



Universidad de Jaén

Escuela Politécnica
Superior de Jaén

Modelado y Predicción Inmediata de Potencia Fotovoltaica Mediante Técnicas de Internet de las Cosas, Aprendizaje Automático y Redes Neuronales

Autor: Guillermo Almonacid Olleros

Director de la tesis: Javier Medina Quero / Gabino Almonacid Puche
Departamento: Informática / Ingeniería Electrónica y Automática

Fecha: 03/06/2024

Licencia CC

RUJJA

Dedicatoria

A Sara, Diego y Pablo, quienes con su generosidad y apoyo incondicional han hecho posible este logro. Su inestimable ayuda y comprensión me han permitido dedicar tiempo y esfuerzo a esta empresa académica, convirtiendo este camino en una experiencia gratificante. A ellos les dedico este trabajo con cariño y agradecimiento.

Agradecimientos

Mi más sincero agradecimiento a mis directores de tesis, Javier Medina y Gabino Almonacid, cuya orientación experta y constante motivación han sido fundamentales en cada etapa de mi doctorado. Su dedicación y sabiduría han enriquecido enormemente esta experiencia académica, guiándome con claridad y paciencia.

Igualmente, deseo reconocer la valiosa ayuda de mi tutora Macarena Espinilla, cuya guía y consejo han sido imprescindibles para el desarrollo de este trabajo.

No puedo dejar de agradecer a mis compañeros de investigación y trabajo de campo, Juan Ignacio Fernández y Aurora Polo, cuya colaboración y aportaciones han sido cruciales para alcanzar los objetivos planteados.

Este trabajo ha sido financiado parcialmente por la *Cátedra ELAND para Energías Renovables* de la Universidad de Jaén y por el gobierno español a través del proyecto RTI2018-098979-A-I00.

Contenido

Dedicatoria	III
Agradecimientos	V
Índice de figuras	X
Índice de tablas	XII
Resumen	XV
1. Introducción	1
1.1. Motivación	1
1.2. Objetivos	4
2. Trabajos Relacionados	9
2.1. Sistemas fotovoltaicos	9
2.2. Modelos predictivos	12
3. Modelos de Aprendizaje Automático y Aprendizaje Profundo para la Predicción Inmediata de Potencia Fotovoltaica	19
3.1. Introducción	19

3.2. Proyecto Ópera: plataforma de procesamiento digital de datos para el modelado y control del comportamiento de sistemas fotovoltaicos	22
3.2.1. Métodos algebraicos para el modelado de sistemas fotovoltaicos	23
3.2.2. Arquitectura técnica de la plataforma	25
3.2.3. Resultados experimentales	28
3.3. Adquisición de datos en tiempo real mediante dispositivos IoT	31
3.4. Metodología para el procesamiento de datos	37
3.4.1. Segmentación y agregación de series temporales para modelos basados en el análisis e interpretación de datos	38
3.4.2. Definición manual de características y múltiples ventanas temporales	40
3.4.3. Extracción automática de características con modelos de aprendizaje profundo	42
3.5. Evaluación de resultados	44
3.5.1. Configuración experimental	45
3.5.2. Resultados	49
3.5.3. Discusión	54
4. Aprendizaje por Transferencia para la Predicción de Potencia Fotovoltaica	57
4.1. Introducción	57
4.2. Metodología para el procesamiento de datos	60
4.2.1. Recopilación de datos de fuentes heterogéneas	61
4.2.2. Segmentación de datos	63
4.2.3. Inicialización de pesos y aprendizaje afinado entre dominios	66

4.3.	Evaluación del caso de estudio	69
4.3.1.	Resultados	69
4.3.2.	Discusión	74
4.3.3.	Limitaciones del caso de estudio	76
5.	Modelos de Redes Adversarias Generativas para la Predicción y el Control de Potencia Fotovoltaica	79
5.1.	Introducción	79
5.2.	Metodología para el procesamiento de datos	81
5.2.1.	Preprocesamiento de datos y modelo de aprendizaje profundo de referencia	82
5.2.2.	Predicción y discriminación de potencia fotovoltaica mediante modelos CGAN de regresión y clasificación	83
5.2.3.	Predicción y rectificación de potencia fotovoltaica mediante modelos CGAN de regresión y dopaje	86
5.3.	Análisis de Resultados Experimentales	88
5.3.1.	Modelo CGAN de regresión y clasificación	90
5.3.2.	Modelo CGAN de regresión y dopaje	93
6.	Conclusiones	99
6.1.	Modelos de Aprendizaje Automático y Aprendizaje Profundo para la Predicción Inmediata de Potencia Fotovoltaica	99
6.2.	Aprendizaje por Transferencia para la Predicción de Potencia Fotovoltaica	100
6.3.	Modelos de Redes Adversarias Generativas para la Predicción y el Control de Potencia Fotovoltaica	101

Índice de figuras

3.1. Diagrama esquemático de la arquitectura de la plataforma Ópera . . .	26
3.2. Aplicación web Ópera para la monitorización del GFV UniVer	27
3.3. Tensión DC del GFV (UniVer) en un día representativo. Línea roja: valores pronosticados por el modelo Araujo, línea azul: valores reales.	29
3.4. Corriente DC (azul, roja) y potencia (verde, amarilla) del GFV (Uni- Ver) en un día representativo. Línea roja y verde: valores pronostica- dos por el modelo Araujo, línea azul y amarilla: valores reales.	30
3.5. Dos vistas del generador fotovoltaico del proyecto UniVer	32
3.6. Sensores de radiación y temperatura	36
3.7. Microprocesador y sensores de corriente y tensión	37
3.8. Arquitectura del módulo IoT para la adquisición de datos en tiempo real	37
3.9. Ejemplo de flujos de datos de sensores, segmentación y agregación mediante ventanas temporales deslizantes.	40
3.10. Ubicación del dispositivo fotovoltaico y módulo IoT en la Universidad de Jaén y fotografías de la implementación.	46

3.11. Muestras de 2 días de la verdad fundamental de la potencia fotovoltaica en comparación con las predicciones de: i) el modelo Araujo, ii) Regresión lineal, iii) k-NN, iv) Random Forest, v) SVM, vi) 3CNN+2LSTM.	53
4.1. Registro de ejemplo del archivo de datos UniVer: tiempo (timestamp), irradiancia global (G.I), temperatura ambiente (Tamb), temperatura del módulo (Tmod), potencia de salida (PA).	61
4.2. Fila de ejemplo del archivo de datos de SolarTech: tiempo, potencia de salida (PV_Power), temperatura ambiente (T_air), irradiancia global horizontal (G_h), irradiancia global medida (G_tilt), velocidad del viento (W_s), dirección del viento (W_d).	63
4.3. Segmentación del flujo de datos de un sensor definida por ventanas temporales deslizantes.	66
4.4. a) Aprendizaje básico, b) Aprendizaje externo y c) Aprendizaje afinado	68
4.5. Historial de entrenamiento y validación del modelo que muestra la función de pérdida (MSE) para los escenarios de aprendizaje básico y afinado en el contexto de SolarTech.	73
5.1. Modelo CGAN de regresión y clasificación	86
5.2. Regresión CGAN con modelo de dopaje y discriminador	88
5.3. Precisión en la predicción de muestras reales y falsas y NRMSE por épocas y nivel de distorsión en la regresión + clasificación CGAN. . .	94
5.4. Error relativo durante las épocas en la etapa de aprendizaje entre los tres componentes: generador, doper y discriminador.	97

Índice de tablas

3.1. Parámetros Medidos del GFV	23
3.2. Variables Calculados	24
3.3. Variables recogidas por el sistema de adquisición de datos	34
3.4. Configuración de las arquitecturas basadas en redes neuronales	44
3.5. Configuración del modelo de Regresión Lineal	47
3.6. Configuración del modelo Random Forest	47
3.7. Configuración del modelo k-NN	48
3.8. Configuración del modelo SVM	48
3.9. Métricas de error de Araujo	50
3.10. Métricas de error de modelos con características predefinidas por ex- pertos y diferentes enfoques de ventanas temporales	51
3.11. Métricas de tiempo de modelos basados en características predefinidas por expertos	51
3.12. Métricas de error para enfoques de aprendizaje profundo	52
3.13. Métricas de tiempo para modelos de aprendizaje profundo	52

3.14. Resumen de métricas de error para las mejores configuraciones de características predefinidas por expertos y enfoques de aprendizaje profundo	52
4.1. Configuración del modelo de aprendizaje profundo	69
4.2. Error normalizado por tipo de aprendizaje en el contexto de SolarTech	70
4.3. Error normalizado por tipo de aprendizaje en el contexto de UniVer .	71
4.4. Error absoluto por tipo de aprendizaje en SolarTech	71
4.5. Error absoluto por tipo de aprendizaje en UniVer	71
4.6. Velocidad de aprendizaje por escenario en el contexto de SolarTech .	72
4.7. Tiempo de entrenamiento por escenario en el contexto de SolarTech .	73
4.8. Error normalizado por tipo de aprendizaje en el contexto de India . .	74
4.9. Error absoluto por tipo de aprendizaje en el contexto de India	74
5.1. Resumen de métricas de error para el modelo base CNN+LSTM (promedio de las últimas 10 épocas)	90
5.2. Resumen de métricas de error en el generador: RMSE (W) y NRMSE en el modelo CGAN de regresión y clasificación (media de las 10 últimas épocas)	92
5.3. Resumen de precisión (Prc.) en el discriminador: precisión de muestras falsas y muestras reales en el modelo CGAN de regresión y clasificación (promedio de las últimas 10 épocas).	93
5.4. Resumen de métricas de error del modelo de regresión de doping para predicción instantánea y rectificación de generación de potencia de salida (promedio de las últimas 10 épocas)	95

Modelado y Predicción Inmediata de Potencia Fotovoltaica Mediante Técnicas de Internet de las Cosas, Aprendizaje Automático y Redes Neuronales

Guillermo Almonacid Olleros

Resumen

La presente tesis se enmarca en la exploración y aplicación de tecnologías de vanguardia en el campo de la energía solar fotovoltaica, centrándose en el análisis y predicción de la generación de energía eléctrica. Con el aumento de la demanda de energía sostenible, la optimización de sistemas fotovoltaicos se ha convertido en una prioridad para garantizar un suministro de energía eficiente y confiable. En este trabajo, se presenta la investigación y desarrollo de técnicas de vanguardia que incluyen el uso de Internet de las Cosas (IoT), el Aprendizaje Profundo (*Deep Learning*) y las Redes Generativas Adversarias (GANs) para abordar los desafíos en la generación de energía fotovoltaica. En este caso, afrontando aspectos relacionados con la optimización del funcionamiento de los sistemas fotovoltaicos y la predicción de generación de energía, así como en la gestión de estos activos para optimizar su rendimiento económico y energético.

Durante la elaboración del presente trabajo se ha desarrollado un sistema de adquisición de datos inalámbrico y versátil que utiliza sensores IoT para monitorizar el comportamiento de sistemas fotovoltaicos. Esta metodología ofrece ventajas sustanciales en términos de facilidad de uso, eficiencia de costes y estandarización de la captura de datos. Se ha demostrado que la incorporación de estos sensores

en una herramienta integral de gestión de Operación y Mantenimiento (O&M) es fundamental para mejorar la monitorización y la eficiencia de estos sistemas.

El contenido principal de esta tesis se centra en el análisis de las técnicas de aprendizaje profundo en la manipulación de datos de sensores en sistemas fotovoltaicos. El uso de redes neuronales recurrentes, específicamente Long Short-Term Memory (LSTM), ha demostrado ser una estrategia eficaz para el análisis y predicción de la generación de energía en función de datos de sensores. Además, la combinación de LSTM con Redes Neuronales Convolucionales (CNNs) ha proporcionado mayor rendimiento en la predicción de la generación de energía en sistemas fotovoltaicos, especialmente en el pronóstico a corto plazo. Estos modelos de aprendizaje profundo han demostrado su capacidad para capturar dependencias temporales y patrones espaciales en los datos fotovoltaicos, y se han destacado como herramientas esenciales para mejorar la precisión en la predicción de la generación de energía.

Además, se ha explorado el concepto de aprendizaje por transferencia en el contexto del análisis y procesamiento de datos de sensores en sistemas fotovoltaicos. La propuesta incluye un enfoque que, basado en el aprendizaje por transferencia, aborda las similitudes y diferencias entre los datos de sensores en sistemas fotovoltaicos de diferentes regiones. Además, se ha presentado un método común de homogeneización y segmentación para el procesamiento del flujo de datos de sensores, lo que ha permitido abordar la variabilidad en la recopilación de datos entre sistemas fotovoltaicos. Este enfoque ha demostrado su eficacia en la mejora de la precisión de la predicción de la generación de energía y la gestión de desviaciones en sistemas fotovoltaicos.

Finalmente, se han aplicado Redes Generativas Adversarias (GANs) en la generación de datos sintéticos para mejorar la robustez de los modelos de predicción en

sistemas fotovoltaicos. Las GANs han demostrado ser valiosas en la generación de datos de energía sintética, la ampliación de conjuntos de datos de entrenamiento y la mejora de la generalización de los modelos. Además, se ha explorado la aplicación de GANs en la reducción de la distorsión durante la generación de datos, lo que ha permitido mejorar la calidad de los datos y la precisión en la predicción de la generación de energía. Este enfoque innovador ha demostrado ser efectivo en la predicción de desviaciones en la generación de energía en sistemas fotovoltaicos.

En resumen, esta tesis aborda los desafíos en la gestión de la generación de energía fotovoltaica a través de la aplicación de tecnologías de vanguardia, como IoT, Aprendizaje Profundo y Redes Generativas Adversarias. Se han propuesto y desarrollado metodologías y enfoques que mejoran significativamente la monitorización, predicción y gestión de la generación de energía en sistemas fotovoltaicos, contribuyendo al avance en la eficiencia y fiabilidad de la energía solar como fuente de energía sostenible.

Capítulo 1

Introducción

1.1. Motivación

La generación de energía solar fotovoltaica, una tecnología que convierte la luz solar directamente en electricidad, ha experimentado un crecimiento y desarrollo sin precedentes en las últimas décadas. Este avance se ve claramente reflejado en los hitos alcanzados recientemente: en 2022, la capacidad de potencia fotovoltaica instalada a nivel mundial superó el umbral de 1 teravatio (TW), llegando a 1,5 TW a finales de 2023. Las proyecciones indican que esta tendencia ascendente se mantendrá en los próximos años, con expectativas de alcanzar los 2 TW en 2025 y 3,5 TW en 2027 [Europe, 2023]. Estos números son especialmente significativos considerando que la mayoría de esta capacidad está concentrada en grandes plantas fotovoltaicas.

Este notable crecimiento no solo refleja los avances tecnológicos en el campo de la energía solar, sino también un cambio en la percepción y políticas globales hacia fuentes de energía más sostenibles y respetuosas con el medio ambiente. La

transición hacia la energía solar fotovoltaica es parte de un esfuerzo mundial más amplio para combatir el cambio climático y reducir la dependencia de los combustibles fósiles, que son recursos finitos y generadores de importantes emisiones de gases de efecto invernadero. La inversión en energía solar, por lo tanto, no solo representa un avance tecnológico, sino también un compromiso político y social hacia un futuro más sostenible.

En este contexto, la gestión efectiva de la operación y el mantenimiento (O&M, *Operation & Maintenance*) de estos sistemas fotovoltaicos se ha vuelto crucial en la industria solar [Gonzalo et al., 2020]. La monitorización y el mantenimiento adecuados son esenciales para garantizar un funcionamiento eficiente y la longevidad de estos sistemas. Sin embargo, gestionar y mantener eficazmente un gran número de instalaciones fotovoltaicas dispersas geográficamente es un desafío logístico y técnico de proporciones significativas.

La motivación detrás de esta investigación surge de la posibilidad de aplicar nuevas tecnologías relacionadas con la monitorización inteligente para mejorar la gestión de los datos de funcionamiento de los sistemas, con el objetivo de optimizar la eficiencia de la generación de energía solar, mejorar la precisión en el pronóstico y detectar problemas de instalación en tiempo real [Prasanna Rani et al., 2023].

La implementación de estas tecnologías avanzadas en la gestión de sistemas fotovoltaicos no solo aumenta la eficiencia y fiabilidad de la generación de energía solar, sino que también contribuye a una integración más efectiva de esta fuente renovable en la matriz energética global. Esto incluye desde la predicción más precisa de la generación de energía, teniendo en cuenta variables ambientales y climáticas, hasta el mantenimiento predictivo de los equipos, maximizando así la eficiencia y la vida útil de estas instalaciones.

Tradicionalmente, se han utilizado expresiones algebraicas aproximadas basadas en las leyes físicas y los parámetros eléctricos de las células solares, junto con los datos de ingeniería de los dispositivos que conforman el sistema fotovoltaico, para construir modelos de funcionamiento estándar [Araujo and Sánchez, 1982].

Si bien estos modelos permiten una buena predicción de funcionamiento y generación de energía de los sistemas fotovoltaicos, adolecen, sin embargo, de fuertes limitaciones para la generalización son modelos dependientes de los parámetros de los dispositivos y la configuración del sistema. Esta característica, además de particularizar su uso a cada instalación, tiene el inconveniente de la necesidad de conocer numerosos datos sobre dispositivos, topología, etc. haciendo su uso complejo y difícil de automatizar.

Adicionalmente, los sensores convencionales, cableados, para toma de datos ofrecen poca flexibilidad, complican la instalación y dificultan el conocimiento del estado y funcionamiento del sistema en tiempo real. El aprovechamiento de la tecnología de Internet de las Cosas (IoT, *Internet of Things*) [Hossein Motlagh et al., 2020] y el aprendizaje automático (*Machine Learning*) se presenta como una solución prometedora para abordar estos desafíos.

La recopilación de datos en tiempo real de sistemas fotovoltaicos distribuidos a través de sensores IoT proporciona información valiosa sobre el rendimiento y el estado de cada instalación [Kumar et al., 2018a]. Estos datos permiten la detección temprana de posibles problemas y la ejecución de acciones de mantenimiento de manera oportuna, maximizando la eficiencia y el tiempo de actividad de las plantas fotovoltaicas. La creciente disponibilidad de datos en tiempo real ha revolucionado la forma en que se gestionan los sistemas fotovoltaicos [Hernandez et al., 2022]. La información generada por los sensores proporciona una visión detallada de las

condiciones en el sitio, lo que permite a los operadores tomar decisiones informadas sobre el mantenimiento y la operación de las instalaciones.

Por otro lado, aprovechando los últimos avances en aprendizaje automático y redes neuronales, se puede adoptar un enfoque diferente mediante el uso de regresores para construir modelos que aprendan a partir de datos sobre el comportamiento real del sistema durante un período relevante y utilicen la predicción de series temporales para monitorizar el rendimiento [Hammoudi et al., 2022]. Los enfoques de aprendizaje automático ofrecen la ventaja de modelar de manera independiente de los parámetros de implementación y configuración del sistema fotovoltaico, que están fuertemente influenciados por la ubicación y las condiciones ambientales.

Además, la generación de datos sintéticos mediante redes adversarias puede enriquecer los conjuntos de datos de entrenamiento, aumentar la robustez de los modelos [Zhang et al., 2018], e incluso detectar posibles situaciones de fraude energético.

En resumen, el presente trabajo se desarrolla en un contexto en el que los datos y la tecnología son una parte importante de la transición energética. El enfoque se centra en la aplicación de técnicas de IoT, aprendizaje automático y redes neuronales para mejorar la monitorización y la predicción de la generación de energía en sistemas fotovoltaicos. Estas técnicas permiten a los operadores tomar decisiones más precisas y oportunas, lo que se traduce en un mejor rendimiento y una mayor confiabilidad de las plantas fotovoltaicas.

1.2. Objetivos

La hipótesis principal de esta tesis radica en la premisa de que la integración de sensores y modelos inteligentes permiten mejorar la monitorización y supervisión

de los sistemas fotovoltaicos, incluyendo capacidades de predicción, adaptación a contextos ambientales y desviación ante errores con alta precisión en tiempo real.

Para validar esta hipótesis, se definen los siguientes objetivos:

1. Desarrollo de un sistema de monitorización IoT

Se pretende diseñar y desarrollar un sistema de adquisición de datos inalámbrico que utilice sensores IoT para recopilar datos en tiempo real sobre la radiación solar, la temperatura ambiente y otros parámetros relevantes. Este sistema de monitorización proporcionará datos precisos y en tiempo real que servirán como base para la predicción de la generación de energía.

2. Desarrollo de modelos de aprendizaje automático para la predicción de generación de energía solar fotovoltaica

El segundo objetivo es desarrollar y evaluar modelos base de aprendizaje automático, como regresión lineal, k vecinos más cercanos (k -NN, *k-Nearest Neighbors*), máquinas de soporte vectorial (SVM, *Support Vector Machine*) y bosques aleatorios (*Random Forests*), para predecir la generación de energía solar. Estos modelos se basarán en características definidas manualmente y ventanas temporales múltiples para capturar patrones complejos en los datos de generación de energía.

3. Implementación de redes neuronales profundas para la predicción de generación de energía solar fotovoltaica

Este objetivo recoge la investigación y aplicación de modelos de Aprendizaje Profundo, en particular, las redes neuronales recurrentes (RNN, *Recurrent Neural Network*) como las redes de memoria a corto y largo plazo (LSTM, *Long*

Short-Term Memory) y las redes neuronales convolucionales (CNN, *Convolutional Neural Network*) para el análisis de datos de sensores en sistemas fotovoltaicos. Estas técnicas permitirán capturar dependencias temporales y patrones espaciales en los datos, lo que facilitará la predicción precisa de la generación de energía.

4. Evaluación del rendimiento en diversos contextos de generación de energía solar fotovoltaica

Se busca explorar el concepto de transferencia de aprendizaje para adaptar modelos de predicción entre diferentes sistemas fotovoltaicos. Estas técnicas permitirán evaluar el rendimiento de los modelos en diferentes contextos de generación de energía solar, incluyendo datos de ubicaciones geográficas variadas y condiciones climáticas diversas. Se explorará la transferencia de conocimiento entre estos contextos y se analizará cómo afecta al rendimiento de los modelos.

5. Uso de redes adversarias para la generación de datos sintéticos

Se desea investigar y aplicar las Redes Adversarias Generativas (GAN, *Generative Adversarial Network*) para la generación de datos sintéticos y la mejora de la robustez de los modelos de predicción. Estas técnicas permitirán aumentar el tamaño de los conjuntos de datos de entrenamiento y mejorar la generalización de los modelos, reduciendo las discrepancias entre los datos generados y los datos reales.

6. Validación experimental y comparativa

Se propone llevar a cabo experimentos en sistemas fotovoltaicos reales para validar la eficacia de las técnicas desarrolladas, comparando estas técnicas con

los enfoques tradicionales de monitorización y predicción de la generación de energía.

Capítulo 2

Trabajos Relacionados

2.1. Sistemas fotovoltaicos

La tecnología fotovoltaica ha alcanzado un notable grado de madurez y los sistemas fotovoltaicos son ya una alternativa real, bien desarrollada, para la generación de energía eléctrica [Europe, 2023]. Sin embargo, su relativa juventud -la mayoría de las instalaciones, especialmente las de gran tamaño, no llevan operando más de dos décadas [Europe, 2016] [Europe, 2019]- implica que no existe una experiencia dilatada en cuanto a su operación y mantenimiento (O&M), y en otras tareas claves para una gestión eficaz de estos activos. Muchas de las herramientas y procesos mas utilizados todavía para llevar a cabo la gestión de este tipo de activos, así como las tareas de Operación y Mantenimiento (O&M), hacen poco uso de las ultimas tecnologías de la información y gestión de datos, tales como BD (Big Data), DL (Deep Learning), BI (Business Intelligence), gestión digital de activos (DAM), etc. [Bermejo et al., 2019].

Para estimar y predecir la generación de energía y analizar el comportamiento de

de los sistemas fotovoltaicos, sigue siendo frecuente la utilización de modelos clásicos basados en la física de las células solares. Estos modelos permiten obtener expresiones analíticas para calcular los parámetros eléctricos y de comportamiento de un sistema fotovoltaico [Luque and Hegedus, 2011]. De este tipo, hay numerosos modelos que ofrecen diferentes enfoques, precisión y grados de dificultad [Rus-Casas et al., 2014a] [Almeida et al., 2017a] [De la Parra et al., 2017] [Fuentes et al., 2007a].

Entre estos modelos clásicos basados en la física del dispositivo, para referencia y validación del desarrollo de un modelo basado en inteligencia artificial, se ha elegido el modelo de *Araujo* con la consideración del Factor de Forma (FF) constante ¹ [Fuentes et al., 2007a]. El modelo de *Araujo* es un modelo estándar muy utilizado, el cual combina gran precisión con una baja complejidad de formulación matemática [Luque and Hegedus, 2011] [Fuentes et al., 2007a]; adicionalmente, este modelo sólo necesita como entrada la medida de pocas variables: tensión y corriente de la célula, irradiación solar y temperatura ambiente [Araujo and Sanchez, 1982].

En los últimos años se han publicado incipientes trabajos relacionados con el uso de nuevas tecnologías para monitorización y estimación del comportamiento de sistemas fotovoltaicos [Bermejo et al., 2019] [Daliento et al., 2017] [Raza et al., 2017] [Fuentes et al., 2014]. En los inicios de este trabajo ya se presentó una plataforma analítica inicialmente orientada a la gestión digital de la O&M, recogida en la sección "3.2 - Proyecto Ópera: plataforma de procesamiento digital de datos para el modelado y control del comportamiento de sistemas fotovoltaicos". Sin embargo, y hasta dónde se ha podido investigar, hay muy pocos trabajos orientados hacia la gestión digital de activos fotovoltaicos; además, o son muy parciales, o dedica-

¹El factor de forma (FF) es un parámetro de calidad de las células solares. Se define como el cociente de la potencia máxima que puede entregar la célula entre el producto de su tensión de circuito abierto por la corriente de cortocircuito; su valor máximo es 1.

dos sólo a algún aspecto de la cuestión: gestión de energía, gestión de los datos de monitorización, smart grid, etc. [Nicolai et al., 2019]; o por el contrario, grandes y costosas plataformas desarrolladas en el ámbito de las grandes compañías del sector energético, e.g. el Centro de Operación de Renovables CORE de Iberdrola ².

La toma de datos para monitorización y control del comportamiento de los sistemas fotovoltaicos tradicionalmente ha sido llevada a cabo con sensores conectados a sistemas de adquisición de datos. Estos equipos suelen ser costosos, ofrecen poca flexibilidad y posibilidades limitadas de conectividad vía internet. La monitorización de sistemas fotovoltaicos está experimentando un cambio significativo en la forma en la que se recopilan los datos operativos. Recientemente se han presentado varios trabajos que proponen el uso de la conectividad IoT en la monitorización del comportamiento de los sistemas fotovoltaicos [Fuentes et al., 2014] [Kumar et al., 2018b][Daliento et al., 2017].

La incorporación de estos sensores en una herramienta integral de gestión de O&M ha permitido desarrollar un sistema de captura de datos con sensores inalámbricos altamente versátil y de fácil operación, que ofrece ventajas significativas en cuanto a facilidad de uso, eficiencia de costes y estandarización de la captura de datos [Raza et al., 2017] [Jayaprakash et al., 2019] [Spanias, 2018]. En la literatura relevante se han presentado varias propuestas basadas en el paradigma IoT en sistemas fotovoltaicos. Por ejemplo, en [Martín-Lopo et al., 2020], se realiza una revisión de la literatura sobre plataformas de energía IoT dirigidas a usuarios finales, donde se considera la selección de plataformas, la construcción de nuevas plataformas de energía y, finalmente, la comparación de plataformas. En [Shapsough et al., 2020], se

²<https://www.iberdrolaespana.com/sostenibilidad/innovacion/core-centro-operacion-energias-renovables-toledo>

presenta el diseño e implementación de un sistema de monitorización solar basado en IoT para instalaciones solares a gran escala y distribuidas en ciudades inteligentes. En [Gutierrez et al., 2020], se propone un sistema de seguimiento solar que permite aumentar la eficiencia de los sistemas fotovoltaicos. Este sistema ejecuta un algoritmo de seguimiento en el servicio web Firebase y permite el intercambio de datos con dicho servicio a través de una placa de desarrollo NodeMCU, que tiene un módulo Wi-Fi integrado. Finalmente, en [Karbhari and Nema, 2020], se analiza y discute el uso de paradigmas de IoT y aprendizaje automático para sistemas de monitorización de plantas de energía solar de próxima generación.

2.2. Modelos predictivos

En la intersección de las tecnologías de Internet de las Cosas y los métodos de aprendizaje automático para el análisis de flujos de datos de sensores, se han desarrollado y probado con éxito diversas técnicas especializadas en el procesamiento de señales y análisis de regresión. Destaca de forma sensible la definición de ventanas temporales simples y múltiples para segmentar y fusionar información temporal de los flujos de datos de sensores [Banos et al., 2015, Espinilla et al., 2018a], ajustando el tamaño de las ventanas que pueden ser asimétricas [López Medina et al., 2019, Medina-Quero et al., 2018], permitiendo agregar los datos de corto a largo plazo y enriqueciendo así las características de los flujos de sensores.

La utilización de modelos de aprendizaje profundo (*Deep Learning*) en series temporales ha marcado un campo de investigación prolífico [Gamboa, 2017], destacando el uso de Redes Neuronales Recurrentes de Memoria a Largo y Corto Plazo (LSTM, *Long Short-Term Memory*) [Hochreiter and Schmidhuber, 1997]. Estas redes LSTM

han demostrado ser adecuadas para modelar dependencias temporales y hacer predicciones a partir de datos de sensores [Wang et al., 2019, Quero et al., 2018], con resultados alentadores en reconocimiento de actividades [Medina-Quero et al., 2018] y estimación del consumo energético en edificios [Mocanu et al., 2016]. Además, el modelado de características espaciales en series temporales mediante Redes Neuronales Convolucionales (CNNs) [Wang et al., 2019, Qiu et al., 2017] ha logrado resultados prometedores en reconocimiento de voz [Abdel-Hamid et al., 2014] y clasificación de gases [Peng et al., 2018], combinadas con modelos LSTM [Huang and Kuo, 2018].

En el ámbito de la energía solar fotovoltaica, la integración de técnicas de aprendizaje profundo para la predicción y manipulación de la producción de energía es un dominio cada vez más explorado [Massaoudi et al., 2021a, Gaviria et al., 2022]. Los modelos de aprendizaje automático desarrollados desde datos supervisados han superado el rendimiento de modelos analíticos en la predicción inmediata de energía [Rajagukguk et al., 2020, Massaoudi et al., 2021b]. Las redes LSTM ha demostrado resultados alentadores en términos de rendimiento [Wang et al., 2021].

En el contexto de la predicción de energía, las redes neuronales recurrentes (RNNs) y las redes neuronales convolucionales (CNNs) han demostrado su eficacia en la captura de dependencias temporales y patrones espaciales en los datos fotovoltaicos [Ahn and Park, 2021] [Park and Ahn, 2019] [He et al., 2022] [Song et al., 2023] [Vlaminck et al., 2022] [Aprillia et al., 2020]. También se han sugerido modelos híbridos que combinan el análisis tradicional de series temporales (LSTM) con el aprendizaje profundo, lo que ha mejorado la precisión en la predicción a corto y largo plazo de la energía [Akhter et al., 2022] [Agga et al., 2022] [Tovar et al., 2020] [Lim et al., 2022] [Abdel-Nasser and Mahmoud, 2019]. Además, en otras áreas de la energía fotovoltaica, se ha propuesto el uso del aprendizaje automático para diagnos-

ticar la detección de fallas en los sistemas fotovoltaicos [Mellit et al., 2018], la conexión a la red de arquitecturas fotovoltaicas [Kazem et al., 2019], o los defectos en los módulos [Mellit and Kalogirou, 2021], utilizando imágenes aéreas [Ding et al., 2018] e imágenes infrarrojas [Akram et al., 2020], entre otros.

El aprendizaje mediante transferencia es una metodología utilizada para aplicar modelos de aprendizaje a partir de patrones reconocidos desde un dominio dado a otro dominio no conocido. Dentro de esta amplia definición, el aprendizaje homogéneo por transferencia se refiere en concreto al modelado predictivo en el que el dominio y el espacio de características de entrada son iguales [Pan and Yang, 2009]. Por otro lado, el aprendizaje por transferencia heterogéneo tiene como objetivo alinear los datos de entrada entre diferentes dominios fuente y objetivo que se representan en espacios de características diferentes [Pan and Yang, 2009]. Este trabajo se sitúa entre estos dos enfoques: es homogéneo en la medida en que el espacio de características de entrada de las señales para estimar la generación de energía de salida es el mismo; sin embargo, la velocidad de recopilación y la representación de los datos entre sistemas fotovoltaicos difieren. Para abordar esto, en este trabajo se propone un método de segmentación común para procesar el flujo de datos de sensores.

Los resultados del aprendizaje por transferencia son variados. Por ejemplo, los enfoques basados en instancias seleccionan muestras del dominio fuente para utilizarlas en el dominio objetivo y mejorar así el rendimiento del entrenamiento [Asgarian et al., 2018]. Los enfoques basados en parámetros se centran en transferir conocimientos mediante la compartición de estructuras desde un espacio latente del dominio fuente [Yao and Doretto, 2010]. Este tipo de aprendizaje por transferencia está relacionado con los enfoques de aprendizaje profundo, ya que los pesos que configuran los núcleos

aprendidos de las capas de la red describen hipotéticamente patrones comunes del dominio objetivo. Por lo tanto, a pesar de utilizar pesos inicializados a cero o al azar, la red aprendida se utiliza como punto de partida para el aprendizaje, y se procede con el aprendizaje afinado del dominio objetivo. Estos enfoques han proporcionado resultados alentadores en la reducción del tiempo de aprendizaje y el aumento del rendimiento, sobre todo en problemas de visión artificial [Kieffer et al., 2017].

En la predicción de series temporales, el uso del aprendizaje profundo y el aprendizaje por transferencia [Fawaz et al., 2018] se ha aplicado con éxito en diferentes escenarios, como la predicción de la temperatura en interiores [Romeu et al., 2013]. En el contexto de los sistemas fotovoltaicos, las capacidades de aprendizaje por transferencia han sido evaluadas recientemente [Zhou et al., 2020] con resultados prometedores. Los modelos LSTM se han propuesto para reducir los errores de predicción con la inicialización previa de pesos en un estudio de caso limitado a la misma región (China), con despliegue de la misma planta y el uso de la misma segmentación de datos en diferentes días para la evaluación.

Por otra parte, las Redes Generativas Adversarias (GANs) [Aggarwal et al., 2021], las cuales incluyen un generador y un discriminador entrenados en un proceso competitivo, han demostrado ser eficaces en la generación de datos sintéticos de alta calidad, mejorando la robustez de los sistemas [Lupion et al., 2022, Tarekegn et al., 2020, Singh and Raza, 2021, Chen et al., 2022]. Concretamente en sistemas de potencia, las GANs han sido utilizadas para la generación de datos de potencia sintéticos [Yilmaz and Korn, 2022, Romero et al., 2023, Tang et al., 2018], ampliando conjuntos de datos de entrenamiento [Luo et al., 2019, Nie et al., 2021, Lu et al., 2021] y mejorando la robustez de modelos de predicción [Wu et al., 2022, Lu et al., 2023]. La aplicación de redes adversarias para reducir la distorsión durante la generación

de datos, como se propone en este trabajo, representa un enfoque novedoso para mejorar la calidad de los datos, la predicción y la estimación de la desviación en la producción fotovoltaica [Huang et al., 2022, Li et al., 2022].

Sin embargo, el uso de Redes Adversarias en problemas de regresión para lidiar con generaciones falsas no identificables y corregir variaciones no ha sido explorado en profundidad. En un primer enfoque general, la evaluación de Redes Adversarias en problemas de regresión ha mostrado un rendimiento no competitivo en conjuntos de datos del mundo real [Aggarwal et al., 2019a]. No obstante, ha demostrado una alta adaptabilidad a problemas generales sin requerir conocimiento humano experto.

En este trabajo, se presenta una estrategia utilizando una red adversaria especializada para abordar la generación de potencia fotovoltaica y la adaptación del problema de regresión. Se extiende el concepto de generación de datos sintéticos [Yilmaz and Korn, 2022] a la manipulación de la salida de potencia fotovoltaica a través de una red que comprende un generador, un modelo de dopaje y un discriminador, colaborando dentro de una arquitectura superior de Red Adversaria condicionada por un nivel de ruido. Se utiliza una función de pérdida de desviación a lo largo del entrenamiento para asegurar un proceso de aprendizaje equilibrado para ambos componentes. El generador se enfoca en construir un modelo de regresión, mientras que el modelo de dopaje y el discriminador son entrenados para generar una variación y corregirla, respectivamente. Para diseñar este modelo de regresión de CGAN adecuado, se adopta un enfoque de entrenamiento similar al propuesto en [Aggarwal et al., 2019b].

A pesar de estos notables avances, es importante enfatizar que la integración del Aprendizaje Profundo y las Redes Adversarias en sistemas fotovoltaicos no ha explorado todas las aplicaciones potenciales, tales como, evaluar y modelar el ren-

dimiento de discriminadores robustos, tratar estimaciones no observables o estimar desviaciones en las predicciones de potencia de salida. Los modelos presentados en este estudio contribuyen a cerrar una brecha en la literatura explorando enfoques innovadores para mejorar la precisión en la predicción de la potencia de salida, la calidad de los datos y la gestión de las desviaciones a través de la sinergia de estas tecnologías de vanguardia.

Capítulo 3

Modelos de Aprendizaje

Automático y Aprendizaje

Profundo para la Predicción

Inmediata de Potencia

Fotovoltaica

3.1. Introducción

En el sector de la energía solar fotovoltaica, y en particular en el ámbito de la Operación y Mantenimiento (O&M, *Operation & Maintenance*) de generadores fotovoltaicos, la variable crítica que rige una gestión eficiente del sistema es la energía generada por el mismo. Esta magnitud representa la cantidad de electricidad con la que el sistema efectivamente contribuye a la red. Sin embargo, dado que la energía

generada por el sistema no es una magnitud que se pueda medir instantáneamente, se recurre a la medición y estimación de la potencia instantánea como un dato clave para monitorizar y gestionar el rendimiento de estos generadores fotovoltaicos.

La potencia instantánea se convierte así en un parámetro fundamental, ya que su predicción inmediata (o *nowcasting*) posibilita realizar una estimación de la cantidad de energía que debería estar generando el sistema. Esta estimación puede ser comparada con la energía real generada, medida a la salida del propio sistema fotovoltaico. El resultado de esta comparación permite verificar el correcto funcionamiento del generador fotovoltaico o, en caso contrario, la identificación de problemas con las correspondientes tareas de mantenimiento para devolver al sistema a un estado óptimo. Si además esta estimación de la potencia y el cotejo con el valor medido se realiza en tiempo real, se posibilita la pronta toma de decisiones y la implementación de estrategias efectivas para garantizar un funcionamiento óptimo, maximizando así la contribución de energía del sistema fotovoltaico a la red eléctrica.

En este capítulo se expone la evaluación exhaustiva realizada de las capacidades de los modelos basados en datos para la predicción inmediata de la potencia producida por un generador fotovoltaico. La información necesaria para alimentar estos modelos proviene tanto del funcionamiento del sistema fotovoltaico como de las condiciones ambientales en el propio sistema o en sus cercanías. En concreto, los datos utilizados para este trabajo son recogidos por un módulo de Internet de las Cosas, el cual recibe los datos medidos por el inversor del sistema fotovoltaico y por un conjunto de sensores ambientales instalados en el sistema. En este contexto, se evalúan dos tipos de modelos en función de su enfoque:

1. Modelos de aprendizaje automático, en los cuales la identificación de carac-

terísticas en el conjunto de datos y la definición múltiples ventanas temporales deben ser realizadas de manera manual.

2. Modelos de aprendizaje profundo basados en redes neuronales, los cuales se distinguen por su capacidad de extraer automáticamente características del conjunto de datos durante el proceso de entrenamiento.

Para cada uno de estos dos enfoques, se estudian varias configuraciones de segmentación y agregación de las series de datos mediante ventanas temporales. Los resultados tras la ejecución de estas diferentes configuraciones son comparados con un modelo analítico de referencia, en este caso el modelo de Araujo cuyo diseño y evaluación se detalla en la siguiente sección "3.2 - Proyecto Ópera: plataforma de procesamiento digital de datos para el modelado y control del comportamiento de sistemas fotovoltaicos".

En esta comparación se observa una mejora significativa para un subconjunto de los modelos estudiados, tanto en términos de error en la estimación como en el tiempo de aprendizaje del modelo. Por lo tanto, se confirma con este experimento un avance importante en este área de conocimiento mediante el uso de técnicas de aprendizaje automático y aprendizaje profundo con el objetivo de monitorizar la energía generada por un sistema fotovoltaico y poder así verificar su buen funcionamiento o posibles errores.

3.2. Proyecto Ópera: plataforma de procesamiento digital de datos para el modelado y control del comportamiento de sistemas fotovoltaicos

La necesidad imperante de desarrollar nuevas herramientas, aplicar las últimas tecnologías y concebir nuevos conceptos emerge como una respuesta ineludible para posibilitar el mantenimiento predictivo y la detección temprana de daños y errores, a fin de mantenerse a la vanguardia, anticipar eventos y realizar diagnósticos eficientes, con miras a lograr una gestión óptima, tanto técnica como económicamente, de los grandes sistemas fotovoltaicos. Todo esto debe llevarse a cabo de manera simple y económica para el cliente.

El propósito del Proyecto Ópera radica en el desarrollo de una plataforma digital de bajo coste, fundamentada en tecnologías de computación en la nube pública para el procesamiento digital de grandes volúmenes de datos (*Big Data*), y destinada a la gestión de la operación y el mantenimiento (*O&M*) en plantas fotovoltaicas de gran escala. El alcance de la plataforma pretende cubrir el proceso en su totalidad, es decir, desde la captura de datos, su almacenamiento y procesamiento, hasta la generación de cuadros de mando, informes, alertas y análisis predictivos.

La implementación de la plataforma ha ido evolucionando desde un prototipo básico de extremo a extremo de las capacidades descritas anteriormente hasta una plataforma plenamente funcional de alta disponibilidad y que incorpora funcionalidades avanzadas. Además, esta plataforma se encuentra actualmente dando soporte al sistema fotovoltaico de 200kW instalado en la Universidad de Jaén (Proyecto UniVer). Los resultados que se presentan en la secciones posteriores proceden de la

implementación original, los cuales permitieron la validación de del funcionamiento del prototipo desarrollado y el modelado del comportamiento del UniVer.

3.2.1. Métodos algebraicos para el modelado de sistemas fotovoltaicos

La monitorización de un Generador Fotovoltaico (GFV) debe llevarse a cabo conforme a la Norma Europea IEC 61724 [IEC, 1998]. En la tabla 3.1, se muestran los parámetros medidos para la supervisión del GFV de acuerdo con esta directriz. Es relevante destacar que, de acuerdo con la norma, el intervalo de muestreo (ΔT) de estas variables es de un minuto, y el intervalo de registro (valores promedio) es de 10 minutos.

Tabla 3.1: Parámetros Medidos del GFV

Parámetro	Símbolo	Unidad
Irradiancia en la superficie PV	GI	$\text{W}\cdot\text{m}^{-2}$
Temperatura ambiente	Tam	$^{\circ}\text{C}$
Corriente de salida del GFV	IA	A
Tensión de salida del GFV	VA	V
Potencia de salida del GFV	PA	kW
Corriente CA (a la red)	ITU	A
Tensión de red	VU	V
Potencia CA (a la red)	PTU	kW

A partir de estas variables medidas, y con los parámetros nominales del GFV en condiciones estándar de prueba (STC), se puede obtener otros parámetros calculados y ratios de pérdidas y rendimiento, útiles para evaluar el comportamiento del sistema fotovoltaico. Los más interesantes para este estudio se muestran en la Tabla 3.2. Estas variables se calculan en un período de tiempo τ . Normalmente puede ser una hora, un día, un mes, un año (es decir, $\tau = h, d, m, a$).

Tabla 3.2: Variables Calculados

Parámetro	Símbolo	Unidad
Irradiación global en la superficie PV	HI, τ	$\text{kWh} \cdot \text{m}^{-2} \cdot \tau^{-1}$
Energía neta del GFV (CC)	EA, τ	kWh
Energía neta a la red (CA)	$ETUN, \tau$	kWh
Eficiencia de Balance de Sistema	η_{BOS}	Adimensional
Rendimiento del GFV	YA	$\text{h} \cdot \tau^{-1}$
Rendimiento Final	Yf	$\text{h} \cdot \tau^{-1}$
Rendimiento de Referencia	Yr	$\text{h} \cdot \tau^{-1}$
Ratio de Rendimiento	PR	Adimensional
Pérdidas de Captura	LC	$\text{h} \cdot \tau^{-1}$
Pérdidas de Balance de Sistema	LBOS	$\text{h} \cdot \tau^{-1}$

Por otro lado, es necesario comparar esos parámetros y ratios medidos con sus equivalentes teóricos para poder implementar un método de diagnóstico efectivo y una detección temprana de daños y fallas con el fin de procurar un mantenimiento predictivo. Para lograr este propósito, es necesario definir un modelo de comportamiento para los sistemas fotovoltaicos. Existe una gran cantidad de métodos o modelos para prever el comportamiento teórico de los sistemas fotovoltaicos, con enfoques y complejidades muy diferentes [Almeida et al., 2017b, Rus-Casas et al., 2014b, Parra et al., 2017].

En este caso, no es necesaria una precisión especial, por lo que es suficiente con un método algebraico simple para corregir desde STC a cualquier condición de funcionamiento. Se ha elegido el llamado "Punto de Máxima Potencia Aproximado" (AMPP) y se ha asumido que el factor de llenado (FF, *Fill Factor*) permanece constante en todas las condiciones de funcionamiento. Este modelo aproxima la corriente y el voltaje del GFV en el punto de máxima potencia tanto desde la corriente de cortocircuito como desde el voltaje de circuito abierto utilizando las expresiones de Araujo & Sánchez [Araujo and Sánchez,] y Nofuentes y colaborado-

res [Fuentes et al., 2007b]. Estudios anteriores han demostrado que este modelo es lo suficientemente preciso, mientras que una mayor complejidad no sería beneficiosa en términos de mejora de la precisión [Fuentes et al., 2007b, Drif et al., 2007a].

Así, se puede calcular la previsión teórica de las variables y parámetros descritos anteriormente, y establecer una comparación, en tiempo real, entre los valores reales (monitorizados y calculados) y los teóricos previstos (del modelo), con el fin de establecer alertas y advertencias para un análisis de rendimiento en tiempo real, diagnóstico y mantenimiento predictivo.

Más allá de la implementación de este método algebraico, en versiones posteriores de la plataforma y gracias a los trabajos desarrollados en esta tesis, se han incluido modelos de inteligencia artificial en el software de la plataforma, los cuales incorporar mejoras y eficiencias en el modelado, predicción y control del comportamiento del sistema.

3.2.2. Arquitectura técnica de la plataforma

La solución técnica de la implementación original de esta plataforma distribuida de datos está basada en una arquitectura lambda [Yamato et al., 2016, Lakhe, 2016] como se representa en la figura 3.1. Las principales ventajas de esta arquitectura son la escalabilidad del sistema en términos de almacenamiento y potencia de procesamiento (es decir, la capacidad de aumentar el tamaño fácilmente mediante la adición de nodos a la plataforma) y la capacidad de procesar datos tanto en flujos de datos en tiempo real como por lotes.

Los flujos de datos de los sistemas fotovoltaicos y fuentes externas son enviados a un sistema de almacenamiento de flujos de publicación/suscripción. Desde allí, los

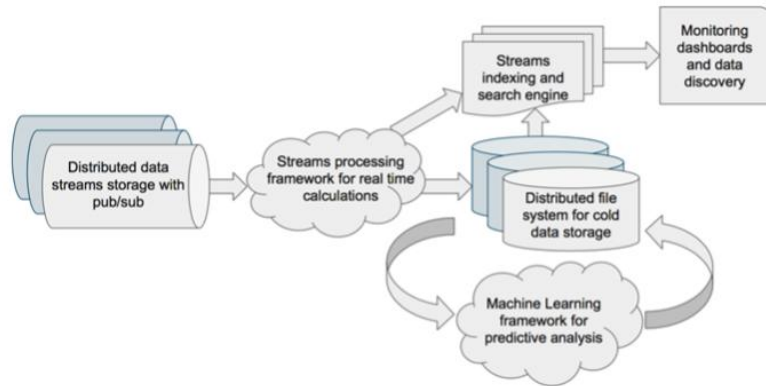


Figura 3.1: Diagrama esquemático de la arquitectura de la plataforma Ópera

datos son procesados a través de la capa de tiempo real de la plataforma mediante un marco de procesamiento de flujos de datos para implementar todos los cálculos de los indicadores de rendimiento, valores calculados y valores teóricos. El flujo de datos enriquecidos es posteriormente indexado en un motor de búsqueda para permitir la monitorización en tiempo real del sistema. Un componente de visualización proporciona los paneles de control de monitorización donde se publican los indicadores y gráficos de rendimiento, permitiendo también el descubrimiento de datos para los usuarios finales a través de consultas detalladas.

El componente central de la capa por lotes de la plataforma es un sistema de archivos distribuido optimizado para el almacenamiento de grandes cantidades de datos. El flujo de datos se dirige hacia este sistema de archivos para persistir los datos y almacenar información histórica de los sistemas. Estos datos históricos son la entrada para el entrenamiento de modelos de análisis predictivo con el objetivo de identificar y anticipar problemas de rendimiento en los sistemas fotovoltaicos. Los modelos se ejecutan con los datos actuales en lotes programados (frecuencia a determinar, por ejemplo, semanal o dos veces al día) para generar las alertas o

informes necesarios para el O&M proactivo, y los resultados también se indexan en el motor de búsqueda para la visualización de las métricas en los paneles de control de gestión.

En versiones posteriores de la plataforma se ha movido el alojamiento a la nube y se ha incluido una base de datos de series temporales para optimizar el almacenado y recuperación de datos indexados por fecha, lo cual ha permitido la ejecución del proceso de inferencia también en tiempo real en vez de por lotes. Por último, también se ha incorporado una interfaz para el intercambio de datos con aplicaciones (API) y sistemas de mensajería para una mayor integración de la plataforma con el ecosistema de herramientas digitales del sector. Esto ha permitido por ejemplo calcular los beneficios económicos generados por el sistema gracias a la integración con los precios del mercado eléctrico proporcionados vía API por Red Eléctrica Española.



Figura 3.2: Aplicación web Ópera para la monitorización del GFV UniVer

3.2.3. Resultados experimentales

Los resultados presentados en esta sección presentan de forma detallada la implementación del prototipo básico de la plataforma digital descrita anteriormente utilizando el método algebraico de Araujo. Para validar su comportamiento y funcionamiento, el prototipo fue puesto a prueba en el generador fotovoltaico perteneciente al Proyecto UniVer [Vidal et al., 2001].

El sistema fotovoltaico del proyecto UniVer ha sido estudiado durante mucho tiempo. Se dispone de una gran cantidad de datos de operación y es bien conocido su rendimiento y comportamiento [Drif et al., 2007a, Pérez et al., 2003]. Por lo tanto, es un proyecto muy adecuado para experimentar y validar esta plataforma y diferentes modelos de predicción. Sin embargo, ha sido necesario implementar un nuevo mecanismo de recolección de datos para obtener los datos en tiempo real de las variables operativas del sistema (Tabla 3.1), para lo cual se han utilizado sensores que permiten la adquisición de datos mediante la tecnología de Internet de las Cosas, tal y como se detalla en la siguiente sección 3.3.

En las figuras 3.3 y 3.4 se muestran algunas imágenes representativas de los parámetros de predicción utilizando el método de Araujo comparados con los valores operativos reales del GFV del Proyecto UniVer.

Los resultados de este experimento validan la arquitectura y configuración del prototipo desarrollado, así como el uso del método algebraico de Araujo para la estimación de la energía producida por un sistema fotovoltaico.

De esta forma, se puede confirmar que el uso de nuevas tecnologías de la información y las aplicaciones de datos, pueden ser claramente muy útiles en la operación y mantenimiento de plantas fotovoltaicas de gran tamaño, y su introducción en la

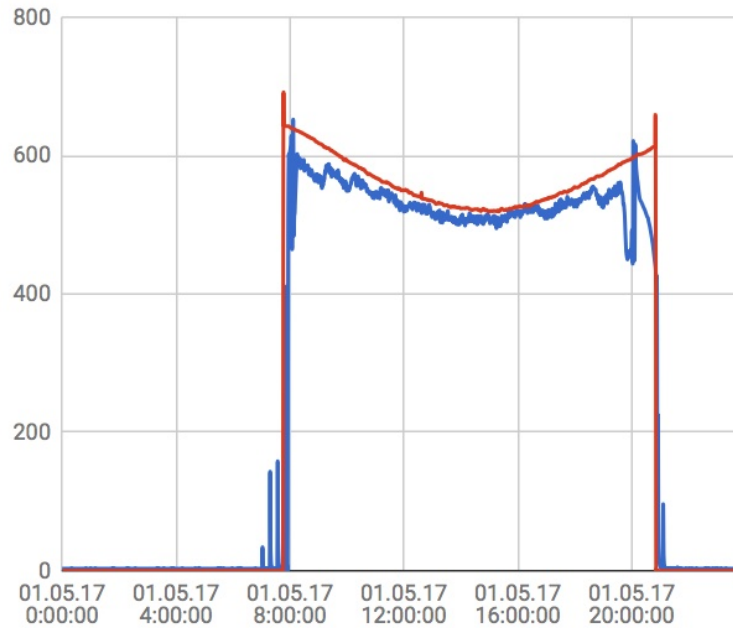


Figura 3.3: Tensión DC del GFV (UniVer) en un día representativo. Línea roja: valores pronosticados por el modelo Araujo, línea azul: valores reales.

industria fotovoltaica debe permitir alcanzar una gestión óptima de estos sistemas. Para ello, es necesario establecer un equipo de trabajo interdisciplinario capaz de llevar a cabo tareas en ambos campos: energía solar fotovoltaica y tecnologías de la información.

Uno de los principales objetivos para mejorar la gestión de los sistemas fotovoltaicos es reducir los gastos operativos, por lo que estas nuevas herramientas deben tener un bajo costo. Esto se puede conseguir con servicios en la nube pública, los cuales permiten un escalado de los servicios para adaptar el consumo de recursos a la demanda de los procesos, a la vez que se benefician de la economía de escala. Además, deben ser fáciles de usar para todo tipo de clientes que no necesariamente tienen que ser profesionales en los sectores fotovoltaico o de tecnologías de la información, para lo cual es de gran valor construir interfaces de usuario de tipo cuadro de

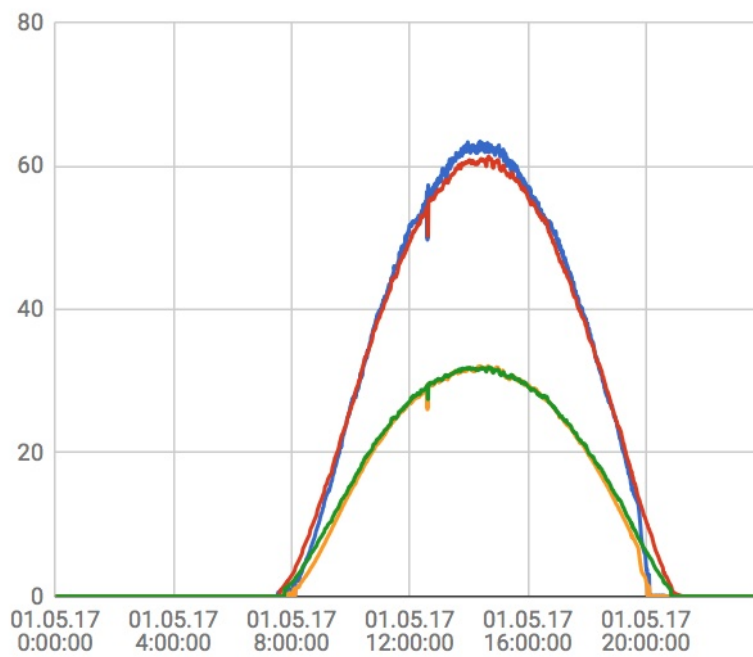


Figura 3.4: Corriente DC (azul, roja) y potencia (verde, amarilla) del GFV (UniVer) en un día representativo. Línea roja y verde: valores pronosticados por el modelo Araujo, línea azul y amarilla: valores reales.

mandos basados en herramientas de inteligencia de negocio (*Business Intelligence*), aislando al usuario del proceso de almacenamiento y procesamiento de datos.

3.3. Adquisición de datos en tiempo real mediante dispositivos IoT

El módulo inspirado bajo el paradigma IoT encargado de recoger los datos para este trabajo pertenece al proyecto Ópera, la iniciativa descrita en la sección anterior, encargada de desarrollar una plataforma digital por un equipo interdisciplinario que abarca las áreas de las tecnologías de la información y la comunicación (TIC), la energía fotovoltaica y la tecnología electrónica. Esta plataforma ha sido diseñada para proporcionar servicios de gestión de operación y mantenimiento (O&M) para instalaciones de energía renovable en base al conocimiento y datos del Proyecto UniVer.

El generador fotovoltaico desarrollado por el proyecto UniVer, como se muestra en la figura 3.5, es un sistema fotovoltaico estándar conectado a la red, de tamaño mediano, que ha estado funcionando durante los últimos veinte años en el Campus de la Universidad de Jaén [Drif et al., 2007b]. Los módulos fotovoltaicos están compuestos por 60 células solares de silicio multicristalino con una eficiencia del 18.34 % y una superficie de 156.75 x 156.75 mm². El generador fotovoltaico está compuesto por 220 de estos módulos, con una topología de 20 módulos en serie por 11 módulos en paralelo. La potencia total del generador fotovoltaico es de 59.4kW en condiciones de prueba estándar (STC, *Standard Test Conditions*), es decir, 1000 W/m² de irradiancia normal sobre las células, temperatura de célula de 25°C y espectro solar

AM1.5.



Figura 3.5: Dos vistas del generador fotovoltaico del proyecto UniVer

El factor crítico de éxito de esta solución basada en Internet de las Cosas para la optimización de la Operación y Mantenimiento de generadores fotovoltaicos es su capacidad para monitorizar a bajo coste la energía generada por el sistema. La energía E_T es el producto final de cada generador eléctrico y se calcula como la integral de la potencia instantánea P en un período de tiempo T : $E_T = \int_T P \cdot dt$. La potencia eléctrica medida a la salida del sistema es la variable instantánea recogida por el módulo de adquisición de datos, la cual se usa como variable objetivo de los modelos de aprendizaje automático para realizar la predicción inmediata del comportamiento del sistema fotovoltaico. Esta salida depende principalmente de la radiación solar, cuya magnitud, la irradiancia solar identificada por la variable G , está definida por la densidad de potencia incidente en una superficie medida en vatios por metro cuadrado (W/m^2). La temperatura ambiente, la temperatura en el módulo y las especificaciones del generador fotovoltaico (GFV) son las otras entradas para este sistema de recopilación de datos que monitoriza el rendimiento del sistema.

Como ya se ha comentado anteriormente, la monitorización del GFV debe realizarse siguiendo la Norma Europea IEC 61724 [IEC, 2017b]. De acuerdo con esta

norma, las variables que se han medido para realizar este experimento se muestran en la tabla 3.3. A partir de estos datos medidos y con las especificaciones nominales del GFV en STC, se calculan los parámetros derivados y métricas relacionadas con las pérdidas en el rendimiento energético, como son:

- Irradiación global en la superficie del generador fotovoltaico.
- Energía neta del generador fotovoltaico durante un período de tiempo concreto.
- Relación de rendimiento (PR, *Performance Ratio*).
- Producción y pérdidas de energía.

Estos parámetros son útiles para evaluar el comportamiento del sistema fotovoltaico y muy valiosos para el diagnóstico de errores y el análisis descriptivo de la operación. Todos ellos están bien definidos analíticamente y conceptualmente en [IEC, 2017b], y su función, significado y utilidad también se describen en [Klise et al., 2018], [Blaesser, G; Munro, 1996] y [Blaesser and Zaaiman, 1991]. Este trabajo se centra en la predicción de la generación de energía de salida, que está directamente relacionada con las métricas analíticas sobre el comportamiento del sistema fotovoltaico.

La adquisición y envío de información ambiental y de generación de energía en tiempo real a la plataforma digital Ópera es posible gracias al desarrollo e implementación de una integración original de sensores ambientales y de suministro de energía. Este sistema está compuesto por un conjunto de sensores basados en tecnología IoT y controlados por un microprocesador que envía los datos a través de una red inalámbrica. Estos sensores miden los datos de funcionamiento del GFV y las variables ambientales mostradas en la tabla 3.3, necesarias para monitorizar y estimar la operación del GFV de acuerdo con la norma [IEC, 2017a].

Tabla 3.3: Variables recogidas por el sistema de adquisición de datos

Parámetro	Símbolo	Unidad
Irradiancia en la superficie fotovoltaica	G_I	$W \cdot m^{-2}$
Temperatura ambiente	T_{am}	$^{\circ}C$
Temperatura del módulo	T_{mod}	$^{\circ}C$
Corriente de salida del GFV	I_A	A
Tensión de salida del GFV	V_A	V
Potencia de salida del GFV	P_A	W

La unidad central del módulo IoT consiste en un microcontrolador programable de la plataforma *Arduino*. Este dispositivo de placa estándar incluye, además del microprocesador, un acondicionador de datos de entrada, una interfaz de red de comunicación y otras interfaces de visualización. Este módulo de IoT se encarga de recopilar los datos fotovoltaicos y ambientales y enviar la información a la plataforma Ópera, alojada en servidores en la nube pública, a través de una conexión a Internet (por ejemplo, cableada, vía WiFi o módem). El módulo se alimenta mediante una fuente de alimentación estándar o mediante un panel solar complementado con una batería.

Los sensores ambientales conectados al microcontrolador *Arduino* recogen las siguientes variables:

- Irradiancia solar en la superficie fotovoltaica.
- Temperatura del módulo.
- Temperatura ambiente.

El sensor de irradiancia solar es una célula solar de silicio calibrada (certificado de calibración del CIEMAT, el Centro de Investigación Español en Energía, Medio Ambiente y Tecnología), con una salida analógica de 0 a 5V que corresponde a un

rango de radiación de 0 a 1250W/m². Los sensores de la temperatura ambiental y de temperatura de las células son sondas Pt100 de cuatro hilos, también con una salida analógica de 0 a 5V, que corresponde a un rango de temperatura de -20 a 130° C. Estos dos sensores, junto con la circuitería de interfaz correspondiente, están incluidos en una unidad comercial fabricada por *Atersa S.L.* (www.atersa.com), como se muestra en la figura 3.6.

Los sensores ambientales se instalan cerca de los paneles fotovoltaicos y se alimentan mediante su propio mini-módulo solar. Estos dispositivos envían los datos medidos al microprocesador *Arduino* utilizando el protocolo *Zigbee*. Este protocolo permite la comunicación inalámbrica directa entre los dispositivos y la placa *Arduino* en áreas abiertas [Ferreira et al., 2013] [Ferreira et al., 2013], lo que es inherente en la implementación de sistemas fotovoltaicos. La elección del protocolo *Zigbee* se basa en resultados experimentales previos que demuestran su mayor eficiencia en términos de consumo de energía en comparación con otras tecnologías inalámbricas populares, como *Wi-Fi* y *Bluetooth* [Siekkinen et al., 2012].

Además de los sensores ambientales, el módulo IoT también recoge la información del generador fotovoltaico correspondiente a los siguientes parámetros:

- Tensión de salida, voltaje generado por el sistema fotovoltaico medido en voltios (V).
- Intensidad de corriente de salida, cantidad de corriente eléctrica del GFV medida en amperios (A),

Estos dos datos permiten obtener la potencia instantánea de salida multiplicando la tensión de salida y la intensidad de corriente. El resultado de esta operación es válido ya que los datos son valores instantáneos. En el caso objeto de este estu-



Figura 3.6: Sensores de radiación y temperatura

dio, la salida del GFV es de corriente continua, por lo que esta forma de obtener la potencia también es válida para valores medios durante un período de tiempo. Alternativamente, se puede instalar un sensor de potencia de salida, como un medidor de potencia o un analizador de red, para obtener cierta redundancia en los datos medidos y capturar además algunos parámetros eléctricos de salida secundarios adicionales mediante este dispositivo adicional.

En la figura 3.7 se muestran los sensores de voltaje y corriente, junto con la unidad de microprocesador utilizada para recoger los datos de operación del generador fotovoltaico del proyecto UniVer. La figura 3.8 muestra un diagrama esquemático de la arquitectura de adquisición de datos.



Figura 3.7: Microprocesador y sensores de corriente y tensión

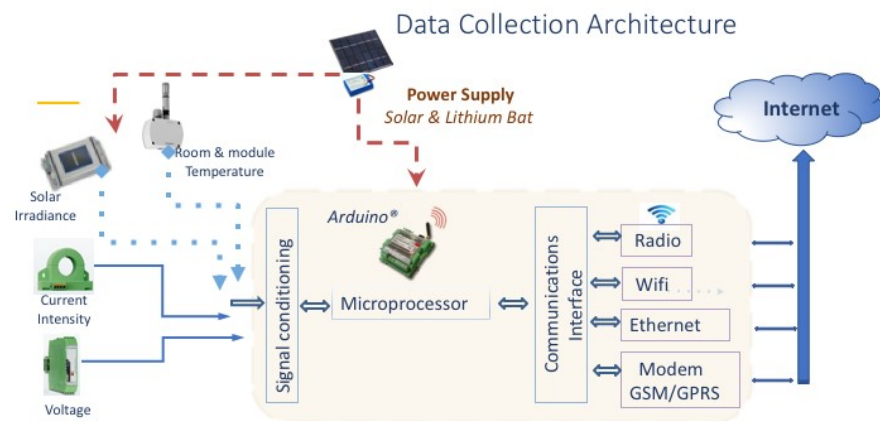


Figura 3.8: Arquitectura del módulo IoT para la adquisición de datos en tiempo real

3.4. Metodología para el procesamiento de datos

Con el fin de realizar predicciones en tiempo real de la potencia producida por un generador fotovoltaico a partir de la información de sensores ambientales, resulta esencial definir una metodología para el procesamiento, segmentación y modelado de las variables asociadas a un sistema fotovoltaico. Esta metodología permite que estos parámetros puedan ser utilizados como señales de entrada (o características) y de salida (o variables objetivo) de los diferentes algoritmos de regresión planteados en este trabajo.

Los enfoques que se evalúan en este experimento se agrupan principalmente en dos tipos:

- Modelos de aprendizaje automático, donde las características son diseñadas manualmente y calculadas mediante múltiples ventanas temporales.
- Modelos de aprendizaje profundo basados en redes neuronales, donde las características son extraídas de forma automática por el propio modelo.

En los apartados siguientes de esta sección, se desarrollan los siguientes aspectos:

1. La segmentación básica con ventanas temporales deslizantes para flujos de datos de sensores en un modelo basado en el análisis de datos.
2. El modelado de características diseñadas manualmente y múltiples ventanas temporales.
3. La configuración de modelos de aprendizaje profundo con extracción automática de características.

3.4.1. Segmentación y agregación de series temporales para modelos basados en el análisis e interpretación de datos

Siguiendo una definición formal, un sensor s recopila datos en tiempo real en forma de un par $\bar{s}_i = \{s_i, t_i\}$, donde s_i representa una medición dada y t_i corresponde a la marca de tiempo asociada a dicha medición. Por lo tanto, el flujo de datos del sensor fuente s está definido por $\overline{S}_s = \{\bar{s}_0, \dots, \bar{s}_i\}$, mientras que un valor dado en una marca de tiempo t_i está definido por $S_s(t_i) = s_i$.

En este experimento, el comportamiento y la producción de energía del sistema fotovoltaico están descritos por los siguientes cinco flujos de datos:

- Irradiancia en la superficie fotovoltaica G_I .
- Temperatura ambiente T_{am} .
- Corriente de salida del generador fotovoltaico I_A .
- Tensión de salida del generador fotovoltaico V_A .
- Potencia de salida del generador fotovoltaico P_A .

Una vez identificados los flujos de datos que forman parte del modelo, se diseña para ellos un conjunto de ventanas temporales deslizantes definidas por el tamaño de la ventana de un intervalo de tiempo $W_w = [W_w^-, W_w^+]$ [Banos et al., 2014]. Estas ventanas deslizantes segmentan las muestras del flujo de datos de un sensor \overline{S}_s y agregan los valores \overline{s}_i mediante una función de agregación dada $T_t(S_s, W_w, t^*)$:

$$T_t(S_s, W_w, t^*) = \bigcup_{s_i}^{\overline{s}_i} s_i, t_i \in [t^* - W_w^-, t^* - W_w^+]$$

El valor de agregación de esta ecuación define una característica dada T_t de los sensores S_s en un tiempo actual t^* . En la figura 3.9, se ilustra este proceso de segmentación y agregación de datos mediante ventanas temporales deslizantes con algunos ejemplos visuales de flujos de datos.

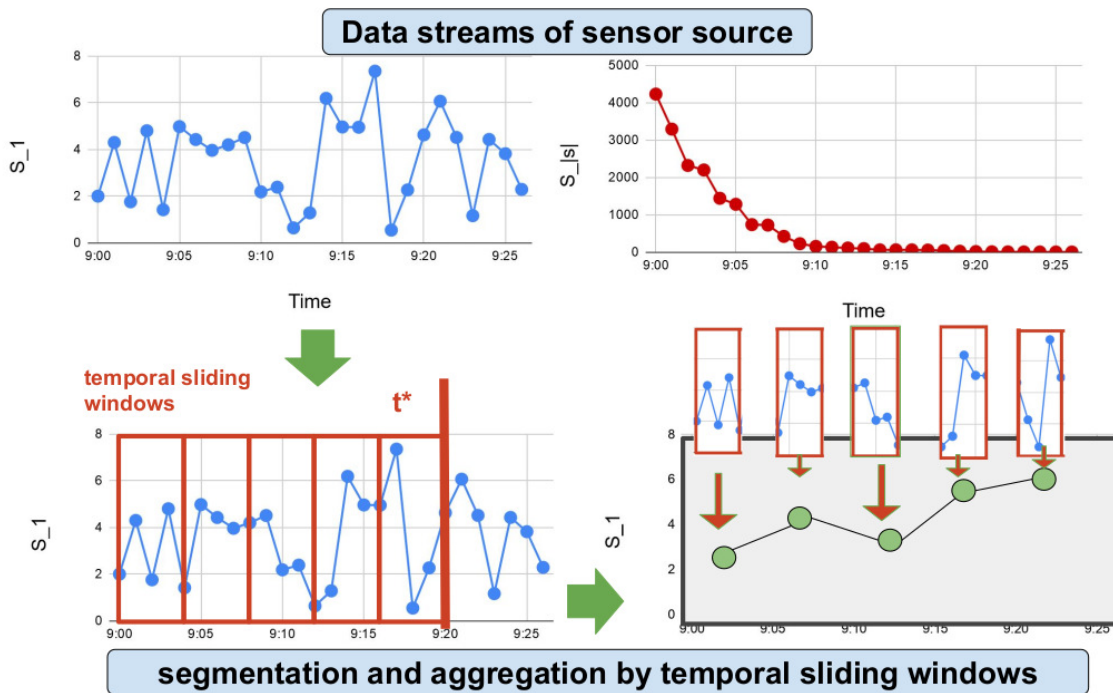


Figura 3.9: Ejemplo de flujos de datos de sensores, segmentación y agregación mediante ventanas temporales deslizantes.

3.4.2. Definición manual de características y múltiples ventanas temporales

Una de las opciones para procesar flujos de datos de manera adecuada para el entrenamiento de un regresor implica utilizar una metodología en la cual las características se diseñan manualmente, tomando como base múltiples ventanas temporales deslizantes. En este enfoque, un experto en la materia debe definir una función de agregación para procesar los flujos de datos provenientes de los sensores. De este modo se genera un vector de características que se utiliza para el aprendizaje de un modelo basado en datos durante la etapa de entrenamiento.

Dentro del amplio espectro de modelos de aprendizaje automático, la categoría elegida para este experimento son los modelos de regresión, ya que su eficiencia per-

mite la ejecución en microprocesadores de las tareas de aprendizaje y evaluación en tiempo real bajo entornos de computación distribuida [López Medina et al., 2019]. En este trabajo se evalúa por tanto un enfoque de características diseñadas manualmente para estos regresores [Cruciani et al., 2019], donde las funciones de agregación y las múltiples ventanas de diferentes tamaños son definidas manualmente por expertos. En concreto, se opta por la siguiente configuración para los modelos de regresión:

- Funciones de agregación T_t basadas en métricas estadísticas, como máximo, mínimo, promedio y desviación estándar. Se opta en esta configuración por estas funciones ya que se ha demostrado en estudios previos que son características relevantes para describir flujos de datos de sensores [Espinilla et al., 2018b].
- Segmentación y fusión de información temporal de flujos de datos de sensores con las siguientes configuraciones de ventanas, para agregar datos de plazos más cortos a más largos enriqueciendo las características de los flujos de datos de sensores. El tamaño de la ventana también está definido según los criterios de un experto.
 - Ventana única
 - Múltiples ventanas [Banos et al., 2015]
 - Ventanas incrementales [López Medina et al., 2019]
- Clasificación utilizando diferentes regresores eficientes, los cuales presentan un bajo tiempo de aprendizaje y pocos requerimientos de entrenamiento:
 - Regresión lineal

- K vecinos más cercanos (k-NN, *k-Nearest Neighbours*)
- Máquinas de soporte vectorial (SVM, *Support Vector Machine*)
- Bosques aleatorios (*Random Forest*).

Por lo tanto, partiendo de un conjunto de sensores de entrada $S = \{S_1, \dots, S_s, \dots, S_{|S|}\}$, un conjunto de tamaños de ventana $W = \{W_1, \dots, W_w, \dots, W_{|W|}\}$ y un conjunto de funciones de agregación $T = \{T_1, \dots, T_t, \dots, T_{|T|}\}$, se define un número total de características $|S| \times |W| \times |T|$ que describen los flujos de datos de sensores S para cada punto de tiempo t^* [Espinilla et al., 2018b]. Dado que el modelo aquí descrito se basa en un enfoque supervisado basado en datos, las características que describen los flujos de datos de sensores están asociadas para cada punto de tiempo t^* con el sensor objetivo para el que se desea realizar la predicción S^* , el cual no está incluido en los sensores de entrada $S \cap S^* = \emptyset$:

$$T_1(S_1, W_1, t^*), \dots, T_t(S_s, W_w, t^*), \dots, T_{|T|}(S_{|S|}, W_{|W|}, t^*) \rightarrow S^*(t^*)$$

3.4.3. Extracción automática de características con modelos de aprendizaje profundo

Una segunda opción en cuanto a la metodología consiste en el uso de modelos de aprendizaje profundo para la predicción de la generación de energía en un sistema fotovoltaico. A diferencia de la propuesta anterior basada en regresores clásicos, los modelos de aprendizaje profundo no requieren características diseñadas manualmente y simplemente se aplica un preprocesamiento para calcular una secuencia homogénea de datos entre las diferentes frecuencias de muestreo de las fuentes de

datos. En este enfoque, se define una segmentación de señal mínima mediante ventanas temporales deslizantes con un tamaño de ventana de corta duración, definida por una granularidad temporal mínima Δ . Los datos originales se promedian con $\bigcup = \mu$ para cada ventana temporal de corta duración dentro de un segmento.

Siguiendo esta metodología, se obtiene una secuencia de datos para cada sensor cuyo tamaño de secuencia es el mismo para todas las fuentes S_s , y que están relacionados con el sensor objetivo a pronosticar S^* para cada momento actual t^* bajo un enfoque de ventana deslizante:

$$S^*(t^*) \rightarrow \begin{cases} \mu(S_1, [0, \Delta], t^*), \mu(S_1, [\Delta, 2\Delta], t^*), \dots, \mu(S_1, [\Delta|W| - \Delta, \Delta|W|], t^*) \\ \dots \\ \mu(S_{|S|}, [0, \Delta], t^*), \mu(S_{|S|}, [\Delta, 2\Delta], t^*), \dots, \mu(S_{|S|}, [\Delta|W| - 1, \Delta|W|], t^*) \end{cases}$$

Una vez definidos los datos de entrada y salida del modelo de aprendizaje profundo, se proponen para este experimento dos arquitecturas de redes neuronales que se han mostrado como configuraciones adecuadas para secuenciar series temporales en enfoques de ventana deslizante en otros trabajos previos [Selvin et al., 2017]:

- 2LSTM: dos capas de memoria a corto-largo plazo (LSTM, *Long Short-Term Memory*). Esta arquitectura previamente se ha identificado como una configuración adecuada para predecir la carga de energía [Marino et al., 2016].
- 3CNN+2LSTM: tres capas de redes convolucionales neuronales (CNN, *Convolutional Neural Network*) se integran primero como extractores de características espaciales. A continuación, dos capas de LSTM modelan las dependencias temporales de CNN. La combinación de redes híbridas CNN-LSTM se ha se-

leccionado debido a que proporciona resultados alentadores en el modelado de la generación de energía [Kim and Cho, 2018].

En la tabla 3.4, se incluyen los parámetros y capas para cada modelo propuesto.

Tabla 3.4: Configuración de las arquitecturas basadas en redes neuronales

2LSTM	3CNN+2LSTM
LSTM(32 units)	2 kernels x 16 filters
dropout(0.25)	Re-Lu
LSTM(32 units)	2 kernels x 32 filters
dropout(0.25)	Re-Lu
connected(1 unit)	2 kernels x 64 filters
activation function:Re-Lu	Re-Lu
loss function:MAE	dropout(0.25)
	LSTM(32 units)
	dropout(0.25)
	LSTM(32 units)
	dropout(0.25)
	connected(1 unit)
	activation function:Re-Lu
	loss function:MAE

3.5. Evaluación de resultados

Con el fin de valorar estas metodologías propuestas para la predicción inmediata de la potencia generada por un sistema fotovoltaico, se llevan a cabo los diferentes experimentos siguiendo configuraciones concretas para los parámetros de las arquitecturas definidas. Estos experimentos se realizan sobre datos reales de sistemas fotovoltaicos, permitiendo así obtener resultados que puedan ser comparados y extraer conclusiones relevantes acerca del uso de estos modelos.

3.5.1. Configuración experimental

La configuración experimental y los resultados descritos en esta sección pertenecen a un caso de estudio desarrollado en la Universidad de Jaén, donde se encuentran instalados el sistema fotovoltaico UniVer y la plataforma Ópera. El módulo IoT que recopila los datos fotovoltaicos en tiempo real dentro de la plataforma Ópera estuvo activo desde el 9 de junio hasta el 23 de noviembre de 2019, generando un conjunto de datos de 168 días de duración. La ubicación del dispositivo fotovoltaico y del módulo IoT en el campus de la Universidad de Jaén es (latitud: 37.787253, longitud: -3.776258), como se muestra en la figura 3.10.

En el marco experimental, gracias a los sensores instalados en el sistema fotovoltaico tal como se describe en la sección "3.3 - Adquisición de datos en tiempo real mediante dispositivos IoT", se dispone para este trabajo de series temporales de datos correspondientes a las siguientes variables:

- Irradiancia solar.
- Temperatura ambiente.
- Temperatura del módulo.
- Corriente de salida.
- Voltaje de salida.

La potencia de salida del sistema fotovoltaico, objetivo de la predicción mediante el modelo de aprendizaje automático, se obtiene utilizando la corriente de salida y el voltaje de salida de acuerdo con la siguiente ecuación: $P = V \cdot I$.

A continuación, se detallan las diferentes configuraciones para los dos enfoques principales de la metodología propuesta.



Figura 3.10: Ubicación del dispositivo fotovoltaico y módulo IoT en la Universidad de Jaén y fotografías de la implementación.

1. Características definidas manualmente con múltiples ventanas temporales. Se evalúa el rendimiento de modelos de regresión de aprendizaje automático, los

cuales utilizan características predefinidas por expertos y múltiples ventanas temporales. La configuración para los modelos de regresión lineal, bosques aleatorios, K vecinos mas cercanos y máquinas de soporte vectorial se muestran respectivamente en las tablas 3.5, 3.6, 3.7 y 3.8.

Tabla 3.5: Configuración del modelo de Regresión Lineal

Parámetro	Valor
Fit intercept	True
Copy X	True
Normalize	False

Tabla 3.6: Configuración del modelo Random Forest

Parámetro	Valor
bootstrap	True
ccp alpha	0.0
criterion	mse
max depth	None
max features	auto
max leaf nodes	None
max samples	None
min impurity decrease	0.0
min impurity split	None
min samples leaf	1
min samples split	2
min weight fraction leaf	0.0
n estimators	100
oob score	False
random state	None

Para cada uno de estos cuatro modelos, se definen y evalúan tres configuraciones de ventanas temporales deslizantes:

- $T = 10min$, una sola ventana temporal de 10 minutos.
- $T = 30min$, tres ventanas temporales de 10 minutos cada una.

ra $\Delta = 5m$. Esta configuración genera un total de 24,031 muestras para el aprendizaje del modelo.

- $\Delta = 5m$, definido por una secuencia de datos de noventa minutos cuya longitud de secuencia es $|W| = 18$, $W = \{[0m, 4m], [5m, 9m], \dots, [85m, 89m]\}$ para $\Delta = 5m$. Esta configuración genera un total de 48,062 muestras para el aprendizaje del modelo.

Para predecir la generación de energía de salida a partir de los datos ambientales recopilados en el sistema fotovoltaico, se comparan las predicciones con la verdad fundamental en las pruebas utilizando una validación cruzada de treinta particiones. Cabe destacar que los datos ambientales de las fuentes fotovoltaicas son normalizados utilizando el método de máximos y mínimos en una etapa previa de pre-procesamiento de datos.

Tanto los datos como los modelos de aprendizaje están disponibles de forma pública para la comunidad científica en el siguiente repositorio de *GitHub*: <https://github.com/galmonacid/opera/>.

3.5.2. Resultados

Para la obtención y comparación de los resultados de los diferentes modelos de predicción con sus respectivas configuraciones, se utiliza la potencia de salida del sistema fotovoltaico recopilada por el módulo IoT como representación de la verdad fundamental. La potencia estimada por cada modelo utiliza como datos de entrada únicamente datos de los sensores ambientales. La predicción se compara frente a la verdad fundamental para el rango de tiempo completo utilizando las siguientes métricas de error:

- Error Absoluto Medio (MAE, *Mean Absolute Error*)
- Raíz del Error Cuadrático Medio (RMSE, *Root Mean Square Error*)
- Coeficiente de determinación (R^2)

Con la configuración de validación cruzada de treinta particiones, también se calcula el tiempo de aprendizaje y el tiempo de evaluación para evaluar el consumo de recursos computacionales de los modelos.

El modelo de Araujo proporciona, mediante el método analítico estándar, una estimación inicial de la potencia del sistema fotovoltaico que se utiliza de base para la comparación posterior de los modelos de predicción. Las métricas de error de este modelo de referencia se muestran en la tabla 3.9:

Tabla 3.9: Métricas de error de Araujo

Modelo	RMSE (W)	MAE (W)	R^2
Araujo	641.36	354.81	0.9947

El siguiente enfoque evaluado en el experimento son los modelos de aprendizaje automático de regresión. Estos modelos se basan en características predefinidas por expertos, las cuales son calculadas a partir de los flujos de datos provenientes de los sensores instalados en el sistema fotovoltaico. Los resultados para estos métodos se muestran en la tabla 3.10 en términos de métricas de error (RMSE, MAE y R^2).

Para completar este experimento desde un punto de vista de consumo en términos de recursos de computación, se incluye en la tabla 3.11 una comparación que considera tanto el tiempo de aprendizaje como el tiempo de evaluación para los modelos basados en características predefinidas por expertos.

Tabla 3.10: Métricas de error de modelos con características predefinidas por expertos y diferentes enfoques de ventanas temporales

Modelo	Tamaños de ventana deslizante	RMSE (W)	MAE (W)	R ²
Regresión Lineal	10min	637.20	425.74	0.9948
	10min+10min+10min	590.27	375.11	0.9955
	10min+20min+60min	537.37	323.04	0.9963
k-NN	10min	466.76	229.62	0.9972
	10min+10min+10min	536.11	249.29	0.9963
	10min+20min+60min	528.40	253.13	0.9964
Bosques Aleatorios	10min	410.44	201.91	0.9978
	10min+10min+10min	375.49	183.49	0.9982
	10min+20min+60min	360.13	173.47	0.9983
SVM	10min	4474.17	2794.36	0.7421
	10min+10min+10min	4593.69	2835.49	0.7281
	10min+20min+60min	4410.17	2653.65	0.7493

Tabla 3.11: Métricas de tiempo de modelos basados en características predefinidas por expertos

Modelo	Tamaño ventana deslizante	T. aprendizaje (ms)	T. evaluación (ms)
Regresión Lineal	10min	11.46	2.10
	10min+10min+10min	36.67	2.72
	10min+20min+60min	37.25	2.81
k-NN	10min	86.52	13.00
	10min+10min+10min	213.55	51.50
	10min+20min+60min	196.54	52.11
Bosques Aleatorios	10min	22743.35	42.89
	10min+10min+10min	69790.30	45.54
	10min+20min+60min	73499.40	42.03
SVM	10min	30912.46	322.88
	10min+10min+10min	48459.79	846.61
	10min+20min+60min	49373.02	857.88

El tercer y último enfoque evaluado en este experimento es el basado en aprendizaje profundo con redes neuronales. Con el propósito de comparar los resultados con el método analítico de Araujo y los modelos basados en características predefinidas por expertos, se proporciona una comparación en términos de métricas de

error (RMSE, MAE y R^2) en la tabla 3.12, así como del tiempo de aprendizaje y evaluación en la tabla 3.13.

Tabla 3.12: Métricas de error para enfoques de aprendizaje profundo

Modelo	Segmentación	RMSE (W)	MAE (W)	R^2
2LSTM	5min	2393.75	618.70	0.9262
	10min	706.57	376.76	0.9936
3CNN + 2LSTM	5min	2384.14	583.11	0.9271
	10min	531.08	274.87	0.9964

Tabla 3.13: Métricas de tiempo para modelos de aprendizaje profundo

Modelo	Segmentación	T. aprendizaje (ms)	T. evaluación (ms)
2LSTM	10min	222657.12	6951.65
3CNN + 2LSTM	10min	197593.52	5627.01

Finalmente, como resumen de los resultados de los diferentes modelos, en la tabla 3.14 se incluye una comparación entre los diferentes enfoques: Araujo, el modelo de aprendizaje profundo de mejor rendimiento (3CNN+2LSTM) y el mejor regresor entre los enfoques basados en características predefinidas por expertos (bosques aleatorios con un tamaño de ventana de 90 minutos).

Tabla 3.14: Resumen de métricas de error para las mejores configuraciones de características predefinidas por expertos y enfoques de aprendizaje profundo

Modelo	RMSE (W)	MAE (W)	R^2
Araujo	641.36	354.81	0.9947
3CNN+2LSTM	531.08	274.87	0.9964
Bosques Aleatorios	360.13	173.47	0.9983

Con el objetivo de proporcionar una representación visual de los resultados conseguidos en la predicción inmediata de la potencia generada por el sistema fotovoltaico, en la figura 3.11 se presenta la gráfica de potencia para una muestra de 2 días

de duración. En dicha representación se presentan las salidas de cada uno de los regresores evaluados junto con la verdad fundamental.

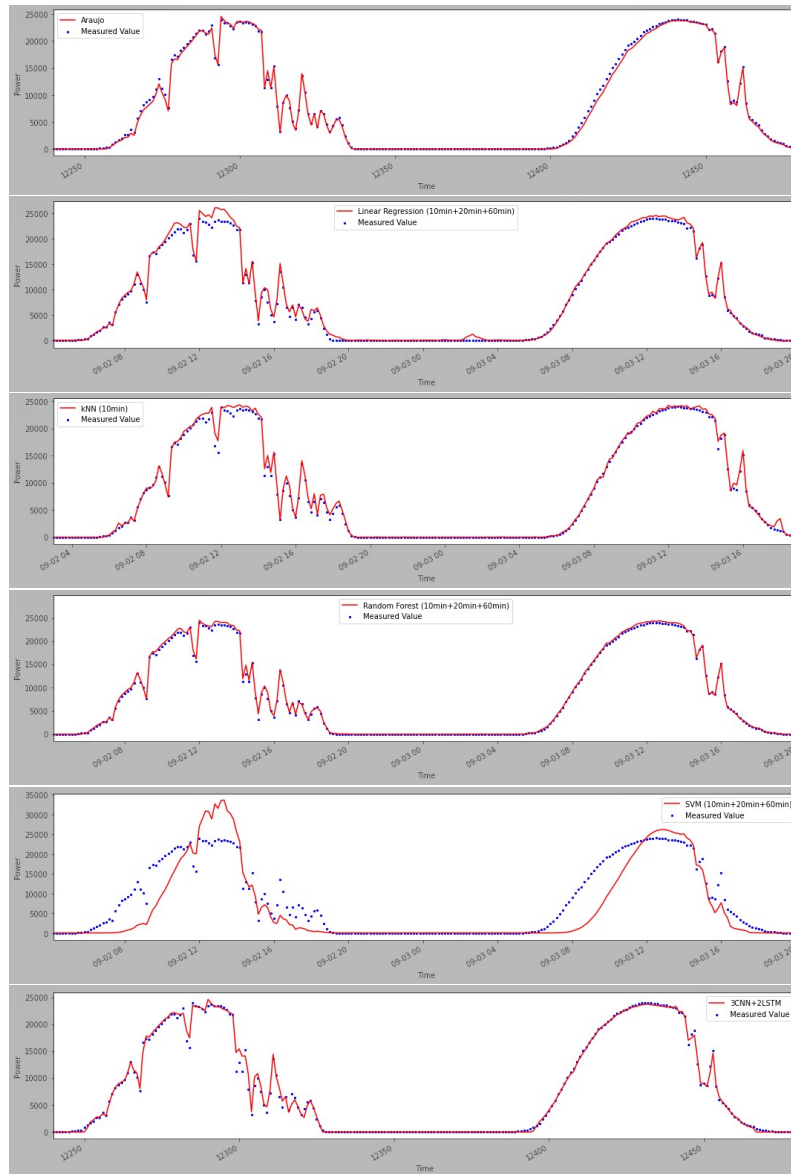


Figura 3.11: Muestras de 2 días de la verdad fundamental de la potencia fotovoltaica en comparación con las predicciones de: i) el modelo Araujo, ii) Regresión lineal, iii) k-NN, iv) Random Forest, v) SVM, vi) 3CNN+2LSTM.

3.5.3. Discusión

En este trabajo se describe inicialmente un módulo basado en la tecnología de Internet de las Cosas (IoT) y desarrollado e implementado en el contexto del Proyecto Ópera, destinado a la recopilación de datos generados por sensores ambientales y datos de generación de energía del sistema fotovoltaico. Posteriormente, con el propósito de evaluar el comportamiento del sistema fotovoltaico y monitorizar su rendimiento, se desarrolla una metodología y una configuración para la predicción inmediata de la potencia fotovoltaica a partir de los datos de irradiancia y temperatura obtenidos por los sensores ambientales instalados en el sistema.

Para llevar a cabo esta tarea, se proponen dos enfoques diferentes para modelos de aprendizaje automático:

- Regresores clásicos de aprendizaje automático, con características diseñadas manualmente y múltiples ventanas temporales.
- Modelos de aprendizaje profundo, con extracción automática de características.

Ambos enfoques presentan resultados alentadores en cuanto a la predicción de la generación de energía del sistema fotovoltaico, basándose en los datos recopilados de sensores ambientales. Sin embargo, destaca el modelo basado en características diseñadas manualmente y múltiples ventanas temporales debido a su menor tiempo de aprendizaje y a la obtención de mejores resultados. Se observan específicamente los siguientes aspectos destacados:

- El uso de múltiples ventanas temporales desequilibradas disminuye el error de la predicción.

- El modelo de bosque aleatorio se posiciona como el regresor más eficaz.
- El modelo de k vecinos más cercanos (k-NN) proporciona un excelente equilibrio entre el tiempo de aprendizaje y los resultados en la predicción.

Se recomienda especialmente el uso de k-NN para la predicción de la generación de energía en sistemas fotovoltaicos que emplean enfoques basados en computación distribuida conocidos como *Edge* o *Fog Computing*, donde los miniprocesadores pueden llevar a cabo el proceso de aprendizaje de datos de manera eficiente en poco tiempo, utilizando pocos recursos computacionales y un bajo consumo de energía.

En el contexto de los enfoques de aprendizaje profundo, el uso de modelos que incluyen redes neuronales convolucionales (CNN) y de memoria a corto-largo plazo (LSTM) demuestra un error menor en la predicción de la generación de energía en comparación con el modelo algebraico de Araujo. Este hecho se debe a la extracción automática de características generada por las redes convolucionales, lo que resume los patrones clave para predecir la generación de energía de salida. Este enfoque supone una mejora notable en comparación con el uso exclusivo de redes de memoria a corto-largo plazo.

Además, se observa que el rendimiento del modelo de aprendizaje profundo con segmentación de diez minutos mejora en comparación con la segmentación de cinco minutos. Esto es debido a que la segmentación a corto plazo duplica el número de variables de entrada en la secuencia de muestras, y la mayor complejidad de los datos reduce el rendimiento de la predicción. Sin embargo, es importante destacar que el modelo basado en características diseñadas manualmente con ventanas temporales desequilibradas supera el rendimiento de los enfoques de aprendizaje profundo y el modelo analítico de Araujo, destacándose como el modelo líder según los resultados

presentados en este trabajo.

Capítulo 4

Aprendizaje por Transferencia para la Predicción de Potencia Fotovoltaica

4.1. Introducción

El concepto de aprendizaje por transferencia se define como el proceso mediante el cual se facilita el aprendizaje en una tarea nueva a través de la transferencia de conocimientos de una tarea relacionada previamente aprendida [Torrey and Shavlik, 2010]. En el ámbito de las metodologías basadas en datos, el aprendizaje supervisado ofrece modelos donde típicamente los datos de entrenamiento y validación se recopilan bajo las mismas condiciones. Estos datos son utilizados para el cálculo de la representación de características y de la segmentación de datos [Pan, 2020]. Sin embargo, en muchos contextos reales, las condiciones de aprendizaje varían entre dominios, lo cual presenta un obstáculo para el uso de modelos de aprendizaje automático

estándar [Polo-Rodriguez et al., 2020]. El aprendizaje por transferencia responde a este desafío de evaluar las capacidades de modelos preentrenados en diferentes contextos.

El uso del aprendizaje profundo en series temporales se ha convertido en un campo de investigación prolífico [Gamboa, 2017]. Este campo abarca redes temporales o de secuencia, como las redes de memoria a corto-largo plazo (LSTM, *Long Short-Term Memory*) [Hochreiter and Schmidhuber, 1997], y también considera características espaciales en series temporales mediante redes neuronales convolucionales (CNN, *Convolutional Neural Network*) [Wang et al., 2019, Qiu et al., 2017]. Estos modelos de aprendizaje profundo tienen la capacidad de transferir conocimiento de forma directa debido a la plasticidad de las redes neuronales [Andina et al., 2009]. Estas redes neuronales pueden ponderar las sinapsis artificiales desarrolladas en etapas de aprendizaje previas en otros dominios [Polo-Rodriguez et al., 2020]. De esta manera, el preentrenamiento de modelos basados en CNN y LSTM demuestra ser eficaz en la inicialización del conocimiento, proporcionando resultados alentadores y reduciendo significativamente el tiempo de aprendizaje requerido para el aprendizaje [Erhan et al., 2010].

El experimento recogido en este capítulo consiste en los siguientes pasos:

- Adquisición de datos del sistema fotovoltaico UniVer, ubicado en la Universidad de Jaén (España). Para este propósito, se utiliza un módulo de Internet de las cosas (IoT, *Internet of Things*) diseñado en el marco del proyecto Ópera para la captura de datos procedentes de sensores ambientales y de la energía generada de un sistema fotovoltaico, como se detalla en la sección "3.3 - Adquisición de datos en tiempo real mediante dispositivos IoT.

- Obtención de datos procedentes del Laboratorio SolarTech, perteneciente a la Universidad Politécnica de Milán (Italia). Estos datos son propuestos con el fin de comparar las predicciones en un contexto diferente al correspondiente al UniVer.

- Recopilación de datos de una planta fotovoltaica situada en India, los cuales están disponibles en la base de datos de *Kaggle* (<https://www.kaggle.com/anikannal/solar-power-generation-data>). La inclusión de estos datos permite comparar las predicciones en una región climática del mundo diferente de los dos sistemas anteriores.

- Desarrollo de una metodología de segmentación homogénea de datos, propuesta para agregar la información de la radiación solar y de la temperatura recibidas en las secuencias de datos de sensores originadas en diferentes contextos.

- Construcción de dos modelos de aprendizaje profundo basados en redes neuronales convolucionales (CNN) y de memoria a corto-largo plazo (LSTM), propuestos para realizar una predicción inmediata de la potencia fotovoltaica en tiempo real. Se evalúan las capacidades de transferencia de estimación de energía entre España, Italia e India utilizando diferentes configuraciones de aprendizaje, comparando los resultados ofrecidos por modelos entrenados desde cero y modelos preentrenados.

4.2. Metodología para el procesamiento de datos

Con el objetivo de estructurar adecuadamente este experimento, se define una metodología que permita la evaluación de las capacidades de transferencia de aprendizaje de diversos modelos para la predicción de la energía generada por sistemas fotovoltaicos en diferentes contextos. Esta metodología recoge los siguientes aspectos fundamentales:

- Identificación de conjuntos de datos diferentes ámbitos y el análisis de su estructura y contenido. Se realiza un análisis de los conjuntos de datos correspondientes a tres dominios distintos: el Proyecto UniVer de la Universidad de Jaén, el Laboratorio SolarTech de la Universidad Politécnica de Milán y la planta solar ubicada en la región de India, cuyos datos se obtienen de la base de datos de Kaggle. Este análisis comprende una evaluación detallada de la estructura y el contenido de dichos conjuntos de datos.
- Definición de un método de segmentación para homogeneizar la entrada de los diferentes dominios. Se desarrolla un proceso de segmentación de forma que permita evaluar conjuntos de datos provenientes de diferentes de diferentes regiones geográficas del mundo, permitiendo así una comparación significativa.
- Diseño y ejecución de modelos CNN y LSTM. Se configuran modelos de aprendizaje profundo para la predicción de la generación de energía fotovoltaica, y se comparan los resultados de los modelos entrenados desde cero frente a modelos a los cuales se les realiza un aprendizaje afinado en el dominio objetivo tras haber sido preentrenados en otro dominio diferente.

4.2.1. Recopilación de datos de fuentes heterogéneas

Los datos para este experimento se han recopilado de tres fuentes diferentes. La primera fuente es el sistema fotovoltaico del proyecto UniVer en la Universidad de Jaén. Este conjunto de datos se ha generado capturando las señales descritas a continuación con una frecuencia de 30 segundos, desde junio hasta noviembre de 2019. Los datos recopilados de UniVer son los siguientes:

- Registro de tiempo, expresado en horario central europeo (CET).
- Irradiancia global en el plano del panel fotovoltaico, medida en vatios por metro cuadrado (W/m^2) por un dispositivo Internet de las Cosas (IoT) instalado en el generador fotovoltaico (GFV).
- Temperatura ambiente, medida en grados centígrados ($^{\circ}C$) por un dispositivo IoT instalado en el GFV.
- Temperatura del módulo, también medida en grados centígrados ($^{\circ}C$) por un dispositivo IoT instalado en el GFV.
- Potencia de salida del GFV, medida en vatios (W) en el inversor del sistema.

En la figura 4.1 se muestra un registro de ejemplo del fichero de datos recogido del sistema UniVer.

	timestamp	G_I	Tamb	Tmod	Pa
9731	2019-08-15 16:40:00+00:00	468.3333333	35.14444444	42.64444444	12970.13333

Figura 4.1: Registro de ejemplo del archivo de datos UniVer: tiempo (timestamp), irradiancia global (G.I), temperatura ambiente (Tamb), temperatura del módulo (Tmod), potencia de salida (PA).

El segundo conjunto de datos incluye la producción de energía fotovoltaica medida en el laboratorio SolarTech de la Universidad Politécnica de Milán, Italia. El conjunto de datos abarca un año entero, desde enero hasta diciembre de 2017, y está compuesto por las variables listadas a continuación, con una resolución temporal de un minuto:

- Registro de tiempo expresados en horario central europeo (CET).
- Potencia de salida del módulo fotovoltaico medida en vatios (W).
- Temperatura ambiente, medida en grados centígrados ($^{\circ}\text{C}$) por la estación meteorológica descrita en la página web del Laboratorio SolarTech (<http://www.solartech.polimi.it/instrumentation/>).
- Irradiancia global horizontal, medida en vatios por metro cuadrado (W/m^2).
- Irradiancia global en el plano del panel solar (30°), medida en vatios por metro cuadrado (W/m^2).
- Velocidad del viento, medida en metros por segundo (m/s).
- Dirección del viento, medida en grados ($^{\circ}$), asumiendo 0° al este, positivo al sur.

En la Figura 4.2 se muestra una fila de muestra del archivo de datos de SolarTech.

Los conjuntos de datos de ambos sistemas presentan algunas similitudes que nos permiten llevar a cabo el experimento utilizando las mismas señales en el modelo de aprendizaje profundo para ambos contextos. Sin embargo, para que se puedan extraer con seguridad conclusiones significativas de los resultados, los conjuntos de

	Time	PV_Power	T_air	G_h	G_tilt	W_s	W_d
226851	07-Jun-2017 12:50:00	206.490966796875	24.58	970	1014	2.39	230.7

Figura 4.2: Fila de ejemplo del archivo de datos de SolarTech: tiempo, potencia de salida (PV_Power), temperatura ambiente (T_air), irradiancia global horizontal (G_h), irradiancia global medida (G_tilt), velocidad del viento (W_s), dirección del viento (W_d).

datos deben ser homogéneos. Se toman por tanto las siguientes medidas para lograr esa homogeneización:

- Selección de las mismas señales como entrada para el modelo: temperatura ambiente e irradiancia global sobre el plano para ambos conjuntos de datos.
- Selección de la misma señal como objeto de la predicción: potencia generada por los sistemas fotovoltaicos.
- Identificación del período del año común en los conjuntos de datos para evitar problemas de estacionalidad. En este caso, la superposición es de junio a noviembre de 2019 en el conjunto de datos de UniVer y de junio a noviembre de 2017 en el conjunto de datos de SolarTech.
- Agregación de los datos utilizando la media de 10 minutos y definición de una segmentación como se describe a continuación en la sección 4.2.2 - Segmentación de datos.

4.2.2. Segmentación de datos

La metodología para llevar a cabo este estudio del aprendizaje por transferencia incluye el desarrollo de un proceso de segmentación donde, de forma análoga a la descripción de la sección "3.4.1 - Segmentación y agregación de series temporales

para modelos basados en el análisis e interpretación de datos”, y siguiendo una definición formal, un sensor s recopila datos en tiempo real en forma de un par $\bar{s}_i = \{s_i, t_i\}$, donde s_i representa una medición dada y t_i la marca de tiempo. Así, el flujo de datos del sensor fuente s está definido por $\bar{S}_s = \{\bar{s}_0, \dots, \bar{s}_i\}$ y un valor dado en una marca de tiempo t_i por $S_s(t_i) = s_i$. En este capítulo, la irradiancia G_I , la temperatura ambiente T_{am} y la potencia P_A proporcionan tres flujos de datos que describen el comportamiento y la producción de energía del sistema fotovoltaico.

Para homogeneizar los datos recopilados por los diferentes sensores, se definen varias ventanas temporales deslizantes simétricas. Estas ventanas están definidas por su tamaño de ventana correspondiente a un intervalo de tiempo $W_w = [W_w^-, W_w^+]$, segmentando las muestras de un flujo de datos del sensor dado \bar{S}_s y agregando los valores \bar{s}_i mediante una función de agregación $T_t(S_s, W_w, t^*)$:

$$T_t(S_s, W_w, t^*) = \bigcup_{s_i}^{\bar{s}_i} s_i, t_i \in [t^* - W_w^-, t^* - W_w^+]$$

Por lo tanto, la agregación \bigcup_{s_i} de los datos del sensor s_i aplicada sobre un corto intervalo de tiempo $W_w = [W_w^-, W_w^+]$ representa el valor relevante en un punto dado en el tiempo t^* .

Usando la agregación de datos, la segmentación de señal se define mediante varias ventanas temporales deslizantes de corta duración, que están definidas por la granularidad temporal Δ . Los datos en la ventana temporal se agregan mediante la función de agregación. Concretamente, los datos dentro de cada ventana temporal se promedian $\bigcup = \mu$ para cada ventana temporal de corta duración dentro del segmento definido en el intervalo de tiempo. La media proporciona una función de agregación sólida para homogeneizar fuentes de sensores crudos heterogéneas en caso

de que sean proporcionadas por diferentes tasas de recopilación de datos.

Así, se obtiene una secuencia de datos para el flujo de datos de cada sensor, cuyo tamaño de secuencia es el mismo para todas las fuentes S_s :

$$S^*(t^*) \rightarrow \begin{cases} S^1(t^*) = T_t(S_1, [0, \Delta], t^*) \rightarrow T_t(S_1, [\Delta, 2\Delta], t^*), \dots, \\ \qquad \qquad \qquad \rightarrow T_t(S_1, [\Delta|W| - \Delta, \Delta|W|], t^*) \\ \dots \\ S^{|S|}(t^*) = T_t(S_{|S|}, [0, \Delta], t^*) \rightarrow T_t(S_{|S|}, [\Delta, 2\Delta], t^*), \dots, \\ \qquad \qquad \qquad \rightarrow T_t(S_{|S|}, [\Delta|W| - 1, \Delta|W|], t^*) \end{cases}$$

Cabe destacar que los diferentes flujos de datos procedentes de los sensores $S_1, \dots, S^{|S|}$ están:

- Alineados en el mismo punto de tiempo t^*
- Segmentados en ventanas temporales deslizantes homogéneas $W_w = [W_w^-, W_w^+] = \{[0, \Delta], \dots, [\Delta, 2\Delta]\}$
- Representados por un vector de características de tamaño único que agrega valores anteriores para el flujo de datos de cada uno de los sensores $S^1(t^*)$.

En la figura 4.3, se describe una representación visual de la segmentación del flujo de datos de un sensor definida por ventanas temporales deslizantes para homogeneizar diferentes frecuencias de recopilación de datos de fuentes heterogéneas.

En este trabajo, la segmentación de las señales de entrada y salida del sistema fotovoltaico se desarrolla para la temperatura ambiente y la irradiancia global como las señales de entrada, y para la potencia de salida como variable objetivo. Este proceso de agregación y segmentación de datos es clave en el enfoque de aprendi-

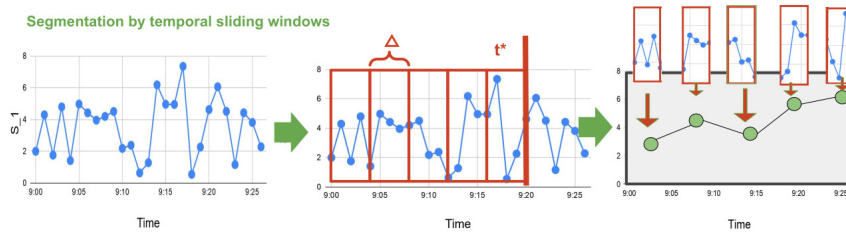


Figura 4.3: Segmentación del flujo de datos de un sensor definida por ventanas temporales deslizantes.

zaje por transferencia, ya que las tasas de muestreo suelen ser diferentes entre los dominios. En los dos contextos de este trabajo, los datos de entrada se configuran de la siguiente manera:

- Conjunto de datos UniVer. Las señales se agregan desde la muestra original de treinta segundos hasta un promedio de diez minutos y luego se segmentan en una ventana deslizante de noventa minutos.
- Conjunto de datos SolarTech. Las señales se agregan desde la muestra original de un minuto hasta un promedio de diez minutos y luego se segmentan en una ventana deslizante de noventa minutos.

Como resultado de esta etapa de segmentación, los datos originalmente heterogéneos de ambos dominios tienen una representación homogénea de las señales de entrada y salida.

4.2.3. Inicialización de pesos y aprendizaje afinado entre dominios

El objetivo de este experimento, como ya se ha introducido anteriormente en este capítulo, es evaluar las capacidades del aprendizaje por transferencia en modelos

de aprendizaje profundo orientados a la predicción de energía fotovoltaica en diferentes dominios. Para realizar esta evaluación de forma exhaustiva, se debe buscar respuesta a las siguientes preguntas:

- ¿Qué precisión tienen los modelos de aprendizaje entrenados en su propio dominio (región climática) para estimar la generación de energía de los sistemas fotovoltaicos?
- ¿Cuál es la diferencia de rendimiento al predecir la generación de energía en un dominio no conocido con un modelo que se ha entrenado en otro generador fotovoltaico, tipo de instalación o zona climática?
- ¿Es efectiva la estructura de los modelos de aprendizaje profundo basados en redes neuronales para transferir y optimizar los pesos en la estimación de energía de sistemas fotovoltaicos de un clima a otro?

Para abordar estas preguntas, después de haber obtenido la representación homogénea de los datos en la sección anterior, se evalúan tres combinaciones de aprendizaje para los modelos, que consisten en:

- a. Aprendizaje básico dentro de cada dominio, donde se realiza el aprendizaje y la evaluación de forma aislada a otros dominios. Este aprendizaje estándar proporciona una referencia del rendimiento del modelo para cada dominio.
- b. Aprendizaje externo entre dominios, donde el aprendizaje se realiza en un dominio dado, pero la evaluación se lleva a cabo en otro diferente. El rendimiento de este aprendizaje complejo describe la transferibilidad de las estructuras de aprendizaje profundo cuando son aplicadas en diferentes dominios desconocidos para el modelo.

- c. Aprendizaje afinado en el dominio objetivo a partir de una red neuronal preentrenada en otro contexto. En esta configuración de aprendizaje, el rendimiento puede mejorar o empeorar con respecto al valor de referencia (aprendizaje básico). Si hay una mejora, los patrones y núcleos calculados en las diferentes capas del modelo de aprendizaje profundo proporcionan información general sobre la generación de energía, a la vez que se pueden optimizar con datos locales para generar un modelo más robusto.

En la figura 4.4 se detallan la configuración de datos y entrenamiento de modelos en los escenarios de aprendizaje básico, externo y afinado.

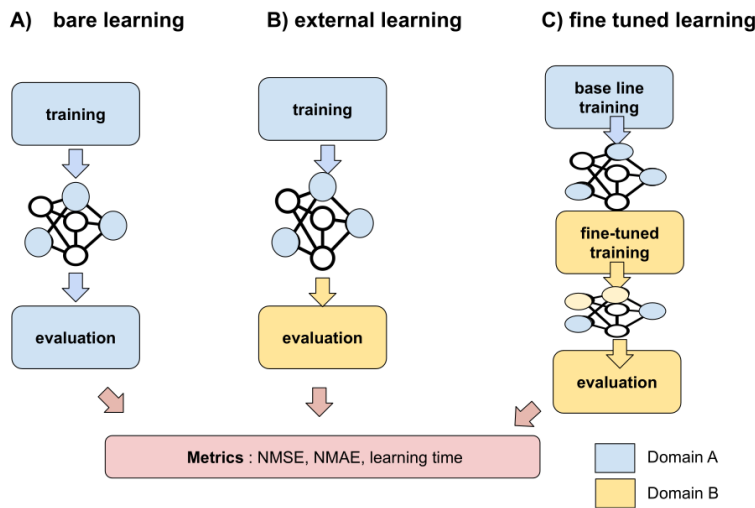


Figura 4.4: a) Aprendizaje básico, b) Aprendizaje externo y c) Aprendizaje afinado

En la tabla 4.1, se describe la configuración del modelo de aprendizaje. Primero, se integran tres capas de CNN como extractores de características espaciales. A continuación, dos capas de LSTM modelan las dependencias temporales a partir de la CNN. Esta combinación de redes híbridas CNN-LSTM se ha seleccionado porque ha proporcionado resultados alentadores en la modelización de la generación de

energía de sistemas fotovoltaicos, como se describe en el capítulo "3 - Modelos de Aprendizaje Automático y Aprendizaje Profundo para la Predicción Inmediata de Potencia Fotovoltaica".

Tabla 4.1: Configuración del modelo de aprendizaje profundo

3CNN+2LSTM
2 kernels x 16 filters
Re-Lu
2 kernels x 32 filters
Re-Lu
2 kernels x 64 filters
Re-Lu
dropout(0.25)
LSTM(32 units)
dropout(0.25)
LSTM(32 units)
dropout(0.25)
connected(1 unit)
activation function:Re-Lu
loss function:MAE

4.3. Evaluación del caso de estudio

En esta sección, se describen los resultados de la implementación de las capacidades de transferencia de aprendizaje de modelos de aprendizaje profundo con el propósito de realizar y afinar una predicción de la generación de energía producida por sistemas fotovoltaicos en diferentes contextos.

4.3.1. Resultados

Las capacidades de transferencia de aprendizaje se estudian primero en el contexto del Laboratorio SolarTech siguiendo las diferentes combinaciones de dominios

definidas en la sección anterior "4.2.3 - Inicialización de pesos y aprendizaje afinado entre dominios". El modelo CNN+LSTM, también descrito en la sección indicada anteriormente, se entrena y valida inicialmente en el conjunto de datos de SolarTech bajo un escenario de aprendizaje básico. Posteriormente, se entrena en el conjunto de datos de UniVer y se valida en el conjunto de datos de SolarTech para evaluar el rendimiento del escenario de aprendizaje externo. Finalmente, el modelo entrenado en el conjunto de datos de UniVer se optimiza mediante un nuevo entrenamiento en el conjunto de datos de SolarTech para finalmente validarlo con los datos del sistema SolarTech, en el denominado escenario de aprendizaje afinado.

Los resultados de estos tres escenarios se comparan utilizando las métricas de error Error Cuadrático Medio Normalizado (NMSE, *Normalised Mean Square Error*) y Error Absoluto Medio Normalizado (NMAE, *Normalised Mean Absolute Error*). Esta comparación de resultados se muestra en la tabla 4.2.

Tabla 4.2: Error normalizado por tipo de aprendizaje en el contexto de SolarTech

Aprendizaje	Pre-entrenamiento	Entrenamiento	Validación	NMSE	NMAE
Básico	N/A	SolarTech	SolarTech	$0,09 \cdot 10^{-2}$	$1,50 \cdot 10^{-2}$
Externo	N/A	UniVer	SolarTech	$0,34 \cdot 10^{-2}$	$3,22 \cdot 10^{-2}$
Afinado	UniVer	SolarTech	SolarTech	$0,06 \cdot 10^{-2}$	$0,94 \cdot 10^{-2}$

A continuación, el mismo modelo y escenarios se ponen a prueba en el contexto del sistema fotovoltaico UniVer. El escenario de aprendizaje básico se lleva cabo realizando el entrenamiento y la validación de modelo en el conjunto de datos de UniVer. Para el escenario de aprendizaje externo, el modelo se entrena con los datos pertenecientes al sistema SolarTech y se valida sobre los datos del UniVer. Y finalmente, para el escenario de ajuste fino, el modelo previamente entrenado en SolarTech se optimiza y valida en el conjunto de datos de UniVer. La tabla 4.3 muestra

los resultados del experimento en este contexto utilizando las mismas métricas de error que en el contexto anterior (NMSE y NMAE).

Tabla 4.3: Error normalizado por tipo de aprendizaje en el contexto de UniVer

Aprendizaje	Pre-entrenamiento	Entrenamiento	Validación	NMSE	NMAE
Básico	N/A	UniVer	UniVer	$0,15 \cdot 10^{-2}$	$1,93 \cdot 10^{-2}$
Externo	N/A	SolarTech	UniVer	$5,81 \cdot 10^{-2}$	$3,56 \cdot 10^{-2}$
Afinado	SolarTech	UniVer	UniVer	$0,10 \cdot 10^{-2}$	$1,63 \cdot 10^{-2}$

Para obtener una comprensión del rendimiento del modelo en los diferentes escenarios al intentar realizar la predicción de la generación de energía de salida para cada uno de los sistemas, en las tablas 4.4 y 4.5 se muestran las métricas de error desnormalizadas Raíz del Error Cuadrático Medio (RMSE, *Root Mean Square Error*) y Error Medio Absoluto (MAE, *Mean Absolute Error*). También se incluye en las tablas el valor máximo de potencia generada por el sistema fotovoltaico observado en los conjuntos de datos para cada dominio.

Tabla 4.4: Error absoluto por tipo de aprendizaje en SolarTech

Aprend.	Preentr.	Entr.	Val.	P Max(W)	RMSE(W)	MAE(W)
Básico	N/A	SolarTech	SolarTech	242.68	7.28	3.64
Externo	N/A	UniVer	SolarTech	242.68	14.15	7.81
Afinado	UniVer	SolarTech	SolarTech	242.68	5.94	2.28

Tabla 4.5: Error absoluto por tipo de aprendizaje en UniVer

Aprend.	Preentr.	Entr.	Val.	P Max(W)	RMSE(W)	MAE(W)
Básico	N/A	UniVer	UniVer	28472.30	1102.73	549.52
Externo	N/A	SolarTech	UniVer	28472.30	6862.95	1013.61
Afinado	SolarTech	UniVer	UniVer	28472.30	900.37	464.10

Junto con las métricas de error descritas hasta ahora en esta sección, también se estudian las métricas de tiempo para las diferentes combinaciones, con el fin de

evaluar el rendimiento del modelo desde una perspectiva de complejidad temporal.

En primer lugar, para evaluar la mejora en la velocidad de aprendizaje de un modelo afinado con respecto a un escenario de aprendizaje básico, se compara la función de pérdida durante las etapas de entrenamiento y validación del modelo en ambas situaciones. La métrica utilizada para esta comparación es el número de épocas (pasadas del modelos sobre los datos) necesarias hasta alcanzar un valor concreto de la función de pérdida, en este caso el error cuadrático medio normalizado (NMSE).

Esta comparación de la velocidad de aprendizaje se realiza sobre el conjunto de datos de SolarTech, estableciendo dos umbrales de diferentes órdenes de magnitud de la función de pérdida: $NMSE < 0,01$ y $NMSE < 0,001$. En la tabla 4.6 se presentan los resultados de este experimento.

Tabla 4.6: Velocidad de aprendizaje por escenario en el contexto de SolarTech

Aprendizaje	Preentr.	Entr.	Validación	$NMSE < 0,01$ (epoch)	$NMSE < 0,001$ (epoch)
Básico	N/A	SolarTech	SolarTech	7	39
Afinado	UniVer	SolarTech	SolarTech	1	7

Con el fin de proporcionar una representación visual de las diferencias en la velocidad de aprendizaje, en la figura 4.5 se muestra la mejora de la métrica de error NMSE en función del número de épocas, durante el entrenamiento y la validación del modelo para los aprendizajes básico y afinado en el contexto de SolarTech.

También desde una perspectiva de complejidad temporal, se incluye en este experimento un estudio sobre el tiempo de aprendizaje del modelo. Para este fin, se compara el tiempo total que conlleva completar el entrenamiento del modelo en los escenarios de aprendizaje básico y aprendizaje afinado sobre el conjunto de datos de SolarTech. La tabla 4.7 muestra los resultados de esta comparación.

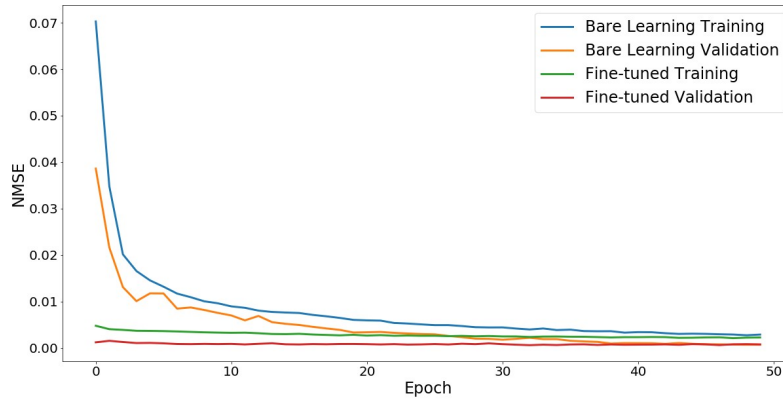


Figura 4.5: Historial de entrenamiento y validación del modelo que muestra la función de pérdida (MSE) para los escenarios de aprendizaje básico y afinado en el contexto de SolarTech.

Tabla 4.7: Tiempo de entrenamiento por escenario en el contexto de SolarTech

Aprendizaje	Preentrenamiento	Entrenamiento	Validación	Tiempo de Entrenamiento (s)
Básico	N/A	SolarTech	SolarTech	81.82
Afinado	UniVer	SolarTech	SolarTech	40.43

Como último ejercicio en este experimento, y con el objetivo de ampliar la comprensión de la capacidad de transferencia de aprendizaje del modelo de aprendizaje profundo, se pretende evaluar en qué medida el conocimiento puede transferirse a contextos más dispares. Para ello, también se evalúa un tercer conjunto de datos, el cual presenta condiciones muy diferentes a los dos conjuntos de datos homogéneos utilizados para los experimentos anteriores:

- Es un conjunto de datos generado por una planta de energía solar en la India, una región climática diferente a los dos sistemas fotovoltaicos anteriores (España e Italia).
- El conjunto de datos abarca treinta y un días durante los meses de mayo y junio, en contraste con la duración de seis meses de los conjuntos de datos UniVer y SolarTech.

- La frecuencia de recopilación de datos para las señales de entrada y salida del sistema fotovoltaico es de quince minutos, por lo que la ventana de segmentación es de 135 minutos.

Los resultados del experimento en este tercer contexto se presentan en la tabla 4.8, la cual muestra las métricas de error normalizadas, así como en la tabla 4.9, la cual incluye las métricas de error desnormalizadas junto con el valor máximo de potencia del sistema fotovoltaico.

Tabla 4.8: Error normalizado por tipo de aprendizaje en el contexto de India

Aprendizaje	Pre-entrenamiento	Entrenamiento	Validación	NMSE	NMAE
Básico	N/A	India	India	$0,34 \cdot 10^{-2}$	$3,08 \cdot 10^{-2}$
Externo	N/A	UniVer	India	$0,45 \cdot 10^{-2}$	$3,67 \cdot 10^{-2}$
Afinado	UniVer	India	India	$0,11 \cdot 10^{-2}$	$1,68 \cdot 10^{-2}$

Tabla 4.9: Error absoluto por tipo de aprendizaje en el contexto de India

Aprendizaje	Preentr.	Entr.	Validación	Potencia Máx. (W)	RMSE(W)	MAE(W)
Básico	N/A	India	India	14312.71	843.50	440.41
Externo	N/A	UniVer	India	14312.71	910.23	483.28
Afinado	UniVer	India	India	14312.71	484.06	241.00

4.3.2. Discusión

Analizando los resultados del experimento, se extraen las siguientes conclusiones acerca del uso de este enfoque de aprendizaje por transferencia de conocimiento para la predicción inmediata de la potencia generada por sistemas fotovoltaicos en diferentes dominios.

Es importante comenzar resaltando que el modelo de aprendizaje profundo se reutiliza en los diferentes contextos sin tener que cambiar ninguna configuración o

parámetros específicos del modelo en sí para cada contexto. Esto puede ser considerado como una de las mayores ventajas de este enfoque comparado con los modelos de monitorización tradicionales, los cuales típicamente se basan en la configuración del modelo con los parámetros físicos del sistema fotovoltaico.

Destaca también la capacidad del modelo CNN+LSTM para proporcionar buenos resultados en el escenario de aprendizaje externo, logrando realizar una predicción de la potencia generada por el sistema SolarTech con un margen de error reducido después de haber sido solamente entrenado con el conjunto de datos del sistema UniVer. Este margen de error es sólo 0,25 puntos porcentuales mayor que el error base calculado con el modelo entrenado en el sistema SolarTech (escenario de aprendizaje básico).

Los mejores resultados se obtienen al aplicar el aprendizaje fino. El modelo se preentrena con el conjunto de datos UniVer y luego se optimiza con datos locales, lo que resulta en una predicción de la generación de energía de la planta SolarTech con el margen de error más pequeño de todas las combinaciones incluidas en el experimento. Este margen de error es 0,03 puntos porcentuales más pequeño que el error base obtenido en el aprendizaje básico.

En el contexto de UniVer, se obtiene un margen de error mayor para el modelo en el escenario de aprendizaje externo (modelo entrenado con el conjunto de datos SolarTech y validado sobre el conjunto de datos del UniVer), con un aumento de 5,66 puntos porcentuales en el error cuadrático medio normalizado. En el escenario de aprendizaje fino, el margen de error es menor que el calculado en el escenario de aprendizaje básico, con un error cuadrático medio normalizado 0,05 puntos porcentuales más pequeño que el error base.

También se observa una mejora en la velocidad de aprendizaje y en el tiempo

de entrenamiento en este escenario de aprendizaje fino. Solo se necesita una época con el modelo preentrenado, pero se requieren siete épocas con el modelo básico para lograr el mismo rendimiento de $NMAE < 0,01$. Para alcanzar una métrica de error más pequeña, $NMAE < 0,001$, el modelo afinado solo necesita siete épocas en comparación con las treinta y nueve del escenario de aprendizaje básico. En términos de tiempo total de aprendizaje, solo se necesitan 40,43 segundos para entrenar el modelo en el escenario de aprendizaje fino en comparación con los 81,82 segundos en el escenario de aprendizaje básico.

4.3.3. Limitaciones del caso de estudio

Los resultados descritos en este trabajo demuestran una mejora en el uso de la transferencia de conocimientos en la predicción de la generación de energía de sistemas fotovoltaicos. Sin embargo, se identifican algunas limitaciones que podrían ser abordadas en futuros trabajos:

1. Impacto de las diferencias climáticas: es importante señalar que la optimización entre climas extremadamente diferentes podría no ser tan efectiva mediante la transferencia de conocimientos. Aunque el aprendizaje desde un clima mediterráneo (España) y la optimización para la India muestra una mejora en el rendimiento, se requiere una evaluación más profunda y datos adicionales para comprender completamente el impacto de las diferencias climáticas en la transferencia de conocimientos.

2. Granularidad de los datos: la agregación y segmentación de los datos originales en intervalos de diez minutos es un factor clave para homogeneizar los datos. Esto significa que los conjuntos de datos con una tasa de recolección de datos más alta no pueden ser evaluados siguiendo esta metodología. Sería interesante explorar cómo

abordar datos con una granularidad diferente en futuras investigaciones.

3. Calidad de los datos procedentes de sensores: los conjuntos de datos utilizados en este trabajo se basan en sensores de alta calidad y arquitecturas bien desarrolladas. Esto puede contribuir a la estabilidad de los resultados y patrones entre los conjuntos de datos. Sin embargo, el desafío se presenta cuando se trabaja con datos recopilados por dispositivos de baja resolución o menor calidad. En futuros estudios, podría ser necesario abordar cómo la transferencia de conocimientos se comporta en situaciones donde la calidad de los datos es variada.

Estas limitaciones proporcionan áreas de enfoque para investigaciones futuras y destacan la necesidad de evaluar cuidadosamente las condiciones y restricciones en las que se puede aplicar la transferencia de conocimientos en la predicción de la generación de energía fotovoltaica en diferentes contextos.

Capítulo 5

Modelos de Redes Adversarias Generativas para la Predicción y el Control de Potencia Fotovoltaica

5.1. Introducción

En este capítulo se exploran las redes adversarias [Creswell et al., 2018] para evaluar su eficacia a la hora de abordar el control y la predicción de energía de en sistemas fotovoltaicos [Huang et al., 2022, Li et al., 2022]. Las redes adversarias han sido propuestas anteriormente en otros trabajos con el objetivo principal de generar datos ausentes o aumentar los datos existentes [Ahmed et al., 2020]. A pesar de estos trabajos iniciales y el considerable potencial de las redes adversarias para crear discriminadores robustos y complementar datos con estimaciones falsas no identificables, su incorporación sigue siendo relativamente poco explorada en este campo.

Otra cuestión clave es identificar el papel de las redes adversarias generativas condicionales (CGAN, *Conditional Generative Adversarial Network*) en un problema de regresión como la predicción inmediata de potencia fotovoltaica, donde los modelos de referencia ofrecen un alto rendimiento según la verdad fundamental de los modelos basados en datos, como se demuestra en el capítulo 3, Modelos de Aprendizaje Automático y Aprendizaje Profundo para la Predicción Inmediata de Potencia Fotovoltaica. En este trabajo, se estudian diferentes arquitecturas ad-hoc de redes adversarias para abordar los siguientes problemas:

- a. Crear discriminadores y evaluar su rendimiento a la hora de diferenciar estimaciones falsas o desviaciones en la generación de energía.
- b. Integrar y evaluar modelos condicionales de dopaje para alternar predicciones no observables con datos reales.
- c. Mejorar el rendimiento en la estimación de la generación de energía de los sistemas fotovoltaicos con el modelo general de redes adversarias.

El primer modelo desarrollado en este trabajo presenta una arquitectura de redes adversarias que incluye dos componentes fundamentales: un discriminador, el cual se encarga de realizar una tarea de clasificación, y un generador, caracterizado principalmente por estar controlado por un nivel de distorsión. Esta arquitectura de doble vía de procesamiento de datos busca mejorar la precisión en la estimación de energía, minimizando las distorsiones en el intento de generar muestras auténticas de energía. El segundo modelo, un enfoque de redes adversarias basado en algoritmos de regresión, adopta una perspectiva novedosa al incorporar mecanismos de dopaje para optimizar el proceso de generación de datos de energía. De forma genérica, am-

bos modelos buscan lograr un equilibrio entre muestras de energía reales y generadas a través de un proceso de discriminación de suma cero.

5.2. Metodología para el procesamiento de datos

El procedimiento definido en este trabajo para el desarrollo y evaluación de modelos de redes adversarias generativas condicionales (CGAN) para la predicción inmediata de la potencia fotovoltaica y la discriminación entre muestras reales y falsas, consiste en los siguientes tres pasos:

- a. En primer lugar, se evalúa como rendimiento de referencia el modelo de aprendizaje profundo basado en redes neuronales convolucionales y redes de memoria a corto-largo plazo (CNN+LSTM) descrito en el capítulo "3 - Modelos de Aprendizaje Automático y Aprendizaje Profundo para la Predicción Inmediata de Potencia Fotovoltaica".
- b. En segundo lugar, se propone un modelo de redes adversarias que integra un discriminador encargado de realizar la tarea de clasificación entre muestras reales y falsas. Esta arquitectura también incluye un modelado del error de distorsión en el generador.
- c. Por último, se describe un modelo de redes adversarias de regresión, el cual está basado en el dopaje de los datos producidos por el modelo y la discriminación de la desviación de la energía de salida.

5.2.1. Preprocesamiento de datos y modelo de aprendizaje profundo de referencia

El modelo de aprendizaje profundo utilizado como referencia para la predicción inmediata de potencia fotovoltaica presentado previamente en el capítulo "3 - Modelos de Aprendizaje Automático y Aprendizaje Profundo para la Predicción Inmediata de Potencia Fotovoltaica", demuestra un nivel elevado de rendimiento en comparación con otros modelos de regresión.

El modelo de referencia está orientado a aprendizaje con datos, donde dos conjuntos de datos procedentes de sensores ambientales (irradiancia en la superficie fotovoltaica y temperatura ambiente) constituyen los flujos de entrada S dentro de una ventana temporal de W pasos, configurando así una matriz de entrada de $S \times W$. El flujo de datos del sensor fuente s se define como $\overline{S}_s = \overline{s}_0, \dots, \overline{s}_i$ y un valor dado en un instante de tiempo t_i se denota como $S_s(t_i) = s_i$. La irradiancia en la superficie fotovoltaica G_I , la temperatura ambiente T_{am} y la potencia generada por el sistema fotovoltaico P_A proporcionan tres flujos de datos que describen el comportamiento del sistema fotovoltaico y, en última instancia, dictaminan la producción de energía del sistema. Al igual que en capítulos anteriores, la segmentación de estos datos se define mediante varias ventanas temporales, que se definen como intervalos de tiempo a corto plazo $W_w = [W_w^-, W_w^+]$ sobre un flujo de datos del sensor \overline{S}_s que agrega los valores \overline{s}_i utilizando una función de agregación $T_t(S_s, W_w, t) = \bigcup_{s_i}^{\overline{s}_i} s_i, t_i \in [t^- W_w^-, t^* - W_w^+]$.

En cuanto a la arquitectura del modelo de aprendizaje profundo, en primer lugar se integran tres capas de redes convolucionales de una dimensión (CNN 1D) como extractores de características espaciales. En segundo lugar, se utilizan dos capas de

redes de memoria a corto-largo plazo (LSTM) para modelar las dependencias temporales a partir de las características generadas por las capas de CNN. Se incluyen capas de eliminación de neuronas (o *Dropout*) para evitar el sobreajuste en las capas de LSTM. Por último, se utilizan tres capas densas para modelar la potencia fotovoltaica.

En base al pequeño error entre la potencia real y la estimada en este modelo de referencia, los modelos CGAN propuestos a continuación tienen como objetivo buscar una mejora en el rendimiento y, al mismo tiempo, introducir mecanismos de dopaje y discriminación en la señal de salida.

5.2.2. Predicción y discriminación de potencia fotovoltaica mediante modelos CGAN de regresión y clasificación

A partir del modelo de aprendizaje profundo de referencia descrito en la sección anterior, se desarrolla un modelo CGAN que se estructura en dos componentes principales:

- Un generador, encargado de producir una estimación de la potencia generada por un sistema fotovoltaico basado en un nivel de distorsión ajustable.
- Un discriminador, que intenta realizar una clasificación binaria entre muestras reales correspondientes a los datos de potencia del sistema fotovoltaico y muestras falsas, producidas por el generador.

La innovación principal en este enfoque, en comparación con modelos más simples de generador y clasificación binaria [Huang et al., 2022, Li et al., 2022, Polo-Rodriguez et al., 2020] radica en la conexión del ruido estándar del generador con un nivel de distorsión

para dopar la generación de energía de salida. El modelo puede aprender este nivel de distorsión gracias al juego de equilibrio entre el generador y el discriminador. Para lograr esto, se permite ajustar el generador con un nivel de distorsión, cuyos valores de ruido cercanos o iguales a cero están relacionados con la generación de energía de salida real, mientras que los valores más cercanos a un umbral α están relacionados con muestras dopadas y falsificadas.

En primer lugar, se implementa el generador utilizando el modelo CNN+LSTM de referencia descrito, cuyas entradas de sensor $S \times W$ son las mediciones del sensor S del sistema fotovoltaico en W pasos temporales, junto con una señal de entrada de ruido N . La salida del modelo (predicción inmediata de potencia fotovoltaica) es extremadamente cercana a los datos reales recogidos del sistema fotovoltaico. Por esta razón, la potencia estimada no puede utilizarse como muestras falsas sin disminuir el rendimiento de la discriminación [Polo-Rodriguez et al., 2023]. Para resolver esta limitación, las salidas reales y no dopadas se configuran en este experimento como muestras verdaderas, mientras que las muestras dopadas mediante ruido aditivo del nivel de distorsión se configuran como muestras falsas para la entrada del modelo discriminador.

A continuación se integra el discriminador, que recibe las muestras (reales y dopadas) así como la entrada del sensor (sin nivel de distorsión), con el objetivo de discernir si la señal recibida correspondiente a la potencia fotovoltaica es real o falsa. La salida del discriminador se configura como una clasificación binaria con entropía cruzada binaria. En última instancia, un modelo de redes adversarias conecta tanto las entradas como las salidas del modelo generador para desarrollar una estructura común para el aprendizaje en dos etapas (que se describe en la sección 5.3) para equilibrar la siguiente función:

$$E_{x,y}[\log(D(x, y))] + E_{x,z}[\log(1 - D(x, G(x, z)))]$$

Donde:

- x representa los datos de entrada de los sensores ambientales
- y corresponde a los datos reales de potencia recibidos del inversor del sistema fotovoltaico
- $D(x, y)$ describe la clasificación del discriminador de los muestras auténticas y dados unos datos de entrada de los sensores ambientales x en verdaderos (1) o falsos (0).
- $G(x, z)$ representa la potencia estimada por el generador para la muestra x con un nivel de distorsión z
- $D(x, G(x, z))$ describe la clasificación del discriminador de los datos procedentes del generador en verdaderos (1) o falsos (0).

La función se deriva de la entropía cruzada entre las distribuciones de los datos reales y los generados con distorsión, la cual es aprendida para ser maximizada por el discriminador y minimizada por el generador. Este enfoque de modelo CGAN permite que el generador y el discriminador trabajen en conjunto para mejorar la precisión de la estimación y la capacidad de discernir entre datos o estimaciones reales y estimaciones falsificadas. El objetivo es lograr un equilibrio óptimo entre la generación realista y la discriminación efectiva.

En la figura 5.1, se presenta una descripción visual de los componentes y conexiones que componen este modelo CGAN.

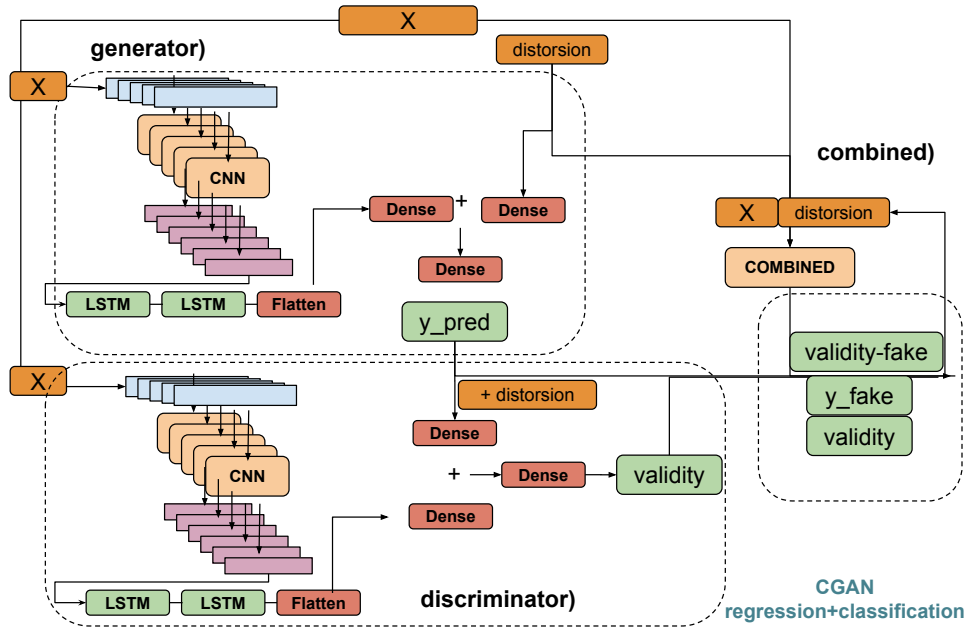


Figura 5.1: Modelo CGAN de regresión y clasificación

5.2.3. Predicción y rectificación de potencia fotovoltaica mediante modelos CGAN de regresión y dopaje

En esta sección, se describe un modelo de regresión basado en CGAN que incluye un nuevo componente encargado de añadir un dopaje a la estimación de potencia fotovoltaica. Este componente adicional aprende a incluir un incremento positivo o negativo en la señal, generando así un modelo de dopaje para intentar engañar al discriminador. Por otro lado, el discriminador aprende a determinar el aumento o disminución en el modelo de dopaje para conseguir rectificar el engaño.

El interés de esta composición se basa en el equilibrio de suma cero, lo que posibilita un aprendizaje no guiado de un pequeño aumento o disminución de la regresión estimada en el modelo de dopaje para engañar al discriminador. El discriminador también aprende cómo rectificar una estimación real o falsa de la potencia

fotovoltaica en este juego entre los componentes del modelo.

De esta manera, este modelo de redes adversarias está compuesto por tres componentes:

- El generador: este componente proporciona una estimación de la potencia fotovoltaica $G(x, z)$ basada en los datos de entrada x y una señal de ruido z .
- El modelo de dopaje: basado en esa estimación, este componente calcula un incremento o decremento $\Delta(x, G(x, z))$ utilizando un nivel de distorsión de ruido sobre la estimación de la potencia del generador.
- El discriminador: por último, el discriminador rectifica la señal del modelo de dopaje con respecto a la verdad fundamental y , minimizando la siguiente función:

$$E_{x,y,z}[|y - D(x, \Delta(x, G(x, z)))|]$$

Esta diferencia $E_{x,y,z}$ entre las muestras dopadas y reales está vinculada al aumento en el modelo de dopaje y la disminución del discriminador en un juego equilibrado de suma cero.

En la figura 5.2, se describe la conexión entre los componentes de este modelo, que incluye el generador, el modelo de dopaje y el discriminador. Cada uno de estos componentes se basa en la configuración de redes neuronales CNN+LSTM descrita anteriormente en el capítulo "3 - Modelos de Aprendizaje Automático y Aprendizaje Profundo para la Predicción Inmediata de Potencia Fotovoltaica" para extraer características espaciales y temporales con el objetivo de ajustar o rectificar los valores de potencia fotovoltaica.

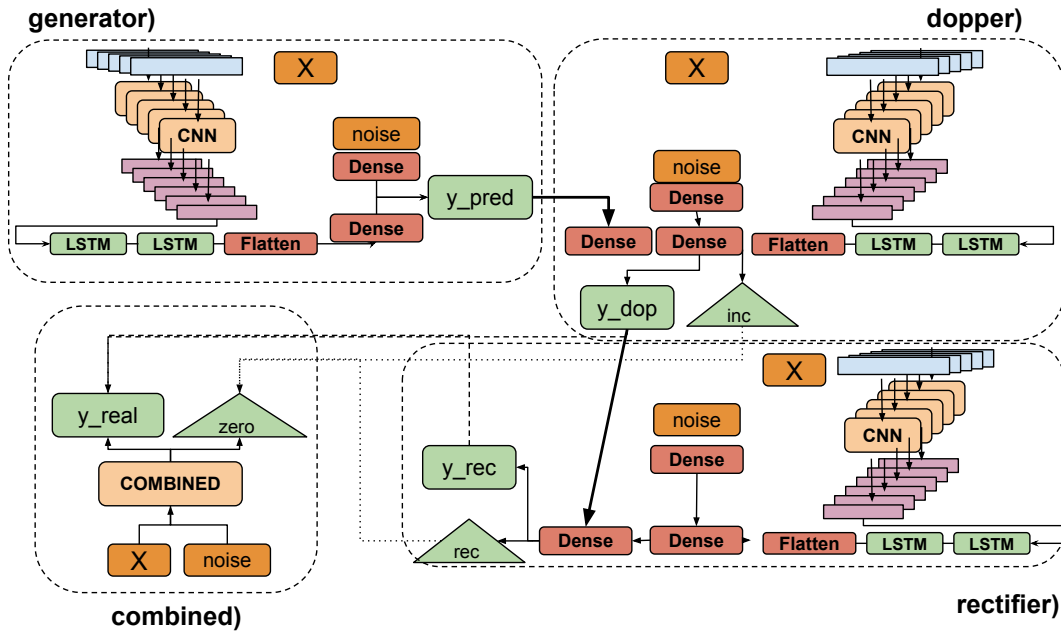


Figura 5.2: Regresión CGAN con modelo de dopaje y discriminador

5.3. Análisis de Resultados Experimentales

Una vez definidos los modelos CGAN y sus componentes, se lleva a cabo la ejecución del experimento sobre tres conjuntos de datos procedentes de sistemas fotovoltaicos. Los resultados del experimento presentados a continuación permiten evaluar el rendimiento del modelo en cuanto a la estimación y la discriminación de muestras falsas en la producción de energía fotovoltaica.

Las señales de entrada de los modelos son la temperatura ambiente y la irradiancia global, mientras que la salida es la potencia generada por el sistema fotovoltaico. En primer lugar, se aplica el preprocesamiento en los tres conjuntos de datos tal como se describe en la sección "5.2.1 - Preprocesamiento de datos y modelo de aprendizaje profundo de referencia":

- Conjunto de datos del sistema UniVer: las señales se agregan desde la muestra

original de 30 segundos hasta un promedio de 10 minutos y luego se segmentan en una ventana deslizante de 90 minutos ($W = 9$). El conjunto de datos está compuesto por un total de 24.031 muestras.

- Conjunto de datos del sistema SolarTech: las señales se agregan desde la muestra original de 1 minuto hasta un promedio de 10 minutos y luego se segmentan en una ventana deslizante de 90 minutos ($W = 9$). El conjunto de datos está compuesto por un total de 17.136 muestras.
- Conjunto de datos del sistema fotovoltaico ubicado en India: las señales de la muestra original son de 15 minutos, y se desarrolla una ventana deslizante de 90 minutos ($W = 6$). El conjunto de datos está compuesto por un total de 3.264 muestras.

La definición del tamaño de ventana de 90 minutos se basa en el estudio realizado con diferentes configuraciones en el capítulo "3 - Modelos de Aprendizaje Automático y Aprendizaje Profundo para la Predicción Inmediata de Potencia Fotovoltaica". En cuanto a la configuración de los diferentes procesos de entrenamiento de los modelos CGAN, se establecen unos parámetros de 60 épocas y 64 lotes para poder comparar resultados homogéneos.

Antes de describir el rendimiento del CGAN, se presentan los resultados del modelo base de aprendizaje profundo para predecir la energía generada por los sistemas fotovoltaicos. En la tabla 5.1, se presentan los resultados utilizando las métricas de error correspondientes a la raíz del error cuadrático medio (RMSE, *Root Mean Squared Error*) y raíz del error cuadrático medio normalizado (NRMSE, *Normalized Root Mean Squared Error*) para cada conjunto de datos de los sistemas fotovoltaicos: UniVer, SolarTech e India.

Tabla 5.1: Resumen de métricas de error para el modelo base CNN+LSTM (promedio de las últimas 10 épocas)

Contexto	RMSE (W)	NRMSE
SolarTech	6.18	$2,49 \cdot 10^{-2}$
UniVer	610.85	$2,65 \cdot 10^{-2}$
India	397.79	$2,78 \cdot 10^{-2}$

5.3.1. Modelo CGAN de regresión y clasificación

En esta sección se presentan los resultados obtenidos al aplicar el modelo CGAN capaz de realizar tareas de regresión y clasificación. Este modelo se enfoca en la ejecución de tareas de predicción inmediata, identificación de dopaje y discernimiento de estimaciones que podrían ser potencialmente falsas. El pseudocódigo descrito en el algoritmo 1 refleja la implementación de este modelo, cuya composición se detalla a continuación:

- **Generador:** este componente incorpora un nivel de distorsión basado en una distribución uniforme aleatoria $U(0, \alpha)$. Consta de dos capas densas con 128 y 64 unidades respectivamente. Estas capas están conectadas al nivel de distorsión, permitiendo la combinación de características procedentes del modelo base (CNN+LSTM) y generando una salida pseudo-estocástica.
- **Discriminador:** en este módulo se reciben las entradas provenientes del sistema fotovoltaico, así como la estimación generada por el generador. De forma similar al generador, el discriminador también cuenta con dos capas densas, cada una con 128 y 64 unidades. Su función principal es combinar características del modelo base (CNN+LSTM) y realizar una estimación mediante clasificación binaria. Esta estimación determina si la estimación generada y las entradas son reales o falsas.

- Combinado: este componente es responsable de recibir la entrada del sistema fotovoltaico, la estimación generada por el generador y el nivel de distorsión. Su tarea consiste en estimar, mediante clasificación binaria, si tanto la estimación generada como las entradas son reales o falsas, basándose en la información combinada de todas las fuentes.

Algorithm 1 Entrenamiento de CGAN de regresión+clasificación

```

1:  $N \leftarrow \text{len}(X)$ 
2:  $\text{batch\_size} \leftarrow N/\text{batchs}$ 
3: for  $e \in [0, \text{epochs}]$  do
4:   for  $b \in [0, \text{batchs}]$  do
5:      $\text{distorsion}_\alpha \leftarrow U(0, \alpha)$ 
6:      $i = \text{ranInd}(\text{batch\_size}, N)$  ▷ Muestras aleatorias
7:      $x, y_{\text{real}} \leftarrow X[i], Y[i]$ 
8:      $y_{\text{pred}} = \text{generator.predict}(x, \text{distorsion}_\alpha)$ 
9:      $\text{discriminator.train}([x, y_{\text{real}}], \text{true})$  ▷ Entrenamiento basado en
muestras reales
10:     $\text{discriminator.train}([x, y_{\text{real}} + \text{distorsion}_\alpha], \text{false})$ 
11:     $\text{discriminator.train}([x, y_{\text{pred}}], \text{true})$  ▷ Entrenamiento basado en
muestras dopadas
12:     $\text{discriminator.train}([x, y_{\text{real}} + \text{distorsion}_\alpha], \text{false})$ 
13:     $i = \text{ranInd}(\text{batch\_size}, N)$  ▷ Muestras para el modelo combinado
14:     $x, y_{\text{real}} \leftarrow X[i], Y[i]$ 
15:     $y_{\text{pred}} = \text{generator.predict}(x, \text{distorsion}_\alpha)$ 
16:     $\text{combined.train}([0, x], [y_{\text{real}}, \text{false}])$ 
17:     $\text{combined.train}([\text{distorsion}_\alpha, x], [y_{\text{pred}} + \text{distorsion}_\alpha, \text{true}])$  ▷
Entrenamiento basado en el nivel de distorsión
18:     $\text{combined.train}([y_{\text{real}} - y_{\text{pred}}, x], [y_{\text{pred}}, \text{true}])$  ▷ Entrenamiento para
relacionar la distorsión con lo real y la predicción
19:   end for
20: end for

```

Con el propósito de llevar a cabo una evaluación exhaustiva de las capacidades de este modelo, se definen los siguientes aspectos:

- Se seleccionan tres valores de nivel de distorsión, denotados como $\text{distorsion}_\alpha =$

$\{0,075, 0,15, 0,25\}$, con el fin de estimar el desempeño tanto del generador como del clasificador.

- Se calculan las métricas de error correspondientes a la raíz del error cuadrático medio (RMSE) y su versión normalizada (NRMSE) durante las últimas 10 épocas del proceso de validación del modelo.
- Se evalúa la precisión del discriminador con respecto a las muestras reales.
- De igual manera, se evalúa la eficacia del discriminador frente a muestras falsas generadas por el generador con el nivel de distorsión máximo α .

Los resultados obtenidos se muestran en las tablas 5.2 y 5.3, las cuales reflejan las métricas de error y la precisión de las muestras reales y falsas, respectivamente. En la Figura 5.3, se representa la evolución de la precisión en la predicción de muestras reales y falsas, así como el NRMSE a lo largo de las distintas épocas y niveles de distorsión.

Tabla 5.2: Resumen de métricas de error en el generador: RMSE (W) y NRMSE en el modelo CGAN de regresión y clasificación (media de las 10 últimas épocas)

	$\alpha=0.075$		$\alpha=0.15$	$\alpha=0.25$
	RMSE	NRMSE	NRMSE	NRMSE
Solartech	6.05	$2,48 \cdot 10^{-2}$	$4,84 \cdot 10^{-2}$	$4,23 \cdot 10^{-2}$
UniVer	529.83	$1,86 \cdot 10^{-2}$	$2,21 \cdot 10^{-2}$	$3,78 \cdot 10^{-2}$
India	368.55	$2,56 \cdot 10^{-2}$	$3,18 \cdot 10^{-2}$	$4,78 \cdot 10^{-2}$

Con base en los datos de evaluación, se pueden identificar los siguientes patrones en las fases de entrenamiento y evaluación del modelo:

- La mejor configuración para construir un generador que se asemeje más a las muestras reales se logra utilizando un valor bajo de alfa. Esta configuración

Tabla 5.3: Resumen de precisión (Prc.) en el discriminador: precisión de muestras falsas y muestras reales en el modelo CGAN de regresión y clasificación (promedio de las últimas 10 épocas).

	$\alpha=0.075$		$\alpha=0.15$		$\alpha=0.25$	
	Prc. falsas	Prc. reales	Prc. falsas	Prc. reales	Prc. falsas	Prc. reales
SolarTech	0.57	0.91	0.91	0.88	0.92	0.88
UniVer	0.73	0.71	0.89	0.67	0.94	0.85
India	0.55	0.90	0.64	0.76	0.80	0.85

conduce a una mejora en el desempeño del modelo de aprendizaje profundo de referencia.

- La configuración óptima para construir un discriminador capaz de distinguir de manera robusta entre muestras reales y falsas se caracteriza por un valor elevado de alfa. Esta configuración muestra un rendimiento prometedor en los conjuntos de datos SolarTech y UniVer. Sin embargo, el conjunto de datos de India muestra resultados más variables durante las épocas de entrenamiento, debido a su tamaño reducido y a la presencia de ruido e imprecisiones en las mediciones.

5.3.2. Modelo CGAN de regresión y dopaje

En esta sección se presentan los resultados obtenidos al aplicar el modelo CGAN de regresión, incluyendo el componente de dopaje, sobre los conjuntos de datos de los sistemas disponibles (SolarTech, UniVer e India), con el propósito de realizar una predicción instantánea de la potencia generada dichos sistemas y una discriminación de posibles estimaciones dopadas o falsas.

A continuación se detallan los componentes de este modelo, cuya implementación se recoge en forma de pseudocódigo en el algoritmo 2:

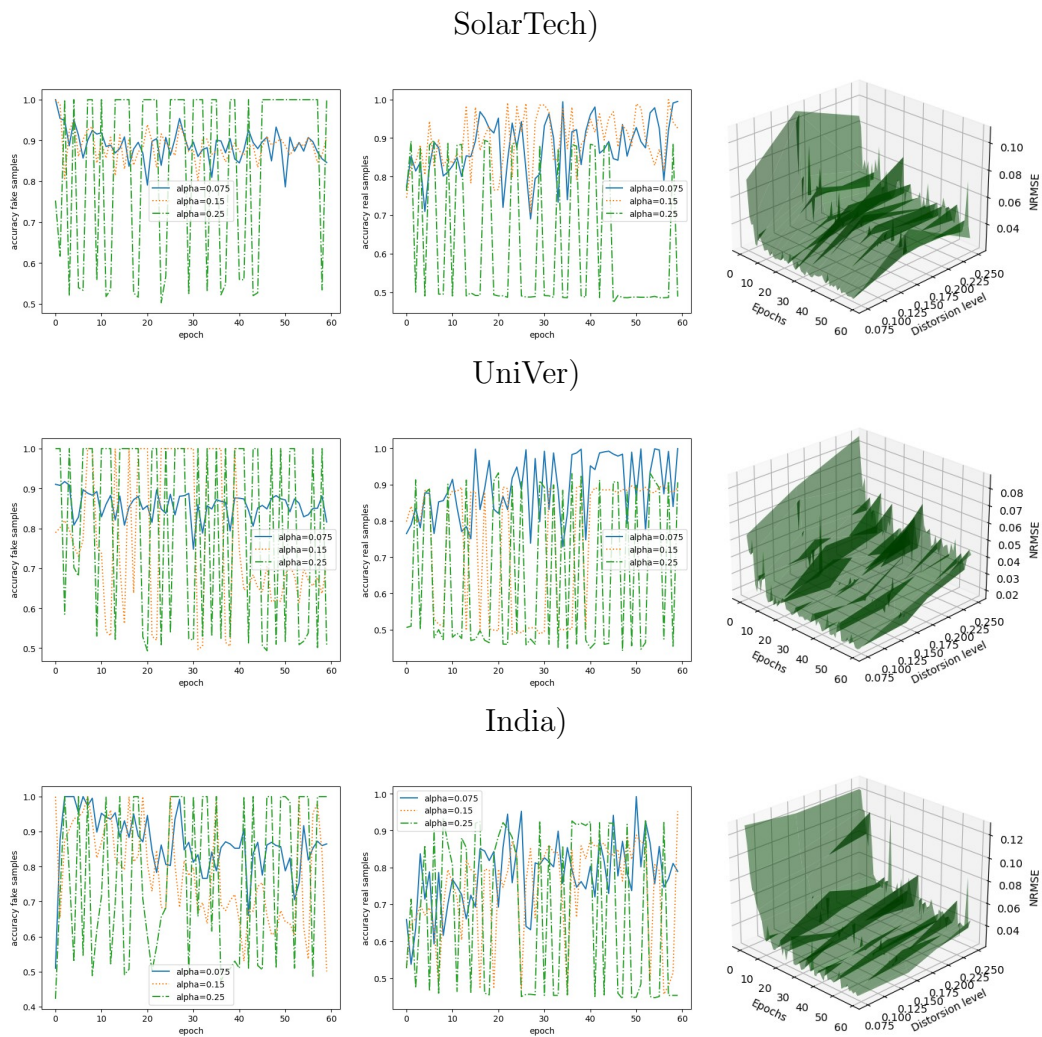


Figura 5.3: Precisión en la predicción de muestras reales y falsas y NRMSE por épocas y nivel de distorsión en la regresión + clasificación CGAN.

- **Generador:** este componente incorpora una entrada de ruido basada en una distribución uniforme aleatoria $U(0, 1)$. Dispone de dos capas densas, con 128 y 64 unidades respectivamente, que están conectadas al nivel de distorsión. Esto permite la combinación de características provenientes del modelo base (CNN+LSTM) y la generación de una salida pseudo-estocástica.
- **Modelo de dopaje (o *doper*):** este componente recibe las entradas procedentes

del sistema fotovoltaico y la estimación generada por el generador, mientras que su salida corresponde a una estimación de potencia de salida dopada. De forma similar al generador, incluye una entrada de ruido basada en una distribución uniforme aleatoria $U(0, 1)$.

- Rectificador: recibe las entradas de los sistemas fotovoltaicos y la estimación de potencia de salida del *doper*. Dos capas densas (128+64 unidades) están conectadas a la predicción para combinar características del modelo base (CNN+LSTM) y estimar la desviación del modelo de dopaje.
- Combinado: este componente recibe la entrada del sistema fotovoltaico, la estimación arrojada por el generador y el nivel de distorsión. Su tarea principal es estimar la potencia de salida real y una desviación nula, ya que se entrena con datos reales.

A continuación, en la tabla 5.3.2, se presentan los resultados en forma de métricas de error para las últimas 10 épocas de los conjuntos de datos, así como el error relativo (RE, *Relative Error*) entre los tres componentes: generador (gen), *doper* (dop) y discriminador (dis). En la figura 5.4, se presenta la progresión del error relativo medido a lo largo de las diferentes épocas durante la etapa de aprendizaje.

Contexto	RMSE (W)	NRMSE (W)	RE Gen-Dop	RE Dop-Dis	RE Gen-Dis
SolarTech	5.55	$2,28 \cdot 10^{-2}$	$3,41 \cdot 10^{-3}$	$3,38 \cdot 10^{-3}$	$0,29 \cdot 10^{-3}$
UniVer	513.38	$1,80 \cdot 10^{-2}$	$2,96 \cdot 10^{-3}$	$5,33 \cdot 10^{-3}$	$2,27 \cdot 10^{-3}$
India	388.63	$1,72 \cdot 10^{-2}$	$12,37 \cdot 10^{-3}$	$12,86 \cdot 10^{-3}$	$0,17 \cdot 10^{-3}$

Tabla 5.4: Resumen de métricas de error del modelo de regresión de doping para predicción instantánea y rectificación de generación de potencia de salida (promedio de las últimas 10 épocas)

Algorithm 2 CGAN de regresión y dopaje

```

1:  $N \leftarrow \text{len}(X)$ 
2:  $\text{batch\_size} \leftarrow N/\text{batchs}$ 
3: for  $e \in [0, \text{epochs}]$  do
4:   for  $b \in [0, \text{batchs}]$  do
5:      $i = \text{ranInd}(\text{batch\_size}, N)$  ▷ Muestras aleatorias
6:      $x, y_{\text{real}} \leftarrow X[i], Y[i]$ 
7:      $\text{ruido} \leftarrow U(0, 1)$ 
8:      $y_{\text{pred}} = \text{generator.predict}(x, \text{ruido})$  ▷ Predicción del generador
9:      $\delta_y = \text{doper.predict}(x, \text{ruido}, y_{\text{pred}})$  ▷ Salida del doper
10:     $\text{discriminator.train}([x, y_{\text{pred}} + \delta_y], -\delta_y)$  ▷ Aprendizaje de desviación
    positiva
11:     $\text{discriminator.train}([x, y_{\text{pred}} - \delta_y], +\delta_y)$  ▷ Aprendizaje de desviación
    negativa
12:     $i = \text{ranInd}(\text{batch\_size}, N)$  ▷ Muestras para el modelo combinado
13:     $x, y_{\text{real}} \leftarrow X[i], Y[i]$ 
14:     $\text{combinado.train}([\text{ruido}, x], [y_{\text{real}}, 0])$  ▷ Aprendizaje de desviación cero
15:   end for
16: end for

```

En función de los datos de evaluación obtenidos, se pueden destacar los siguientes aspectos clave acerca del rendimiento de este modelo:

- El aumento ficticio en la estimación del *doper* disminuye gradualmente durante el proceso de aprendizaje, a través de un juego de suma cero entre el *doper* y el rectificador. Se observa que estos dos componentes tienen patrones prácticamente opuestos entre sí, como se puede apreciar en la figura 5.4.
- En los conjuntos de datos de UniVer y SolarTech, se registra una notable mejora en la estimación de la potencia de salida al utilizar el modelo de regresión CGAN propuesto, en comparación con el modelo base y el modelo de clasificación CGAN. Esta mejora se atribuye a la capacidad de minimizar el aumento o disminución del modelo de dopaje y a la robustez del discriminador, lo que supera el rendimiento del generador en la predicción de la potencia de salida

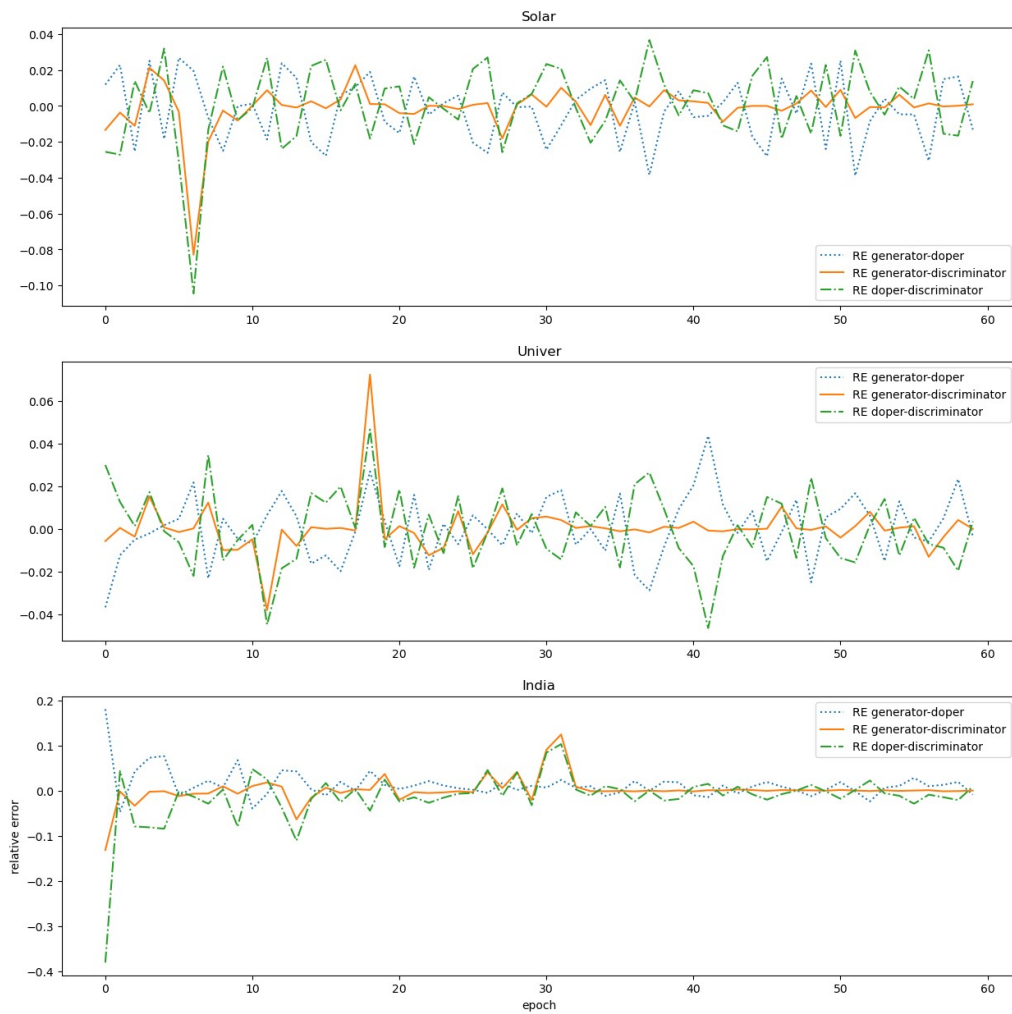


Figura 5.4: Error relativo durante las épocas en la etapa de aprendizaje entre los tres componentes: generador, doper y discriminador.

instantánea.

- Al igual que en el caso del modelo de clasificación CGAN, el conjunto de datos de India muestra resultados más variables durante las épocas de aprendizaje debido a su tamaño más reducido y la presencia de ruido y datos inciertos. Aunque los resultados son superiores al modelo base, no logran superar al modelo de clasificación CGAN en este conjunto de datos.

Capítulo 6

Conclusiones

6.1. Modelos de Aprendizaje Automático y Aprendizaje Profundo para la Predicción Inmediata de Potencia Fotovoltaica

En el contexto de esta tesis se ha desarrollado un prototipo completo, *end-to-end*, de una plataforma para el seguimiento y gestión de plantas fotovoltaicas que incorpora tecnologías novedosas en este sector como adquisición de datos con dispositivos IoT e inteligencia artificial para el pronóstico en tiempo real del comportamiento de la planta.

El módulo de captación de datos IoT se ha desarrollado utilizando un microprocesador Arduino junto con sensores de bajo coste instalados que recopilan datos ambientales y de energía del sistema fotovoltaico. En cuanto a la predicción del comportamiento del sistema se han desarrollado modelos basados en datos con dos enfoques de aprendizaje automático: i) características diseñadas manualmente con

múltiples ventanas temporales y ii) modelos de aprendizaje profundo, CNN+LSTM, k-NN y Random Forest que ofrecen un mejor rendimiento en comparación con el modelo analítico estándar de Araujo. En el caso de CNN+LSTM, la ventaja de este modelo de aprendizaje profundo es la falta de intervención humana en la definición de características. El rendimiento de k-NN es notable, con un tiempo de aprendizaje notablemente bajo y capacidades de integración en sistemas embebidos. Finalmente, *Random Forest* con ventanas temporales incrementales tuvo el mejor rendimiento en términos de métricas de error. El aprendizaje profundo mejoró las técnicas clásicas de aprendizaje por extracción de características orientadas por conocimiento.

Un avance potencial en esta línea de trabajo consistiría en un análisis en profundidad del diagnóstico, tipología y patrones de falla en los sistemas fotovoltaicos para predecir estos eventos mediante modelos de aprendizaje automático.

Este trabajo sienta las bases para la implementación de sistemas avanzados de monitorización y gestión de sistemas fotovoltaicos, a la vez que ofrece oportunidades para futuras investigaciones en el campo de la generación de energía renovable y la predicción de eventos en sistemas fotovoltaicos.

6.2. Aprendizaje por Transferencia para la Predicción de Potencia Fotovoltaica

En este trabajo, se ha evaluado la transferencia de aprendizaje de un modelo de aprendizaje profundo, construido con capas convolucionales y de memoria a corto-largo plazo, con el propósito de predecir la generación de energía fotovoltaica en dos dominios diferentes: el sistema fotovoltaico UniVer en la Universidad de Jaén

y el Laboratorio SolarTech en la Universidad Politécnica de Milán. Los resultados destacan la eficacia de este enfoque al trabajar con fuentes heterogéneas.

Utilizando el escenario de aprendizaje básico de cada uno de los dominios como referencia, se mide un margen de error mayor en el escenario de aprendizaje externo, es decir, al entrenar el modelo en un dominio y validar el modelo en el otro. Sin embargo, el modelo mostró un mejor rendimiento en el escenario de aprendizaje afinado que en la referencia, obteniendo un margen de error menor cuando se utiliza un modelo preentrenado de un contexto diferente para ejecutar el entrenamiento y la validación en el contexto local. Por último, también se observa una mejora al medir el tiempo de aprendizaje del modelo de aprendizaje afinado en comparación con el modelo de aprendizaje básico.

Los resultados de este trabajo respaldan la integración de un nuevo componente en la plataforma Ópera para incluir la transferencia de aprendizaje en sistemas en tiempo real bajo demanda. De esta manera, se puede desarrollar un método de agrupación para estimar la estacionalidad durante largos períodos de tiempo y seleccionar grupos de patrones, donde los modelos de predicción instantánea se cargarán según el contexto meteorológico de cada día.

6.3. Modelos de Redes Adversarias Generativas para la Predicción y el Control de Potencia Fotovoltaica

En este trabajo se ha explorado la evaluación del potencial de los modelos de aprendizaje profundo y redes adversarias para la predicción instantánea, el dopaje y

la rectificación de la generación de energía de salida en sistemas fotovoltaicos. Se han introducido dos modelos novedosos para abordar estos objetivos. En primer lugar, se ha propuesto un modelo adversario que incorpora un discriminador basado en clasificación y control de distorsión dentro del generador. Este enfoque se ha diseñado para mejorar la generación de datos de energía de salida minimizando las distorsiones mientras se producen muestras realistas. En segundo lugar, se presentó un modelo adversario basado en regresión, centrado en optimizar el proceso de generación de energía de salida al considerar mecanismos de doping y discriminar las desviaciones de la energía de salida esperada en un equilibrio de suma cero.

Al integrar redes adversarias, los modelos presentados ofrecen soluciones potenciales para detectar datos falsos y mejorar la precisión en la predicción instantánea de generación de energía, mejorando la autenticidad de los datos generados y mitigando las desviaciones de rendimiento. Además, ampliar la evaluación a otros campos de la ingeniería, conjuntos de datos del mundo real y diferentes condiciones ambientales podría proporcionar una comprensión más completa de los modelos propuestos, su robustez y generalización.

En resumen, este trabajo abre nuevas vías para aprovechar las Redes Adversarias y mejorar la precisión y fiabilidad de la generación de energía de salida en sistemas fotovoltaicos. Los modelos propuestos muestran resultados prometedores en predicción instantánea, mejora de datos y gestión de desviaciones, allanando el camino para continuar la investigación y la innovación en el campo de la optimización de la energía renovable.

Capítulo 7

Trabajos Futuros

La integración del aprendizaje automático en la generación de energía solar se presenta como una prometedora vía de investigación y desarrollo con múltiples aplicaciones potenciales. A continuación, se profundiza en las áreas de investigación previamente mencionadas y se introducen nuevos enfoques y tecnologías que podrían ser explorados:

- Identificación de fallos en el generador fotovoltaico y generación de alarmas:
 - a. Avanzar en técnicas de procesamiento de señales y aprendizaje automático para analizar datos de vibración y sonido de los generadores fotovoltaicos, lo que podría ayudar en la detección temprana de problemas mecánicos antes de que afecten la eficiencia general del sistema.
 - b. Desarrollar plataformas de simulación digital que modelen el comportamiento de los generadores fotovoltaicos bajo diversas condiciones operativas y ambientales. Esto permitiría realizar pruebas virtuales de escenarios de fallo y ajustar los sistemas de alarma de acuerdo con los hallazgos.

-
- c. Integrar tecnologías de visión por computadora para inspeccionar visualmente los paneles solares utilizando dispositivos de visión ambiental de alta resolución o termal. Esto podría automatizar la detección de fallos superficiales, como suciedad acumulada o daños físicos en los paneles.
- Detección de fraude:
 - a. Aplicar técnicas avanzadas de minería de datos para descubrir patrones irregulares en los datos de medición que sugieran manipulación o robo de energía. Esto podría incluir el análisis de comportamientos atípicos en comparación con patrones de consumo históricos.
 - b. Desarrollar algoritmos de inteligencia artificial que se adapten dinámicamente a nuevas formas de fraude mediante técnicas de aprendizaje continuo y transferencia de conocimientos entre distintos sistemas y regiones.
 - Computación difusa en los módulos de IoT:
 - a. Investigar la integración de la lógica difusa con tecnologías emergentes de inteligencia artificial en el borde, que permitan realizar el procesamiento de datos directamente en los dispositivos IoT sin necesidad de comunicarse constantemente con un centro de datos central.
 - b. Diseñar y testear nuevos modelos de control basados en lógica difusa que no solo respondan a las fluctuaciones del entorno, sino que también aprendan de ellas para mejorar su eficiencia y precisión en la toma de decisiones.
 - Optimización de la producción de energía y gestión de la red:
-

- a. Desarrollar modelos predictivos que integren datos meteorológicos para prever la generación solar diaria y ajustar de manera proactiva la operación y mantenimiento de las plantas solares.
- b. Investigar nuevas formas de almacenamiento de energía solar que puedan ser optimizadas mediante algoritmos de aprendizaje automático, mejorando la estabilidad de la red y la disponibilidad de energía durante picos de demanda o condiciones climáticas adversas.
- c. Explorar la aplicación de técnicas de aprendizaje profundo para prever y gestionar la demanda energética a nivel local y regional, adaptando la producción de energía solar en tiempo real para satisfacer las necesidades específicas de la red eléctrica.

Estos enfoques no solo fortalecen las bases de la investigación en el campo de la energía solar y aprendizaje automático, sino que también abren nuevas vías para garantizar una operación más eficiente, segura y sostenible de los sistemas de generación de energía solar en el futuro.

Bibliografía

- [IEC, 1998] (1998). Photovoltaic system performance monitoring. guidelines for measurement, data exchange and analysis. Technical Report IEC 61724:1998, European Standard.
- [IEC, 2017a] (2017a). European standard iec 61724:2017. photovoltaic system performance monitoring. guidelines for measurement, data exchange and analysis.
- [Abdel-Hamid et al., 2014] Abdel-Hamid, O., Mohamed, A.-r., Jiang, H., Deng, L., Penn, G., and Yu, D. (2014). Convolutional neural networks for speech recognition. *IEEE/ACM Transactions on audio, speech, and language processing*, 22(10):1533–1545.
- [Abdel-Nasser and Mahmoud, 2019] Abdel-Nasser, M. and Mahmoud, K. (2019). Accurate photovoltaic power forecasting models using deep lstm-rnn. *Neural computing and applications*, 31:2727–2740.
- [Agga et al., 2022] Agga, A., Abbou, A., Labbadi, M., El Houm, Y., and Ali, I. H. O. (2022). Cnn-lstm: An efficient hybrid deep learning architecture for predicting short-term photovoltaic power production. *Electric Power Systems Research*, 208:107908.

- [Aggarwal et al., 2021] Aggarwal, A., Mittal, M., and Battineni, G. (2021). Generative adversarial network: An overview of theory and applications. *International Journal of Information Management Data Insights*, 1(1):100004.
- [Aggarwal et al., 2019a] Aggarwal, K., Kirchmeyer, M., Yadav, P., Keerthi, S. S., and Gallinari, P. (2019a). Conditional generative adversarial networks for regression. *ArXiv190512868 Cs Stat.(10)*, 133:142–146.
- [Aggarwal et al., 2019b] Aggarwal, K., Kirchmeyer, M., Yadav, P., Keerthi, S. S., and Gallinari, P. (2019b). Regression with conditional gan. *arXiv preprint arXiv:1905.12868*.
- [Ahmed et al., 2020] Ahmed, R., Sreeram, V., Mishra, Y., and Arif, M. (2020). A review and evaluation of the state-of-the-art in pv solar power forecasting: Techniques and optimization. *Renewable and Sustainable Energy Reviews*, 124:109792.
- [Ahn and Park, 2021] Ahn, H. K. and Park, N. (2021). Deep rnn-based photovoltaic power short-term forecast using power iot sensors. *Energies*, 14(2):436.
- [Akhter et al., 2022] Akhter, M. N., Mekhilef, S., Mokhlis, H., Almohaimeed, Z. M., Muhammad, M. A., Khairuddin, A. S. M., Akram, R., and Hussain, M. M. (2022). An hour-ahead pv power forecasting method based on an rnn-lstm model for three different pv plants. *Energies*, 15(6):2243.
- [Akram et al., 2020] Akram, M. W., Li, G., Jin, Y., Chen, X., Zhu, C., and Ahmad, A. (2020). Automatic detection of photovoltaic module defects in infrared images with isolated and develop-model transfer deep learning. *Solar Energy*, 198:175–186.

- [Almeida et al., 2017a] Almeida, M. P., Muñoz, M., de la Parra, I., and Perpiñán, O. (2017a). Comparative study of pv power forecast using parametric and non-parametric pv models. *Solar Energy*, 155:854–866.
- [Almeida et al., 2017b] Almeida, M. P., Muñoz, M., de la Parra, I., and Perpiñán, O. (2017b). Comparative study of PV power forecast using parametric and non-parametric PV models. *Sol. Energy*.
- [Andina et al., 2009] Andina, D., Alvarez-Vellisco, A., Aleksandar, J., and Fombellida, J. (2009). Artificial metaplasticity can improve artificial neural networks learning. *Intelligent Automation & Soft Computing*, 15(4):683–696.
- [Aprillia et al., 2020] Aprillia, H., Yang, H.-T., and Huang, C.-M. (2020). Short-term photovoltaic power forecasting using a convolutional neural network–salp swarm algorithm. *Energies*, 13(8):1879.
- [Araujo and Sanchez, 1982] Araujo, G. and Sanchez, E. (1982). Analytical expressions for the determination of the maximum power point and the fill factor of a solar cell. *Solar Cells*, 5(4):377–386.
- [Araujo and Sánchez,] Araujo, G. L. and Sánchez, E. Analytical expressions for the determination of the maximum power point and the fill factor of a solar cell.
- [Araujo and Sánchez, 1982] Araujo, G. L. and Sánchez, E. (1982). Analytical expressions for the determination of the maximum power point and the fill factor of a solar cell. *Sol. Cells*.

- [Asgarian et al., 2018] Asgarian, A., Sobhani, P., Zhang, J. C., Mihailescu, M., Sibilia, A., Ashraf, A. B., and Taati, B. (2018). A hybrid instance-based transfer learning method. *arXiv preprint arXiv:1812.01063*.
- [Banos et al., 2015] Banos, O., Galvez, J.-M., Damas, M., Guillen, A., Herrera, L.-J., Pomares, H., Rojas, I., Villalonga, C., Hong, C. S., and Lee, S. (2015). Multiwindow fusion for wearable activity recognition. In *International Work-Conference on Artificial Neural Networks*, pages 290–297. Springer.
- [Banos et al., 2014] Banos, O., Galvez, J.-M., Damas, M., Pomares, H., and Rojas, I. (2014). Window size impact in human activity recognition. *Sensors*, 14(4):6474–6499.
- [Bermejo et al., 2019] Bermejo, J. F., Fernández, J. F. G., Polo, F. O., and Márquez, A. C. (2019). A review of the use of artificial neural networks models for energy and reliability prediction. a study for the solar pv, hydraulic and wind energy sources. *Appl. Sci*, 9:1844.
- [Blaesser and Zaiman, 1991] Blaesser, G. and Zaiman, W. (1991). On-Site Power Measurements on Large PV Arrays. In *Tenth E.C. Photovolt. Sol. Energy Conf.*
- [Blaesser, G; Munro, 1996] Blaesser, G; Munro, D. (1996). Guidelines for the assessment of photovoltaic plants.
- [Chen et al., 2022] Chen, Y., Yang, X.-H., Wei, Z., Heidari, A. A., Zheng, N., Li, Z., Chen, H., Hu, H., Zhou, Q., and Guan, Q. (2022). Generative adversarial networks in medical image augmentation: A review. *Computers in Biology and Medicine*, 144:105382.

- [Creswell et al., 2018] Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., and Bharath, A. A. (2018). Generative adversarial networks: An overview. *IEEE signal processing magazine*, 35(1):53–65.
- [Cruciani et al., 2019] Cruciani, F., Vafeiadis, A., Nugent, C., Cleland, I., McCullagh, P., Votis, K., Giakoumis, D., Tzovaras, D., Chen, L., and Hamzaoui, R. (2019). Comparing cnn and human crafted features for human activity recognition. IEEE.
- [Daliento et al., 2017] Daliento, S., Chouder, A., Guerriero, P., Pavan, A. M., Mellit, A., Moeini, R., and Tricoli, P. (2017). Monitoring, diagnosis, and power forecasting for photovoltaic fields: A review.
- [De la Parra et al., 2017] De la Parra, I., Muñoz, M., Lorenzo, E., García, M., Marcos, J., and Martínez-Moreno, F. (2017). Pv performance modelling: A review in the light of quality assurance for large pv plants. *Renewable and Sustainable Energy Reviews*, 78:780–797.
- [Ding et al., 2018] Ding, S., Yang, Q., Li, X., Yan, W., and Ruan, W. (2018). Transfer learning based photovoltaic module defect diagnosis using aerial images. In *2018 International Conference on Power System Technology (POWERCON)*, pages 4245–4250. IEEE.
- [Drif et al., 2007a] Drif, M., Pérez, P., Aguilera, J., Almonacid, G., Gomez, P., de la Casa, J., and Aguilar, J. (2007a). Univer Project. A grid connected photovoltaic system of 200 kW_p at Jaén University. Overview and performance analysis. *Sol. Energy Mater. Sol. Cells*, 91(8):670–683.

- [Drif et al., 2007b] Drif, M., Pérez, P., Aguilera, J., Almonacid, G., Gomez, P., De la Casa, J., and Aguilar, J. (2007b). Univer project. a grid connected photovoltaic system of 200kwp at jaén university. overview and performance analysis. *Solar Energy Materials and Solar Cells*, 91(8):670–683.
- [Erhan et al., 2010] Erhan, D., Courville, A., Bengio, Y., and Vincent, P. (2010). Why does unsupervised pre-training help deep learning? In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 201–208. JMLR Workshop and Conference Proceedings.
- [Espinilla et al., 2018a] Espinilla, M., Medina, J., Hallberg, J., and Nugent, C. (2018a). A new approach based on temporal sub-windows for online sensor-based activity recognition. *Journal of Ambient Intelligence and Humanized Computing*, pages 1–13.
- [Espinilla et al., 2018b] Espinilla, M., Medina, J., Salguero, A., Irvine, N., Donnelly, M., Cleland, I., and Nugent, C. (2018b). Human activity recognition from the acceleration data of a wearable device. which features are more relevant by activities? In *Multidisciplinary Digital Publishing Institute Proceedings*, volume 2 (9), page 1242.
- [Europe, 2016] Europe, S. P. (2016). O&m best practice guidelines. *Solar Power Europe*.
- [Europe, 2019] Europe, S. P. (2019). Global market outlook for solar power/2019–2023. *Solar Power Europe, Brussels, Belgium, Tech. Rep.*
- [Europe, 2023] Europe, S. P. (2023). Global market outlook for solar power/2023–2027. *Solar Power Europe, Brussels, Belgium, Tech. Rep.*

- [Fawaz et al., 2018] Fawaz, H. I., Forestier, G., Weber, J., Idoumghar, L., and Muller, P.-A. (2018). Transfer learning for time series classification. In *2018 IEEE international conference on big data (Big Data)*, pages 1367–1376. IEEE.
- [Ferreira et al., 2013] Ferreira, H. G. C., Canedo, E. D., and De Sousa, R. T. (2013). Iot architecture to enable intercommunication through rest api and upnp using ip, zigbee and arduino. In *2013 IEEE 9th international conference on wireless and mobile computing, networking and communications (WiMob)*, pages 53–60. IEEE.
- [Fuentes et al., 2007a] Fuentes, M., Nofuentes, G., Aguilera, J., Talavera, D., and Castro, M. (2007a). Application and validation of algebraic methods to predict the behaviour of crystalline silicon pv modules in mediterranean climates. *Solar Energy*, 81(11):1396–1408.
- [Fuentes et al., 2007b] Fuentes, M., Nofuentes, G., Aguilera, J., Talavera, D. L., and Castro, M. (2007b). Application and validation of algebraic methods to predict the behaviour of crystalline silicon PV modules in Mediterranean climates. *Sol. Energy*.
- [Fuentes et al., 2014] Fuentes, M., Vivar, M., Burgos, J. M., Aguilera, J., and Vacas, J. A. (2014). Design of an accurate, low-cost autonomous data logger for PV system monitoring using Arduino™ that complies with IEC standards. *Sol. Energy Mater. Sol. Cells*, 130.
- [Gamboa, 2017] Gamboa, J. C. B. (2017). Deep learning for time-series analysis. *arXiv preprint arXiv:1701.01887*.

- [Gaviria et al., 2022] Gaviria, J. F., Narváez, G., Guillen, C., Giraldo, L. F., and Bressan, M. (2022). Machine learning in photovoltaic systems: A review. *Renewable Energy*.
- [Gonzalo et al., 2020] Gonzalo, A. P., Marugán, A. P., and Márquez, F. P. G. (2020). Survey of maintenance management for photovoltaic power systems. *Renewable and Sustainable Energy Reviews*, 134:110347.
- [Gutierrez et al., 2020] Gutierrez, S., Rodrigo, P. M., Alvarez, J., Acero, A., and Montoya, A. (2020). Development and testing of a single-axis photovoltaic sun tracker through the internet of things. *Energies*, 13(10):2547.
- [Hammoudi et al., 2022] Hammoudi, Y., Idrissi, I., Boukabous, M., Zerguit, Y., and Bouali, H. (2022). Review on maintenance of photovoltaic systems based on deep learning and internet of things. *Indones. J. Electr. Eng. Comput. Sci*, 26(2):1060–1072.
- [He et al., 2022] He, Y., Gao, Q., Jin, Y., and Liu, F. (2022). Short-term photovoltaic power forecasting method based on convolutional neural network. *Energy Reports*, 8:54–62.
- [Hernandez et al., 2022] Hernandez, G. O., Ariolli, D. G., Paez, P. E., and Chiantore, P. (2022). Trends and innovations in photovoltaic operations and maintenance. *Progress in Energy*, 4(4):042002.
- [Hochreiter and Schmidhuber, 1997] Hochreiter, S. and Schmidhuber, J. (1997). Lstm can solve hard long time lag problems. In *Advances in neural information processing systems*, pages 473–479.

- [Hossein Motlagh et al., 2020] Hossein Motlagh, N., Mohammadrezaei, M., Hunt, J., and Zakeri, B. (2020). Internet of things (iot) and the energy sector. *Energies*, 13(2):494.
- [Huang and Kuo, 2018] Huang, C.-J. and Kuo, P.-H. (2018). A deep cnn-lstm model for particulate matter (pm_{2.5}) forecasting in smart cities. *Sensors*, 18(7):2220.
- [Huang et al., 2022] Huang, X., Li, Q., Tai, Y., Chen, Z., Liu, J., Shi, J., and Liu, W. (2022). Time series forecasting for hourly photovoltaic power using conditional generative adversarial network and bi-lstm. *Energy*, 246:123403.
- [IEC, 2017b] IEC (2017b). IEC TS 61724-1. In *Photovolt. Syst. Perform. – Part 1 Monit.*
- [Jayaprakash et al., 2019] Jayaprakash, M., Kavitha, D., Ramkumar, M. S., Balachander, K., and Krishnan, M. S. (2019). Achieving efficient and secure data acquisition for cloud-supported internet of things in grid connected solar, wind and battery systems. *Mathematical & Computational Forestry & Natural Resource Sciences*, 11(1).
- [Karbhari and Nema, 2020] Karbhari, G. and Nema, P. (2020). Iot & machine learning paradigm for next generation solar power plant monitoring system. *International Journal of Advanced Science and Technology*, 29 (3):6894–6902.
- [Kazem et al., 2019] Kazem, H. A., Yousif, J., Chaichan, M. T., and Al-Waeli, A. H. (2019). Experimental and deep learning artificial neural network approach for evaluating grid-connected photovoltaic systems. *International Journal of Energy Research*, 43(14):8572–8591.

- [Kieffer et al., 2017] Kieffer, B., Babaie, M., Kalra, S., and Tizhoosh, H. R. (2017). Convolutional neural networks for histopathology image classification: Training vs. using pre-trained networks. In *2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pages 1–6. IEEE.
- [Kim and Cho, 2018] Kim, T.-Y. and Cho, S.-B. (2018). Predicting the household power consumption using cnn-lstm hybrid networks. In *International Conference on Intelligent Data Engineering and Automated Learning*, pages 481–490. Springer.
- [Klise et al., 2018] Klise, K. A., Stein, J. S., and Cunningham, J. (2018). Application of IEC 61724 Standards to Analyze PV System Performance in Different Climates.
- [Kumar et al., 2018a] Kumar, N. M., Atluri, K., and Palaparthi, S. (2018a). Internet of things (iot) in photovoltaic systems. In *2018 national power engineering conference (NPEC)*, pages 1–4. IEEE.
- [Kumar et al., 2018b] Kumar, N. M., Atluri, K., and Palaparthi, S. (2018b). Internet of Things (IoT) in Photovoltaic Systems. In *2018 Natl. Power Eng. Conf. NPEC 2018*.
- [Lakhe, 2016] Lakhe, B. (2016). *Lambda Architecture for Real-time Hadoop Applications*. Practical Hadoop Migration.
- [Li et al., 2022] Li, F., Zheng, H., and Li, X. (2022). A novel hybrid model for multi-step ahead photovoltaic power prediction based on conditional time series generative adversarial networks. *Renewable Energy*, 199:560–586.

- [Lim et al., 2022] Lim, S.-C., Huh, J.-H., Hong, S.-H., Park, C.-Y., and Kim, J.-C. (2022). Solar power forecasting using cnn-lstm hybrid model. *Energies*, 15(21):8233.
- [López Medina et al., 2019] López Medina, M. Á., Espinilla, M., Paggeti, C., and Medina Quero, J. (2019). Activity recognition for iot devices using fuzzy spatio-temporal features as environmental sensor fusion. *Sensors*, 19(16):3512.
- [Lu et al., 2021] Lu, S., Ma, R., Sirojan, T., Phung, B., and Zhang, D. (2021). Light-weight transfer nets and adversarial data augmentation for photovoltaic series arc fault detection with limited fault data. *International Journal of Electrical Power & Energy Systems*, 130:107035.
- [Lu et al., 2023] Lu, T., Wang, C., Cao, Y., and Chen, H. (2023). Photovoltaic power prediction under insufficient historical data based on dendrite network and coupled information analysis. *Energy Reports*, 9:1490–1500.
- [Luo et al., 2019] Luo, Z., Cheng, S., and Zheng, Q. (2019). Gan-based augmentation for improving cnn performance of classification of defective photovoltaic module cells in electroluminescence images. In *IOP Conference Series: Earth and Environmental Science*, volume 354, page 012106. IOP Publishing.
- [Lupion et al., 2022] Lupion, M., Polo-Rodriguez, A., Medina-Quero, J., Sanjuan, J. F., and Ortigosa, P. M. (2022). On the limits of conditional generative adversarial neural networks to reconstruct the identification of inhabitants from iot low-resolution thermal sensors. *Expert Systems with Applications*, 203:117356.
- [Luque and Hegedus, 2011] Luque, A. and Hegedus, S. (2011). *Handbook of photovoltaic science and engineering*. John Wiley & Sons.

- [Marino et al., 2016] Marino, D. L., Amarasinghe, K., and Manic, M. (2016). Building energy load forecasting using deep neural networks. In *IECON 2016-42nd Annual Conference of the IEEE Industrial Electronics Society*, pages 7046–7051. IEEE.
- [Martín-Lopo et al., 2020] Martín-Lopo, M. M., Boal, J., and Sánchez-Miralles, Á. (2020). A literature review of iot energy platforms aimed at end users. *Computer Networks*, 171:107101.
- [Massaoudi et al., 2021a] Massaoudi, M., Chihi, I., Abu-Rub, H., Refaat, S. S., and Oueslati, F. S. (2021a). Convergence of photovoltaic power forecasting and deep learning: State-of-art review. *IEEE Access*, 9:136593–136615.
- [Massaoudi et al., 2021b] Massaoudi, M., Chihi, I., Abu-Rub, H., Refaat, S. S., and Oueslati, F. S. (2021b). Convergence of photovoltaic power forecasting and deep learning: State-of-art review. *IEEE Access*.
- [Medina-Quero et al., 2018] Medina-Quero, J., Zhang, S., Nugent, C., and Espinilla, M. (2018). Ensemble classifier of long short-term memory with fuzzy temporal windows on binary sensors for activity recognition. *Expert Systems with Applications*, 114:441–453.
- [Mellit and Kalogirou, 2021] Mellit, A. and Kalogirou, S. (2021). Artificial intelligence and internet of things to improve efficacy of diagnosis and remote sensing of solar photovoltaic systems: Challenges, recommendations and future directions. *Renewable and Sustainable Energy Reviews*, 143:110889.

- [Mellit et al., 2018] Mellit, A., Tina, G. M., and Kalogirou, S. A. (2018). Fault detection and diagnosis methods for photovoltaic systems: A review. *Renewable and Sustainable Energy Reviews*, 91:1–17.
- [Mocanu et al., 2016] Mocanu, E., Nguyen, P. H., Gibescu, M., and Kling, W. L. (2016). Deep learning for estimating building energy consumption. *Sustainable Energy, Grids and Networks*, 6:91–99.
- [Niccolai et al., 2019] Niccolai, A., Grimaccia, F., and Leva, S. (2019). Advanced asset management tools in photovoltaic plant monitoring: Uav-based digital mapping. *ENERGIES*, 12(24).
- [Nie et al., 2021] Nie, Y., Zamzam, A. S., and Brandt, A. (2021). Resampling and data augmentation for short-term pv output prediction based on an imbalanced sky images dataset using convolutional neural networks. *Solar Energy*, 224:341–354.
- [Pan, 2020] Pan, S. J. (2020). Transfer learning. *Learning*, 21:2–1.
- [Pan and Yang, 2009] Pan, S. J. and Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359.
- [Park and Ahn, 2019] Park, N. and Ahn, H. K. (2019). Multi-layer rnn-based short-term photovoltaic power forecasting using iot dataset. In *2019 AEIT international annual conference (AEIT)*, pages 1–5. IEEE.
- [Parra et al., 2017] Parra, I. D. L., Muñoz, M., Lorenzo, E., García, M., Marcos, J., and Martínez-Moreno, F. (2017). Pv performance modelling: A review in the

light of quality assurance for large pv plants. *Renewable and Sustainable Energy Reviews*, 78:780–797.

[Peng et al., 2018] Peng, P., Zhao, X., Pan, X., and Ye, W. (2018). Gas classification using deep convolutional neural networks. *Sensors (Switzerland)*.

[Pérez et al., 2003] Pérez, P., Aguilera, J., Almonacid, G., Vidal, P., De La Casa, J., and Muñoz, J. (2003). Project univer (universidad verde). 200 kWp grid connected PV system at Jaén University Campus. Final conclusion after four operation years. In *Proc. 3rd World Conf. Photovolt. Energy Convers.*, volume C, pages 2310–2313.

[Polo-Rodriguez et al., 2023] Polo-Rodriguez, A., Almonacid-Olleros, G., Almonacid, G., Nugent, C., and Medina-Quero, J. (2023). Discriminating deceptive energy generation of photovoltaic systems by deep learning and adversarial networks. In *International Conference on Ubiquitous Computing and Ambient Intelligence*. Springer.

[Polo-Rodriguez et al., 2020] Polo-Rodriguez, A., Cruciani, F., Nugent, C. D., and Medina, J. (2020). Domain adaptation of binary sensors in smart environments through activity alignment. *IEEE Access*, 8:228804–228817.

[Prasanna Rani et al., 2023] Prasanna Rani, D., Suresh, D., Rao Kapula, P., Mohammad Akram, C., Hemalatha, N., and Kumar Soni, P. (2023). Iot based smart solar energy monitoring systems. *Materials Today: Proceedings*, 80:3540–3545. SI:5 NANO 2021.

- [Qiu et al., 2017] Qiu, Z., Yao, T., and Mei, T. (2017). Learning spatio-temporal representation with pseudo-3d residual networks. In *proceedings of the IEEE International Conference on Computer Vision*, pages 5533–5541.
- [Quero et al., 2018] Quero, J. M., Medina, M. Á. L., Hidalgo, A. S., and Espinilla, M. (2018). Predicting the urgency demand of copd patients from environmental sensors within smart cities with high-environmental sensitivity. *IEEE Access*, 6:25081–25089.
- [Rajagukguk et al., 2020] Rajagukguk, R. A., Ramadhan, R. A., and 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.
- [Raza et al., 2017] Raza, M., Aslam, N., Le-Minh, H., Hussain, S., Cao, Y., and Khan, N. M. (2017). A critical analysis of research potential, challenges, and future directives in industrial wireless sensor networks. *IEEE Communications Surveys & Tutorials*, 20(1):39–95.
- [Romero et al., 2023] Romero, H. F. M., Hernández-Callejo, L., Rebollo, M. Á. G., Cardeñoso-Payo, V., Gómez, V. A., Bello, H. J., Moyo, R. T., and Aragonés, J. I. M. (2023). Synthetic dataset of electroluminescence images of photovoltaic cells by deep convolutional generative adversarial networks. *Sustainability*, 15(9):7175.
- [Romeu et al., 2013] Romeu, P., Zamora-Martínez, F., Botella-Rocamora, P., and Pardo, J. (2013). Time-series forecasting of indoor temperature using pre-trained deep neural networks. In *International conference on artificial neural networks*, pages 451–458. Springer.

- [Rus-Casas et al., 2014a] Rus-Casas, C., Aguilar, J., Rodrigo, P., Almonacid, F., and Pérez-Higueras, P. (2014a). Classification of methods for annual energy harvesting calculations of photovoltaic generators. *Energy Conversion and Management*, 78:527–536.
- [Rus-Casas et al., 2014b] Rus-Casas, C., Aguilar, J., Rodrigo, P., Almonacid, F., and Pérez-Higueras, P. (2014b). Classification of methods for annual energy harvesting calculations of photovoltaic generators. *Energy Conversion and Management*, 78:527–536.
- [Selvin et al., 2017] Selvin, S., Vinayakumar, R., Gopalakrishnan, E., Menon, V. K., and Soman, K. (2017). Stock price prediction using lstm, rnn and cnn-sliding window model. In *2017 international conference on advances in computing, communications and informatics (icacci)*, pages 1643–1647. IEEE.
- [Shapsough et al., 2020] Shapsough, S., Takroui, M., Dhaouadi, R., and Zualkernan, I. (2020). An iot-based remote iv tracing system for analysis of city-wide solar power facilities. *Sustainable Cities and Society*, page 102041.
- [Siekkinen et al., 2012] Siekkinen, M., Hienkari, M., Nurminen, J. K., and Nieminen, J. (2012). How low energy is bluetooth low energy? comparative measurements with zigbee/802.15. 4. In *2012 IEEE wireless communications and networking conference workshops (WCNCW)*, pages 232–237. IEEE.
- [Singh and Raza, 2021] Singh, N. K. and Raza, K. (2021). Medical image generation using generative adversarial networks: A review. *Health informatics: A computational perspective in healthcare*, pages 77–96.

- [Song et al., 2023] Song, H., Al Khafaf, N., Kamoona, A., Sajjadi, S. S., Amani, A. M., Jalili, M., Yu, X., and McTaggart, P. (2023). Multitasking recurrent neural network for photovoltaic power generation prediction. *Energy Reports*, 9:369–376.
- [Spanias, 2018] Spanias, A. S. (2018). Solar energy management as an Internet of Things (IoT) application. In *2017 8th Int. Conf. Information, Intell. Syst. Appl. IISA 2017*.
- [Tang et al., 2018] Tang, R., Leong, P. H., Dore, J., and Vassallo, A. (2018). Generating residential pv production and electricity consumption scenarios via generative adversarial networks. In *Proc. Asia-Pac. Solar Res. Conf.*, pages 4–6.
- [Tarekegn et al., 2020] Tarekegn, G. B., Juang, R.-T., Lin, H.-P., Munaye, Y. Y., and Adege, A. B. (2020). Reduce fingerprint construction for positioning iot devices based on generative adversarial nets. In *2020 International Conference on Pervasive Artificial Intelligence (ICPAI)*, pages 23–28. IEEE.
- [Torrey and Shavlik, 2010] Torrey, L. and Shavlik, J. (2010). Transfer learning. In *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*, pages 242–264. IGI global.
- [Tovar et al., 2020] Tovar, M., Robles, M., and Rashid, F. (2020). Pv power prediction, using cnn-lstm hybrid neural network model. case of study: Temixco-morelos, méxico. *Energies*, 13(24):6512.
- [Vidal et al., 2001] Vidal, P. G., Almonacid, G., Pérez, P. J., and Aguilera, J. (2001). Measures used to protect people exposed to a PV generator: 'Univer project'. *Prog. Photovoltaics Res. Appl.*, 9(1):57–67.

- [Vlaminck et al., 2022] Vlaminck, M., Heidbuchel, R., Philips, W., and Luong, H. (2022). Region-based cnn for anomaly detection in pv power plants using aerial imagery. *Sensors*, 22(3):1244.
- [Wang et al., 2019] Wang, J., Chen, Y., Hao, S., Peng, X., and Hu, L. (2019). Deep learning for sensor-based activity recognition: A survey. *Pattern Recognition Letters*, 119:3–11.
- [Wang et al., 2021] Wang, Y., Chen, Y., Liu, H., Ma, X., Su, X., and Liu, Q. (2021). Day-ahead photovoltaic power forecasting using convolutional-lstm networks. In *2021 3rd Asia Energy and Electrical Engineering Symposium (AEEES)*, pages 917–921. IEEE.
- [Wu et al., 2022] Wu, Y., Wang, B., Song, X., and Zou, J. (2022). Controllable photovoltaic scenario generation via mixup-based deep generative networks. In *2022 International Conference on Cyber-Physical Social Intelligence (ICCSI)*, pages 692–697. IEEE.
- [Yamato et al., 2016] Yamato, Y., Kumazaki, H., and Fukumoto, Y. (2016). Proposal of lambda architecture adoption for real time predictive maintenance. In *2016 Fourth International Symposium on Computing and Networking (CANDAR)*.
- [Yao and Doretto, 2010] Yao, Y. and Doretto, G. (2010). Boosting for transfer learning with multiple sources. In *2010 IEEE computer society conference on computer vision and pattern recognition*, pages 1855–1862. IEEE.
- [Yilmaz and Korn, 2022] Yilmaz, B. and Korn, R. (2022). Synthetic demand data generation for individual electricity consumers: Generative adversarial networks (gans). *Energy and AI*, 9:100161.

- [Zhang et al., 2018] Zhang, C., Kuppannagari, S. R., Kannan, R., and Prasanna, V. K. (2018). Generative adversarial network for synthetic time series data generation in smart grids. In *2018 IEEE international conference on communications, control, and computing technologies for smart grids (SmartGridComm)*, pages 1–6. IEEE.
- [Zhou et al., 2020] Zhou, S., Zhou, L., Mao, M., and Xi, X. (2020). Transfer learning for photovoltaic power forecasting with long short-term memory neural network. In *2020 IEEE International Conference on Big Data and Smart Computing (BigComp)*, pages 125–132. IEEE.