TESLA Revista Científica, ISSN 2796-9320
Vol. 6 Núm. 2 (2026), e772

Área: Ingeniería
Artículo de Revisión Sistemática

Redes neuronales profundas para el pronóstico de generación fotovoltaica: revisión sistemática de arquitecturas, validación y generalización

Deep Neural Networks for Photovoltaic Generation Forecasting: A Systematic Review of Architectures, Validation, and Generalization

Miguel Ángel Tisalema Sisa [1][0009-0001-7650-7436] Gabriela Elizabeth Revelo Salgado [1][0000-0001-5238-223X]

[1] Universidad Politécnica Estatal del Carchi (UPEC), Ecuador.
miguel.tisalema@upec.edu.ec gabriela.revelo@upec.edu.ec

CITA EN APA:

Tisalema Sisa, M. Á., & Revelo Salgado, G. E. (2026). Redes neuronales profundas para el pronóstico de generación fotovoltaica: revisión sistemática de arquitecturas, validación y generalización. *Tesla Revista Científica*, *6*(2). https://tesla.puertomaderoeditorial.com.ar/index.php/tesla/article/view/772

Recibido: 2026-11-10
Aceptado: 2026-11-19
Publicado: 2026-11-24

TESLA
Revista Científica
ISSN: 2796-9320

Los contenidos de este artículo están bajo una licencia de Creative Commons Attribution 4.0 International (CC BY 4.0)
Los autores conservan los derechos morales y patrimoniales de sus obras.
The contents of this article are under a Creative Commons Attribution 4.0 International (CC BY 4.0) license. The authors retain the moral and patrimonial rights of their works.

Resumen: Introducción: Las redes neuronales profundas se han aplicado ampliamente al pronóstico de generación fotovoltaica; sin embargo, la heterogeneidad de datos, horizontes, variables y estrategias de validación dificulta la comparación directa entre arquitecturas y limita la generalización de los resultados. Objetivo: Sintetizar críticamente la evidencia sobre arquitecturas de aprendizaje profundo aplicadas al pronóstico de generación fotovoltaica, considerando los horizontes de predicción, las variables de entrada, las estrategias de validación, el riesgo de fuga de información y la evidencia de generalización. Métodos: Se realizó una revisión sistemática conforme a PRISMA 2020 en Scopus, IEEE Xplore y Web of Science, con estudios publicados entre 2019 y 2026. Se identificaron 2.130 registros; tras la deduplicación y el proceso de selección, 314 estudios fueron incluidos en la síntesis. La evidencia se organizó según arquitectura, horizonte, variables y fuerza de validación. Resultados: El corpus estuvo dominado por estudios de forecasting y solo 2 de 314 abordaron la optimización como objetivo primario. Nueve estudios presentaron validación temporal independiente o externa y seis reunieron criterios de evidencia prioritaria. No se identificó una arquitectura universalmente superior; las ventajas observadas dependieron del horizonte, los datos y el protocolo de validación. Conclusiones: La credibilidad del pronóstico fotovoltaico depende principalmente de la separación temporal adecuada, el control del data leakage, la comparabilidad de métricas y la validación externa. La evidencia internacional puede orientar nuevas aplicaciones, pero requiere validación específica en cada contexto.

Palabras Clave: aprendizaje profundo; pronóstico de potencia fotovoltaica; LSTM; Transformer; validación de series temporales; revisión sistemática.

Abstract: Introduction: Deep neural networks have been widely applied to photovoltaic generation forecasting; however, heterogeneity in datasets, forecasting horizons, input variables, and validation strategies limits direct comparisons across architectures and constrains the generalizability of reported results. Objective: To critically synthesize the evidence on deep learning architectures applied to photovoltaic generation forecasting, considering forecasting horizons, input variables, validation strategies, risk of information leakage, and evidence of generalization. Methods: A systematic review following PRISMA 2020 was conducted in Scopus, IEEE Xplore, and Web of Science, including studies published between 2019 and 2026. A total of 2,130 records were identified; after deduplication and study selection, 314 studies were included in the synthesis. Evidence was organized according to architecture, forecasting horizon, input variables, and validation strength. Results: The corpus was dominated by forecasting studies, while only 2 of 314 studies addressed optimization as a primary objective. Nine studies reported independent temporal or external validation, and six met priority-evidence criteria. No architecture was found to be universally superior; reported advantages depended on forecasting horizon, dataset characteristics, and validation protocol. Conclusions: The credibility of photovoltaic forecasting depends mainly on appropriate temporal separation, control of data leakage, comparability of performance metrics, and external validation. International evidence can guide new applications, but model performance should be validated within each specific context.

Keywords: deep learning; photovoltaic power forecasting; LSTM; Transformer; time series validation; systematic review.

1 INTRODUCCIÓN

La generación fotovoltaica depende de la irradiancia solar, la temperatura, la nubosidad y otras variables atmosféricas que introducen no linealidad, estacionalidad y fluctuaciones abruptas en la potencia producida (Husein et al., 2024). La variabilidad de estas condiciones, tanto a lo largo del día como entre estaciones, dificulta la anticipación del comportamiento de los sistemas fotovoltaicos, especialmente cuando se producen cambios rápidos en las condiciones meteorológicas. Estas características representan un desafío para la programación operativa y la integración de la generación solar en las redes eléctricas, particularmente cuando el horizonte de pronóstico comprende minutos, horas o un día de anticipación. Las diferencias entre la generación prevista y la observada pueden afectar la planificación de los recursos energéticos, la gestión del almacenamiento y las decisiones relacionadas con el equilibrio entre oferta y demanda. En este contexto, la precisión del pronóstico constituye un componente importante para apoyar las decisiones de operación, almacenamiento, control y gestión energética, así como para facilitar la integración de la energía fotovoltaica en los sistemas eléctricos (Wedhasari & Castro, 2026).

Los modelos de aprendizaje profundo permiten representar relaciones temporales y no lineales sin depender exclusivamente de supuestos paramétricos, lo que ha favorecido su aplicación al pronóstico de la generación fotovoltaica. Entre las arquitecturas más utilizadas se encuentran las redes de memoria a largo y corto plazo (LSTM) y las unidades recurrentes controladas (GRU), diseñadas para modelar dependencias temporales y reconocer patrones en datos secuenciales. Por su parte, las redes neuronales convolucionales (CNN) permiten extraer características locales de las series temporales y de otras representaciones de los datos, mientras que los mecanismos de atención y las arquitecturas Transformer ofrecen alternativas para modelar relaciones entre observaciones situadas en diferentes momentos de la secuencia (Rajagukguk et al., 2020; Yu et al., 2024). A estas aproximaciones se suman modelos generativos, redes neuronales de grafos y arquitecturas híbridas que combinan diferentes estructuras de aprendizaje profundo con técnicas de descomposición de señales u optimizadores metaheurísticos. Estas combinaciones buscan aprovechar las capacidades complementarias de los modelos para representar la variabilidad y las relaciones complejas que caracterizan la producción fotovoltaica (Husein et al., 2024; Yu et al., 2024). Entre las primeras aplicaciones se encuentran el uso de LSTM para el pronóstico en seguidores solares activos (Kasburg & Stefenon, 2019), el pronóstico horario a partir de información meteorológica (Lee & Kim, 2019) y los esquemas LSTM-BP que incorporan múltiples variables atmosféricas (Hua et al., 2019). Estos antecedentes muestran la diversidad de enfoques desarrollados para abordar el pronóstico fotovoltaico, aunque la elección de una arquitectura debe considerar las características de los datos disponibles, el horizonte de predicción y las condiciones de evaluación.

La evolución reciente de los modelos de aprendizaje profundo para el pronóstico fotovoltaico se orienta a capturar dependencias espacio-temporales, cuantificar la incertidumbre de las predicciones y mejorar la capacidad de generalización entre instalaciones con diferentes condiciones ambientales. En esta

línea se encuentran modelos LSTM con mecanismos de atención y optimización bayesiana (Yang et al., 2019), redes CNN multientrada (Huang & Kuo, 2019), arquitecturas Transformer para horizontes ultracortos (Tian et al., 2022), enfoques de transferencia multitarea (Schreiber & Sick, 2022) y modelos probabilísticos o híbridos que permiten obtener intervalos de predicción además de estimaciones puntuales (Gu et al., 2023; Herrera-Casanova et al., 2024). Estas aproximaciones responden a diferentes necesidades del pronóstico, como identificar patrones temporales complejos, integrar múltiples fuentes de información meteorológica y representar la incertidumbre asociada con las fluctuaciones de la generación solar. No obstante, la incorporación de arquitecturas más complejas no garantiza por sí misma un mejor desempeño predictivo, ya que los resultados también dependen de la calidad de los datos, las estrategias de preprocesamiento y las condiciones de entrenamiento y evaluación (Khouili et al., 2025). Por otra parte, las diferencias climáticas, estacionales y geográficas entre instalaciones pueden limitar la aplicación directa de modelos desarrollados en otros contextos, lo que ha motivado el estudio de estrategias de transferencia y adaptación entre sitios (Bak et al., 2025). Estas consideraciones muestran que la evolución de las arquitecturas debe examinarse junto con las condiciones en las que se evalúa su desempeño y la evidencia disponible sobre su capacidad de generalización.

Las revisiones publicadas sobre el pronóstico fotovoltaico mediante inteligencia artificial y aprendizaje profundo han documentado una amplia variedad de arquitecturas, horizontes de predicción, variables de entrada y métricas de evaluación (Mellit et al., 2020; Rajagukguk et al., 2020; Başaran et al., 2020; Iheanetu, 2022; Wu et al., 2022; Chandel et al., 2023; Husein et al., 2024; Yu et al., 2024). Sin embargo, esta diversidad también dificulta establecer comparaciones entre los resultados, ya que los modelos suelen evaluarse con conjuntos de datos, períodos de observación y procedimientos metodológicos diferentes. Incluso cuando dos estudios utilizan las mismas métricas, como el error cuadrático medio (RMSE) o el error absoluto medio (MAE), sus resultados no son directamente comparables si difieren las características de los datos o las condiciones de evaluación (Khouili et al., 2025). Otro aspecto relevante es la forma en que se dividen los datos para entrenar y evaluar los modelos. En el pronóstico de series temporales, las particiones deben respetar el orden cronológico, pues el uso de información futura durante el entrenamiento o el preprocesamiento puede producir estimaciones demasiado optimistas del desempeño predictivo (Sesay, 2026). A estas dificultades se añade la limitada información que algunos estudios ofrecen sobre el comportamiento de los modelos fuera de las instalaciones o períodos utilizados para su desarrollo. Por ello, además de conocer qué arquitecturas se han utilizado y qué resultados reportan, interesa examinar cómo se validaron esos resultados, qué riesgos de fuga de información presentan los procedimientos empleados y hasta qué punto existe evidencia que permita valorar su generalización a otros contextos.

Aunque la literatura disponible muestra avances importantes en el pronóstico de la generación fotovoltaica mediante aprendizaje profundo, las diferencias en los datos utilizados, los horizontes de predicción y los procedimientos de evaluación dificultan establecer qué resultados pueden considerarse

suficientemente sólidos y reproducibles. La capacidad de un modelo para obtener errores bajos en un conjunto de datos no garantiza que mantenga ese desempeño en otros períodos, instalaciones o condiciones climáticas, por lo que resulta necesario distinguir entre la precisión reportada y la evidencia disponible sobre su generalización (Bak et al., 2025; Husein et al., 2024). A partir de estas consideraciones, la pregunta principal de investigación fue: ¿Qué arquitecturas de aprendizaje profundo se han utilizado para el pronóstico de la generación fotovoltaica y en qué condiciones de datos, horizonte y validación se ha evaluado su desempeño?

El objetivo de esta revisión fue sintetizar críticamente la evidencia sobre las arquitecturas de aprendizaje profundo aplicadas al pronóstico fotovoltaico, considerando los horizontes de predicción, las variables de entrada, las estrategias de validación, el riesgo de fuga de información y la evidencia de generalización.

2 METODOLOGÍA

2.1 Diseño y guía de reporte

Se realizó una revisión sistemática de la literatura siguiendo las recomendaciones de PRISMA 2020 (Page et al., 2021). El proceso de identificación, deduplicación, cribado, evaluación de elegibilidad, extracción de información y síntesis se documentó mediante matrices de seguimiento que permiten reconstruir las decisiones adoptadas en cada etapa. El protocolo de trabajo no fue registrado previamente en un repositorio externo, por lo que no existe un registro público prospectivo de los procedimientos establecidos antes de iniciar la revisión. Esta circunstancia se reconoce como una limitación para la transparencia y la verificación de posibles modificaciones metodológicas durante el desarrollo del estudio.

2.2 Fuentes de información y estrategia de búsqueda

La búsqueda bibliográfica se realizó en Scopus, IEEE Xplore y Web of Science Core Collection. Las consultas se efectuaron hasta agosto de 2026 y se consideraron artículos de revista publicados entre 2019 y 2026, según los registros disponibles en cada base de datos. La búsqueda se restringió a publicaciones en inglés, sin establecer limitaciones relacionadas con el acceso abierto. Se combinaron términos asociados con el aprendizaje profundo, como *deep learning*, *deep neural networks*, CNN, RNN, LSTM, GRU y Transformer, con expresiones relacionadas con los sistemas fotovoltaicos y el pronóstico o la optimización de potencia, energía, generación o producción. Se recuperaron 749 registros de Scopus, 553 de IEEE Xplore y 828 de Web of Science, que sumaron 2.130 registros antes de eliminar duplicados.

La estructura general de búsqueda combinó tres grupos de términos: (deep learning OR deep neural network OR CNN OR RNN OR LSTM OR GRU OR Transformer) AND (photovoltaic OR PV) AND (forecast* OR predict* OR optimiz*).

Esta estructura se adaptó a la sintaxis y los campos de búsqueda de cada base de datos, incluidos TITLE, TITLE-ABS-KEY y TS, según correspondiera, y se aplicaron los filtros de año, idioma y tipo documental. Las ecuaciones completas utilizadas en Scopus, IEEE Xplore y Web of Science, con sus respectivos campos, operadores y restricciones, se presentan en la Tabla suplementaria S2.

2.3 Deduplicación y cribado

Los registros recuperados se integraron en una base maestra que conservó la información sobre su procedencia. La identificación de duplicados se realizó mediante el DOI y el título normalizado. Tras eliminar 669 registros duplicados, quedaron 1.461 registros únicos para el cribado por título y resumen. En esta etapa se aplicaron criterios previamente definidos de inclusión y exclusión, registrando el código correspondiente y la justificación de cada decisión. Se excluyeron 417 registros y los 1.044 restantes pasaron a la etapa de recuperación de texto completo y evaluación de elegibilidad.

Para el cribado se establecieron cuatro criterios de inclusión: IN1, artículos de revista publicados en inglés e indexados dentro del período 2019–2026; IN2, estudios relacionados directamente con sistemas fotovoltaicos; IN3, utilización del aprendizaje profundo o de redes neuronales profundas como método central o integrado; e IN4, estudios cuyo objetivo incluyera el pronóstico, la predicción o la optimización de la potencia, energía, generación o producción fotovoltaica.

2.4 Elegibilidad a texto completo

La evaluación de elegibilidad se realizó mediante cinco criterios: FT1, estudio primario o empírico con datos de sistemas fotovoltaicos; FT2, utilización de una arquitectura de aprendizaje profundo como método central o integrado; FT3, análisis de potencia, energía, generación o producción fotovoltaica como variable de resultado; FT4, presentación de resultados cuantitativos de desempeño predictivo u optimización; y FT5, disponibilidad de información metodológica suficiente para la extracción y síntesis de los resultados.

De los 1.044 informes que pasaron a la etapa de recuperación y elegibilidad, 351 llegaron a evaluarse mediante el texto completo o la información documental disponible. De estos, 37 fueron excluidos por no cumplir los criterios establecidos y 314 se incluyeron en la síntesis.

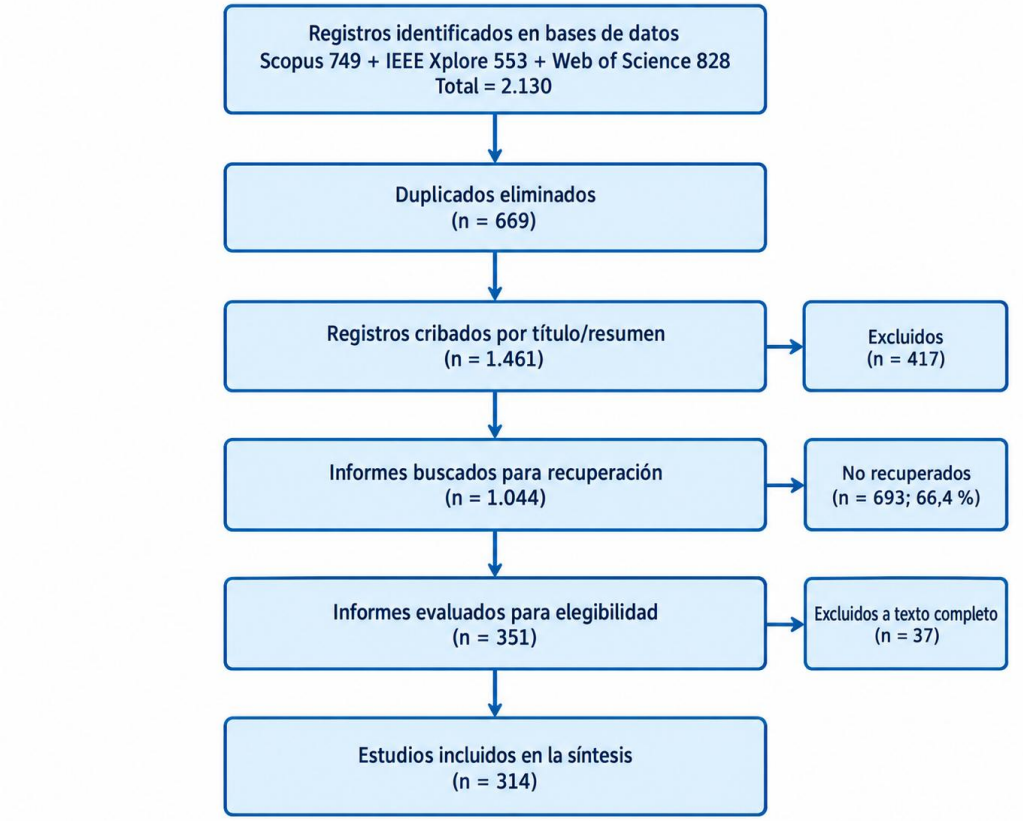

Figura 1. Diagrama de flujo PRISMA 2020 reconciliado de la revisión sistemática.

Nota. La trazabilidad disponible no distingue de forma exhaustiva entre no recuperación documental y ausencia de evaluación completa; por ello, los 693 casos se interpretan de forma conservadora como informes que no llegaron a evaluación a texto completo.

El cribado por título y resumen, la evaluación de elegibilidad y la extracción inicial de datos estuvieron a cargo del primer autor. Posteriormente, el segundo autor verificó las decisiones de inclusión, la información extraída y la clasificación metodológica V0–V4. Las discrepancias identificadas durante esta verificación se resolvieron mediante la revisión conjunta de los estudios correspondientes.

2.5 Extracción de datos

La matriz final de extracción incluyó 314 estudios y 28 campos principales, que recogieron información bibliográfica, ubicación geográfica, características del sistema fotovoltaico, fuente y período de los datos, variables de entrada, variable objetivo, horizonte de predicción, arquitectura de aprendizaje profundo, hiperparámetros, procedimientos de entrenamiento y validación, modelos de comparación, métricas de desempeño (RMSE, MAE, MAPE y R^2, cuando correspondía), resultados principales y limitaciones reportadas. También se registró información relacionada con el posible riesgo de fuga de información, la disponibilidad de código y datos, y la evidencia de generalización o aplicación de los modelos en nuevos contextos. Cuando la información no estuvo disponible o no pudo verificarse a partir de las fuentes consultadas, se registró como no especificada o pendiente de verificación, sin completar los campos mediante inferencias.

Para facilitar la identificación y consulta de los estudios incluidos, la Tabla suplementaria S1 presenta el corpus bibliográfico de la revisión, mientras que la Tabla S2 contiene las estrategias de búsqueda

utilizadas en cada base de datos. La Tabla S3 proporciona las referencias que permiten identificar individualmente los estudios analizados. El registro RS1461 se recuperó dentro del período de búsqueda establecido, aunque posteriormente recibió una asignación bibliográfica correspondiente a 2027. Se mantuvo en el corpus y se identificó expresamente esta particularidad en el material suplementario.

2.6 Solidez metodológica y fuerza de validación

La evaluación metodológica distinguió entre la información disponible en los artículos y la solidez de los procedimientos utilizados para estimar el desempeño predictivo. La completitud documental se consideró un indicador de transparencia del reporte, pero no se utilizó por sí sola como medida de calidad científica. Para examinar los estudios se desarrolló un marco adaptado al pronóstico fotovoltaico, tomando como referencia los dominios de PROBAST+AI relacionados con los datos, los predictores, la variable de resultado y el análisis, e incorporando consideraciones específicas de la validación de series temporales (Moons et al., 2025). Esta adaptación se utilizó como orientación conceptual y no constituye una aplicación formal de PROBAST+AI, instrumento desarrollado para evaluar modelos de predicción en el ámbito de la salud.

La evaluación comprendió cuatro dimensiones: D1, descripción del conjunto de datos, período de observación, resolución temporal, ubicación y representatividad; D2, disponibilidad de los predictores en el momento del pronóstico y posibles fuentes de fuga de información durante la normalización, selección de variables o descomposición de señales; D3, definición de la variable objetivo y del horizonte de predicción; y D4, independencia de los datos de prueba, procedimientos de validación, comparadores, métricas de desempeño y evaluación externa. Estas dimensiones permitieron examinar las condiciones bajo las cuales se obtuvieron los resultados, sin asumir que la ausencia de información equivalía necesariamente a una deficiencia del procedimiento empleado.

La fuerza de validación se clasificó en cinco categorías: V4, evaluación externa entre instalaciones o sitios diferentes; V3, validación temporal independiente o mediante origen móvil (*rolling-origin*); V2, partición de entrenamiento y prueba o validación cruzada sin información suficiente sobre el respeto del orden cronológico; V1, división aleatoria de los datos; y V0, procedimiento de validación no especificado o insuficientemente descrito. La clasificación se basó en la evidencia metodológica disponible en cada estudio. Cuando no fue posible establecer cómo se realizó la validación, se asignó la categoría V0, sin interpretar automáticamente la falta de información como evidencia de un riesgo alto de sesgo.

2.7 Estrategia de síntesis

La síntesis de los resultados consideró las diferencias en la solidez metodológica de los estudios, evitando otorgar el mismo peso interpretativo a resultados obtenidos mediante procedimientos de validación distintos. Para ello, la evidencia se organizó en tres niveles: A, evidencia prioritaria, correspondiente a estudios clasificados como V3 o V4, con resultados cuantitativos extraíbles y sin señales

explícitas de alto riesgo metodológico; B, evidencia de apoyo, integrada por estudios V2 con evaluación cuantitativa; y C, evidencia contextual o con incertidumbre metodológica, que comprendió estudios V0 o V1 y aquellos con información insuficiente para valorar adecuadamente sus resultados.

Las comparaciones se organizaron considerando la arquitectura de aprendizaje profundo, el horizonte de predicción, las variables de entrada y la fuerza de validación. También se examinaron las características de los procedimientos de evaluación y la evidencia disponible sobre robustez y generalización. Debido a las diferencias entre estudios en las escalas de medición, las capacidades instaladas, los horizontes de pronóstico, los conjuntos de datos y los procedimientos de normalización, los valores de RMSE, MAE y MAPE no se combinaron en una estimación global. Por estas razones, se realizó una síntesis descriptiva y crítica de los resultados, sin metaanálisis de desempeño predictivo.

2.8 Consideraciones para la generalización y aplicabilidad en nuevos contextos

La posibilidad de aplicar los resultados a nuevos contextos se examinó de manera cualitativa, considerando la disponibilidad de las variables de entrada, la correspondencia de los horizontes de predicción, las características de los sistemas fotovoltaicos y la solidez de los procedimientos de validación. Se distinguió entre la evidencia de generalización temporal, relacionada con el desempeño en períodos diferentes a los utilizados durante el entrenamiento, y la generalización geográfica, referida a la evaluación de modelos en instalaciones o ubicaciones distintas. Este análisis tuvo un propósito interpretativo y no se utilizó como una escala formal de validación externa.

2.9 Sesgo de no recuperación y disponibilidad documental

De los 1.044 informes que pasaron a la etapa de recuperación y evaluación de elegibilidad, 693 (66,4 %) no llegaron a ser evaluados. Esta proporción constituye una posible fuente de sesgo de selección, ya que no puede descartarse que los estudios no evaluados presentaran características o resultados diferentes de los incluidos en la síntesis. La matriz de trazabilidad no permite establecer de manera exhaustiva la causa correspondiente a cada uno de estos 693 casos, por lo que no se asignaron retrospectivamente motivos específicos de no recuperación o exclusión. Esta limitación se documentó expresamente y se realizó un análisis exploratorio basado en los metadatos y la procedencia de los registros disponibles, sin considerarlo una reconstrucción completa de los informes no evaluados.

3 RESULTADOS

3.1 Fuerza de la evidencia y riesgo metodológico

La evaluación metodológica permitió distinguir entre el número de estudios disponibles y la solidez de la evidencia que aportan. De los 314 estudios incluidos, únicamente nueve (2,9 %) documentaron procedimientos de validación temporal independiente o externa entre instalaciones (V3–V4). De ellos, seis cumplieron simultáneamente los criterios establecidos para la evidencia prioritaria (Tier A): validación

sólida, resultados cuantitativos extraíbles y ausencia de señales explícitas de alto riesgo metodológico. Los tres estudios restantes se mantuvieron como evidencia contextual, debido a que la información disponible no permitió reconstruir con suficiente precisión determinados aspectos del protocolo de evaluación, las métricas reportadas o la independencia de los datos de prueba.

La evaluación del riesgo metodológico mostró que 300 estudios quedaron clasificados como «no informado» y 14 presentaron «algunas preocupaciones». La categoría «no informado» no implica que los procedimientos empleados fueran incorrectos, sino que la documentación disponible resultó insuficiente para emitir un juicio fundamentado. Esta distinción es especialmente importante en el pronóstico fotovoltaico, donde la forma de separar los datos de entrenamiento y prueba, el tratamiento de las variables meteorológicas y la disponibilidad de información futura pueden influir en las estimaciones de desempeño. Por tanto, aunque los 314 estudios forman parte de la síntesis, sus resultados no ofrecen el mismo grado de respaldo metodológico para establecer comparaciones entre arquitecturas o valorar su capacidad de generalización.

Tabla 1. Fuerza de validación en los 314 estudios.

Nivel	Definición	n	Lectura
V4	Validación externa entre sitios	6	Mayor capacidad para juzgar generalización.
V3	Validación temporal independiente u origen móvil	3	Respeta la dirección temporal.
V2	Holdout o validación cruzada sin cronología clara	11	Evidencia de apoyo.
V1	División aleatoria	0	Sin casos clasificados exclusivamente como V1 en la matriz final.
V0	Validación insuficientemente documentada	294	No permite ponderación fuerte del error.
Total		314	

3.2 Arquitectura × validación

La distribución de los 314 estudios muestra un predominio de las arquitecturas híbridas y recurrentes, que representan el 40,4 % y el 32,2 % del corpus, respectivamente. Los modelos basados en Transformer o mecanismos de atención constituyeron el 10,8 %, mientras que las arquitecturas exclusivamente convolucionales tuvieron una presencia menor (1,0 %). Sin embargo, al considerar la fuerza de validación, se observan diferencias importantes entre las familias arquitectónicas (Tabla 2).

En los modelos recurrentes, únicamente un estudio presentó validación temporal independiente (V3) y cinco se clasificaron como V2, mientras que los 95 restantes no documentaron con suficiente claridad el procedimiento de validación (V0). Entre los estudios basados en Transformer o atención, dos incorporaron validación externa entre sitios (V4), uno se clasificó como V2 y 31 como V0. Por su parte, las arquitecturas híbridas concentraron cuatro estudios con validación externa (V4) y uno con validación temporal independiente (V3), aunque 118 de los 127 trabajos de esta familia quedaron clasificados como V0.

Estos resultados muestran que las arquitecturas híbridas y recurrentes concentran buena parte de la producción científica, pero la mayoría de sus aplicaciones no aporta información suficiente para valorar con precisión la independencia de los datos de prueba. Aunque los modelos híbridos reúnen el mayor número absoluto de estudios con validación V3–V4, esta diferencia debe interpretarse considerando el tamaño de cada familia y las características de los procedimientos utilizados. La frecuencia de publicación, por sí sola, no permite establecer la superioridad predictiva de una arquitectura. Tampoco puede concluirse que los modelos más complejos ofrezcan necesariamente mejores resultados, ya que su desempeño depende de los datos, los horizontes de predicción y las condiciones de evaluación.

Tabla 2. Familias arquitectónicas según fuerza de validación.

Familia	V4	V3	V2	V0	Total
Recurrente	0	1	5	95	101
Transformer/atención	2	0	1	31	34
Híbrida	4	1	4	118	127
Convolucional	0	0	0	3	3
Otras	0	1	1	47	49
Total	6	3	11	294	314

3.3 Condiciones bajo las que aparecen ventajas

Los seis estudios clasificados como evidencia prioritaria (Tier A) mostraron resultados favorables para diferentes arquitecturas de aprendizaje profundo, aunque las ventajas observadas dependieron del horizonte de predicción, las variables utilizadas y los procedimientos de evaluación. En el pronóstico a corto plazo, Hua et al. (2019) reportaron un error relativo del 1,43 % con una arquitectura LSTM-BP, frente al 4,07 % obtenido mediante BP y el 2,13 % mediante GA-BP, utilizando datos de entrenamiento correspondientes a 2014–2016 y datos de prueba de 2017. Kim et al. (2020), por su parte, obtuvieron menores valores de NMAE y NRMSE con una red neuronal profunda (DNN) que con los modelos SVM y ANN, incorporando imágenes satelitales y un procedimiento de evaluación cronológica. No obstante, este último estudio también examinó particiones aleatorias de los datos, por lo que resulta importante distinguir los resultados correspondientes a cada esquema de validación.

Para horizontes ultracortos, Tian et al. (2022) reportaron reducciones del MAPE del 22,0 % frente a GRU y del 29,1 % frente a DNN mediante una arquitectura Transformer, cuya evaluación incluyó datos de otra región. En el pronóstico con un día de anticipación (*day-ahead*), Tortora et al. (2026) desarrollaron MATNet, un modelo que combina registros históricos de generación fotovoltaica, información y predicciones meteorológicas numéricas (NWP). Su evaluación incluyó análisis de ablación y pruebas *zero-shot* en cinco instalaciones externas, con mejoras respecto a los modelos de comparación utilizados. En este mismo horizonte, Shanker y Kulkarni (2026) reportaron una reducción del MAE del 48,8 % mediante una arquitectura LSTM con autoencoder residual, evaluada entre diferentes sitios. Por su parte, Liu et al. (2026)

examinaron la transferencia de un modelo TA–SH–LSTM entre dos plantas fotovoltaicas sin reentrenamiento, obteniendo valores de R^2 de 0,997 y 0,993, respectivamente.

Aunque estos estudios ofrecen evidencia favorable sobre las arquitecturas evaluadas, sus resultados no permiten establecer una superioridad general de un modelo sobre los demás. Las diferencias entre los conjuntos de datos, los horizontes de predicción, las métricas y las condiciones de validación impiden comparar directamente las magnitudes de error reportadas. La Tabla 3 presenta las características de evaluación, los resultados y las principales limitaciones de los seis estudios.

Tabla 3. Condiciones, resultados y límites de los seis estudios Tier A.

Estudio	Arquitectura y horizonte	Validación	Resultado comparativo	Limitación principal
Hua et al. (2019), RS0022	LSTM-BP; corto plazo	V3: entrenamiento 2014–2016 y prueba 2017	Error relativo de 1,43 %, frente a 4,07 % con BP y 2,13 % con GA-BP	Evaluación en una estación; no demuestra generalización geográfica
Kim et al. (2020), RS0055	DNN con imágenes satelitales; corto plazo	V3: evaluación cronológica	Menores NMAE y NRMSE que SVM y ANN	También utiliza muestreo aleatorio; requiere imágenes satelitales
Tian et al. (2022), RS0196	Transformer; ultracorto plazo	V4: evaluación en otra región	MAPE 22,0 % menor que GRU y 29,1 % menor que DNN	Resultados condicionados por los datos y el protocolo de evaluación
Tortora et al. (2026), RS1309	MATNet (Transformer); day-ahead	V4: evaluación zero-shot en cinco sitios externos	RMSE de 0,0445; mejora relativa aproximada del 65 % frente al mejor comparador externo	Requiere predicciones meteorológicas numéricas; resultados dependientes del conjunto de evaluación
Shanker y Kulkarni (2026), RS1380	LSTM con autoencoder residual; day-ahead	V4: evaluación entre sitios	Reducción del MAE del 48,8 %; R^2 de 0,9467 a 0,9882 en Mumbai	Resultados vinculados a instalaciones y configuraciones específicas
Liu et al. (2026), RS1418	TA–SH–LSTM; corto plazo	V4: transferencia entre plantas	Planta 1: MAE/RMSE de 223,4/424,6 kW y R^2 de 0,997; planta 2 sin reentrenamiento: R^2 de 0,993	Evaluación en dos plantas; falta evidencia en una mayor diversidad de condiciones

3.4 Horizonte, resolución y variables meteorológicas

El pronóstico a corto plazo fue el horizonte más frecuente entre los 314 estudios incluidos, con 129 publicaciones (41,1 %), seguido del largo plazo (33; 10,5 %), el ultracorto plazo (32; 10,2 %) y el pronóstico con un día de anticipación (*day-ahead*) (25; 8,0 %). Sin embargo, en 63 estudios (20,1 %) no fue posible identificar con suficiente claridad el horizonte de predicción, lo que limita la comparación de sus resultados. La distribución por fuerza de validación mostró que únicamente tres de los 129 estudios de corto plazo (2,3 %) documentaron procedimientos V3–V4. En el ultracorto plazo se identificó un estudio con este nivel de validación, mientras que en el pronóstico *day-ahead* se registraron dos. No se identificaron estudios V3–V4 en la categoría de largo plazo (Tabla 4).

Tabla 4. Distribución no solapada del corpus por horizonte y fuerza de validación.

Horizonte o categoría	n	% del corpus	V3–V4
Corto plazo	129	41,1	3
No especificado	63	20,1	1
Largo plazo	33	10,5	0
Ultra corto plazo	32	10,2	1

Horizonte o categoría	n	% del corpus	V3–V4
Day-ahead	25	8,0	2
Otras categorías*	32	10,2	2
Total	314	100	9

**Las categorías se mantuvieron mutuamente excluyentes para evitar el doble conteo detectado en versiones previas.*

Solo dos estudios (0,6 %) tuvieron la optimización como objetivo principal, por lo que los resultados de esta revisión ofrecen mayor información sobre el pronóstico de la generación fotovoltaica que sobre la optimización directa de la producción. Asimismo, las diferencias en la resolución temporal y en la forma de reportar los resultados dificultaron las comparaciones entre investigaciones. Los estudios utilizaron distintas frecuencias de observación, como 5 minutos, 15 minutos, una hora o períodos diarios, sin que la información disponible permitiera establecer comparaciones homogéneas. Por esta razón, las métricas de error deben interpretarse considerando el horizonte de predicción, la escala de la variable objetivo, los procedimientos de normalización y las características de los datos utilizados para la evaluación.

En cuanto a las variables meteorológicas, las entradas más claramente identificadas fueron la irradiancia o radiación solar (54 estudios), los pronósticos meteorológicos numéricos o NWP (45), la nubosidad o información derivada de imágenes (42), la temperatura (25), la humedad (13) y los datos satelitales (11). Estas categorías no fueron mutuamente excluyentes, ya que un mismo estudio podía utilizar varias fuentes de información. La proporción de estudios con validación V3–V4 fue del 18,2 % entre aquellos que incorporaron datos satelitales (2 de 11) y del 12,0 % entre los que utilizaron temperatura (3 de 25). Sin embargo, estas diferencias no permiten atribuir una mayor solidez predictiva a determinadas variables, puesto que los estudios también difieren en las arquitecturas, los horizontes de predicción, las instalaciones y los procedimientos de evaluación. Para determinar la contribución específica de una variable sería necesario comparar modelos bajo las mismas condiciones de entrenamiento y prueba, mediante análisis de ablación u otros procedimientos equivalentes.

3.5 Consideraciones sobre generalización y aplicabilidad

La evidencia revisada muestra que la posibilidad de aplicar los modelos de aprendizaje profundo en nuevos contextos depende de las características de los datos disponibles, las condiciones de operación de los sistemas fotovoltaicos y los procedimientos utilizados para evaluar su desempeño. Entre los aspectos más relevantes se encuentran la disponibilidad de variables meteorológicas comparables, la correspondencia de los horizontes de predicción, la resolución temporal de los registros y las diferencias entre instalaciones. Estas condiciones deben considerarse al interpretar los resultados obtenidos en otros sitios, ya que un modelo con errores reducidos en una instalación no necesariamente mantendrá el mismo desempeño en períodos o condiciones ambientales diferentes.

La evaluación de la generalización también estuvo limitada por la información metodológica disponible. De los 314 estudios incluidos, únicamente nueve documentaron procedimientos de validación temporal independiente o externa entre instalaciones (V3–V4), y seis cumplieron los criterios de evidencia prioritaria (Tier A). Algunos de estos trabajos incorporaron evaluaciones entre regiones o plantas fotovoltaicas, lo que permitió examinar el comportamiento de los modelos fuera de los datos utilizados para su desarrollo. Sin embargo, estas evaluaciones se realizaron bajo condiciones específicas y no permiten asumir que los resultados sean transferibles a cualquier instalación. Por ello, la literatura revisada proporciona referencias útiles para seleccionar arquitecturas y diseñar procedimientos de evaluación, pero la aplicación de estos modelos en nuevos contextos requiere comprobar su desempeño con datos representativos de las condiciones locales.

3.6 Sesgo por no recuperación/evaluación a texto completo

De los 1.044 informes que pasaron a la etapa de recuperación y evaluación de elegibilidad, 351 (33,6 %) pudieron evaluarse, mientras que 693 (66,4 %) no llegaron a esta etapa. Esta proporción representa una limitación importante para la cobertura de la revisión, ya que no puede descartarse que los informes no evaluados presentaran características metodológicas o resultados diferentes de los estudios incluidos. La información disponible tampoco permite establecer de manera exhaustiva las razones por las que estos informes no llegaron a evaluarse. Por tanto, aunque 314 de los 351 estudios evaluados (89,5 %) cumplieron los criterios de inclusión, esta proporción no debe interpretarse como representativa de los 1.044 informes inicialmente considerados para elegibilidad.

El análisis exploratorio de metadatos mostró diferencias en la representación de los registros según su procedencia. Aproximadamente el 30,8 % de los candidatos vinculados a Scopus y el 30,7 % de los vinculados a Web of Science pudieron emparejarse con el corpus final, frente al 19,0 % de los procedentes de IEEE Xplore. Sin embargo, estas proporciones deben interpretarse con cautela, debido a que un mismo registro podía estar presente en más de una base de datos y el conjunto de metadatos utilizado para el análisis permitió emparejar únicamente 296 de los 314 estudios finalmente incluidos. En consecuencia, los resultados sugieren posibles diferencias de representación entre las fuentes consultadas, pero no permiten determinar las causas de la falta de evaluación ni establecer si los informes no evaluados habrían modificado los hallazgos de la revisión.

4 DISCUSIÓN

4.1 Evidencia sobre arquitecturas de aprendizaje profundo para forecasting fotovoltaico

La frecuencia con la que se utiliza una arquitectura de aprendizaje profundo no constituye, por sí misma, evidencia de un mejor desempeño predictivo. Los modelos recurrentes, como LSTM y GRU, mantienen una presencia importante en la literatura sobre pronóstico fotovoltaico y continúan utilizándose como modelos de referencia, especialmente en horizontes de corto plazo. Sin embargo, los resultados de

esta revisión muestran que las ventajas observadas dependen de las condiciones en las que se evalúa cada arquitectura, entre ellas el horizonte de predicción, las variables de entrada, las características de los datos y los procedimientos de validación.

Entre los estudios clasificados como evidencia prioritaria se identificaron resultados favorables para diferentes configuraciones. Las arquitecturas LSTM-BP mostraron ventajas frente a sus comparadores bajo una separación temporal de los datos (Hua et al., 2019), mientras que la incorporación de imágenes satelitales permitió mejorar el desempeño de una DNN en el pronóstico a corto plazo (Kim et al., 2020). Para horizontes ultracortos, Tian et al. (2022) reportaron resultados favorables mediante Transformer e incluyeron una evaluación con datos de otra región. En el pronóstico con un día de anticipación, las arquitecturas que combinaron información histórica, variables meteorológicas y predicciones meteorológicas numéricas también mostraron resultados prometedores, incluidos procedimientos de evaluación entre instalaciones (Tortora et al., 2026; Shanker & Kulkarni, 2026). Asimismo, Liu et al. (2026) documentaron la transferencia de un modelo entre dos plantas fotovoltaicas sin reentrenamiento.

Estas diferencias muestran que ninguna de las familias arquitectónicas examinadas puede considerarse superior en todos los escenarios. Aunque los modelos híbridos concentraron el mayor número absoluto de estudios con validación V3–V4, su predominio en el corpus y las diferencias entre los procedimientos de evaluación impiden atribuirles una ventaja general. La selección de arquitecturas recurrentes, Transformer o híbridas debe responder a las características del problema de pronóstico y a la evidencia obtenida mediante comparaciones realizadas bajo condiciones equivalentes.

4.2 Validación, data leakage y generalización como determinantes de la credibilidad

Los procedimientos de validación son fundamentales para interpretar la credibilidad de los resultados reportados en los estudios de pronóstico fotovoltaico. Una división aleatoria de las observaciones puede distribuir patrones temporales y condiciones meteorológicas similares entre los conjuntos de entrenamiento y prueba, mientras que una separación cronológica permite evaluar el comportamiento del modelo frente a observaciones posteriores a las utilizadas durante su desarrollo. La validación externa entre instalaciones aporta información adicional sobre la capacidad de aplicar un modelo en sistemas o ubicaciones diferentes. Sin embargo, estos procedimientos evalúan aspectos distintos y ninguno garantiza, por sí solo, la ausencia de problemas metodológicos.

En el análisis de series temporales, la validación debe respetar la disponibilidad cronológica de la información utilizada para generar cada predicción. Los esquemas de origen móvil (*rolling-origin*) permiten reproducir evaluaciones sucesivas en las que el modelo se entrena con información anterior al período de prueba, por lo que constituyen una alternativa apropiada para examinar el desempeño predictivo en distintos momentos (Hyndman & Athanasopoulos, 2021). Desde esta perspectiva, un error reducido

obtenido mediante una partición aleatoria no proporciona necesariamente evidencia más sólida que un resultado obtenido con una validación temporal independiente.

Otro aspecto relevante es la fuga de información (*data leakage*), que puede producirse cuando el entrenamiento, el preprocesamiento o la selección de modelos incorporan información que no estaría disponible en el momento de realizar el pronóstico. Este riesgo debe considerarse durante la normalización, la selección de variables, el ajuste de hiperparámetros y la aplicación de procedimientos de descomposición de señales (Yang et al., 2024). En particular, las técnicas VMD, CEEMDAN y wavelet requieren una descripción suficientemente detallada de su implementación para establecer si las transformaciones utilizadas respetaron la separación entre entrenamiento y prueba. La utilización de estas técnicas no implica necesariamente fuga de información, pero una mejora en las métricas resulta difícil de valorar cuando el procedimiento no está claramente documentado.

La interpretación de los errores también depende del horizonte de predicción, la resolución temporal y la escala de las variables. Los valores de RMSE, MAE o MAPE obtenidos en diferentes estudios no pueden compararse directamente cuando existen diferencias importantes en las condiciones de evaluación. Una predicción realizada con cinco minutos de anticipación responde a un problema distinto de un pronóstico *day-ahead*, mientras que las diferencias en potencia instalada, normalización y conjuntos de prueba pueden modificar considerablemente la magnitud de los errores. Por ello, las comparaciones entre arquitecturas requieren considerar estas características antes de atribuir las diferencias observadas al modelo utilizado.

Las variables meteorológicas constituyen otro componente relevante del desempeño predictivo. La irradiancia, la temperatura, la humedad, la nubosidad y los pronósticos meteorológicos numéricos aportan información sobre las condiciones que afectan la generación fotovoltaica. No obstante, los resultados de esta revisión no permiten establecer una jerarquía general de importancia entre estas variables, debido a las diferencias entre los conjuntos de datos y las configuraciones examinadas. Su contribución específica debería evaluarse mediante comparaciones controladas y análisis de ablación que mantengan constantes las demás condiciones de entrenamiento y prueba.

4.3 Limitaciones de la evidencia disponible

La interpretación de los resultados está condicionada por varias limitaciones de la evidencia disponible. La principal dificultad metodológica fue la escasa información sobre los procedimientos de validación: 294 de los 314 estudios quedaron clasificados como V0, debido a que la documentación disponible no permitió establecer con suficiente claridad cómo se evaluó el desempeño predictivo. Esta clasificación no demuestra que los modelos se hayan desarrollado mediante procedimientos incorrectos, pero limita la confianza con la que pueden compararse sus resultados. Las diferencias en los horizontes de

predicción, las resoluciones temporales, las variables de entrada y las métricas utilizadas también dificultaron la identificación de ventajas consistentes entre arquitecturas.

La revisión presenta, además, una limitación importante de cobertura. De los 1.044 informes que pasaron a la etapa de recuperación y evaluación de elegibilidad, 693 no llegaron a evaluarse. Aunque se realizó un análisis exploratorio de su procedencia mediante metadatos, la información disponible no permite reconstruir exhaustivamente las causas de esta situación ni determinar qué resultados habrían aportado esos documentos. Por tanto, no puede descartarse que la evidencia no evaluada hubiera modificado la distribución de arquitecturas, las características metodológicas observadas o las conclusiones de la síntesis.

Otra limitación corresponde a la escasa representación de investigaciones centradas en la optimización directa de la producción fotovoltaica. Solo dos estudios tuvieron este objetivo como propósito principal, por lo que la evidencia reunida permite discutir con mayor fundamento los modelos de pronóstico que las estrategias de optimización. Asimismo, la heterogeneidad de los datos, los horizontes y las métricas impidió realizar una síntesis cuantitativa global del desempeño.

Estas limitaciones restringen el alcance de las comparaciones entre arquitecturas, pero no impiden identificar aspectos metodológicos relevantes para futuras investigaciones. La documentación de los procedimientos de preprocesamiento, la separación temporal de los datos, la utilización de comparadores adecuados y la evaluación en períodos o instalaciones diferentes son elementos necesarios para interpretar con mayor fundamento el desempeño de los modelos. La contribución de esta revisión reside tanto en caracterizar las arquitecturas utilizadas como en examinar la solidez de la evidencia que respalda sus resultados.

4.4 Implicaciones para nuevas aplicaciones y contextos locales

Los resultados de esta revisión ofrecen orientaciones para el desarrollo y la evaluación de modelos de aprendizaje profundo destinados al pronóstico fotovoltaico en diferentes condiciones de operación. La selección de una arquitectura debería considerar la disponibilidad y calidad de los registros históricos, las variables meteorológicas accesibles, la resolución temporal, el horizonte de predicción y las características del sistema fotovoltaico. Estas condiciones influyen tanto en la construcción del modelo como en la posibilidad de interpretar su desempeño y evaluar su aplicación en otros períodos o instalaciones.

A partir de la evidencia examinada, una estrategia razonable para futuras investigaciones consiste en establecer primero modelos de referencia, como persistencia o arquitecturas recurrentes convencionales, y compararlos con configuraciones progresivamente más complejas. La incorporación de variables como irradiancia, temperatura, humedad, nubosidad, pronósticos meteorológicos numéricos o información satelital debería responder a su disponibilidad y pertinencia para el horizonte estudiado. Esta propuesta no representa una jerarquía de variables demostrada por la revisión, sino una forma de organizar

comparaciones que permitan identificar qué información y qué componentes del modelo aportan mejoras bajo condiciones equivalentes.

La evaluación debería contemplar una separación cronológica entre entrenamiento y prueba, procedimientos de preprocesamiento que eviten utilizar información futura y métricas adecuadas para la escala y el horizonte de predicción. Cuando los datos lo permitan, la validación mediante origen móvil o en períodos independientes puede aportar información sobre la estabilidad temporal del desempeño. La evaluación en instalaciones diferentes permitiría examinar, además, hasta qué punto los modelos conservan sus resultados frente a cambios en las condiciones meteorológicas y las características de los sistemas fotovoltaicos.

Estas recomendaciones adquieren especial importancia ante la creciente utilización de arquitecturas híbridas y Transformer, cuya complejidad puede dificultar la interpretación de las mejoras reportadas cuando no se dispone de comparaciones controladas. La investigación futura debería prestar mayor atención a la reproducibilidad de los procedimientos, la disponibilidad de datos y código, los análisis de ablación y la evaluación externa. De esta manera, sería posible avanzar desde la comparación de errores obtenidos en conjuntos de datos particulares hacia una evaluación más consistente de la capacidad predictiva y de generalización de los modelos.

5. CONCLUSIONES

La revisión sistemática de 314 estudios publicados entre 2019 y 2026 mostró un predominio de investigaciones centradas en el pronóstico de la generación fotovoltaica, mientras que únicamente dos estudios abordaron la optimización directa como objetivo principal. Las arquitecturas híbridas y recurrentes fueron las más frecuentes, aunque su mayor presencia en la literatura no se tradujo necesariamente en evidencia más sólida sobre su desempeño. Los modelos LSTM y GRU, así como sus variantes, continúan siendo utilizados en escenarios de corto plazo, mientras que las arquitecturas Transformer e híbridas mostraron resultados favorables en determinadas condiciones de predicción ultracorta y con un día de anticipación (*day-ahead*). Sin embargo, las diferencias entre horizontes, variables de entrada, conjuntos de datos y procedimientos de evaluación impiden establecer una superioridad general de una arquitectura sobre las demás.

Uno de los principales hallazgos fue la limitada documentación de procedimientos de validación temporal independiente o externa. De los 314 estudios incluidos, únicamente nueve (2,9 %) presentaron validación V3–V4 y seis cumplieron los criterios establecidos para la evidencia prioritaria (Tier A). Esta situación restringe la posibilidad de comparar los resultados predictivos y valorar su generalización a otros períodos o instalaciones, sin que ello implique necesariamente que los modelos restantes hayan sido desarrollados mediante procedimientos incorrectos. La evaluación de modelos de pronóstico requiere respetar la disponibilidad temporal de la información y prevenir posibles fugas de datos durante el

preprocesamiento, la selección de variables y el ajuste de hiperparámetros. Asimismo, la contribución de las variables meteorológicas y de la potencia histórica debe examinarse mediante comparaciones realizadas bajo condiciones equivalentes, en lugar de asumir que determinadas combinaciones ofrecen ventajas en todos los escenarios.

Los hallazgos proporcionan una base para orientar futuras investigaciones sobre pronóstico fotovoltaico, particularmente en la selección de arquitecturas, el diseño de procedimientos de validación y la evaluación de su capacidad de generalización. No obstante, el alcance de las conclusiones está condicionado por la heterogeneidad metodológica de los estudios y por la falta de evaluación de 693 de los 1.044 informes que pasaron a la etapa de recuperación y elegibilidad, lo que representa una posible fuente de sesgo de selección.

La evidencia disponible respalda la necesidad de desarrollar comparaciones reproducibles, utilizar modelos de referencia adecuados y evaluar el desempeño en períodos o instalaciones diferentes. Estos aspectos resultan fundamentales para determinar en qué condiciones las mejoras predictivas reportadas pueden considerarse suficientemente sólidas y aplicables a nuevos sistemas fotovoltaicos.

DISPONIBILIDAD DE DATOS Las matrices de extracción, trazabilidad, evaluación metodológica y síntesis se conservaron como archivos de trabajo de la revisión. El material suplementario identifica los 314 estudios incluidos (Tabla S1), documenta las estrategias de búsqueda por base de datos (Tabla S2) y proporciona una cita trazable para cada estudio incluido (Tabla S3). Su eventual publicación deberá respetar las condiciones de uso y derechos de las bases de datos de origen.

ASPECTOS ÉTICOS Al tratarse de una revisión sistemática de literatura sin intervención ni recolección de datos de participantes humanos o animales, no se realizó intervención experimental.

REGISTRO DEL PROTOCOLO El protocolo no fue registrado en un repositorio público; esta limitación se declara de forma explícita.

MATERIAL SUPLEMENTARIO Tabla S1: identificación bibliográfica de los 314 estudios incluidos y campos resumidos de extracción. Tabla S2: estrategias de búsqueda congeladas por base de datos. Tabla S3: citas trazables de los 314 estudios incluidos.

FINANCIACIÓN
Los autores no recibieron financiación específica para el desarrollo de la presente revisión sistemática.

CONFLICTO DE INTERESES
Los autores declaran que no existe conflicto de intereses relacionado con esta investigación.

CONTRIBUCIÓN DE AUTORÍA
De acuerdo con la taxonomía CRediT, las contribuciones de autoría se detallan en la siguiente matriz.

Participar activamente en:	*Autor 1.*	*Autor 2.*
Conceptualización	X	X
Análisis formal	X	X
Adquisición de fondos	X	X
Investigación	X	X
Metodología	X	X
Administración del proyecto	X	X
Recursos	X	X
Redacción –borrador original	X	X
Redacción –revisión y edición	X	X
La discusión de los resultados	**X**	**X**
Revisión y aprobación de la versión final del trabajo.	**X**	**X**

REFERENCIAS BIBLIOGRÁFICAS

Bak, S., Choi, S., Yang, D., Kim, D., Rho, H., & Lee, K. (2025). Transfer learning for photovoltaic power forecasting across regions using large-scale datasets. IEEE Access, 13, 136175–136190. https://doi.org/10.1109/ACCESS.2025.3591040

Başaran, K., Bozyiğit, F., Siano, P., Taşer, P. Y., & Kılınç, D. (2020). Systematic literature review of photovoltaic output power forecasting. IET Renewable Power Generation, 14(19), 3961–3973. https://doi.org/10.1049/iet-rpg.2020.0351

Chandel, S. S., Gupta, A., Chandel, R., & Tajjour, S. (2023). Review of deep learning techniques for power generation prediction of industrial solar photovoltaic plants. Solar Compass, 8, 100061. https://doi.org/10.1016/j.solcom.2023.100061

Gu, B., Li, X., Xu, F., Yang, X., Wang, F., & Wang, P. (2023). Forecasting and uncertainty analysis of day-ahead photovoltaic power based on WT-CNN-BiLSTM-AM-GMM. Sustainability, 15(8), 6538. https://doi.org/10.3390/su15086538

Herrera-Casanova, R., Conde, A., & Santos-Pérez, C. (2024). Hour-ahead photovoltaic power prediction combining BiLSTM and Bayesian optimization algorithm, with bootstrap resampling for interval predictions. Sensors, 24(3), 882. https://doi.org/10.3390/s24030882

Hua, C., Zhu, E., Kuang, L., & Pi, D. (2019). Short-term power prediction of photovoltaic power station based on long short-term memory-back-propagation. International Journal of Distributed Sensor Networks, 15(10), 1550147719883134. https://doi.org/10.1177/1550147719883134

Huang, C.-J., & Kuo, P.-H. (2019). Multiple-input deep convolutional neural network model for short-term photovoltaic power forecasting. IEEE Access, 7, 74822–74834. https://doi.org/10.1109/ACCESS.2019.2921238

Husein, M., Gago, E. J., Hasan, B., & Pegalajar, M. C. (2024). Towards energy efficiency: A comprehensive review of deep learning-based photovoltaic power forecasting strategies. Heliyon, 10(13), e33419. https://doi.org/10.1016/j.heliyon.2024.e33419

Hyndman, R. J., & Athanasopoulos, G. (2021). Forecasting: Principles and practice (3rd ed.). OTexts. https://otexts.com/fpp3/tscv.html

Iheanetu, K. J. (2022). Solar photovoltaic power forecasting: A review. Sustainability, 14(24), 17005. https://doi.org/10.3390/su142417005

Kasburg, C., & Stefenon, S. F. (2019). Deep learning for photovoltaic generation forecast in active solar trackers. IEEE Latin America Transactions, 17(12), 2004–2010. https://doi.org/10.1109/TLA.2019.9011546

Khouili, O., Hanine, M., Louzazni, M., López Flores, M. Á., García Villena, E., & Ashraf, I. (2025). Evaluating the impact of deep learning approaches on solar and photovoltaic power forecasting: A systematic review. Energy Strategy Reviews, 59, 101735. https://doi.org/10.1016/j.esr.2025.101735

Kim, M., Song, H., & Kim, Y. (2020). Direct short-term forecast of photovoltaic power through a comparative study between COMS and Himawari-8 meteorological satellite images in a deep neural network. Remote Sensing, 12(15), 2357. https://doi.org/10.3390/rs12152357

Lee, D., & Kim, K. (2019). Recurrent neural network-based hourly prediction of photovoltaic power output using meteorological information. Energies, 12(2), 215. https://doi.org/10.3390/en12020215

Liu, H., Du, C., Guo, R., Luo, Y., Cui, Y., Zi, J., He, L., & Cao, Y. (2026). Short-term PV power forecasting with temporal-attention LSTM and successive-halving hyperparameter search. Electronics, 15(5), 1019. https://doi.org/10.3390/electronics15051019

Mellit, A., Massi Pavan, A., Ogliari, E., Leva, S., & Lughi, V. (2020). Advanced methods for photovoltaic output power forecasting: A review. Applied Sciences, 10(2), 487. https://doi.org/10.3390/app10020487

Moons, K. G. M., Damen, J. A. A., Kaul, T., Hooft, L., Andaur Navarro, C., Dhiman, P., Beam, A. L., Van Calster, B., Celi, L. A., Denaxas, S., Denniston, A. K., Ghassemi, M., Heinze, G., Kengne, A. P., Maier-Hein, L., Liu, X., Logullo, P., McCradden, M. D., Liu, N., ... van Smeden, M. (2025). PROBAST+AI: An updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. BMJ, 388, e082505. https://doi.org/10.1136/bmj-2024-082505

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71

Rajagukguk, R. A., Ramadhan, R. A. A., & 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. https://doi.org/10.3390/en13246623

Schreiber, J., & Sick, B. (2022). Multi-task autoencoders and transfer learning for day-ahead wind and photovoltaic power forecasts. Energies, 15(21), 8062. https://doi.org/10.3390/en15218062

Sesay, M. M. (2026). Leakage-aware benchmarking of temporal data preparation for hourly energy forecasting. Energy Conversion and Management: X, 30, 101791. https://doi.org/10.1016/j.ecmx.2026.101791

Shanker, M., & Kulkarni, V. (2026). Residual-enhanced hybrid deep learning framework for day-ahead PV power forecasting with cross-site generalization. Next Energy, 11, 100601. https://doi.org/10.1016/j.nxener.2026.100601

Tian, F., Fan, X., Wang, R., Qin, H., & Fan, Y. (2022). A power forecasting method for ultra-short-term photovoltaic power generation using Transformer model. Mathematical Problems in Engineering, 2022, 9421400. https://doi.org/10.1155/2022/9421400

Tortora, M., Conte, F., Natrella, G., & Soda, P. (2026). MATNet: Multi-level fusion transformer-based model for day-ahead PV generation forecasting. Frontiers in Artificial Intelligence, 9, 1891798. https://doi.org/10.3389/frai.2026.1891798

Wedhasari, T., & Castro, R. (2026). A critical review and strategic roadmap of PV power forecasting (2016–2026): Addressing temporal leakage and operational integration gaps. Energies, 19(12), 2937. https://doi.org/10.3390/en19122937

Wu, Y.-K., Huang, C.-L., Phan, Q.-T., & Li, Y.-Y. (2022). Completed review of various solar power forecasting techniques considering different viewpoints. Energies, 15(9), 3320. https://doi.org/10.3390/en15093320

Yang, T. G., Li, B., & Xun, Q. (2019). LSTM-attention-embedding model-based day-ahead prediction of photovoltaic power output using Bayesian optimization. IEEE Access, 7, 171471–171484. https://doi.org/10.1109/ACCESS.2019.2954290

Yang, X., Li, J., & Jiang, X. (2024). Research on information leakage in time series prediction based on empirical mode decomposition. Scientific Reports, 14, 28362. https://doi.org/10.1038/s41598-024-80018-9

Yu, J., Li, X., Yang, L., Li, L., Huang, Z., Shen, K., Yang, X., Yang, X., Xu, Z., Zhang, D., & Du, S. (2024). Deep learning models for PV power forecasting: Review. Energies, 17(16), 3973. https://doi.org/10.3390/en17163973