronograma de actividades.

4.1. Adquisición de datos






En este estudio, los datos se recolectaron de una estación meteorológica ubicada en la Universidad Autónoma de Querétaro, campus Aeropuerto, situada a una latitud de 20.626098 y una longitud de -100.359066. La estación ha estado monitoreando varias variables meteorológicas, incluyendo la dirección y velocidad del viento, humedad relativa, punto de rocío, precipitación, radiación solar global y temperatura desde diciembre de 2017 hasta junio de 2018. Se recopilieron un total de 28,053 entradas durante este período. El objetivo principal de esta estación es estimar el potencial para la utilización de energía solar y calor residual en la región del estado de Querétaro. La estación meteorológica cuenta con sensores de velocidad, dirección del viento, humedad relativa, presión, radiación solar y precipitación, descritos de manera más clara en la Tabla 4.1.

4.2. Análisis de datos

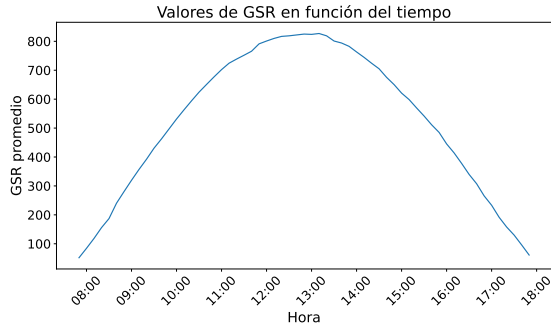
Para ilustrar el comportamiento de la radiación solar en diferentes condiciones meteorológicas, se presentan ejemplos específicos extraídos de la base de datos recolectada. La Fig. 4.3 (a) muestra un día soleado del 10 de enero de 2018. In esta gráfica, se observa un gradual incremento of the radiación solar global (GSR) from 8:00 to alcanzar su punto máximo alrededor del medioda, seguido by una gradual disminución hasta las 18:00. Este patrón es típico de un día soleado donde la intensidad de la radiación solar es más alta debido a la posición del sol en el cielo. La Fig. 4.3 (b), correspondiente al 4 de enero de 2018, presenta un patrón similar con un aumento gradual de la GSR desde las 9:00 hasta las 18:00 horas, reflejando también un da soleado con un pico al medioda.

Por otro lado, para analizar el comportamiento en días nublados, la Fig. 4.4 (a) muestra los datos del 12 de mayo de 2018. En esta gráfica, se presentan valores de GSR significativamente más bajos y con fluctuaciones irregulares a lo largo del día, lo que es característico de días nublados donde la cobertura de nubes reduce la cantidad de radiación solar que llega a la superficie terrestre. La Fig. 4.4 (b) del 10 de junio de 2018 también refleja un día nublado

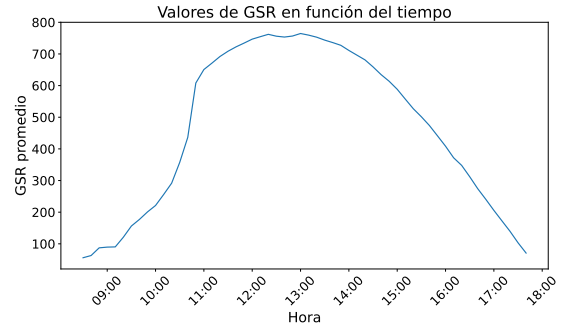
Tabla 4.1. Sensores utilizados en el estudio

Foto	Sensor	Características
	Sensor de velocidad y dirección de viento	Anemómetro ultrasónico bidimensional Tasa de muestreo de 40Hz
	Sensor de humedad relativa y temperatura	Humedad: 0 % a 100 %, Resolución de 0.03 % Temperatura: -40°C a 70°C, Resolución de 0.01°C
	Sensor de presión	Exactitud ± 0.3 hPa a +20°C
	Piranómetro	Mide la irradiación solar, Sensitividad del 12.2 $\mu\text{V}/\text{W}/\text{m}^2$.
	Pluviómetro	Recolección de 200 mm (7,87 pulgadas) de diámetro

con mediciones de GSR desde las 7:00 hasta las 17:00 horas, mostrando variaciones irregulares y sin un patrón claro como en los días soleados. Estas gráficas demuestran claramente

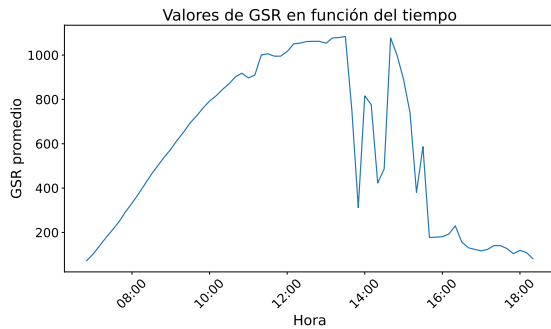


a)

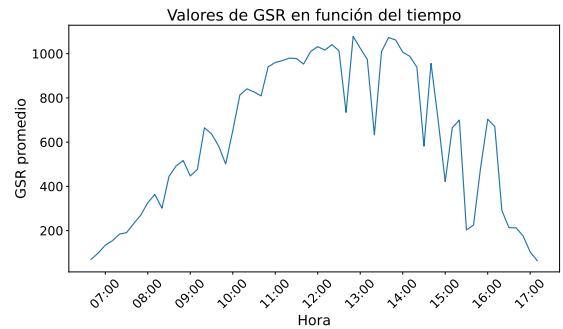


b)

Figura 4.3. Valores de la radiación solar global (GSR) promedio en función del tiempo para días soleados: a) 10 de enero de 2018, b) 4 de enero de 2018.



a)



b)

Figura 4.4. Valores de la radiación solar global (GSR) promedio en función del tiempo para días nublados: a) 12 de mayo de 2018, b) 10 de junio de 2018.

la diferencia en la radiación solar recibida en días soleados frente a días nublados. Los días soleados presentan un patrón de radiación más estable y predecible, con un claro pico al mediodía, mientras que los días nublados muestran una variabilidad considerable debido a la cobertura de nubes. Estos ejemplos de días específicos son parte del análisis detallado de los datos meteorológicos recolectados, proporcionando una comprensión más profunda del comportamiento de la radiación solar bajo diferentes condiciones climáticas.

Basado en los hallazgos de estudios previos (Woldegiyorgis et al., 2022; Dahmani et al., 2024; Guermoui et al., 2022; Zhou et al., 2023), las variables seleccionadas para la modelización mediante aprendizaje automático o profundo incluyen temperatura, humedad relativa, presión atmosférica, radiación solar global, velocidad del viento y dirección del viento. La Tabla 4.2 presenta las métricas estadísticas esenciales tales como media, mínimo, máximo, primer cuartil, mediana, tercer cuartil y desviación estándar (SD) tanto para los parámetros de entrada como de salida utilizados en esta investigación.

Las variables que tuvieron la mayor correlación con la radiación solar global fueron la temperatura y la dirección del viento, mientras que se observó una correlación negativa con

Tabla 4.2. Datos meteorológicos usados en el estudio.

Resumen	T_{avg} (C°)	HR (%)	GSR_{avg} (Wh/m ²)	P_{avg} (hPa)	Prec (mm)	VelVto (m/s)	DirVtoD1 (grados)	DirVtoSD1
media	177.738	560.49	2557.81	807.15	59	3.645	1387.89	149.558
SD	6.4609	242.06	3494.45	5.02	1.57	1.642	883.451	11.8125
min	-695	7.796	0	57.65	0	95	0	0
25 %	13.14	34.93	0	806	0	2.428	73.64	8.57
50 %	17.31	54.03	374	807	0	3.483	89.8	10.88
75 %	22.67	77.54	525.1	809	0	4.735	242.3	16.77
max	35.81	100	1220	814	13	11.2	359.8	102.5

la humedad relativa. Por otro lado, la presión atmosférica, la precipitación, la velocidad del viento y la hora del día no mostraron ninguna correlación significativa con la radiación solar global Fig. 4.5. Estas relaciones se visualizaron además a través de un mapa de correlación, proporcionando una visión completa de la interacción entre los diferentes factores meteorológicos y la distribución de la radiación solar.

4.3. Construcción del modelo

Para la creación de modelos de aprendizaje automático, se utilizó la biblioteca scikit-learn. RandomForestRegressor (RF) empleó 1000 regresores. En el caso de Support Vector Regressor (SVR), se utilizó GridSearchCV para obtener las mejores configuraciones para cada kernel. Con el kernel lineal, se obtuvieron los parámetros $C=10$ y $\text{gamma}=\text{scale}$; con el kernel de función de base radial, se establecieron los parámetros $C=100$ y $\text{gamma}=\text{scale}$; y con el kernel polinómico, se definieron los parámetros $C=10$ y $\text{gamma}=\text{auto}$.

Se emplearon diferentes configuraciones para los modelos de aprendizaje profundo. Para el Multilayer Perceptron (MLP), se implementaron 4 capas ocultas con 9 neuronas de entrada, seguidas de 100, 200, 500 y 200 neuronas en las capas ocultas respectivas, con una neurona de salida. Para ABC-MLP, el algoritmo ABC determinó el número de neuronas en las capas ocultas, que fueron 660, 955, 439 y 644 respectivamente. Ambos modelos utilizaron la función de activación ReLU en las capas ocultas y una función lineal en la capa de salida.

Para el modelo Long Short-Term Memory (LSTM), se emplearon 2 capas LSTM con 50 unidades cada una. Todos los modelos se entrenaron utilizando el optimizador Adam y la función de pérdida de Error Absoluto Medio (MAE) durante 500 épocas.

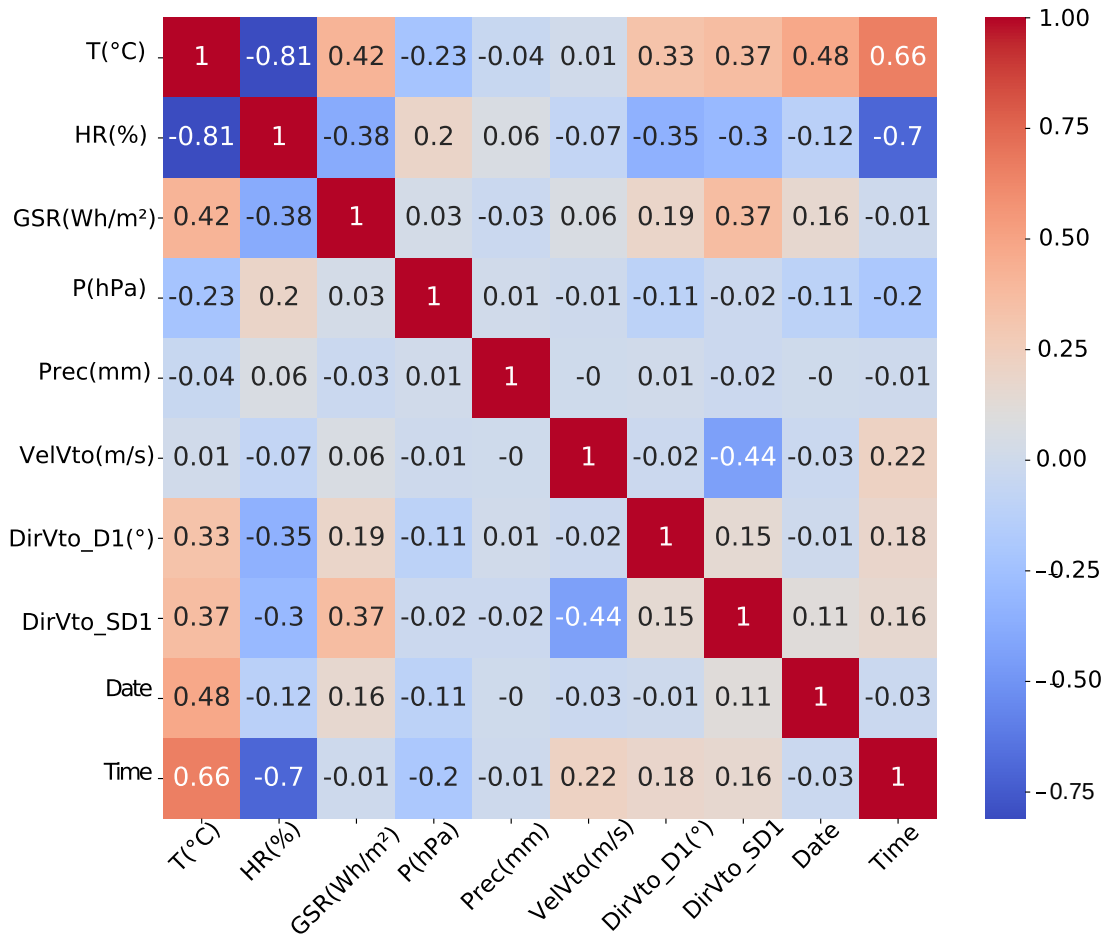


Figura 4.5. Matriz de correlación de las variables climatológicas.

4.4. Optimización de las capas del modelo MLP

La configuración de los hiperparámetros de la red neuronal, específicamente el número de neuronas en cada capa oculta, fue optimizada utilizando el algoritmo 1 colonia artificial de abejas (ABC), creado por (Vasuki, 2020). Este proceso de optimización se centró en encontrar la combinación óptima de neuronas para maximizar la precisión del modelo mientras se minimiza el tiempo de ejecución. La Tabla 4.5 resume los resultados de esta optimización.

El tiempo de ejecución promedio en segundos para diferentes tamaños de colonia (número de abejas) muestra que el tiempo de ejecución aumenta significativamente con el tamaño de la colonia. En particular, una colonia con 10 abejas tiene un tiempo de ejecución menor comparado con colonias más grandes, como las de 50 o 100 abejas. Esto indica que al escoger 10 abejas se logra una ejecución más rápida, lo cual es deseable para aplicaciones que re-

Tabla 4.3. Hiperparámetros usados en los modelos ML

Modelo	Hiperparámetros	Valor
SVR Lineal	C	10
	gamma	scale
SVR Radial basis function	C	100
	gamma	scale
SVR Polinomial	C	10
	gamma	auto
RF	Regresores	1000

quieren predicciones rápidas y eficientes, como se observa en la Fig. 4.6. Además, la relación entre la función objetivo y el tiempo de ejecución para diferentes tamaños de colonia revela que, aunque el tiempo de ejecución aumenta con el tamaño de la colonia, la mejora en la función objetivo no es proporcionalmente significativa. Para una colonia de 10 abejas, se logra un valor de la función objetivo que es competitivo comparado con colonias más grandes, pero con un tiempo de ejecución mucho menor, sugiriendo que se obtiene un buen balance entre la precisión de la predicción y la eficiencia del tiempo de ejecución con una colonia de 10 abejas, como se muestra en la Fig. 4.7. El mapa de calor que muestra en la Fig. 4.8, la función objetivo promedio para diferentes combinaciones de tamaño de colonia y límite, indica que para un tamaño de colonia de 10 y un límite de 5, la función objetivo promedio es la más baja con un promedio de 31.38 comparado con otras combinaciones, indicando que esta configuración no solo es eficiente en términos de tiempo de ejecución, sino que también proporciona una buena precisión en la predicción de la radiación solar global.

Nota. Pseudocódigo del algoritmo de la colonia artificial de abejas (Vasuki, 2020).

El algoritmo ABC exploró el espacio de búsqueda definido por los límites inferior y superior para cada capa oculta, evaluando múltiples configuraciones. Los resultados de esta exploración se utilizaron para seleccionar las neuronas óptimas en cada capa oculta. Los valores obtenidos (454, 227, 1000, y 429 para las capas ocultas 1, 2, 3, y 4 respectivamente)

Algorithm 1 Pseudocódigo del algoritmo de optimización ABC

- 1: **Inicialización**
 - 2: Tamaño de la población de abejas N
 - 3: Defina la función objetivo para el problema $f(X)$
 - 4: Número de etapas M
 - 5: $f(X)$ consta de los componentes de la solución $[X_1, X_2, \dots, X_n]$
 - 6: La mejor solución actual C_{Best} se inicializa a cualquier solución factible
 - 7: Número máximo de iteraciones $MaxIter$
 - 8: $np \leftarrow 1$ ▷ np es la variable para indexar las etapas
 - 9: $iter \leftarrow 1$
 - 10: **while** $iter \leq MaxIter$ **do**
 - 11: **for** $np \leftarrow 1$ **to** M **do**
 - 12: Las abejas vuelan de la colmena y crean soluciones parciales durante el paso hacia adelante
 - 13: Las soluciones parciales se eligen del conjunto disponible en la etapa $s(np)$
 - 14: Las abejas vuelven a la colmena durante el paso hacia atrás y se intercambian información
 - 15: **end for**
 - 16: **if** $S(iter)$ es mejor que C_{Best} **then**
 - 17: $C_{Best} \leftarrow S(iter)$
 - 18: **end if**
 - 19: $iter \leftarrow iter + 1$
 - 20: **end while**
 - 21: La mejor solución actual C_{Best} es la solución óptima global al problema.
-

Tabla 4.4. Hiperparámetros usados en los modelos DL

Modelo	Capa entrada	Capa oculta 1	Capa oculta 2	Capa oculta 3	Capa oculta 4	Capa salida	Fun. activación capas ocultas	Fun. activación capa de salida
MLP	9	100	200	500	200	1	ReLu	Linear
ABC-MLP	9	660	955	439	644	1	ReLu	Linear
Propuesta Selector ABC-MLP	7	454	227	1000	429	1	ReLu	Linear
LSTM	-	50	50	-	-	-	-	-

	Límite inferior	Límite superior	Neuronas optimizadas
Capa oculta 1	1	1000	454
Capa oculta 2	1	1000	227
Capa oculta 3	1	1000	1000
Capa oculta 4	1	1000	429
Variables	Temperatura del aire, Humedad Relativa, Velocidad del viento, Dirección del viento, Año, Mes, Día		
Parámetros ABC	Tamaño colonia: 10, Límite: 5		

Tabla 4.5. Tabla de parámetros de capas ocultas y variables

reflejan una configuración que balancea adecuadamente la capacidad de aprendizaje del modelo y la eficiencia computacional.

Además, el proceso de optimización incluyó la consideración de las variables de entrada: temperatura del aire, humedad relativa, velocidad del viento, dirección del viento, año, mes y día. Estas variables fueron seleccionadas debido a su relevancia en la predicción de la radiación solar global.

El tamaño de la colonia y el límite del algoritmo ABC también fueron configurados para optimizar el desempeño del modelo. Se eligió un tamaño de colonia de 10 y un límite de 5, basados en análisis previos que mostraron que esta configuración proporciona un buen balance entre la precisión del modelo y el tiempo de ejecución, tal como se detalla en los gráficos y análisis presentados.

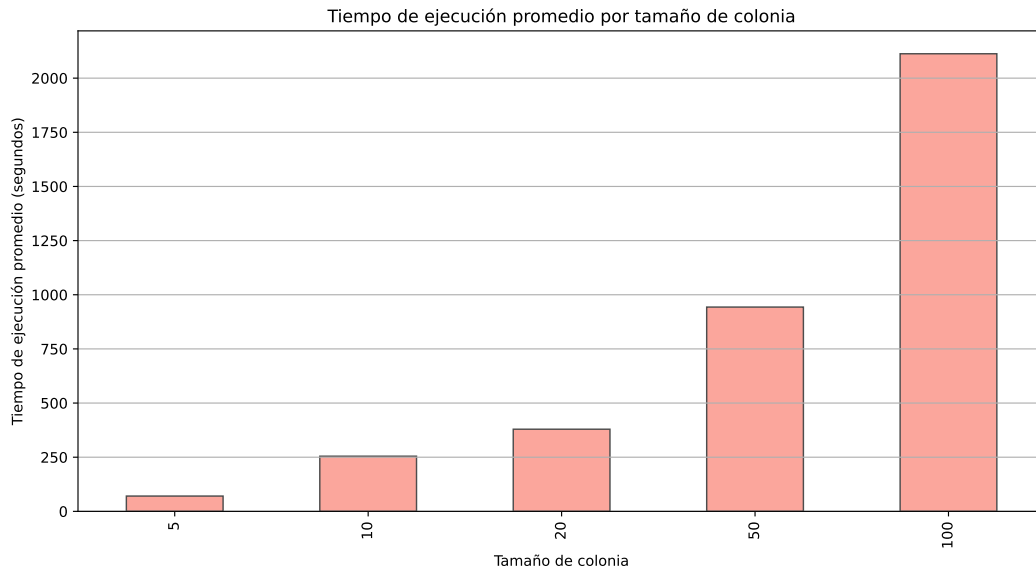


Figura 4.6. Tiempo de ejecución promedio en función del tamaño de la colonia.

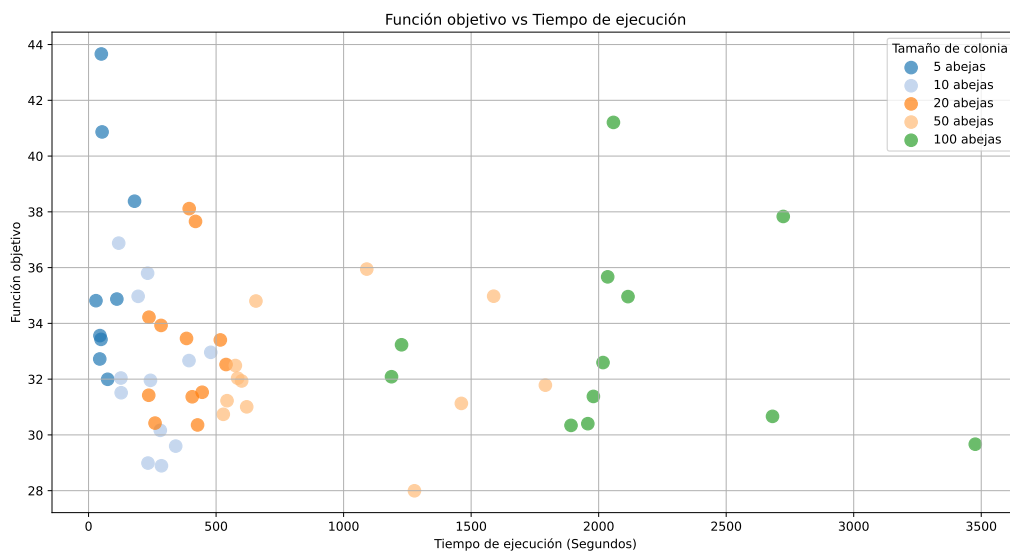


Figura 4.7. Relación entre el tiempo de ejecución y la función objetivo según el tamaño de la colonia.

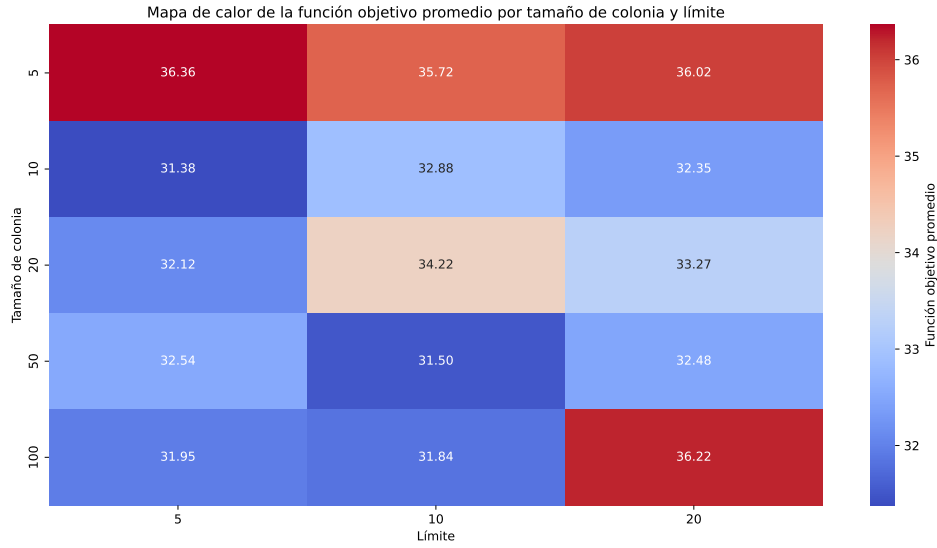


Figura 4.8. Mapa de calor de la función objetivo promedio en función del tamaño de la colonia y el límite.

4.5. Entrenamiento del modelo con base de datos pre-procesada

Los modelos se entrenaron utilizando el 80% de los datos para el entrenamiento y el 20% restante para las pruebas. El proceso de entrenamiento utilizó el optimizador Adam y la función de pérdida de Error Absoluto Medio (MAE) durante 500 épocas. Este proceso permitió el ajuste de los parámetros del modelo para minimizar los errores de predicción y mejorar la precisión.

4.6. Evaluación del Modelo

Se utilizaron varios indicadores estadísticos, como el error absoluto medio (MAE) 4.1, el error cuadrático medio (RMSE) 4.2 y el coeficiente de determinación (R^2) 4.3, para evaluar los modelos. Estas métricas se calcularon para comparar el rendimiento de los modelos en la predicción de la radiación solar global. Los resultados de la evaluación de los modelos se presentan en una tabla comparativa entre modelos de aprendizaje automático y aprendizaje profundo, y se ilustran en los gráficos de regresión.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (4.1)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4.2)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (4.3)$$

Resultados y discusión

5.1. Resultados de la revisión sistemática

En esta sección se muestran los resultados obtenidos de una revisión sistemática para obtener un mejor entendimiento de los modelos híbridos para la predicción de radiación solar global, usando algoritmos de optimización y algún modelo de machine learning o deep learning.

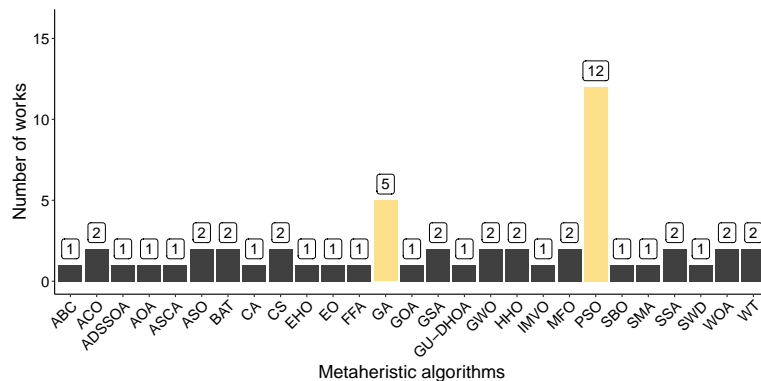


Figura 5.1. El número de trabajos que utilizan los diferentes algoritmos de optimización.

El algoritmo PSO se ha destacado como el optimizador más utilizado en la literatura científica, apareciendo en un total de 12 estudios, lo que representa el 29% de la literatura revisada. Esto resalta su importancia y efectividad para resolver una amplia gama de problemas. Le sigue de cerca el GA, que ha estado presente en 5 ocasiones, representando el 12.1%, confirmando su popularidad y aplicabilidad en varios contextos.

Otros métodos de optimización, como ACO, ASO, BAT, CS, GSA, GWO, HHO, MFO, SSA, WOA y WT, también se han empleado en 2 trabajos cada uno, aunque algunos de ellos solo se mencionaron en sus respectivos artículos. Esta variedad de enfoques refleja la diversidad de herramientas disponibles y la búsqueda continua de métodos efectivos para abordar problemas complejos.

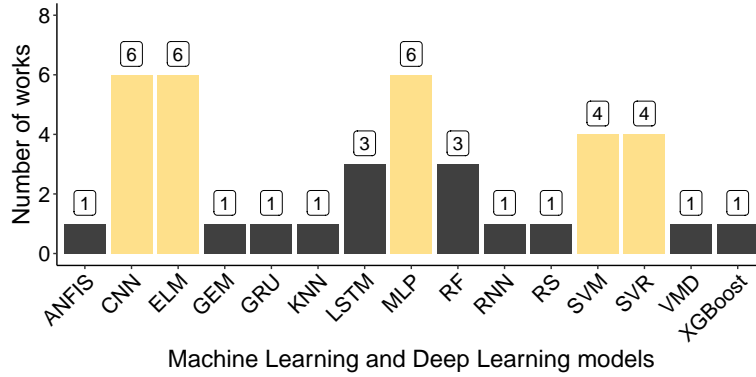


Figura 5.2. El número de estudios que utilizan los diferentes modelos de IA revisados en la literatura.

En el campo de la predicción de SR, se ha observado que CNN, ELM y MLP han surgido como los modelos más utilizados, representando el 43.9% de los trabajos, lo que sugiere su capacidad para capturar patrones complejos en los datos solares.

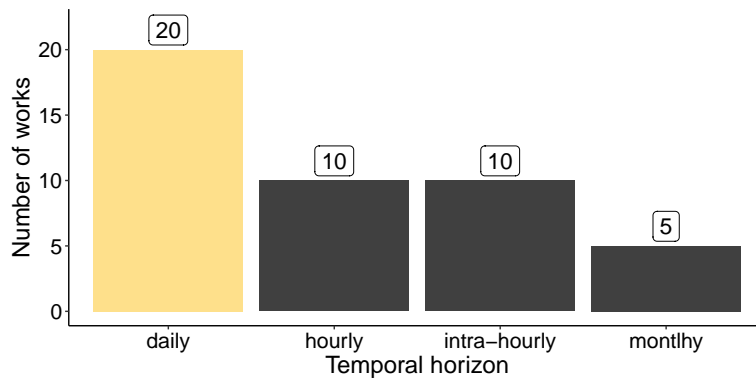


Figura 5.3. El número de trabajos que utilizan los diferentes horizontes temporales revisados en la literatura.

En cuanto a los horizontes temporales utilizados en los estudios, es notable que el período diario ha sido el más frecuentemente referido, con un total de 20 menciones, constituyendo el 48.7% de la literatura revisada. Esto indica un énfasis significativo en la predicción a corto plazo y la optimización de la gestión energética para la producción, el consumo y las operaciones de la red. Este énfasis subraya la importancia de abordar las necesidades operativas inmediatas. Posteriormente, los períodos intra-horario y horario han obtenido 10 menciones, representando el 24.3% de las referencias. Esto sugiere un interés en las predicciones a través de varias escalas de tiempo para una gestión efectiva de la energía y consideraciones de seguridad del sistema. Por el contrario, el período mensual ha recibido comparativamente menos atención, con solo 5 menciones. Esto puede implicar su menor relevancia en ciertos contextos dentro del campo de la investigación de la radiación solar.

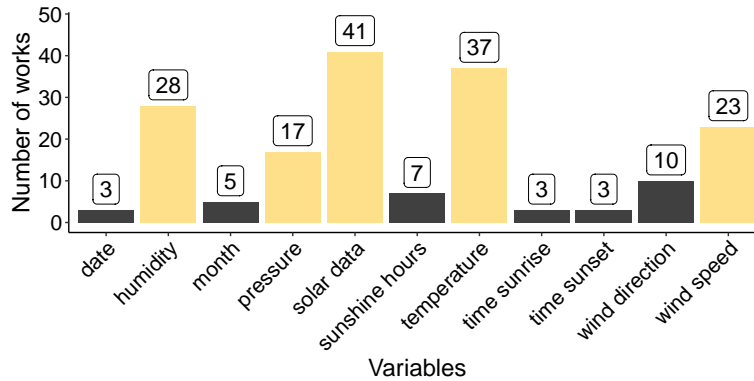


Figura 5.4. El número de estudios que utilizan las diferentes variables meteorológicas.

La distribución de trabajos basada en la utilización de diversas variables meteorológicas, según lo revisado en la literatura, se ha representado en la Figura 5.4. Un análisis de las variables más frecuentemente empleadas revela la inclusión constante de datos solares en todos los estudios, reflejando su papel fundamental en la predicción de SR. La temperatura ha surgido como una variable ampliamente estudiada, mencionada 37 veces, representando el 90 % de los casos, lo que subraya su importancia en los modelos de predicción. Además, la humedad, la velocidad del viento y la presión atmosférica también han surgido de manera prominente, con 28, 23 y 17 menciones, respectivamente, constituyendo el 68.2 %, 56 % y 41.4 % de los casos. Estos hallazgos destacan la importancia de estas variables en la modelización y predicción de SR.

Según la revisión realizada, los algoritmos como MLP, CNN y RNN, incluidos LSTM y GRU, son ampliamente utilizados debido a su efectividad cuando se integran con estrategias de optimización específicas como PSO, FFA y GWO. En este contexto, MLP muestra adaptabilidad cuando se combina con PSO y FFA, debido a su capacidad para optimizar soluciones en predicciones complejas. Mientras tanto, CNN, cuando se integra con AOA y GWO, facilita la identificación de características espaciales y temporales relevantes, permitiendo el descubrimiento de patrones complejos en la predicción de radiación solar. También es importante destacar la efectividad del modelo ADSSOA-LSTM, que logra un RMSE de 0.000388, convirtiéndose en uno de los modelos de mejor rendimiento en la literatura. Además, métodos como GWO, conocidos por su robustez en la optimización, fortalecen la adaptación en entornos dinámicos.

Además, RNN, cuando se combina con algoritmos como ADSSOA y PSO, permite el ajuste preciso de predicciones basadas en patrones a los modelos, mejorando así la adaptabilidad del modelo a nuevas condiciones sin comprometer la precisión. El uso de técnicas de transformación como la Transformada Wavelet (WT) facilita la descomposición de series temporales de radiación solar en componentes más manejables, mejorando la precisión de los modelos predictivos. Según estos enfoques, la diversidad de modelos no solo proporciona adaptabilidad y eficiencia en el procesamiento de datos, sino que también mejora la precisión de las predicciones, lo que impacta significativamente en la gestión de los recursos de energía

renovable.

Al comparar los trabajos del estado del arte con la propuesta Selector ABC-MLP, se observa que nuestro enfoque logra un mejor desempeño para horizontes temporales intrahorarios. Gracias a la optimización de hiperparámetros utilizando el algoritmo ABC, el modelo propuesto alcanza un RMSE de 73.7466 Wh/m², un MAE de 37.1981 Wh/m² y un R² de 0.9409, superando a otros métodos en precisión y eficiencia. La Tabla 5.1 resume esta comparación.

Tabla 5.1. Comparación de trabajos del estado del arte con la propuesta Selector ABC-MLP

Autor	Algoritmo	Comparativa	Horizonte temporal	Optimización	Resultados
(Jadidi et al., 2018)	GA	-	Diario	Selección de características	RMSE 47.6955 Wh/m ² y MAE 24.7772 Wh/m ²
(Vaisakh y Jayabharathi, 2022)	GU-DHOA	CNN, RNN	Intra-horario	Neuronas ocultas	RMSE 212.27 Wh/m ² y MAE 136.51 Wh/m ²
(Y. Zhang et al., 2019)	PSO	-	Intra-horario	Pesos y sesgos	R ² 0.7649–0.9678
(Djeldjli et al., 2023)	FFA	SVM	Diario	Selección de características	R 0.9321 W/m ² , rRMSE % 11.96 y MAPE % 8.26
(Huang et al., 2022)	GOA	ANFIS	Mensual	No mencionado	nMAE 0.545
(Xu et al., 2024)	EO	-	Horario	Pesos y sesgos	RMSE 141.61 W/m ² , MAE 108.07 W/m ² y R ² 0.91
Propuesta	ABC	MLP	Intra-horario	Neuronas ocultas	RMSE 73.7466, MAE 37.1981, R ² 0.9409

5.2. Resultados de la comparativa de modelos

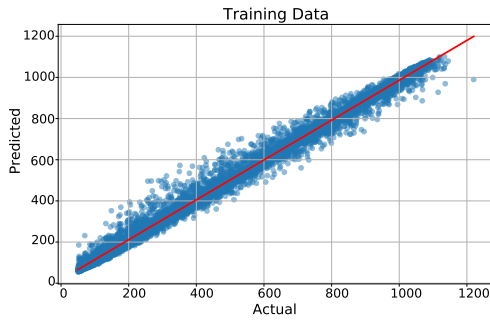
En esta sección se muestran los resultados obtenidos de la comparativa entre modelos de machine learning como SVR y RF, además de modelos de deep learning como LSTM, MLP, ABC-MLP y el modelo propuesto Selector ABC-MLP. La Tabla 5.2 muestra los valores de RMSE, MAE y R², permitiendo una comparación detallada del rendimiento de los modelos.

Es evidente que el modelo propuesto Selector ABC-MLP mostró el mejor rendimiento en todas las métricas, seguido de cerca por MLP estándar y RF. El rendimiento del modelo Selector ABC-MLP y ABC-MLP en la métrica MAE es notable, ya que lograron un buen resultado, debido a que el algoritmo de optimización ABC, tenía como objetivo minimizar el error de MAE. Sin embargo, es importante destacar que los modelos SVR con kernels lineal y polinómico deben descartarse debido a sus valores negativos de R^2 . Esto indica que estos modelos no son adecuados para predecir valores de radiación solar global. La presencia de un valor negativo en R^2 sugiere que el modelo es inadecuado y que su rendimiento es inferior a simplemente usar la media de los valores observados como predicción (Chicco et al., 2021; Nakagawa y Schielzeth, 2013).

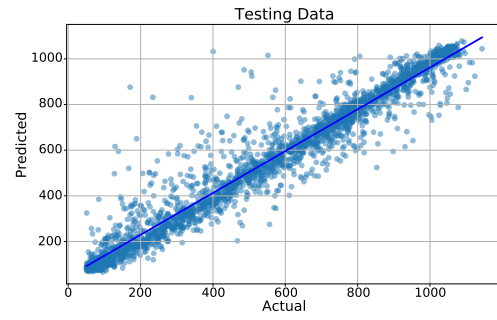
Tabla 5.2. Comparativa entre modelos de machine learning y deep learning.

Modelo	RMSE	MAE	R^2
Propuesta Selector ABC-MLP	73.7466	37.1981	0.9409
MLP	78.5627	40.9764	0.9330
RF	79.5018	47.8121	0.9314
LSTM	89.8786	48.3569	0.9123
SVR-RBF	104.4158	57.3045	0.8816
ABC-MLP	107.4657	41.3594	0.8746
SVR-Linear	958.7575	165.2579	-8.9838
SVR-Poly	42,018.2718	997.7063	-19,174.8420

Las Figuras 5.5, 5.8, 5.9, 5.10 y 5.11 presentan comparaciones gráficas que ilustran el rendimiento de los modelos identificados durante las fases de entrenamiento y prueba del proceso de modelado. Una menor dispersión corresponde a una mayor precisión, resultando en menores errores de pronóstico. Se observó que el modelo óptimo se determinó en función de su capacidad para producir los valores más bajos de Root Mean Squared Error (RMSE), ya que una menor dispersión corresponde a una mayor precisión, resultando en menores errores de pronóstico. A pesar de que el modelo propuesto supera a los modelos MLP, RF, LSTM, ABC-MLP y SVR-RBF en términos de precisión de estimación, es evidente que estos enfoques alternativos surgen como opciones convincentes para una predicción precisa de la radiación solar. La discrepancia entre los datos medidos y pronosticados es significativa para los modelos MLP, SVR-RBF y LSTM (Ver Figuras 5.9, 5.8 y 5.11), mientras que es mínima al usar el modelo propuesto Selector ABC-MLP. En la Figura 5.6 se puede observar que SVR con kernel lineal intenta predecir los datos de entrenamiento, sin éxito y tampoco logra predecir los datos de prueba, lo cual se refleja en el coeficiente R^2 , al igual que SVR con kernel polinomial (Figura 5.7).

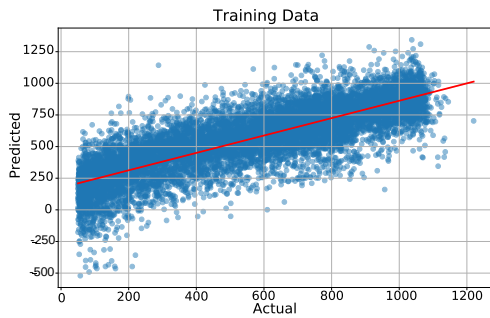


a)

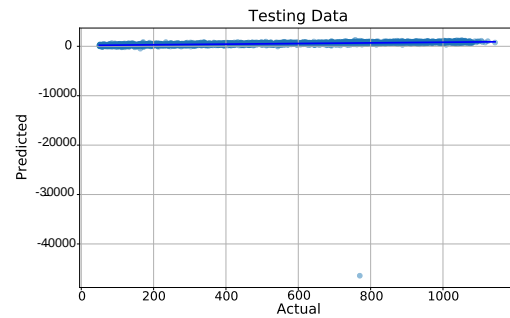


b)

Figura 5.5. Gráfico de regresión que ilustra los valores predichos versus reales del modelo RF en a) datos de entrenamiento, b) datos de prueba.

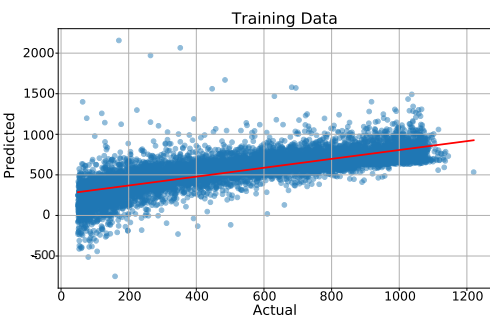


a)

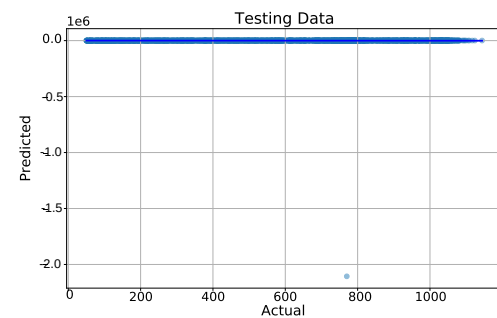


b)

Figura 5.6. Gráfico de regresión que ilustra los valores predichos versus reales del modelo SVR con kernel lineal en a) datos de entrenamiento, b) datos de prueba.

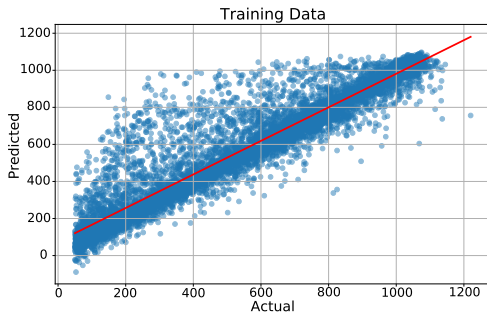


a)

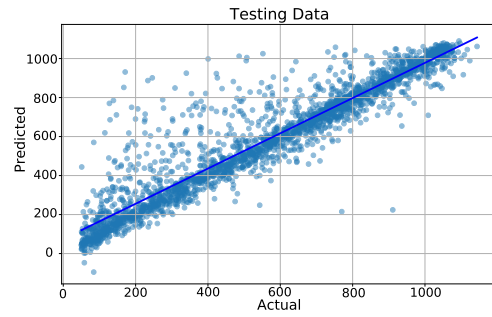


b)

Figura 5.7. Gráfico de regresión que ilustra los valores predichos versus reales del modelo SVR con kernel polinómico en a) datos de entrenamiento, b) datos de prueba.

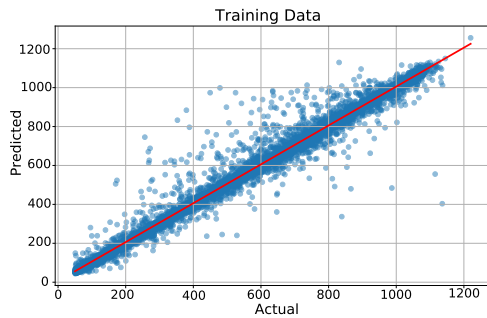


a)

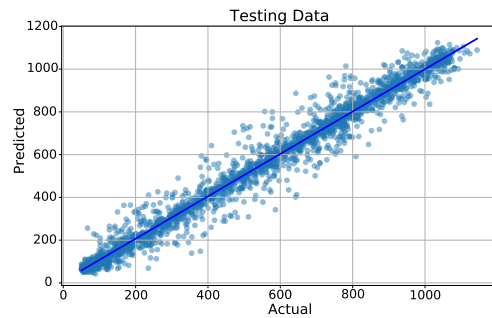


b)

Figura 5.8. Gráfico de regresión que ilustra los valores predichos versus reales del modelo SVR con kernel de función de base radial en a) datos de entrenamiento, b) datos de prueba.

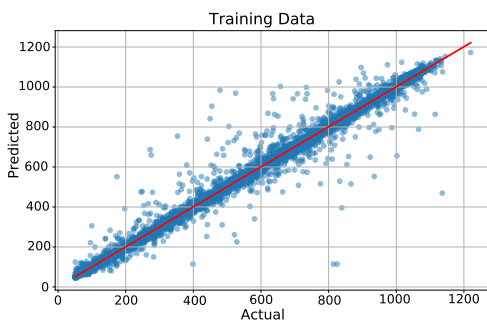


a)

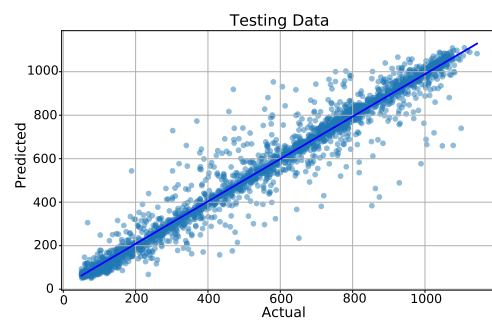


b)

Figura 5.9. Gráfico de regresión que ilustra los valores predichos versus reales del modelo MLP en a) datos de entrenamiento, b) datos de prueba.

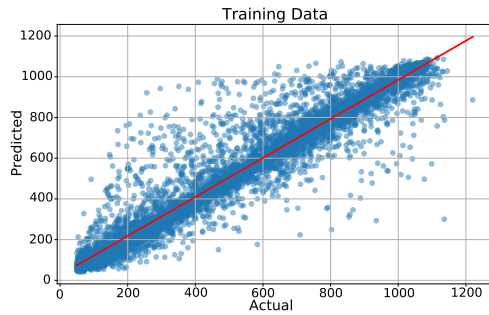


a)

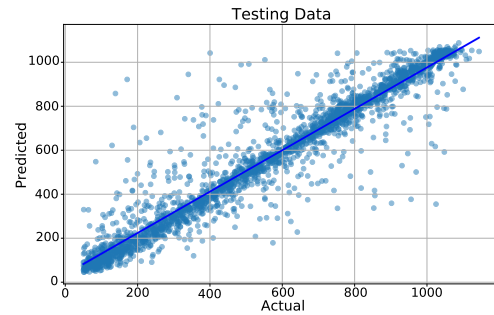


b)

Figura 5.10. Gráfico de regresión que ilustra los valores predichos versus reales del modelo ABC-MLP en a) datos de entrenamiento, b) datos de prueba.

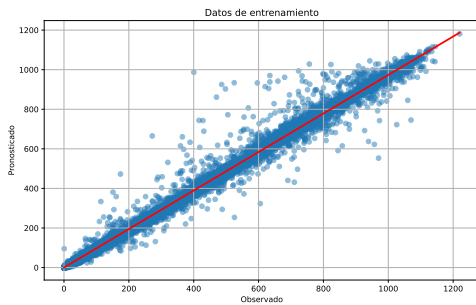


a)

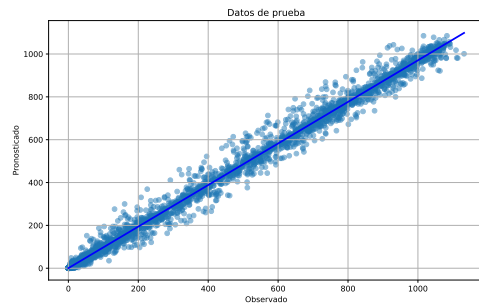


b)

Figura 5.11. Gráfico de regresión que ilustra los valores predichos versus reales del modelo LSTM en a) datos de entrenamiento, b) datos de prueba.



a)



b)

Figura 5.12. Gráfico de regresión que ilustra los valores predichos versus reales del modelo ABC-MLP con selección de características en a) datos de entrenamiento, b) datos de prueba.

Para finalizar, se realiza una comparación en la función de pérdida del modelo propuesto Selector ABC-MLP Figura 5.13 a) y la del modelo MLP estándar Figura 5.13 b), en donde podemos observar que ambas funciones tienen un comportamiento similar, aunque el modelo propuesto logra seguir mejorando a través de más épocas, lo que podría ayudar a tener mejores resultados en cuanto a sobre entrenamiento.

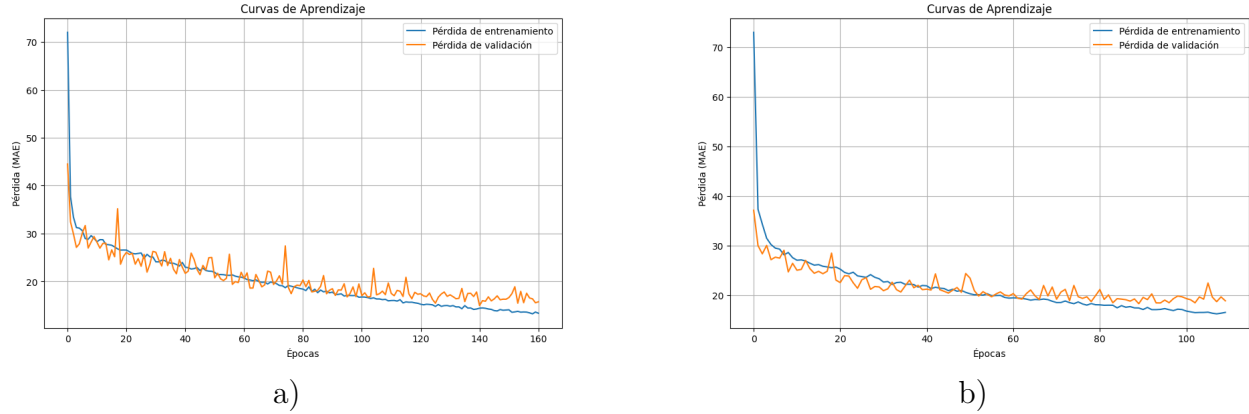


Figura 5.13. Gráfico con las funciones de pérdida de entrenamiento y validación del modelo propuesto Selector ABC-MLP a) y el modelo MLP b).

5.3. Validación con otras bases de datos

Se ha experimentado con el modelo propuesto ABC-MLP utilizando un selector de características, determinando que las variables más importantes eran el promedio de temperatura del aire, humedad relativa, promedio de radiación solar global, velocidad del viento, dirección del viento, mes, día y hora. Con estas variables, se realizó una validación cruzada en tres Estaciones Meteorológicas Automáticas (EMA) en México: una ubicada en Huichapan, Hidalgo (latitud 20.3886, longitud -99.6638), otra en San Miguel de Allende, Guanajuato (latitud 20.8483, longitud -100.8247), y finalmente, en Huimilpan, Querétaro (latitud 20.3902, longitud -100.2836111). La ubicación de estas estaciones se puede observar en la Tabla 5.3 y en la Figura 5.14.

Tabla 5.3. Ubicación EMAs.

EMA	Latitud	Longitud	Altura (m)
UAQ Aeropuerto	20.626098	-100.359066	1970
Huichapan	20.388611	-99.663888	2087
Huimilpan	20.390277	-100.283611	2280
Presa Allende	20.848333	-100.824722	1847

Como se observa en la Tabla 5.4, la estación UAQ Aeropuerto ha mostrado resultados excelentes, con una RMSE de 48.4032 W/m^2 , una MAE de 19.5054 W/m^2 y un R^2 de 0.9812, lo que indica que el modelo es muy preciso. Huichapan y Huimilpan han presentado valores de RMSE de 167.8547 W/m^2 y 181.2048 W/m^2 , respectivamente, mientras que las MAE han sido de 91.0539 W/m^2 y 89.9884 W/m^2 . Aunque estos valores son relativamente altos, mantienen buenos valores de R^2 , sugiriendo que el modelo es razonablemente bueno, pero con

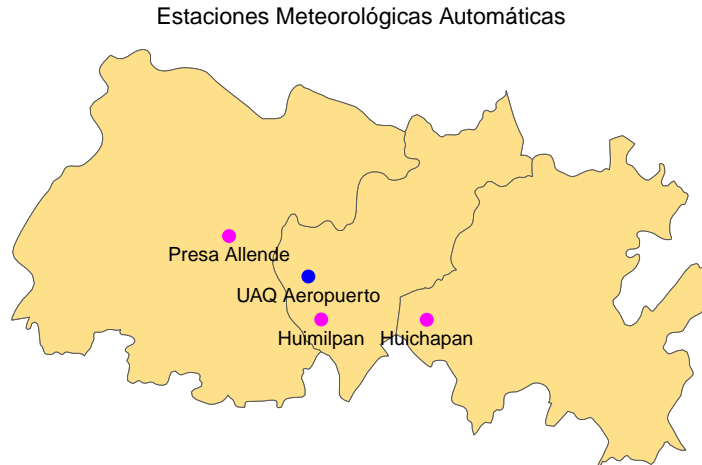


Figura 5.14. Ubicación de EMAs cercanas a la estación meteorológica UAQ Aeropuerto.

Tabla 5.4. Comparativa entre resultados del modelo ABC-MLP con selector de características en diferentes EMAs.

EMA	RMSE	MAE	R ²
UAQ Aeropuerto	73.7466	37.1981	0.9409
Huichapan	167.8547	91.0539	0.8112
Huimilpan	181.2048	89.9884	0.7882
Presa Allende	197.793	108.2048	0.7423

errores absolutos significativos. La estación Presa Allende ha mostrado el peor desempeño, con RMSE y MAE más altos, de 197.793 W/m² y 108.2048 W/m² respectivamente, y un R² de 0.7423, más bajo en comparación con las otras estaciones. No obstante, un R² de 0.7423 sigue siendo aceptable. En conclusión, los resultados del modelo ABC-MLP con selector de características han sido muy buenos para la estación meteorológica de la UAQ Aeropuerto y razonablemente buenos para las otras estaciones.

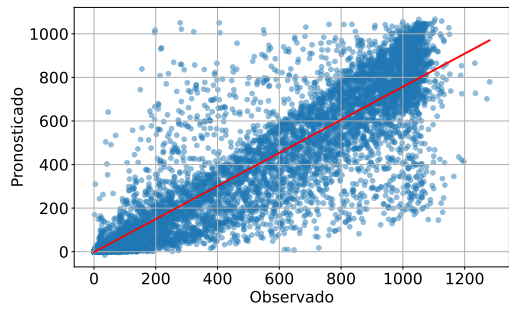


Figura 5.15. Estación Huichapan

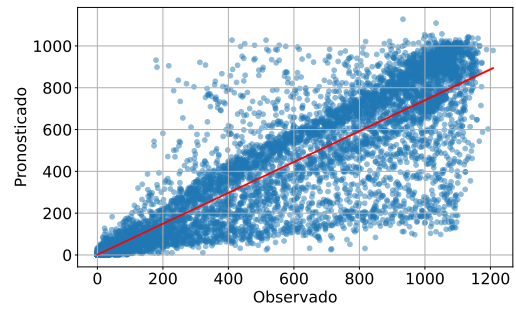


Figura 5.16. Estación Huimilpan

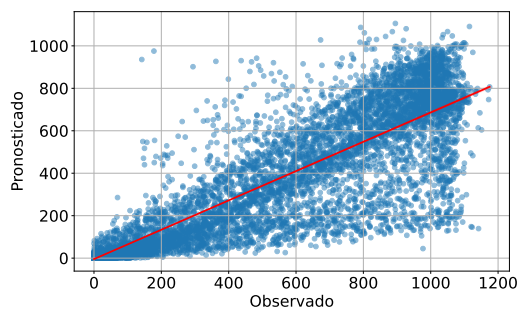


Figura 5.17. Estación Presa Allende

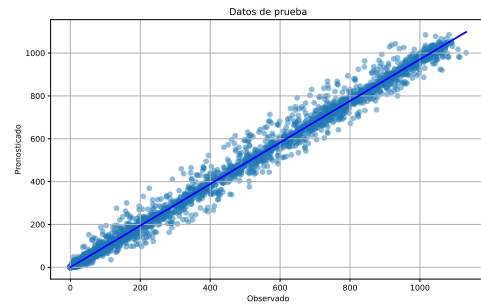


Figura 5.18. Estación UAQ Aeropuerto

Figura 5.19. Gráficos de regresión que ilustran los valores predichos en diferentes estaciones.

Conclusiones

En esta tesis, se ha desarrollado y validado un modelo híbrido para la predicción de la radiación solar global (GSR), combinando el algoritmo de optimización bioinspirado Colonia Artificial de Abejas (ABC) con un Perceptrón Multicapa (MLP). A través de este trabajo, se han alcanzado las siguientes conclusiones significativas, el modelo propuesto mostró una mejora notable en la precisión de la predicción de GSR en comparación con modelos predictivos tradicionales. En particular, se obtuvo un RMSE de 73.7466 W/m^2 y un MAE de 37.1981 W/m^2 en la estación UAQ Aeropuerto, con un R^2 de 0.9409. A partir de los resultados obtenidos y la comparación de las funciones de pérdida entre el modelo propuesto Selector ABC-MLP y el modelo MLP estándar, se puede concluir que la hipótesis planteada ha sido validada. Las curvas de aprendizaje muestran que ambas funciones tienen un comportamiento similar durante las primeras épocas, pero el modelo Selector ABC-MLP sigue mejorando a través de más épocas. Esto indica que el modelo optimizado es más robusto y presenta una mejor capacidad de generalización, reduciendo el sobreentrenamiento.

El uso del algoritmo de optimización por la colonia artificial de abejas ha demostrado ser efectivo en la disminución del sobreentrenamiento al explorar la relación entre la precisión y las técnicas de predicción, considerando las variables en series de tiempo climáticas. Estos resultados confirman que la implementación del modelo Selector ABC-MLP puede ofrecer un mejor desempeño en comparación con un MLP estándar, validando así la hipótesis propuesta. La optimización de los hiperparámetros del MLP mediante el algoritmo ABC permitió una configuración óptima del número de neuronas en las capas ocultas, mejorando la precisión del modelo. La validación cruzada utilizando datos de diversas estaciones meteorológicas automáticas en México demostró la robustez y generalización del modelo ABC-MLP. Este modelo es capaz de adaptarse a diferentes contextos geográficos, manteniendo una alta precisión en la predicción de GSR. En la estación Presa Allende, el modelo optimizado mostró un RMSE de 197.793 W/m^2 , un MAE de 108.2048 W/m^2 y un R^2 de 0.7423, lo que indica áreas de mejora para condiciones climáticas diversas. La selección de características críticas, como la temperatura del aire, humedad relativa, radiación solar global promedio, velocidad y dirección del viento, contribuyó significativamente a la mejora del rendimiento del modelo predictivo. Esta selección permitió capturar de manera más precisa las variables que

afectan la GSR. Los resultados obtenidos destacan la superioridad del modelo propuesto Selector ABC-MLP en la predicción de GSR, proporcionando una herramienta valiosa para el desarrollo de sistemas de energía solar más eficientes y la gestión optimizada de recursos energéticos. Este enfoque puede contribuir significativamente a la toma de decisiones en el sector energético, mejorando la seguridad y eficiencia del suministro energético.

6.1. Trabajos a futuro

Basado en los hallazgos y la discusión en el estudio sobre las técnicas de aprendizaje automático y profundo utilizadas en la predicción de la radiación solar global, se han sugerido varias direcciones para futuras investigaciones que podrían expandir y profundizar el impacto y la aplicación de estas tecnologías:

- La investigación adicional en algoritmos de optimización alternativos para modelos de aprendizaje profundo en la predicción de la radiación solar global puede ser beneficiosa. Por ejemplo, la investigación puede explorar metodologías como la Optimización por Colonia de Hormigas (ACO), el Algoritmo Genético (GA) y la Optimización por Enjambre de Partículas (PSO). Estos algoritmos pueden ofrecer ventajas únicas y potencialmente mejorar el rendimiento predictivo de los modelos.
- Experimentar con variables de entrada adicionales más allá de las consideradas en este estudio puede ser valioso. Factores como las condiciones atmosféricas, la cobertura de nubes y las características geográficas pueden influir en los patrones de radiación solar y pueden mejorar las capacidades predictivas de los modelos.
- Evaluar el rendimiento a largo plazo y la estabilidad del modelo desarrollado durante períodos de tiempo extendidos puede proporcionar información valiosa sobre su fiabilidad y efectividad para aplicaciones prácticas. Los estudios longitudinales que rastrean el rendimiento del modelo durante múltiples años o estaciones pueden ayudar a identificar desafíos potenciales y oportunidades para el refinamiento del modelo.

Anexos



Figura 7.1. Constancia de idioma.



UNIVERSIDAD AUTÓNOMA DE QUERÉTARO
FACULTAD DE INGENIERÍA



THE QUERÉTARO STATE UNIVERSITY THROUGH
THE ENGINEERING FACULTY GRANT THE
PRESENT ACKNOWLEDGMENT TO:

Sergio Alejandro Pérez Rodríguez

FOR THE PARTICIPATION:

CONFERENCE:

Exploring Machine Learning and Deep Learning Models for Global Solar Radiation Prediction: A
Comparative Analysis
May, 2024


Dra. María de la Luz Pérez Rea
PRINCIPAL
ENGINEERING FACULTY


Dr. Gonzalo Macías Bobadilla
GENERAL COORDINATOR CONIIN
ENGINEERING FACULTY



<http://ingenieria.uaq.mx/coniin>

Figura 7.2. Participación en CONiIN XX.



Metaheuristic Algorithms for Solar Radiation Prediction: A Systematic Analysis

SERGIO A. PÉREZ-RODRÍGUEZ¹, JOSÉ M. ÁLVAREZ-ALVARADO²,
JULIO-ALEJANDRO ROMERO-GONZÁLEZ¹, MARCOS AVILES³, (Member, IEEE),
AMERICA EILEEN MENDOZA-ROJAS³, CARLOS FUENTES-SILVA³,
AND JUVENAL RODRÍGUEZ-RESÉNDIZ², (Senior Member, IEEE)

¹Facultad de Informática, Universidad Autónoma de Querétaro, Querétaro 76230, Mexico

²Facultad de Ingeniería, Universidad Autónoma de Querétaro, Querétaro 76010, Mexico

³Universidad Politécnica de Querétaro, El Marqués, Querétaro 76240, Mexico

Corresponding author: José M. Álvarez-Alvarado (jmalvarez@uaq.edu.mx)

ABSTRACT In the contemporary world, where the escalating demand for energy and the imperative for sustainable sources, notably solar energy, have taken precedence, the investigation into solar radiation (SR) has become indispensable. Characterized by its intermittency and volatility, SR may experience considerable fluctuations, exerting a significant influence on energy supply security. Consequently, the precise prediction of SR has become imperative, particularly in the context of the potential proliferation of photovoltaic panels and the need for optimized energy management. Several works in the existing literature review the state of the art in SR prediction, focusing on trends identified using machine learning (ML) or deep learning (DL) techniques. However, there is a gap in the literature regarding the integration of optimization algorithms with ML and DL techniques for SR prediction. This systematic review addresses this gap by studying prediction models for SR that leverage metaheuristic optimization algorithms alongside artificial intelligence (AI) techniques, aiming primarily for maximum prediction accuracy. Metaheuristic algorithms such as Particle Swarm Optimization (PSO) and Genetic Algorithm (GA) have featured in 29% and 12.1% of the analyzed articles, respectively, while intelligent approaches like Convolutional Neural Networks (CNN), Extreme Learning Machine (ELM), and Multilayer Perceptron (MLP) have emerged as the predominant choices, collectively accounting for 43.9% of the studies. Analysis has encompassed studies examining SR across hourly, daily, and monthly intervals, with daily intervals representing 48.7% of the focus. Noteworthy variables including temperature, humidity, wind speed, and atmospheric pressure have surfaced, capturing proportions of 90%, 68.2%, 56%, and 41.4%, respectively, within the reviewed literature.

INDEX TERMS Forecasting, metaheuristics, optimization, solar radiation, time-series.

I. INTRODUCTION

Energy have played a crucial role in a growing society, with solar energy having emerged as a promising and sustainable alternative [1]. The increasing dependence on technology and the scarcity of traditional energy resources have driven the search for solutions, underscoring the importance of accurate prediction of global solar radiation (GSR) for efficient solar energy capture [2]. In many countries, especially those in development, the high cost of GSR measurement instruments

have hindered obtaining precise data. To address this challenge, various SR prediction models have been developed employing a variety of techniques, such as Machine Learning (ML), Deep Learning (DL), and optimization algorithms, utilizing meteorological data and satellite images. While ML models have demonstrated remarkable performance, it has been crucial to highlight that parametric search have become essential in constructing the underlying structure of any ML model. This had been achieved by incorporating optimization methods that had allowed fine-tuning and refining the model parameters, playing a crucial role in its effectiveness and accuracy [3].

The associate editor coordinating the review of this manuscript and approving it for publication was Bo Pu

© 2024 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Figura 7.3. Artículo publicado en revista indexada: Pérez-Rodríguez, S. A., Álvarez-Alvarado, J. M., Romero-González, J. A., Aviles, M., Eileen, M. R. A., Carlos, F. S., & Rodríguez-Reséndiz, J. (2024). Metaheuristic Algorithms for Solar Radiation Prediction: A Systematic Analysis. IEEE Access, doi: 10.1109/ACCESS.2024.3429073.

Bibliografía

- Abbas, F., Zhang, F., Ismail, M., Khan, G., Iqbal, J., Alrefaei, A. F., y Albeshr, M. F. (2023). Optimizing machine learning algorithms for landslide susceptibility mapping along the karakoram highway, gilgit baltistan, pakistan: A comparative study of baseline, bayesian, and metaheuristic hyperparameter optimization techniques. *Sensors*, *23*(15), 6843.
- Alkhatat, G., y Mehmood, R. (2021). A review and taxonomy of wind and solar energy forecasting methods based on deep learning. *Energy and AI*, *4*, 100060.
- Álvarez-Alvarado, J. M., Ríos-Moreno, J. G., Obregón-Biosca, S. A., Ronquillo-Lomelí, G., Ventura-Ramos Jr, E., y Trejo-Perea, M. (2021). Hybrid techniques to predict solar radiation using support vector machine and search optimization algorithms: a review. *Applied Sciences*, *11*(3), 1044.
- Bakali, S. E., Hamid, O., y Gheouany, S. (2023, 8). Day-ahead seasonal solar radiation prediction, combining vmd and stack algorithms. *Clean Energy*, *7*, 911-925. doi: 10.1093/ce/zkad025
- Bamisile, O., Oluwasanmi, A., Ejayi, C., Yimen, N., Obiora, S., y Huang, Q. (2022, 6). Comparison of machine learning and deep learning algorithms for hourly global/diffuse solar radiation predictions. *International Journal of Energy Research*, *46*, 10052-10073. doi: 10.1002/er.6529
- Besharat, F., Dehghan, A. A., y Faghieh, A. R. (2013). Empirical models for estimating global solar radiation: A review and case study. *Renewable and Sustainable Energy Reviews*, *21*, 798-821. Descargado de <https://www.sciencedirect.com/science/article/pii/S1364032112007484> doi: <https://doi.org/10.1016/j.rser.2012.12.043>
- Bischl, B., Binder, M., Lang, M., Pielok, T., Richter, J., Coors, S., ... others (2023). Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *13*(2), e1484.
- Bounoua, Z., Ouazzani Chahidi, L., y Mechaqrane, A. (2021). Estimation of daily global solar radiation using empirical and machine-learning methods: A case study of five moroccan locations. *Sustainable Materials and Technologies*, *28*, e00261. Descargado de <https://>

www.sciencedirect.com/science/article/pii/S2214993721000166 doi: <https://doi.org/10.1016/j.susmat.2021.e00261>

- Brockwell, P. J., y Davis, R. A. (2002). *Introduction to time series and forecasting*. Springer.
- Chaibi, M., Benghoulam, E. M., Tarik, L., Berrada, M., y El Hmaidi, A. (2022). Machine learning models based on random forest feature selection and bayesian optimization for predicting daily global solar radiation [Article]. *International Journal of Renewable Energy Development*, 11(1), 309 – 323. Descargado de <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85123176455&doi=10.14710%2fIJRED.2022.41451&partnerID=40&md5=f02e1446ccfaad5a99a6fb5f18a2de0d> (Cited by: 7; All Open Access, Gold Open Access) doi: 10.14710/IJRED.2022.41451
- Chaibi, M., Benghoulam, E. M., Tarik, L., Berrada, M., y Hmaidi, A. E. (2022). Machine learning models based on random forest feature selection and bayesian optimization for predicting daily global solar radiation. *International Journal of Renewable Energy Development*, 11, 309 – 323. Descargado de <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85123176455&doi=10.14710%2fIJRED.2022.41451&partnerID=40&md5=f02e1446ccfaad5a99a6fb5f18a2de0d> (Cited by: 10; All Open Access, Gold Open Access) doi: 10.14710/IJRED.2022.41451
- Chicco, D., Warrens, M. J., y Jurman, G. (2021). The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation. *Peerj computer science*, 7, e623.
- Dahmani, A., Ammi, Y., y Hanini, S. (2024). A novel non-linear model based on bootstrapped aggregated support vector machine for the prediction of hourly global solar radiation. *Smart Grids and Sustainable Energy*, 9. Descargado de <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85178247865&doi=10.1007%2fs40866-023-00179-w&partnerID=40&md5=855887600dbced3a534a5a803f998269> (Cited by: 0; All Open Access, Bronze Open Access) doi: 10.1007/s40866-023-00179-w
- Djeldjli, H., Benatiallah, D., Tanougast, C., y Benatiallah, A. (2023). Solar radiation forecasting based on ann, svm and a novel hybrid ffa-ann model: A case study of six cities south of algeria. *AIMS Energy*, 12, 62-83. doi: 10.3934/ENERGY.2024004
- Dong, Y., y Jiang, H. (2019). Global solar radiation forecasting using square root regularization-based ensemble [Article]. *Mathematical Problems in Engineering*, 2019. Descargado de <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85066324914&doi=10.1155%2f2019%2f9620945&partnerID=40&md5=123c962bd752c4dab4719714322ce2bb> (Cited by: 11; All Open Access, Gold Open Access, Green Open Access) doi: 10.1155/2019/9620945
- Elgeldawi, E., Sayed, A., Galal, A. R., y Zaki, A. M. (2021). Hyperparameter tuning for machine learning algorithms used for arabic sentiment analysis. En *Informatics* (Vol. 8, p. 79).

- El-Kenawy, E.-S. M., Mirjalili, S., Ghoneim, S. S. M., Eid, M. M., El-Said, M., Khan, Z. S., y Ibrahim, A. (2021). Advanced ensemble model for solar radiation forecasting using sine cosine algorithm and newton's laws. *IEEE Access*, 9, 115750–115765. Descargado de <http://dx.doi.org/10.1109/ACCESS.2021.3106233> doi: 10.1109/access.2021.3106233
- Ensafi, Y., Amin, S. H., Zhang, G., y Shah, B. (2022). Time-series forecasting of seasonal items sales using machine learning – a comparative analysis. *International Journal of Information Management Data Insights*, 2(1), 100058. Descargado de <https://www.sciencedirect.com/science/article/pii/S2667096822000027> doi: <https://doi.org/10.1016/j.jjime.2022.100058>
- Fan, C., Chen, M., Wang, X., Wang, J., y Huang, B. (2021). A review on data preprocessing techniques toward efficient and reliable knowledge discovery from building operational data. *Frontiers in Energy Research*, 9. Descargado de <https://www.frontiersin.org/articles/10.3389/fenrg.2021.652801> doi: 10.3389/fenrg.2021.652801
- Fan, J., Wu, L., Ma, X., Zhou, H., y Zhang, F. (2020, 1). Hybrid support vector machines with heuristic algorithms for prediction of daily diffuse solar radiation in air-polluted regions. *Renewable Energy*, 145, 2034-2045. doi: 10.1016/j.renene.2019.07.104
- Fan, J., Wu, L., Zhang, F., Cai, H., Wang, X., Lu, X., y Xiang, Y. (2018). Evaluating the effect of air pollution on global and diffuse solar radiation prediction using support vector machine modeling based on sunshine duration and air temperature. *Renewable and Sustainable Energy Reviews*, 94, 732-747. Descargado de <https://www.sciencedirect.com/science/article/pii/S1364032118304659> doi: <https://doi.org/10.1016/j.rser.2018.06.029>
- Gensler, A., Henze, J., Sick, B., y Raabe, N. (2016). Deep learning for solar power forecasting — an approach using autoencoder and lstm neural networks. En *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (p. 002858-002865). doi: 10.1109/SMC.2016.7844673
- Ghazvinian, H., Mousavi, S.-F., Karami, H., Farzin, S., Ehteram, M., Hossain, M. S., ... El-Shafie, A. (2019). Integrated support vector regression and an improved particle swarm optimization-based model for solar radiation prediction [Article]. *PLoS ONE*, 14(5). Descargado de <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85066505795&doi=10.1371%2fjournal.pone.0217634&partnerID=40&md5=261a81c04082f7656b172db8fbb7804c> (Cited by: 43; All Open Access, Gold Open Access, Green Open Access) doi: 10.1371/journal.pone.0217634
- Ghimire, S., Bhandari, B., Casillas-Pérez, D., Deo, R. C., y Salcedo-Sanz, S. (2022). Hybrid deep cnn-svr algorithm for solar radiation prediction problems in queensland, australia [Article]. *Engineering Applications of Artificial Intelligence*, 112. Descargado de <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85128175831&doi=10.1016%2fj.engappai.2022.104860&partnerID=40&md5=2048c39a31f83c0704ed1c350876a9e0>

(Cited by: 23; All Open Access, Green Open Access, Hybrid Gold Open Access) doi: 10.1016/j.engappai.2022.104860

Ghimire, S., Deo, R. C., Casillas-Pérez, D., y Salcedo-Sanz, S. (2022). Boosting solar radiation predictions with global climate models, observational predictors and hybrid deep-machine learning algorithms. *Applied Energy*, 316, 119063.

Ghimire, S., Deo, R. C., Casillas-Pérez, D., y Salcedo-Sanz, S. (2022, 6). Boosting solar radiation predictions with global climate models, observational predictors and hybrid deep-machine learning algorithms. *Applied Energy*, 316. doi: 10.1016/j.apenergy.2022.119063

Ghimire, S., Deo, R. C., Casillas-Pérez, D., Salcedo-Sanz, S., Sharma, E., y Ali, M. (2022). Deep learning cnn-lstm-mlp hybrid fusion model for feature optimizations and daily solar radiation prediction [Article]. *Measurement: Journal of the International Measurement Confederation*, 202. Descargado de <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85136630745&doi=10.1016%2fj.measurement.2022.111759&partnerID=40&md5=46708f4a08dcaf6d4ec21fea82f19e27> (Cited by: 17; All Open Access, Green Open Access, Hybrid Gold Open Access) doi: 10.1016/j.measurement.2022.111759

Ghimire, S., Deo, R. C., Raj, N., y Mi, J. (2019). Wavelet-based 3-phase hybrid svr model trained with satellite-derived predictors, particle swarm optimization and maximum overlap discrete wavelet transform for solar radiation prediction [Article]. *Renewable and Sustainable Energy Reviews*, 113. Descargado de <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85068217665&doi=10.1016%2fj.rser.2019.109247&partnerID=40&md5=082ff30f958637d84649a37100a18442> (Cited by: 59) doi: 10.1016/j.rser.2019.109247

Guermoui, M., Arrif, T., Belaid, A., Hassani, S., y Bailek, N. (2024, 3). Enhancing direct normal solar irradiation forecasting for heliostat field applications through a novel hybrid model. *Energy Conversion and Management*, 304. doi: 10.1016/j.enconman.2024.118189

Guermoui, M., Benkacali, S., Gairaa, K., Bouchouicha, K., Boulmaiz, T., y Boland, J. W. (2022, 2). A novel ensemble learning approach for hourly global solar radiation forecasting. *Neural Computing and Applications*, 34, 2983-3005. doi: 10.1007/s00521-021-06421-9

Guermoui, M., Gairaa, K., Boland, J., y Arrif, T. (2021, 4). A novel hybrid model for solar radiation forecasting using support vector machine and bee colony optimization algorithm: Review and case study. *Journal of Solar Energy Engineering, Transactions of the ASME*, 143. doi: 10.1115/1.4047852

Gürel, A. E., Ağbulut, Ü., Bakır, H., Ergün, A., y Yıldız, G. (2023). A state of art review on estimation of solar radiation with various models. *Heliyon*.

Han, C., Phillips, P. C., y Sul, D. (2011). Uniform asymptotic normality in stationary and unit root autoregression. *Econometric Theory*, 27(6), 1117–1151.

- Han, Z., Zhao, J., Leung, H., Ma, K. F., y Wang, W. (2021). A review of deep learning models for time series prediction. *IEEE Sensors Journal*, 21(6), 7833-7848. doi: 10.1109/JSEN.2019.2923982
- Hasan, S. S., Agee, Z. S., Tahir, B. S., y Zeebaree, S. R. (2023). Solar radiation prediction using satin bowerbird optimization with modified deep learning [Article]. *Computer Systems Science and Engineering*, 46(3), 3225 – 3238. Descargado de <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85158834421&doi=10.32604%2fcsse.2023.037434&partnerID=40&md5=d2d5982a0eb9a1af1bd349cad70404ce> (Cited by: 0; All Open Access, Hybrid Gold Open Access) doi: 10.32604/csse.2023.037434
- Hassan, M. A., Al-Ghussain, L., Khalil, A., y Kaseb, S. A. (2022). Self-calibrated hybrid weather forecasters for solar thermal and photovoltaic power plants. *Renewable Energy*, 188, 1120-1140. Descargado de <https://www.sciencedirect.com/science/article/pii/S0960148122002476> doi: <https://doi.org/10.1016/j.renene.2022.02.098>
- Hong, Y.-Y., y Martinez, J. J. F. (2024). Forecasting solar irradiation using convolutional long short-term memory and feature selection of data from neighboring locations. *Sustainable Energy, Grids and Networks*, 38, 101271. Descargado de <https://www.sciencedirect.com/science/article/pii/S2352467723002795> doi: <https://doi.org/10.1016/j.segan.2023.101271>
- Huang, H., Band, S. S., Karami, H., Ehteram, M., wing Chau, K., y Zhang, Q. (2022, 12). Solar radiation prediction using improved soft computing models for semi-arid, slightly-arid and humid climates. *Alexandria Engineering Journal*, 61, 10631-10657. doi: 10.1016/j.aej.2022.03.078
- Ikram, R. M. A., Dai, H.-L., Ewees, A. A., Shiri, J., Kisi, O., y Zounemat-Kermani, M. (2022). Application of improved version of multi verse optimizer algorithm for modeling solar radiation [Article]. *Energy Reports*, 8, 12063 – 12080. Descargado de <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85141872972&doi=10.1016%2fj.egyr.2022.09.015&partnerID=40&md5=cf0f6e915d9ae883084e34733ac4a275> (Cited by: 45; All Open Access, Gold Open Access) doi: 10.1016/j.egyr.2022.09.015
- Irshad, K., Islam, N., Gari, A. A., Algarni, S., Alqahtani, T., y Imteyaz, B. (2023, 6). Arithmetic optimization with hybrid deep learning algorithm based solar radiation prediction model. *Sustainable Energy Technologies and Assessments*, 57, 103165. Descargado de <https://linkinghub.elsevier.com/retrieve/pii/S2213138823001583> doi: 10.1016/j.seta.2023.103165
- Islam, M. M., Nagrial, M., Rizk, J., y Hellany, A. (2021). Solar radiation and wind speed forecasting using deep learning technique. Institute of Electrical and Electronics Engineers Inc. doi: 10.1109/CSDE53843.2021.9718372

- Jadidi, A., Menezes, R., Souza, N. D., y Lima, A. C. D. C. (2018, 10). A hybrid ga-mlpnn model for one-hour-ahead forecasting of the global horizontal irradiance in elizabeth city, north carolina. *Energies*, *11*. doi: 10.3390/en11102641
- Jayalakshmi, N. Y., Shankar, R., Subramaniam, U., Baranilingesan, I., Karthick, A., Stalin, B., ... Ghosh, A. (2021, 5). Novel multi-time scale deep learning algorithm for solar irradiance forecasting. *Energies*, *14*. doi: 10.3390/en14092404
- Jia, Y., Wang, H., Li, P., Su, Y., Wang, F., y Huo, S. (2023, 1). Particle swarm optimization algorithm with gaussian exponential model to predict daily and monthly global solar radiation in northeast china. *Environmental Science and Pollution Research*, *30*, 12769-12784. doi: 10.1007/s11356-022-22934-9
- Karaboga, D., y cols. (2005). *An idea based on honey bee swarm for numerical optimization* (Inf. Téc.). Technical report-tr06, Erciyes university, engineering faculty, computer
- Khafaga, D. S., Alhussan, A. A., Eid, M. M., y El-kenawy, E.-S. M. (2023). Improving solar radiation source efficiency using adaptive dynamic squirrel search optimization algorithm and long short-term memory [Article]. *Frontiers in Energy Research*, *11*. Descargado de <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85165185758&doi=10.3389%2ffenrg.2023.1164528&partnerID=40&md5=ff9967569ad49cbfa93d033f1d9f048a> (Cited by: 0; All Open Access, Gold Open Access) doi: 10.3389/fenrg.2023.1164528
- Khodayar, M., Mohammadi, S., Khodayar, M. E., Wang, J., y Liu, G. (2020, 4). Convolutional graph autoencoder: A generative deep neural network for probabilistic spatio-temporal solar irradiance forecasting. *IEEE Transactions on Sustainable Energy*, *11*, 571-583. doi: 10.1109/TSTE.2019.2897688
- Kohonen, T. (2012). *Self-organization and associative memory* (Vol. 8). Springer Science & Business Media.
- Kolambe, M. (2024, 04). Forecasting the future: A comprehensive review of time series prediction techniques. *Journal of Electrical Systems*, *20*, 575-586. doi: 10.52783/jes.1478
- Kumar, M., Namrata, K., y Kumar, N. (2023a). Data-driven hyperparameter optimized extreme gradient boosting machine learning model for solar radiation forecasting [Article]. *Advances in Electrical and Electronic Engineering*, *20*(4), 549 – 559. Descargado de <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85162649473&doi=10.15598%2faeee.v20i4.4650&partnerID=40&md5=61f12c105a82d1c3717d6e9065e96160> (Cited by: 1; All Open Access, Gold Open Access, Green Open Access) doi: 10.15598/aeee.v20i4.4650
- Kumar, M., Namrata, K., y Kumar, N. (2023b). Data-driven hyperparameter optimized extreme gradient boosting machine learning model for solar radiation forecasting. *Advances in Electrical and Electronic Engineering*, *20*, 549-559. doi: 10.15598/aeee.v20i4.4650

- Kumari, P., y Toshniwal, D. (2021). Deep learning models for solar irradiance forecasting: A comprehensive review. *Journal of Cleaner Production*, 318, 128566.
- Langazane, S. N., y Saha, A. K. (2022). Effects of particle swarm optimization and genetic algorithm control parameters on overcurrent relay selectivity and speed. *IEEE Access*, 10, 4550-4567. doi: 10.1109/ACCESS.2022.3140679
- Lara-Benítez, P., Carranza-García, M., y Riquelme, J. C. (2021). An experimental review on deep learning architectures for time series forecasting. *International Journal of Neural Systems*, 31(03), 2130001. Descargado de <https://doi.org/10.1142/S0129065721300011> (PMID: 33588711) doi: 10.1142/S0129065721300011
- Li, D. H., Chen, W., Li, S., y Lou, S. (2019). Estimation of hourly global solar radiation using multivariate adaptive regression spline (mars) – a case study of hong kong. *Energy*, 186, 115857. Descargado de <https://www.sciencedirect.com/science/article/pii/S0360544219315294> doi: <https://doi.org/10.1016/j.energy.2019.115857>
- Lim, B., y Zohren, S. (2021). Time-series forecasting with deep learning: a survey. *Philosophical Transactions of the Royal Society A*, 379(2194), 20200209.
- Malik, P., Gehlot, A., Singh, R., Gupta, L. R., y Thakur, A. K. (2022). A review on ann based model for solar radiation and wind speed prediction with real-time data. *Archives of Computational Methods in Engineering*, 29(5), 3183–3201.
- Mann, S. F., Pain, R., y Kirchhoff, M. D. (2022). Free energy: a user’s guide. *Biology & Philosophy*, 37(4), 33.
- Mfetoum, I. M., Ngoh, S. K., Molu, R. J. J., Nde Kenfack, B. F., Onguene, R., Naoussi, S. R. D., ... Berhanu, M. (2024). A multilayer perceptron neural network approach for optimizing solar irradiance forecasting in central africa with meteorological insights. *Scientific Reports*, 14(1), 3572.
- Mirbolouki, A., Heddami, S., Parmar, K. S., Trajkovic, S., Mehraein, M., y Kisi, O. (2022). Comparison of the advanced machine learning methods for better prediction accuracy of solar radiation using only temperature data: A case study. *International Journal of Energy Research*, 46, 2709 – 2736. Descargado de <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85116013923&doi=10.1002%2fer.7341&partnerID=40&md5=b4fa2cb1cc12ef6ef32bf50ea488a749> (Cited by: 6; All Open Access, Bronze Open Access) doi: 10.1002/er.7341
- Nadira, T., y Rustam, Z. (2018). Classification of cancer data using support vector machines with features selection method based on global artificial bee colony. En *Aip conference proceedings* (Vol. 2023).
- Nakagawa, S., y Schielzeth, H. (2013). A general and simple method for obtaining r2 from generalized linear mixed-effects models. *Methods in ecology and evolution*, 4(2), 133–142.

- Ng, C. S. W., Nait Amar, M., Jahanbani Ghahfarokhi, A., y Imsland, L. S. (2023). A survey on the application of machine learning and metaheuristic algorithms for intelligent proxy modeling in reservoir simulation. *Computers & Chemical Engineering*, 170, 108107. Descargado de <https://www.sciencedirect.com/science/article/pii/S0098135422004409> doi: <https://doi.org/10.1016/j.compchemeng.2022.108107>
- Pani, A. K., y Nayak, N. (2019). *Forecasting solar irradiance with weather classification and chaotic gravitational search algorithm based wavelet kernel extreme learning machine* (Vol. 9).
- Peng, T., Li, Y., Song, Z., Fu, Y., Nazir, M. S., y Zhang, C. (2023). Hybrid intelligent deep learning model for solar radiation forecasting using optimal variational mode decomposition and evolutionary deep belief network - online sequential extreme learning machine [Article]. *Journal of Building Engineering*, 76. Descargado de <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85164216027&doi=10.1016%2fj.jobe.2023.107227&partnerID=40&md5=6e8c6600b8e81b8d0899d84d8a6c77ce> (Cited by: 1) doi: 10.1016/j.jobe.2023.107227
- Prasad, R., Ali, M., Kwan, P., y Khan, H. (2019, 2). Designing a multi-stage multivariate empirical mode decomposition coupled with ant colony optimization and random forest model to forecast monthly solar radiation. *Applied Energy*, 236, 778-792. doi: 10.1016/j.apenergy.2018.12.034
- Premalatha, M., Naveen, C., y cols. (2018). Analysis of different combinations of meteorological parameters in predicting the horizontal global solar radiation with ann approach: A case study. *Renewable and Sustainable Energy Reviews*, 91, 248-258.
- Rajagukguk, R. A., Ramadhan, R. A., y Lee, H.-J. (2020). A review on deep learning models for forecasting time series data of solar irradiance and photovoltaic power. *Energies*, 13(24), 6623.
- Reddy, K. N., Thillaikarasi, M., Kumar, B. S., y Suresh, T. (2022). A novel elephant herd optimization model with a deep extreme learning machine for solar radiation prediction using weather forecasts [Article]. *Journal of Supercomputing*, 78(6), 8560 - 8576. Descargado de <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85122682242&doi=10.1007%2fs11227-021-04244-y&partnerID=40&md5=be2fd363ea7f690054dca39c40e95096> (Cited by: 3) doi: 10.1007/s11227-021-04244-y
- Sansine, V., Ortega, P., Hissel, D., y Hopuare, M. (2022, 11). Solar irradiance probabilistic forecasting using machine learning, metaheuristic models and numerical weather predictions. *Sustainability (Switzerland)*, 14. doi: 10.3390/su142215260
- Sharma¹, D., y Khurana, P. (2018). Solar power: Challenges, mission and potential of solar power in india. *Target*, 2010, 13.

- Sun, S., Wang, S., Zhang, G., y Zheng, J. (2018, 3). A decomposition-clustering-ensemble learning approach for solar radiation forecasting. *Solar Energy*, *163*, 189-199. doi: 10.1016/j.solener.2018.02.006
- Tao, H., Ewees, A. A., Al-Sulttani, A. O., Beyaztas, U., Hameed, M. M., Salih, S. Q., ... Yaseen, Z. M. (2021). Global solar radiation prediction over north dakota using air temperature: Development of novel hybrid intelligence model. *Energy Reports*, *7*, 136-157. Descargado de <https://www.sciencedirect.com/science/article/pii/S235248472031458X> doi: <https://doi.org/10.1016/j.egyr.2020.11.033>
- Torres, J. F., Hadjout, D., Sebaa, A., Martínez-Álvarez, F., y Troncoso, A. (2021). Deep learning for time series forecasting: a survey. *Big Data*, *9*(1), 3-21.
- Vaisakh, T., y Jayabarathi, R. (2022, 3). Analysis on intelligent machine learning enabled with meta-heuristic algorithms for solar irradiance prediction. *Evolutionary Intelligence*, *15*, 235-254. doi: 10.1007/s12065-020-00505-6
- Vasuki, A. (2020). *Nature-inspired optimization algorithms*. CRC Press.
- Voyant, C., Notton, G., Kalogirou, S., Nivet, M.-L., Paoli, C., Motte, F., y Fouilloy, A. (2017). Machine learning methods for solar radiation forecasting: A review. *Renewable energy*, *105*, 569-582.
- Wang, L., y Shi, J. (2021). A comprehensive application of machine learning techniques for short-term solar radiation prediction. *Applied Sciences*, *11*(13), 5808.
- Woldegiyorgis, T. A., Admasu, A., Benti, N. E., y Asfaw, A. A. (2022). A comparative evaluation of artificial neural network and sunshine based models in prediction of daily global solar radiation of lalibela, ethiopia. *Cogent Engineering*, *9*. doi: 10.1080/23311916.2021.1996871
- Wu, Z., Cui, N., Gong, D., Zhu, F., Li, Y., Xing, L., ... Zha, Y. (2023). Predicting daily global solar radiation in various climatic regions of china based on hybrid support vector machines with meta-heuristic algorithms [Article]. *Journal of Cleaner Production*, *385*. Descargado de <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85144627742&doi=10.1016%2fj.jclepro.2022.135589&partnerID=40&md5=a77f34c054f082d196a45ddc58c752f9> (Cited by: 5) doi: 10.1016/j.jclepro.2022.135589
- Xu, T., Sabzalian, M. H., Hammoud, A., Tahami, H., Gholami, A., y Lee, S. (2024). An innovative machine learning based on feed-forward artificial neural network and equilibrium optimization for predicting solar irradiance. *Scientific Reports*, *14*(1), 2170.
- Xue, X., y Zhou, H. (2019). Soft computing methods for predicting daily global solar radiation [Article]. *Numerical Heat Transfer, Part B: Fundamentals*, *76*(1), 18 - 31. Descargado de <https://www.scopus.com/inward/record.uri>

?eid=2-s2.0-85068799815&doi=10.1080%2f10407790.2019.1637629&partnerID=40&md5=bf24f90990766665435b87ef1ef16dcd (Cited by: 6) doi: 10.1080/10407790.2019.1637629

- Yang, L., y Shami, A. (2020). On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing*, 415, 295–316.
- Yu, X. (2023). Evaluating parallelized support vector regression and nearest neighbor regression with different input variations for estimating daily global solar radiation of the humid subtropical region in china. *International Journal of Low-Carbon Technologies*, 18, 95 – 110. Descargado de <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85160210613&doi=10.1093%2fijlct%2fctad005&partnerID=40&md5=f3944dc43b876a00ea9f7ed24f69f908> (Cited by: 0; All Open Access, Gold Open Access) doi: 10.1093/ijlct/ctad005
- Zeng, A., Chen, M., Zhang, L., y Xu, Q. (2023). Are transformers effective for time series forecasting? En *Proceedings of the aaai conference on artificial intelligence* (Vol. 37, pp. 11121–11128).
- Zhang, C., Hua, L., Ji, C., Shahzad Nazir, M., y Peng, T. (2022). An evolutionary robust solar radiation prediction model based on wt-ceemdan and iaso-optimized outlier robust extreme learning machine [Article]. *Applied Energy*, 322. Descargado de <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85132880447&doi=10.1016%2fj.apenergy.2022.119518&partnerID=40&md5=f18af0b02be8593d76c2cc945ae7e2f4> (Cited by: 19) doi: 10.1016/j.apenergy.2022.119518
- Zhang, Q., Tian, X., Zhang, P., Hou, L., Peng, Z., y Wang, G. (2022). Solar radiation prediction model for the yellow river basin with deep learning. *Agronomy*, 12(5), 1081.
- Zhang, X., y Wei, Z. (2019). A hybrid model based on principal component analysis, wavelet transform, and extreme learning machine optimized by bat algorithm for daily solar radiation forecasting [Article]. *Sustainability (Switzerland)*, 11(15). Descargado de <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85070458783&doi=10.3390%2fsu11154138&partnerID=40&md5=9b80eab2fe1626afe9739ac173f5f829> (Cited by: 20; All Open Access, Gold Open Access, Green Open Access) doi: 10.3390/su11154138
- Zhang, Y., Cui, N., Feng, Y., Gong, D., y Hu, X. (2019, 9). Comparison of bp, pso-bp and statistical models for predicting daily global solar radiation in arid northwest china. *Computers and Electronics in Agriculture*, 164. doi: 10.1016/j.compag.2019.104905
- Zhou, Y., Li, Y., Wang, D., y Liu, Y. (2023, 5). A multi-step ahead global solar radiation prediction method using an attention-based transformer model with an interpretable mechanism. *International Journal of Hydrogen Energy*, 48, 15317-15330. doi: 10.1016/j.ijhydene.2023.01.068

- Zhou, Y., Liu, Y., Wang, D., Liu, X., y Wang, Y. (2021a). A review on global solar radiation prediction with machine learning models in a comprehensive perspective. *Energy Conversion and Management*, 235, 113960. Descargado de <https://www.sciencedirect.com/science/article/pii/S0196890421001369> doi: <https://doi.org/10.1016/j.enconman.2021.113960>
- Zhou, Y., Liu, Y., Wang, D., Liu, X., y Wang, Y. (2021b). A review on global solar radiation prediction with machine learning models in a comprehensive perspective [Review]. *Energy Conversion and Management*, 235. Descargado de <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85102358454&doi=10.1016%2fj.enconman.2021.113960&partnerID=40&md5=c1de4beb8d33c6e6ee0efeeb826bf6b4> (Cited by: 115) doi: [10.1016/j.enconman.2021.113960](https://doi.org/10.1016/j.enconman.2021.113960)
- Zhu, T., Guo, Y., Li, Z., y Wang, C. (2021). Solar radiation prediction based on convolution neural network and long short-term memory. *Energies*, 14(24), 8498.

Índice de tablas

1.1. Trabajos donde se utilizan MLP para la predicción de SR.	6
1.2. Trabajos donde se utilizan CNN para la predicción de SR.	9
1.3. Los siguientes artículos emplean algunos modelos optimizados de redes neuronales recurrentes.	10
1.4. Los siguientes artículos emplean algunos modelos optimizados de ML.	12
4.1. Sensores utilizados en el estudio	30
4.2. Datos meteorológicos usados en el estudio.	32
4.3. Hiperparámetros usados en los modelos ML	34
4.4. Hiperparámetros usados en los modelos DL	36
4.5. Tabla de parámetros de capas ocultas y variables	36
5.1. Comparación de trabajos del estado del arte con la propuesta Selector ABC-MLP	43
5.2. Comparativa entre modelos de machine learning y deep learning.	44
5.3. Ubicación EMAs.	48
5.4. Comparativa entre resultados del modelo ABC-MLP con selector de características en diferentes EMAs.	49

Índice de figuras

1.1.	Número de trabajos que han utilizado diferentes algoritmos de IA para la predicción de la radiación solar global en Scopus.	2
1.2.	Variables que inciden en la predicción de la radiación solar global.	5
1.3.	Informe de publicaciones desde 2018 hasta 2024, presentando los documentos publicados por año en Scopus.	5
2.1.	Técnicas de preprocesamiento de datos.	20
2.2.	Horizontes de predicción para la gestión de energía solar en una red eléctrica.	22
2.3.	Diagrama de flujo del algoritmo de la colonia artificial de abejas (Vasuki, 2020).	24
4.1.	Diagrama de metodología propuesta.	28
4.2.	Cronograma de actividades.	29
4.3.	Valores de la radiación solar global (GSR) promedio en función del tiempo para días soleados: a) 10 de enero de 2018, b) 4 de enero de 2018.	31
4.4.	Valores de la radiación solar global (GSR) promedio en función del tiempo para días nublados: a) 12 de mayo de 2018, b) 10 de junio de 2018.	31
4.5.	Matriz de correlación de las variables climatológicas.	33
4.6.	Tiempo de ejecución promedio en función del tamaño de la colonia.	37
4.7.	Relación entre el tiempo de ejecución y la función objetivo según el tamaño de la colonia.	37
4.8.	Mapa de calor de la función objetivo promedio en función del tamaño de la colonia y el límite.	38
5.1.	El número de trabajos que utilizan los diferentes algoritmos de optimización.	40
5.2.	El número de estudios que utilizan los diferentes modelos de IA revisados en la literatura.	41
5.3.	El número de trabajos que utilizan los diferentes horizontes temporales revisados en la literatura.	41
5.4.	El número de estudios que utilizan las diferentes variables meteorológicas.	42
5.5.	Gráfico de regresión que ilustra los valores predichos versus reales del modelo RF en a) datos de entrenamiento, b) datos de prueba.	45

5.6.	Gráfico de regresión que ilustra los valores predichos versus reales del modelo SVR con kernel lineal en a) datos de entrenamiento, b) datos de prueba. . . .	45
5.7.	Gráfico de regresión que ilustra los valores predichos versus reales del modelo SVR con kernel polinómico en a) datos de entrenamiento, b) datos de prueba. . . .	45
5.8.	Gráfico de regresión que ilustra los valores predichos versus reales del modelo SVR con kernel de función de base radial en a) datos de entrenamiento, b) datos de prueba.	46
5.9.	Gráfico de regresión que ilustra los valores predichos versus reales del modelo MLP en a) datos de entrenamiento, b) datos de prueba.	46
5.10.	Gráfico de regresión que ilustra los valores predichos versus reales del modelo ABC-MLP en a) datos de entrenamiento, b) datos de prueba.	46
5.11.	Gráfico de regresión que ilustra los valores predichos versus reales del modelo LSTM en a) datos de entrenamiento, b) datos de prueba.	47
5.12.	Gráfico de regresión que ilustra los valores predichos versus reales del modelo ABC-MLP con selección de características en a) datos de entrenamiento, b) datos de prueba.	47
5.13.	Gráfico con las funciones de pérdida de entrenamiento y validación del modelo propuesto Selector ABC-MLP a) y el modelo MLP b).	48
5.14.	Ubicación de EMAs cercanas a la estación meteorológica UAQ Aeropuerto.	49
5.15.	Estación Huichapan	50
5.16.	Estación Huimilpan	50
5.17.	Estación Presa Allende	50
5.18.	Estación UAQ Aeropuerto	50
5.19.	Gráficos de regresión que ilustran los valores predichos en diferentes estaciones.	50
7.1.	Constancia de idioma.	53
7.2.	Participación en CONiiN XX.	54
7.3.	Artículo publicado en revista indexada: Pérez-Rodríguez, S. A., Álvarez-Alvarado, J. M., Romero-González, J. A., Aviles, M., Eileen, M. R. A., Carlos, F. S., & Rodríguez-Reséndiz, J. (2024). Metaheuristic Algorithms for Solar Radiation Prediction: A Systematic Analysis. IEEE Access, doi: 10.1109/ACCESS.2024.3429073.	55