



Institución Universitaria

**Desarrollo de metodología para el
modelamiento computacional y
optimización de celdas solares de
perovskita basada en técnicas de
aprendizaje de máquinas**

Yeraldin Alejandra Velez Galvis

Instituto Tecnológico Metropolitano
Facultad de Ingenierías
Medellín, Colombia
2026

Desarrollo de metodología para el modelamiento computacional y optimización de celdas solares de perovskita basada en técnicas de aprendizaje de máquinas

Yeraldin Alejandra Velez Galvis

Ingeniera de Telecomunicaciones

Tesis presentada como requisito parcial para optar al título de:
Magíster en Automatización y Control Industrial

Directores:

Ph.D. Erick Reyes Vera

Ph.D. Esteban González Valencia

Ph.D. Franklin Alexander Sepúlveda Sepúlveda

Línea de Investigación:

Visión Artificial y Fotónica

Grupo de Investigación:

Automática, electrónica y ciencias computacionales

Instituto Tecnológico Metropolitano

Facultad de Ingenierías

Medellín, Colombia

2026

Agradecimientos

Inicialmente, quisiera agradecer a mi madre, Johanna Galvis, por su apoyo incondicional e incalculable amor. A mi abuelita, mis tíos, primos, hermano y amigos, que con sus palabras de aliento y presencia constante me sostuvieron en cada paso de este camino.

Expreso mi profundo agradecimiento a mis colegas y amigos, Sebastián Montoya y Natalia Lara, cuyo apoyo ha sido esencial en mi desarrollo formativo y profesional. A mis compañeros del semillero, por resolver mis dudas, compartir sus conocimientos y escucharme en los momentos de mayor necesidad.

Un reconocimiento especial a mis directores de tesis, el Dr. Erick Reyes Vera y el Dr. Esteban González Valencia, cuya paciencia infinita, dedicación y valiosos consejos hicieron posible la culminación exitosa de este trabajo.

Agradezco también al proyecto “Aprendizaje automático en celdas solares de tercera generación: diseño, simulación y validación experimental” (con código 2022-0724), financiado por MinCiencias-Colombia con recursos administrados por ICETEX-Colombia.

Finalmente, agradezco a mi compañero de vida, quien con su tiempo, ánimo constante y apoyo incondicional me impulsó a crecer personal, profesional y espiritualmente.

Desde el corazón, gracias.

Producción Científica

Artículos de investigación

- Y. Velez-Galvis, E. Gonzalez-Valencia, A. Sepulveda-Sepulveda & N. Gomez-Cardona. Synthetic dataset to study the performance of perovskite solar cell simulations. *F1000Research*. (2025), 14. 961. <https://doi.org/10.12688/f1000research.168996.2>. Categoría Q1.
- Y. Velez-Galvis, E. Gonzalez-Valencia, E. Reyes-Vera & A. Sepulveda-Sepulveda. Comparative Benchmark of Machine Learning Models for Predicting Perovskite Solar Cell Performance. *Solar Energy – Sometido en Revista Categoría Q1*.

Capítulo de libro

- C. Rojas-Rincón, Y. Vélez-Galvis, M. Botero-Londoño, E. Reyes-Vera & A. Sepúlveda. Machine Learning for the Prediction of Perovskite Solar Cell Performance: A Brief Review. *Congress on Research, Development and Innovation in Renewable Energies. Green Energy and Technology*. Springer. (2025). 15-24. https://doi.org/10.1007/978-3-031-88995-0_2

Conferencias internacionales

- Y. Velez-Galvis, J. Perez, E. Reyes-Vera, E. Gonzalez-Valencia, B Medina-Delgado. & A. Sepulveda. Simulation of lead-free perovskite solar cell based on MASnI_3 compared with FAPbI_3 [Participación con póster y publicación de artículo de conferencia]. *Latin American Optics and Photonics Conference*. Puerto Vallarta, Mexico (2024). Tu4A.16. <https://doi.org/10.1364/LAOP.2024.Tu4A.16>
- E. Gonzalez-Valencia, E. Reyes-Vera, Y. Velez-Galvis. & P. Torres. Potential of Bloch surface wave sensors for refractive index sensing [Participación con póster y publicación de artículo de conferencia]. *Latin American Optics and Photonics Conference*. Puerto Vallarta, Mexico (2024). W4A.20. <https://doi.org/10.1364/LAOP.2024.W4A.20>
- J. Florez-Argaez, Y. Velez-Galvis, J. Echeverri-Perez. & N. Gomez-Cardona. A comparative study of MAPbI_3 and FAPbI_3 for enhanced perovskite solar cell electrical properties [Participación oral]. *11a Semana internacional de ciencia, tecnología e innovación*. Virtual (2024).

Resumen

En los últimos años, la urgencia de adoptar fuentes de energía renovable ha aumentado significativamente debido al rápido cambio climático que experimenta nuestro planeta, impulsado por el crecimiento exponencial en las emisiones de gases de efecto invernadero derivadas del uso de combustibles fósiles. Esta necesidad se ha convertido en una preocupación global de gran relevancia, ya que busca garantizar un suministro energético sostenible y reducir el impacto ambiental de las tecnologías basadas en combustibles fósiles.

Como resultado, el desarrollo, la optimización y la mejora de los sistemas de generación de energía renovable se han convertido en desafíos fundamentales para la comunidad científica. La energía solar, en particular, ha experimentado un crecimiento notable en este contexto. Sin embargo, los paneles solares comerciales más comunes (como los monocristalinos o policristalinos de silicio) no logran captar gran parte de la energía solar, alcanzando una eficiencia energética entre el 15 % y el 20 %. Esto significa que una parte significativa de la energía solar no se aprovecha completamente. Además, estos paneles presentan limitaciones de diseño debido a su rigidez mecánica y peso elevado.

En contraste, las celdas solares de perovskita han despertado un gran interés en los últimos años debido a su notable aumento en eficiencia energética. Alcanzando el valor de 26.1 % en prototipos comerciales y un máximo de 37.5 % en pruebas de laboratorio, estas celdas prometen ser candidatas potenciales para la producción comercial de la próxima generación de células solares. Su adopción podría reducir significativamente los costos de electricidad y simplificar los procesos de fabricación.

De allí, que poder generar metodologías computacionales que permitan modelar, optimizar y predecir las características optoelectrónicas de las mismas sea clave en la reducción de costos de esta tecnología, más aún teniendo en cuenta que las técnicas de fabricación en ocasiones demandan el uso de reactivos costosos y cuartos limpios. Es entonces cuando poder optimizar la combinación de materiales y ajustar las dimensiones estructurales de las mismas a partir de simulaciones juega un papel crucial a la hora de impulsar el desarrollo de celdas solares de perovskita. Además, la aparición de técnicas de inteligencia artificial ofrece una perspectiva innovadora en este campo. Estas técnicas pueden ser eficaces para predecir el rendimiento de las celdas solares, lo que brinda a los investigadores una valiosa herramienta para mejorar su eficiencia y estabilidad. Esto es especialmente importante considerando que los parámetros clave de estas celdas son fundamentales para su desempeño.

Palabras clave: Celdas solares de perovskita, Aprendizaje de máquinas, Simulación física, Energía solar, Dispositivos fotovoltaicos.

Abstract

In recent years, the urgency to adopt renewable energy sources has increased significantly due to the rapid climate change our planet is experiencing, driven by the exponential growth in greenhouse gas emissions from the use of fossil fuels. This need has become a global concern of great relevance, as it seeks to ensure a sustainable energy supply and reduce the environmental impact of fossil fuel-based technologies.

As a result, the development, optimization, and improvement of renewable energy generation systems have become fundamental challenges for the scientific community. Solar energy, in particular, has experienced notable growth in this context. However, the most common commercial solar panels (such as monocrystalline or polycrystalline silicon) do not manage to capture a large part of the solar energy, achieving an energy efficiency between 15 % and 20 %. This means that a significant portion of solar energy is not fully utilized. Moreover, these panels have design limitations due to their mechanical rigidity and high weight.

In contrast, perovskite solar cells have garnered significant interest in recent years due to their remarkable increase in energy efficiency. Reaching a value of 26.1 % in commercial prototypes and a maximum of 37.5 % in laboratory tests, these cells promise to be potential candidates for the commercial production of the next generation of solar cells. Their adoption could significantly reduce electricity costs and simplify manufacturing processes. Its adoption could significantly reduce electricity costs and simplify manufacturing processes.

Hence, being able to generate computational methodologies that allow for the modeling, optimization, and prediction of their optoelectronic characteristics is key to reducing the costs of this technology, especially considering that manufacturing techniques sometimes require the use of expensive reagents and clean rooms. It is then that being able to optimize the combination of materials and adjust their structural dimensions through simulations plays a crucial role in driving the development of perovskite solar cells. Moreover, the emergence of artificial intelligence techniques offers an innovative perspective in this field. These techniques can be effective in predicting the performance of solar cells, providing researchers with a valuable tool to improve their efficiency and stability. This is especially important considering that the key parameters of these cells are fundamental to their performance.

Keywords: Perovskite Solar Cells, Machine Learning, Physical Simulation, Solar Energy, Photovoltaic Devices.

Contenido

Agradecimientos	IV
Producción Científica	V
Resumen	VI
Lista de Figuras	XII
Lista de Tablas	XIII
1 Introducción	1
1.1 Motivación	4
1.2 Objetivos	5
1.2.1 Objetivo General	5
1.2.2 Objetivos Específicos	5
1.3 Estructura de la tesis	6
2 Marco Teórico	7
2.1 Celdas solares o dispositivos fotovoltaicos	7
2.2 Estructura y operación de las celdas solares de perovskita	8
2.3 Eficiencia energética y factor de llenado.	11
2.4 Investigación de celdas solares de perovskita en Colombia	13
2.5 Modelamiento computacional de celdas solares	14
2.6 Uso de técnicas de aprendizaje automático para el modelado de celdas solares de perovskita.	18
2.7 Técnicas de aprendizaje automático para predecir el desempeño de celdas solares de perovskita	20
2.7.1 Random Forest Regressor	20
2.7.2 Extreme Gradient Boosting	21
2.7.3 Artificial Neural Network	21
2.7.4 Support Vector Regression	21
2.7.5 Multi-Layer Perceptron	22
2.8 Métricas para evaluar los modelos de regresión	23
2.8.1 Error Absoluto Medio	23
2.8.2 Raíz del Error Cuadrático Medio	23

2.8.3	Coefficiente de Determinación	24
3	Metodología para la simulación de PSCs	25
3.1	Generación de la base de datos	26
3.2	Descripción de los datos	27
3.3	Validación de datos	31
4	Modelos de inteligencia artificial para predicción	33
4.1	Preprocesamiento de la base de datos	34
4.2	Modelos de aprendizaje automático	35
4.2.1	Random Forest (RF)	35
4.2.2	XGBoost	36
4.2.3	Artificial Neural Network (ANN)	37
4.2.4	Multi-Layer Perceptron (MLP)	39
4.2.5	Support Vector Machine (SVM)	39
4.3	Rendimiento predictivo comparativo de los modelos de aprendizaje automático	40
4.4	Importancia de las características e interpretación física	44
4.5	Validación externa mediante datos simulados y experimentales	46
4.5.1	Diseño inverso para la selección de una configuración óptima a partir de predicciones del modelo Random Forest	48
5	Conclusiones	52
5.1	Limitaciones	53
5.2	Trabajo futuro	53
	Bibliografía	54

Lista de Figuras

1-1	Proyección anual de demanda de energía eléctrica (GWh-año) en Colombia. Obtenido del informe “Proyección de demanda de energéticos para el periodo 2022-2036” [1]	2
2-1	Estructura esquemática de una celda solar de perovskita.	9
2-2	Mapa conceptual de los tipos de celdas solares de perovskita.	10
2-3	Estructura de los diferentes tipos de celdas de perovskita: (a) PSC plana tipo n-i-p; (b) PSC mesoporosa tipo n-i-p; (c) PSC plana tipo p-i-n; (d) PSC mesoporosa tipo p-i-n.	11
2-4	Curva J-V, densidad de corriente vs voltaje. Adaptado de: [2]	12
3-1	Estructura de la celda solar de perovskita simulada. Adaptado de [3].	26
3-2	Distribución de datos de las métricas de rendimiento: a) Voc, b) Jsc, c) FF y d) PCE. Adaptado de [3]	30
4-1	Esquema del flujo de trabajo del marco de aprendizaje automático propuesto para predecir el rendimiento de los PSC. El diagrama ilustra la secuencia que va desde la generación del conjunto de datos y el preprocesamiento de los datos basados en SCAPS-1D hasta el entrenamiento del modelo, la evaluación basada en métricas de regresión y el análisis de la importancia de las características.	34
4-2	Comparación de los valores reales generados a partir de simulación y los datos predichos en el modelo de Random Forest.	36
4-3	Comparación de los valores reales generados a partir de simulación y los datos predichos en el modelo de XGBoost.	37
4-4	Comparación de los valores reales generados a partir de simulación y los datos predichos en el modelo de ANN.	38
4-5	Comparación de los valores reales generados a partir de simulación y los datos predichos en el modelo de MLP.	40
4-6	Comparación de los valores reales generados a partir de simulación y los datos predichos en el modelo de SVM.	41
4-7	Importancia de la permutación de las características de entrada más influyentes para predecir Voc, Jsc, FF y PCE con RF.	44

Lista de Tablas

3-1	Rangos de variación de los parámetros geométricos y fotoeléctricos utilizados para la simulación. Adaptado de [3].	28
3-2	Comparación entre las métricas de rendimiento publicadas y simuladas para distintas PSCs. Adaptado de [3]	32
4-1	Métricas de rendimiento para cada modelo de regresión.	42
4-2	Comparación entre valores reportados y predichos para validación externa.	47
4-3	Errores relativos de los valores predichos y los valores reportados para Voc, Jsc, FF y PCE.	47
4-4	Comparativa entre FF y PCE predichas por el modelo de diseño inverso basado en Random Forest y los valores obtenidos mediante la simulación de validación en SCAPS-1D para la configuración óptima.	50

1 Introducción

La creciente conciencia sobre la amenaza del cambio climático y la necesidad de reducir las emisiones de gases de efecto invernadero han dado lugar a un aumento en la demanda de fuentes de energía renovable. No solo está motivada por la obligación de mitigar los impactos ambientales derivados del uso de combustibles fósiles, sino también por la búsqueda de una mayor seguridad energética [4]. De hecho, la transición hacia fuentes de energía renovable se ha convertido en una necesidad global, respaldada por acuerdos internacionales como el Acuerdo de París [5].

En este contexto, la inversión en tecnologías de energía solar [6], eólica [7], hidroeléctrica [8] y otras formas de energías renovables está experimentando un crecimiento significativo en los últimos años [9]. Estas tecnologías no solo son más amigables con el medio ambiente, sino que también ofrecen la ventaja de ser un recurso inagotable, lo que contribuye a la seguridad energética a largo plazo. Las celdas solares, en particular, son una tecnología prometedora para generar energía eléctrica mediante la conversión de la radiación solar en electricidad utilizable, desempeñando un papel esencial en la transición hacia fuentes de energía sostenibles y en la reducción de la dependencia de los combustibles fósiles [8].

A medida que avanza el desarrollo de las tecnologías solares, se vuelven más eficientes, accesibles y versátiles. Esto significa que pueden aplicarse en una amplia gama de entornos y escenarios, desde paneles solares en techos residenciales hasta la construcción de grandes plantas de energía solar a escala industrial [10]. La optimización de las celdas solares se traduce en una mayor eficiencia en la conversión de la luz solar en electricidad, lo que genera un impacto directo en la rentabilidad y la viabilidad de la energía solar como fuente de energía principal. Los avances en materiales, diseño y técnicas de fabricación están impulsando constantemente la mejora de estas tecnologías, lo que a su vez está acelerando su adopción en todo el mundo [11].

Por lo anterior, en los últimos años se ha despertado un fuerte interés por el estudio de celdas solares de perovskita, ya que estas cuentan con altos coeficientes de absorción, banda prohibida (comúnmente conocido como bandgap) ajustable, diferentes materiales en su composición y menores costos de fabricación que las celdas solares en silicio [12, 13]. Este interés se explica también porque las celdas solares de perovskita (PSC, por sus siglas en inglés) han mostrado un ascenso sin precedentes en eficiencia de conversión de potencia, pasando

del 3.8 % reportado en 2009 al 26.95 % en dispositivos individuales y más del 34 % en configuraciones tándem con silicio, según datos del NREL [14]. Este crecimiento, alcanzado en menos de dos décadas, supera ampliamente el ritmo histórico de mejora observado en otras tecnologías fotovoltaicas.

Esto es especialmente relevante en un país como Colombia, donde el consumo promedio de energía eléctrica continúa creciendo, como se observa en la Figura 1-1. Por ello, se considera que una mejora del 1 % en la eficiencia en la energía eléctrica generada con las celdas solares representaría un gran avance para la generación de energía del país y, por ende, para que el sector energético pueda satisfacer la demanda cada vez más creciente por parte de los usuarios [15].

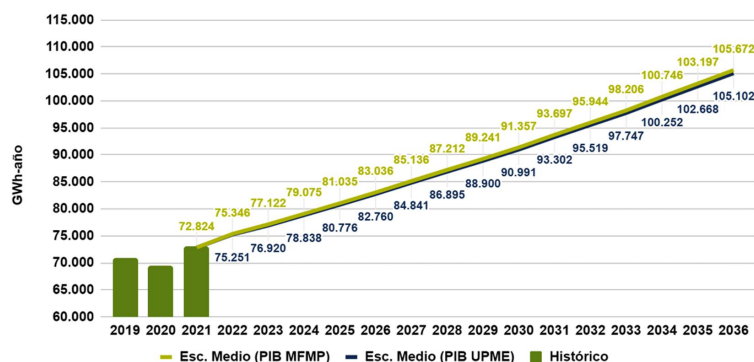


Figura 1-1: Proyección anual de demanda de energía eléctrica (GWh-año) en Colombia. Obtenido del informe “Proyección de demanda de energéticos para el periodo 2022-2036” [1]

En comparación con otras formas de generación de electricidad, como los combustibles fósiles, las celdas solares tienen un impacto medioambiental relativamente menor [16]. La investigación sobre las mismas puede ayudar a reducir aún más el impacto ambiental y convertir las en una fuente de energía más usada, no solo reduciendo la huella de carbono y el impacto en la salud, sino también haciendo uso de un recurso ilimitado como la luz solar.

Por otro lado, es importante remarcar la importancia de invertir recursos para el estudio de las celdas solares de perovskita, puesto que han demostrado tener propiedades ópticas prometedoras, atrayendo la atención de investigadores y empresas que desarrollan dispositivos fotovoltaicos alrededor del mundo, debido a que la perovskita cuenta con altos coeficientes de absorción (propiedad óptica que determina la capacidad que tienen los materiales de absorber fotones). Distintos grupos de investigación buscan aprovechar la mayor cantidad de energía solar posible y convertir las celdas solares de perovskita en dispositivos estables, mitigando los problemas que se presentan cuando se degradan los materiales de las películas

que componen la celda y, como consecuencia, disminuye la eficiencia de la celda [17]. Este representa uno de los desafíos más importantes en la comercialización de los paneles solares de perovskita, puesto que su durabilidad se ve fuertemente reducida. Sin embargo, la evolución de la eficiencia energética y el desempeño de las celdas solares presentadas en los últimos años muestra una mejora significativa en un periodo de tiempo corto, lo que demuestra que es una opción que merece la pena investigar [18–20].

En la actualidad, algunos investigadores han empleado técnicas de aprendizaje de máquinas para identificar patrones de comportamiento y evaluar la influencia que tienen las variables geométricas, ópticas y eléctricas en el rendimiento de las celdas solares de perovskita. De esta forma, se busca predecir si la celda resultante puede dar como resultado un dispositivo con mejor desempeño al generar valores altos de voltaje de circuito abierto (V_{oc}), densidad de corriente de corto circuito, factor de llenado (FF) y eficiencia de conversión de potencia (PCE). La integración de modelos de Machine Learning en el análisis fotovoltaico ha demostrado ser una herramienta poderosa para acelerar la optimización de materiales y dispositivos, permitiendo explorar espacios de diseño de alta dimensionalidad imposibles de abordar únicamente mediante experimentación tradicional [21, 22]. En PSCs, técnicas como Random Forest, Gradient Boosting y redes neuronales profundas han sido utilizadas para predecir métricas de desempeño y para evaluar la sensibilidad de la eficiencia frente a parámetros críticos del dispositivo [13, 23, 24]. Lo anterior puede revolucionar el modelado de las celdas solares de perovskita, pues no solo reduciría costos en la investigación y tiempos de análisis, sino que también crearía la posibilidad de diseñar celdas con eficiencia notable, sin la necesidad de fabricar un gran número de prototipos [25, 26].

En términos generales, la investigación sobre celdas solares de perovskita es fundamental para satisfacer la creciente demanda de energía renovable y reducir los costos de fabricación, el cual es uno de los limitantes más importantes en la actualidad para el uso masivo de paneles solares como fuente alternativa de energía. A mediano plazo se mejorará su eficiencia y estabilidad, lo que permitirá reducir el impacto ambiental generado por los combustibles fósiles y lograr beneficios económicos [27].

Por lo tanto, el presente trabajo se enfocará en desarrollar una metodología computacional que permite evaluar el desempeño de las celdas solares de perovskita al optimizar los parámetros geométricos y fotoeléctricos de las mismas. Estos dispositivos resultan atractivos por sus altas eficiencias energéticas reportadas. Sin embargo, al estar en etapa de desarrollo, la procesabilidad y estabilidad de la eficiencia energética en el tiempo siguen siendo un reto para los investigadores. A pesar de los avances recientes, el estado del arte muestra vacíos importantes, ya que la mayoría de estudios emplean simulaciones aisladas sin integrar análisis sistemático de sensibilidad. De hecho, pocos trabajos exploran simultáneamente múltiples parámetros estructurales y fotoeléctricos, y solo un número reducido combina simulación

numérica con algoritmos de aprendizaje de máquinas para predicción y optimización del desempeño. Además, persiste la falta de metodologías reproducibles que permitan generar conjuntos de datos robustos a partir de modelos físicos confiables y utilizarlos para entrenar modelos predictivos con capacidad real de generalización.

Frente a estas brechas, el aporte central de esta tesis consiste en proponer una metodología integrada que combina: (i) simulaciones numéricas basadas en física mediante SCAPS-1D para modelar el comportamiento óptico y eléctrico de la celda; (ii) generación sistemática de un conjunto de datos que relaciona parámetros geométricos y fotoeléctricos con métricas de desempeño clave; y (iii) desarrollo de modelos de aprendizaje de máquinas para predecir la eficiencia del dispositivo y apoyar la identificación de configuraciones óptimas. Esta integración permite explorar espacios de diseño de alta dimensionalidad, reducir el costo computacional y experimental asociado al análisis tradicional, y establecer un marco metodológico reproducible para futuras investigaciones.

1.1. Motivación

La demanda global de energía eléctrica ha experimentado un crecimiento exponencial debido al aumento en el uso de medios electrónicos para actividades cotidianas, los cuales consumen grandes cantidades de energía. A nivel mundial, para la generación de energía eléctrica se usan principalmente fuentes derivadas de combustibles fósiles, tales como petróleo, carbón y gas, energía nuclear y, en menor medida, fuentes renovables tales como energía solar, eólica e hidráulica. En la actualidad, el uso de fuentes de energía renovables es imperativo para mitigar los impactos medioambientales y en la salud pública. Esto no solo ayudaría a satisfacer la creciente demanda de energía eléctrica y a reducir la contaminación ambiental, sino que también disminuiría los riesgos de futuras crisis energéticas debido a la disminución de las reservas de petróleo o carbón, cuya disponibilidad es cada vez más limitada [10].

Ahora bien, la energía solar es un excelente recurso de fuente renovable para la generación de energía eléctrica. Siendo las celdas solares de silicio costosas de fabricar, las celdas solares de perovskita se perfilan como una posible solución para disminuir los costos y aumentar la eficiencia energética. Sin embargo, a pesar de su rápido crecimiento en eficiencia, las celdas solares de perovskita aún presentan desafíos técnicos significativos relacionados con su estabilidad, la variabilidad de los parámetros de diseño y la compleja interacción entre propiedades ópticas, electrónicas y estructurales. Estos retos dificultan el desarrollo de dispositivos robustos y obstaculizan su escalamiento a procesos industriales.

En este contexto, comprender en profundidad los fenómenos fisicoquímicos que gobiernan su funcionamiento es fundamental para avanzar hacia dispositivos más eficientes y estables. Sin

embargo, los estudios experimentales suelen ser costosos, lentos y dependen de condiciones de fabricación altamente controladas, lo cual limita la exploración sistemática de alternativas de diseño.

Teniendo en cuenta lo anterior, este trabajo propone desarrollar una metodología de simulación que permita analizar y comprender los diferentes fenómenos fisicoquímicos presentes en las celdas solares de perovskita e implementar técnicas de aprendizaje de máquinas para predecir valores óptimos de las características geométricas y fotoeléctricas de los materiales, de forma que se obtenga la mayor eficiencia energética.

La motivación principal de este trabajo de maestría radica en ofrecer una herramienta computacional que reduzca el costo y el tiempo asociados al diseño de nuevas celdas solares, permitiendo explorar con rapidez espacios multidimensionales de parámetros; identificar las variables que más influyen el desempeño; y predecir configuraciones óptimas antes de su fabricación. De esta manera, se contribuye a cerrar brechas críticas del estado del arte al integrar simulaciones basadas en física con modelos predictivos de aprendizaje de máquinas, generando un enfoque sistemático y reproducible que acelera la investigación en tecnologías fotovoltaicas emergentes.

1.2. Objetivos

1.2.1. Objetivo General

Desarrollar una metodología computacional basada en técnicas de aprendizaje de máquinas que permita optimizar el desempeño de celdas solares de perovskita a partir de sus variables geométricas y fotoeléctricas.

1.2.2. Objetivos Específicos

1. Implementar una metodología computacional para modelar numéricamente celdas solares de perovskita en función de sus variables geométricas y fotoeléctricas para comprender su funcionamiento óptico y eléctrico.
2. Desarrollar un algoritmo basado en técnicas de aprendizaje de máquinas para optimizar los parámetros de diseño y predecir el desempeño de las celdas solares de perovskita a partir de sus variables geométricas y fotoeléctricas.
3. Validar y ajustar el modelo computacional utilizando una base de datos que contenga el rendimiento de las celdas solares de perovskita para asegurar la capacidad de este realizando predicciones precisas sobre el desempeño de este tipo de celdas en función de sus características geométricas y fotoeléctricas.

1.3. Estructura de la tesis

En el Capítulo 1, se presenta una introducción general de la temática de este estudio, respecto al desarrollo de celdas solares de perovskita para la generación de energías basadas en fuentes renovables. Se presentan la motivación, los objetivos general y específicos y la estructura de la tesis.

El Capítulo 2 contiene el marco teórico donde se abordan los fundamentos básicos de las celdas solares, su estructura y operación, las métricas de desempeño de las mismas, el modelado computacional para la evaluación del rendimiento del dispositivo para evaluación del desempeño de la celda y el uso de técnicas de aprendizaje de máquinas para potenciar el desarrollo de esta tecnología en la generación de energía eléctrica.

El Capítulo 3 presenta la metodología computacional desarrollada para el diseño y evaluación del rendimiento de las celdas solares de perovskita a partir de la curva J-V y sus métricas de desempeño. Esta metodología fue la base para crear, almacenar y alimentar los resultados contenidos en la base de datos sintética producto de este trabajo.

En el Capítulo 4 se presentan los modelos de aprendizaje de máquinas desarrollados para predecir y optimizar el rendimiento de las celdas solares de perovskita a partir de sus parámetros fotoeléctricos, donde se presenta el proceso de diseño, entrenamiento, validación y evaluación de los modelos.

Finalmente, en el Capítulo 5 se detallan las conclusiones y recomendaciones del presente trabajo.

2 Marco Teórico

2.1. Celdas solares o dispositivos fotovoltaicos

Los dispositivos fotovoltaicos, comúnmente conocidos como celdas solares, son estructuras capaces de convertir la energía proveniente de la radiación solar en energía eléctrica mediante el efecto fotovoltaico. En aplicaciones comerciales, la mayoría de las celdas solares están fabricadas a partir de silicio, cuya eficiencia aumenta a medida que se transforma el material (de amorfo a policristalino y monocristalino) [28]. Una de las principales ventajas de las celdas fotovoltaicas es que no requieren de una fuente de alimentación externa para producir electricidad o de otro tipo de fuente que genere alguna señal eléctrica; solo es necesario que sean iluminadas con radiación solar y que estén conectadas a un sistema de almacenamiento o consumo para generar electricidad de forma continua [29].

Las celdas solares pueden agruparse en módulos y paneles mediante conexiones en serie y paralelo, con el fin de aumentar potencia, corriente o voltaje según la aplicación. Estos paneles pueden alimentar centrales eléctricas al convertir la luz del sol en energía eléctrica para distribuirse en una población, para hacer uso doméstico como sistema secundario de generación e incluso reemplazar la red eléctrica convencional [30]. Los dispositivos fotovoltaicos son ampliamente utilizados para suministrar energía eléctrica en lugares remotos donde las fuentes de energía eléctrica convencionales no están disponibles o su instalación es muy costosa [31]. Como las celdas solares no requieren intervención humana para su funcionamiento, son muy útiles para proporcionar energía a la mayoría de las instalaciones espaciales, desde satélites de comunicaciones o meteorológicos, telescopios y estaciones espaciales. Las celdas solares también se han implementado para cargar productos de consumo común, como juguetes electrónicos, calculadoras portátiles, iluminación en exteriores, relojes de mano o radios portátiles [32].

Las PSCs han surgido como una alternativa prometedora para generar electricidad a partir de la energía solar, por lo que su desarrollo y optimización es de gran importancia para satisfacer la creciente demanda de energía eléctrica a nivel mundial. En la última década han cobrado gran relevancia debido al notable aumento en la eficiencia energética y a su potencial para fabricarse mediante procesos de bajo costo, a diferencia de las celdas tradicionales basadas en silicio [33, 34]. Además, contribuyen reduciendo la emisión de gases de efecto invernadero y otros contaminantes ambientales presentes en los procesos para la generación

de energía a partir de combustibles fósiles [35].

2.2. Estructura y operación de las celdas solares de perovskita

Las perovskitas son materiales sintéticos con estructura cristalina tipo ABX_3 , donde A es el catión orgánico o inorgánico, B es el catión metálico y X es un anión halógeno; comúnmente se usa el cloro (Cl), yodo (I) o bromo (Br). Esta combinación confiere a las perovskitas propiedades fotoeléctricas excepcionales, como bandgap ajustable, altos coeficientes de absorción y movilidad efectiva para el transporte de portadores, características deseables para la conversión fotovoltaica. Sin embargo, la estabilidad operativa del material continúa siendo un reto significativo. La durabilidad limitada, las rutas de degradación aceleradas y la sensibilidad a factores ambientales como humedad, oxígeno y radiación UV siguen siendo los principales obstáculos para su transición hacia aplicaciones industriales [36–38].

En general, las celdas solares cuentan con la misma estructura (sin importar si se usa en una calculadora o en una central eléctrica), la cual consiste en un dispositivo que recibe la radiación solar a través de su capa frontal, compuesta por un revestimiento óptico y antirreflejante, el cual ayuda a minimizar las pérdidas por reflexión. Los fotones incidentes llegan a un material de transporte de electrones (tipo n , con exceso de electrones), donde excitan a los electrones presentes y estos interactúan a su vez con los huecos presentes en la capa de transporte de huecos (capa tipo p , con déficit de electrones), lo que produce una diferencia de potencial. Los fotones tienen energía suficiente para generar en los electrones movimiento, lo cual produce un flujo eléctrico y, como consecuencia, se produce energía eléctrica [2]. En la Figura 2-1, se muestra la estructura típica de una celda solar de perovskita, la cual consiste en una unión de capas con diferentes materiales que cumplen un papel importante y se describen a continuación:

1. **Cubierta de cristal:** Esta capa se ubica en la parte superior de la celda, permite el paso de luz hacia las capas internas y actúa como soporte estructural para la celda.
2. **Contacto frontal transparente:** Esta capa actúa como el electrodo frontal de la celda, compuesto comúnmente por óxido de indio dopado con estaño (ITO) u óxido de estaño dopado con flúor (FTO). Es transparente para permitir el paso de la luz hacia las capas activas de la celda y movilizar los electrones generados hacia el electrodo trasero.
3. **Capa tipo n o de transporte de electrones (ETL, por sus siglas en inglés):** Es una capa con exceso de electrones, la cual está compuesta típicamente por materiales

como TiO_2 , SnO_2 o PCBM. Esta capa facilita la extracción selectiva de electrones y evita el flujo de huecos hacia el electrodo frontal. Su calidad, movilidad electrónica y alineación de bandas influyen directamente en la recombinación y en la eficiencia de la celda.

4. **Capa de perovskita:** Es la capa absorbente y constituye el núcleo del dispositivo. Allí se generan los pares electrón–hueco gracias a su alta absorción óptica, su bandgap ajustable y su excelente difusión de portadores. La composición química y la microestructura de esta capa son determinantes para el rendimiento final.
5. **Capa tipo p o de transporte de huecos (HTL, por sus siglas en inglés):** Materiales como Spiro-OMeTAD, PTAA o CuSCN se utilizan para permitir la extracción selectiva de huecos mientras bloquean a los electrones. Su función es esencial para evitar recombinación interfacial y asegurar un transporte eficiente.
6. **Contacto metálico:** Los huecos generados en la capa HTL se recogen en el contacto trasero, completando el circuito al conectarse con el electrodo frontal y permitiendo el flujo de electrones a través de la celda. Está compuesto comúnmente de cobre, aluminio, plata u oro.

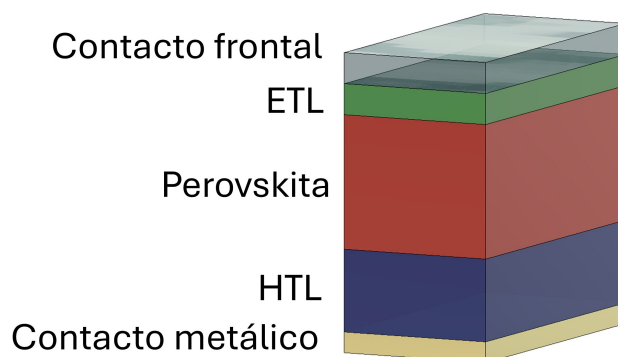


Figura 2-1: Estructura esquemática de una celda solar de perovskita.

Sin embargo, es importante aclarar que no existe una única arquitectura posible. Dependiendo de la secuencia de capas, pueden fabricarse dispositivos tipo n-i-p (estructura convencional) o p-i-n (estructura invertida), donde el orden de las capas ETL y HTL se invierte para optimizar el desempeño bajo condiciones específicas de iluminación, temperatura o técnicas de fabricación. Ambas han demostrado eficiencias máximas de hasta 26.9%, según el registro más reciente de NREL [14]. Adicionalmente, es posible incorporar capas adicionales dependiendo del diseño a implementar o las necesidades específicas del estudio, con el fin de mejorar la estabilidad, reducir la recombinación o aumentar la eficiencia global del dispositivo [39].

En la Figura 2-2 se muestra la clasificación de las dos arquitecturas de las PSC más estudiadas, que se clasifican en celdas mesoporosas y planas. En las PSC mesoporosas, la arquitectura incluye una capa intermedia con estructura porosa, como el dióxido de titanio (TiO_2), que sirve como soporte físico y como guía para la infiltración del material de perovskita. Esta capa proporciona una mayor superficie, lo que favorece la generación y separación de portadores, y contribuye a mejorar la estabilidad mecánica del dispositivo [40]. Se ha demostrado que las PSC mesoporosas alcanzan una eficiencia notablemente alta tanto bajo iluminación interior como en condiciones solares estándar [12].

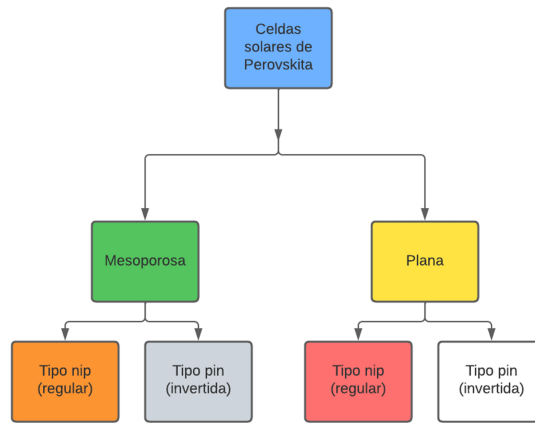


Figura 2-2: Mapa conceptual de los tipos de celdas solares de perovskita.

Ahora bien, las PSC planas tienen una estructura más simple que la de las PSC mesoporosas, con la capa de perovskita ubicada entre la capa de ETL y la capa de transporte de huecos HTL. Su arquitectura reduce la complejidad de fabricación y, al no depender de infiltración en una matriz porosa, permite un mayor control sobre el espesor, la cristalinidad y la uniformidad de la capa absorbente. Estas características las convierten en una plataforma ideal para estudios de simulación físico-eléctrica, análisis térmico y modelamiento computacional, como el que se desarrolla en este trabajo [41, 42]. En la Figura 2-3, se muestran los tipos de PSC planas y mesoposotras, tanto en configuraciones p-i-n como en configuraciones n-i-p. En la figura, el ánodo corresponde al contacto metálico trasero, HTL es la capa transportadora de huecos, ETL es la capa de transporte de electrones y el contacto frontal de la estructura actúa como contacto superior.

Actualmente, los investigadores continúan desarrollando diversas estrategias para mejorar su estabilidad, como el uso de una estructura de doble capa de polímero conductor, o las investigaciones en PSC tipo plana invertida, donde se utiliza una estructura de dispositivo similar a la de las celdas solares orgánicas [43].

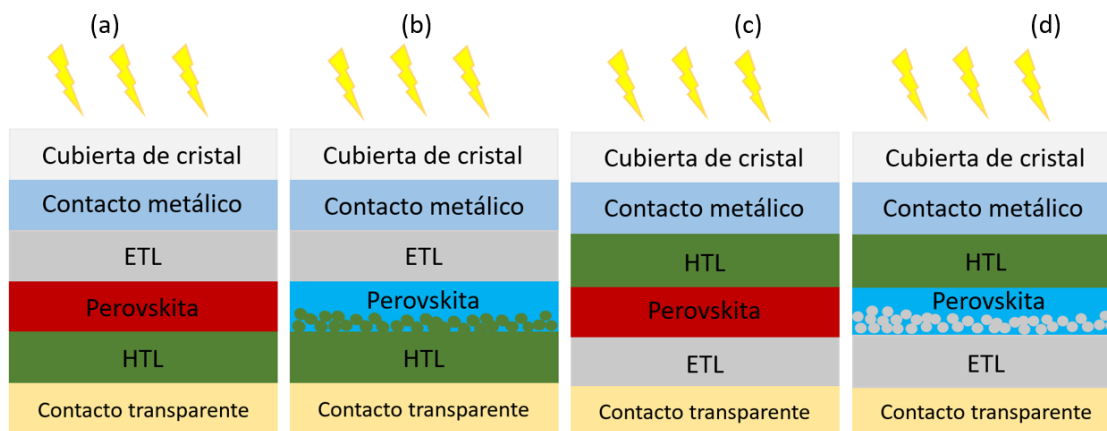


Figura 2-3: Estructura de los diferentes tipos de celdas de perovskita: (a) PSC plana tipo n-i-p; (b) PSC mesoporosa tipo n-i-p; (c) PSC plana tipo p-i-n; (d) PSC mesoporosa tipo p-i-n.

2.3. Eficiencia energética y factor de llenado.

El rendimiento real de las celdas solares de perovskita varía según la estructura, los materiales y el proceso de fabricación del dispositivo. Los récords actuales de eficiencia de las celdas solares de perovskita son del 26.95 % para dispositivos de perovskita de unión simple y del 34.85 % para celdas solares en tándem de perovskita-silicio [14, 44]. El rendimiento de las celdas solares de perovskita ha estado limitado por su estabilidad, que puede verse afectada por la exposición a factores como la humedad, el oxígeno, la luz y el calor, modificando sus propiedades y disminuyendo las características de absorción por las cuales la perovskita ha llamado tanto la atención de la academia y la industria [45]. Las celdas solares mesoporosas de perovskita suelen tener una eficiencia mayor a las celdas solares planas debido a que la superficie de contacto perovskita mesoporosa es más grande y a su cristalinidad en la película de perovskita [46].

Para analizar el desempeño de una celda solar, se debe considerar la curva de densidad de corriente vs voltaje (J-V) mostrada en la Figura 2-4. La densidad de corriente está definida como la relación entre la corriente que circula por el área de la celda, mientras que el voltaje es medido entre el ánodo y el cátodo de la celda. A partir de estos dos valores se pueden calcular parámetros importantes como densidad de corriente de corto circuito (J_{sc}), el voltaje y la densidad de corriente en la potencia máxima (V_{mp}) y (J_{mp}), respectivamente, y voltaje de circuito abierto (V_{oc}).

A partir de estos parámetros, es posible calcular la eficiencia de conversión de potencia (PCE por su siglas en inglés), como se muestra en la ecuación (2-1) [2], donde (P_{IN}) representa a la potencia de energía solar que incide en la celda. La PCE es uno de los parámetros más

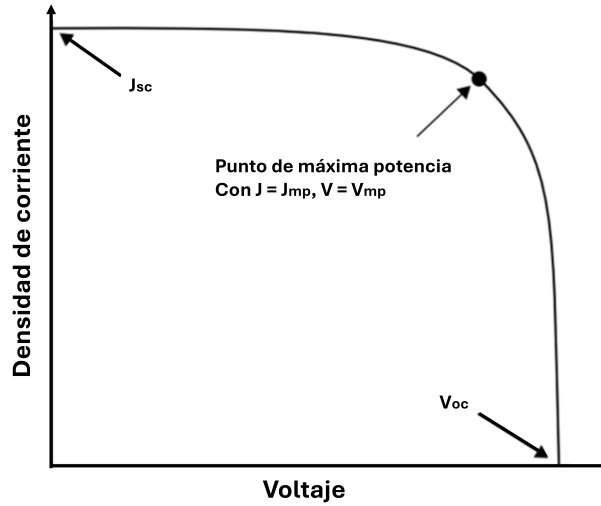


Figura 2-4: Curva J-V, densidad de corriente vs voltaje. Adaptado de: [2]

importantes para estudiar el desempeño de las celdas solares, ya que permite calcular la capacidad de conversión de radiación solar en energía eléctrica.

$$PCE = \frac{J_{mp} \times V_{mp}}{P_{IN}} \quad (2-1)$$

Por otro lado, el factor de llenado (FF) determina la calidad de la celda, calculando la relación entre la potencia máxima y la potencia pico obtenida a partir de la curva J-V, mostrada en la Figura 2-4. El factor de llenado se puede calcular como muestra la ecuación (2-2) [2]:

$$FF = \frac{J_{mp}V_{mp}}{J_{sc}V_{oc}} \quad (2-2)$$

Es importante mencionar que, aunque la PCE se calcula mediante variables eléctricas como se ilustra en la ecuación (2-1), diferentes artículos han encontrado que algunas propiedades intrínsecas del material afectan notablemente la eficiencia energética de la celda. Algunas de estas propiedades son el bandgap, el espesor de las capas y/o la densidad de defectos [47–49]. El bandgap es una propiedad crucial de los semiconductores, que describe la energía mínima necesaria para excitar a un electrón y permitir que se desprenda de su átomo, quedando libre para desplazarse por el material. Esto es de vital importancia, ya que un valor demasiado alto o bajo puede afectar negativamente la absorción de fotones y, en consecuencia, disminuir significativamente la eficiencia energética [50]. Además, los espesores de las capas también juegan un papel clave en la eficiencia energética; una capa demasiado delgada reduce el área de absorción de luz, mientras que una capa demasiado gruesa puede generar pérdidas por recombinación [51]. Finalmente, la densidad de defectos es fundamental, ya que se refiere a la cantidad de defectos químicos o morfológicos derivados de impurezas en las estructuras

cristalinas o pequeñas variaciones superficiales originadas durante la fabricación. Estos defectos pueden reducir la eficiencia de las celdas solares al provocar la pérdida de electrones y disminuir la corriente eléctrica [52]. Otras propiedades de los materiales pueden influir en PCE y se interrelacionan de manera compleja según las condiciones específicas de cada estructura. En este sentido, la optimización de estos dispositivos no se limita solo a ajustar las variables mencionadas, sino que requiere un enfoque integral que considere los aspectos geométricos, fotoeléctricos y de composición de las capas. En última instancia, el éxito en la mejora del desempeño de la celda se basa en la capacidad de combinar adecuadamente las propiedades de los materiales seleccionados con la precisión en los procesos de fabricación.

2.4. Investigación de celdas solares de perovskita en Colombia

En Colombia, la investigación en PSC es emergente y durante los últimos dos años se observan contribuciones concretas tanto en fabricación y optimización experimental de dispositivos como en análisis de datos y aprendizaje de máquinas aplicados a esta tecnología. En conjunto, estos avances muestran que el país ya cuenta con capacidades para abordar el problema desde perspectivas complementarias, incluyendo síntesis de materiales, procesamiento de dispositivos, evaluación ambiental y modelado computacional [53–55].

Desde la Universidad de Antioquia, en el grupo CIDEMAT, reportan desarrollos orientados a la fabricación y mejora del desempeño de PSC. En 2024, Vélez Ramírez desarrolló una dispersión coloidal de nanopartículas de SnO_2 para su uso como capa transportadora de electrones en celdas solares de perovskita tipo p-i-n procesadas en solución. Ese trabajo incluyó síntesis del material, caracterización y fabricación de dispositivos, además de la evaluación de bicapas ETL y electrodos de carbono, alcanzando una eficiencia máxima de 13.14 % [56]. En una línea complementaria, Valderrama y su equipo evaluaron en 2025 el ciclo de vida de una PSC de 1 cm^2 con arquitectura p-i-n meso-superestructurada fabricada en Colombia, y encontraron que la sustitución de FTO por ITO y de Au por Ag reduce de forma importante el impacto ambiental del dispositivo; además, concluyeron que la tecnología podría ser competitiva en emisiones de gases de efecto invernadero frente a celdas comerciales si su vida útil supera los 10 años [57]. Recientemente, evaluaron los desafíos de estabilidad en condiciones reales y, desde la misma universidad, describieron líneas de trabajo asociadas con precursores escalables, deposición *meniscus-guided*, dispositivos flexibles tipo p-i-n y electrodos de carbono, lo que evidencia una agenda activa en estabilidad, escalabilidad y desempeño [58]. En la Universidad Industrial de Santander también se identifican contribuciones recientes al área, especialmente desde enfoques basados en datos y aprendizaje automático. Vélez y sus colaboradores propusieron en 2024 el uso de información mutua para cuantificar la relevancia

de descriptores de síntesis respecto a las variables de desempeño fotovoltaico (V_{oc} , J_{sc} , FF y PCE), encontrando que el ion X es el factor con mayor impacto global, seguido por variables como el bandgap, el espesor de la capa de perovskita y los iones A y B [59]. Posteriormente, se evaluaron en 2025 cinco modelos de inteligencia artificial entrenados con variables categóricas relacionadas con la fabricación de PSC, incluyendo materiales ETL y HTL, perovskita absorbente, proceso de deposición, contacto trasero, anti-solvente y solución precursora. En ese estudio, el modelo Random Forest fue identificado como el de mejor desempeño global para regresión de eficiencia, aunque los autores señalaron que aún existen limitaciones para inferir de manera robusta las condiciones óptimas de fabricación [60]. Por su parte, en la Universidad Francisco de Paula Santander se han publicado trabajos recientes orientados al análisis exploratorio de datos y a la sostenibilidad de las PSC. En 2024, Hernández y su equipo reportaron un estudio de caracterización de celdas solares de perovskita de unión simple mediante análisis exploratorio de datos a partir de variables eléctricas, bandgap y área de la celda, constituyendo una base relevante para posteriores desarrollos en analítica de datos y aprendizaje automático [61]. En el mismo año, analizaron el impacto ambiental de diferentes estructuras moleculares en celdas solares de perovskita y concluyeron que el plomo representa uno de los componentes más críticos desde el punto de vista ambiental, mientras que el estaño emerge como una alternativa de sustitución con una disminución no significativa de PCE [62].

En conjunto, estos antecedentes dejan claro que el desarrollo de PSC en Colombia ha despertado el interés de grupos de investigación multidisciplinarios, que el impacto de sus investigaciones en los últimos años en PSC se está desarrollando en dos frentes: por un lado, la fabricación y optimización de dispositivos para mejorar estabilidad y desempeño; y, por otro, el uso de herramientas de análisis de datos, inteligencia artificial y evaluación ambiental para acelerar el diseño y la selección de configuraciones prometedoras. En ese contexto, la metodología propuesta en esta tesis se alinea con una necesidad real del entorno nacional, al integrar simulación física y aprendizaje de máquinas en una estrategia reproducible para el modelamiento y la optimización de celdas solares de perovskita.

2.5. Modelamiento computacional de celdas solares

El modelamiento de las PSC es un campo de investigación muy activo actualmente, debido a que estas celdas tienen un gran potencial y presentan mejora notable de su desempeño, a través de la optimización de las características que limitan su eficiencia y durabilidad. Además de implementar diferentes diseños y estructuras, es posible evaluarlas y analizarlas matemáticamente sin necesidad de fabricar prototipos de laboratorio, reduciendo en gran medida los costos de investigación y tiempo de implementación, permitiendo obtener el diseño más eficiente antes de continuar con la fase experimental [63].

Existen diferentes métodos para modelar las PSC usando herramientas computacionales, uno de ellos es el análisis del comportamiento de los fenómenos eléctricos, térmicos y ópticos en el dispositivo. Evaluando los efectos de las variaciones en las propiedades de los materiales semiconductores de las celdas (bandgap, afinidad del electrón, longitud de difusión del carrier, permitividad dieléctrica, entre otros) en los parámetros Voc, Jsc, FF y PCE. Diversos investigadores han publicado trabajos en donde han implementado el modelamiento eléctrico y evidenciaron que la longitud de difusión del carrier, bandgap, densidad de defectos y permitividad dieléctrica tienen un efecto directo en la eficiencia de la celda [47, 48, 64–69]. Otros autores han reportado estudios de cómo, al variar los espesores de las capas internas de la celda, es posible optimizar la eficiencia de la PSC [70–74]. Por otro lado, se ha evidenciado en la literatura que el aumento de la temperatura conlleva la degradación de los materiales de la celda y la pérdida de las propiedades eléctricas y ópticas, por lo que se ha propuesto el uso de materiales con propiedades que faciliten la disipación de calor [75, 76]. Otros autores están estudiando la posibilidad de añadir electrodos de materiales orgánicos (como el carbono) en lugar de los contactos metálicos comúnmente utilizados (Au, Al o Ag) [68, 77–80]. Mientras que otros autores proponen una arquitectura específica (tipo n-i-p o p-i-n) para evaluar la combinación de diferentes materiales y, así, determinar cuál ofrece la mayor eficiencia energética, optimizan el espesor de las capas y reportan los valores de Voc, Jsc, FF y PCE obtenidos para cada dispositivo simulado. [64, 81–85]. En los últimos trabajos, se observa una tendencia a desarrollar celdas solares libres de plomo (Pb) en el catión metálico, debido a que es altamente tóxico, por lo cual se busca reemplazar este compuesto por Bismuto (Bi), Germanio (Ge), Antimonio (Sb) o Estaño (Sn). Sin embargo, actualmente las celdas de perovskita con plomo son las que generan mejores eficiencias, por lo que reemplazarlo representa un desafío en la evolución de estos dispositivos [70, 72, 84, 86–89].

En el presente trabajo se hace uso del software SCAPS-1D para la simulación de PSC, el cual permite modelar diferentes tipos de celdas solares, como las de Silicio, Arseniuro de galio, perovskita, entre otras. Siendo una de las herramientas más mencionadas en los artículos de investigación relacionados con las simulaciones de celdas solares de perovskita, debido a su versatilidad y velocidad de cálculo [90]. SCAPS-1D analiza la respuesta eléctrica de las celdas solares de perovskita resolviendo un conjunto acoplado de ecuaciones diferenciales que incluyen la ecuación de Poisson (2-3) y las ecuaciones de continuidad para electrones (2-4) y huecos (2-5).

$$\frac{\partial}{\partial x} \left(\varepsilon_0 \varepsilon_r \frac{\partial \psi}{\partial x} \right) = -q \left(p - n + N_D^+ - N_A^- + \frac{\rho_{def}}{q} \right) \quad (2-3)$$

$$-\frac{\partial J_n}{\partial x} - U_n + G = \frac{\partial n}{\partial t}, \quad (2-4)$$

$$-\frac{\partial J_p}{\partial x} - U_p + G = \frac{\partial p}{\partial t}. \quad (2-5)$$

donde ψ es el potencial electrostático, q es la carga eléctrica, ε_0 y ε_r son la permitividad del vacío y la permitividad relativa, p y n son las concentraciones de huecos y electrones, N_D^+ y N_A^- son las impurezas de carga de tipo donante y aceptor, ρ_{def} es la densidad de defectos, J_n y J_p son las densidades de corriente de electrones y huecos, G es la tasa de generación de pares electrón-hueco, y U_n y U_p son las tasas de recombinación de electrones y huecos.

Para calcular las métricas de rendimiento J_{sc} (2-6), V_{oc} (2-7), FF (2-8) y PCE (2-9), SCAPS-1D utiliza las ecuaciones que se presentan a continuación:

$$J_{sc} = \int_0^{\lambda_{max}} q \phi(\lambda) EQE(\lambda) d\lambda \quad (2-6)$$

$$V_{oc} = \frac{kT}{q} \ln \left(\frac{J_{sc}}{J_0} + 1 \right), \quad (2-7)$$

$$FF = \frac{V_{mp} \cdot J_{mp}}{V_{oc} \cdot J_{sc}}, \quad (2-8)$$

$$PCE = \frac{V_{oc} \cdot J_{sc} \cdot FF}{P_{IN}} \quad (2-9)$$

donde λ es la longitud de onda, $\phi(\lambda)$ es el espectro solar incidente y $EQE(\lambda)$ es la eficiencia cuántica externa, k es la constante de Boltzmann, T es la temperatura absoluta, J_0 es la corriente de saturación inversa, V_{mp} y J_{sc} son el voltaje y la corriente en el punto de máxima potencia, y P_{IN} es la potencia incidente.

SCAPS-1D permite optimizar todas las capas de la celda de perovskita, incluyendo la ETL, la capa absorbente de perovskita, la HTL y los contactos metálicos, explorando todas las posibles combinaciones de espesores de cada capa para determinar los valores óptimos que maximicen la eficiencia. Esto es de gran relevancia, ya que los parámetros fotoeléctricos afectan directamente la absorción de luz y el transporte de carga. El software facilita la evaluación comparativa de diferentes materiales para ETL, la capa de perovskita y HTL. Por otro lado, estudios recientes han utilizado SCAPS-1D para estudiar PCS basadas en plomo (como la MAPbI₃), perovskitas sin plomo (como MASnI₃) [91–93].

En 2016, Huang y sus colaboradores utilizaron SCAPS-1D para estudiar la influencia del espesor de las capas y de la densidad de defectos en la eficiencia energética de la celda, encontrando que el espesor óptimo de capa de perovskita es de 500 nm, generando una celda con eficiencia de 18 % y comprobando que estos dos parámetros afectan directamente el rendimiento de la celda [94]. Otros autores no solo estudian los espesores de las capas y la

densidad de defectos, sino también la concentración de dopaje en una celda con capa de perovskita usando formamidio (FASnI₃), obteniendo una eficiencia de 14.03 %, lo que representa un avance importante al evitar el uso de plomo en la estructura propuesta [95, 96]. Este mismo año se reportó en la literatura un análisis de la influencia de los parámetros anteriores, aunado a la variación de temperatura, donde se evidenció que, al aumentar la temperatura y la densidad de defectos, la eficiencia decrece notablemente [48, 97]. Se han encontrado estudios experimentales que reportan que los aumentos de temperatura y densidad de defectos no solo afectan a la eficiencia, sino que también aceleran la degradación de la celda y hacen que pierda sus propiedades de absorción de luz en pocos días e incluso en cuestión de horas [17, 41]. Otros autores han propuesto el uso de otros materiales para la capa de perovskita en la celda, como es el caso del La₂NiMnO₆, el cual se usa en el diseño de una celda tipo n-i-p empleando diferentes materiales para las capas absorbentes (usan PCBM, ZnO, C60 y WS₂ para la capa ETL y CuSCN, P3HT, PEDOT:PSS, NiO, CuI y CFTS para la capa HTL), buscando qué configuración permite alcanzar el valor más alto de eficiencia luego de optimizar los espesores. Los autores encontraron que la estructura con mejores resultados fue ITO/WS₂/LNMO/CFTS/Au con 20.18 % de eficiencia [83]. Otros autores investigan el Cs₃Bi₂I₉ como capa de perovskita, proponiendo enfoques diferentes para su implementación. Por ejemplo, Banik y sus colaboradores alcanzaron una eficiencia de 20.96 % para una estructura ITO/WS₂/Cs₃Bi₂I₉/NiO/Au, con parámetros optimizados [86]. Sin embargo, otros autores no solo optimizaron los espesores, sino que también se centraron en combinarlos con diferentes materiales en ETL y HTL, dando como resultado que la celda ITO/WS₂/Cs₃Bi₂I₉/Cu₂O/Au presentaba el mejor desempeño con eficiencia del 16.61 % [98]. Adicionalmente, un estudio reciente reportó la simulación y fabricación de un dispositivo basado en Cs₃Bi₂I₉, donde reportan una eficiencia para la celda simulada de 9.64 %, mientras que para la fabricada fue 0.83 %. Esto demuestra que en el proceso de simulación se estimaron valores de densidad de defectos idealizados y notablemente inferiores a los que se presentaron en el prototipo fabricado, generando un desfase alto en los resultados obtenidos [82]. Otros investigadores han reportado eficiencias para la celda simulada de 20.95 % y para la fabricada de 14.15 %, con la estructura FTO/SnO/FAPbI₃/Spiro-OMeTAD:LiTFSI/Ag [99]. Es importante aclarar que esta última contiene plomo y como se estableció previamente, evitar el uso de este compuesto tóxico es uno de los retos actuales en las celdas de perovskita [100]. En el 2025, se presentó un estudio sobre perovskitas basadas en un novedoso material (KSnI₃), en el cual se analiza la influencia del espesor de las capas, la temperatura, la densidad de defectos y la afinidad electrónica en el desempeño del dispositivo, pasando de una eficiencia inicial de 11.3 % a un dispositivo optimizado con eficiencia de 13.46 % [101].

2.6. Uso de técnicas de aprendizaje automático para el modelado de celdas solares de perovskita.

En los últimos años, el uso de técnicas de aprendizaje automático (ML, por sus siglas en inglés) en el desarrollo de PSC se ha perfilado como un método eficiente para analizar su funcionamiento y comprender mejor los fenómenos físicos, químicos y eléctricos que se presentan en estas celdas. Las técnicas de ML prometen ahorrar tiempo y recursos en la fabricación de las celdas de perovskita, aprovechando su capacidad para identificar patrones, optimizar variables de diseño y realizar análisis predictivos [102, 103]. Esto permite evaluar el peso que tiene cada parámetro fotoeléctrico y geométrico en la eficiencia de la celda e identificar patrones de desempeño con el fin de obtener estructuras optimizadas que permitan alcanzar valores máximos de eficiencia [104, 105]. Mediante algoritmos de regresión de ML, se pueden evaluar funciones de costo o pérdida que minimicen los errores entre las predicciones del modelo y los valores de referencia obtenidos mediante simulación computacional. Esto resulta de gran ayuda para determinar los parámetros óptimos de diseño y predecir los valores que resulten en una mayor eficiencia energética [22].

Por otro lado, en aprendizaje automático los hiperparámetros son configuraciones externas del modelo que se definen manualmente antes del entrenamiento como la capa de profundidad, número de árboles, número de neuronas o coeficientes de regularización y, a diferencia de los parámetros internos, no se aprenden directamente a partir de los datos. Dado que estos valores controlan el proceso de ajuste y afectan la capacidad de generalización, suelen seleccionarse mediante estrategias de optimización como la búsqueda por cuadrícula (grid search), la búsqueda aleatoria (random search), la optimización bayesiana (Bayesian optimization) o librerías especializadas como Optuna [21, 106]. Estas técnicas permiten buscar de manera automática la combinación de hiperparámetros que mejoran el rendimiento del algoritmo y aumentan la precisión del modelo [23, 107]. En los últimos años, el uso de ML en el desarrollo de PSC ha despertado un gran interés en la comunidad científica debido a sus prometedores avances en el diseño, optimización y evolución de estos dispositivos [108]. En 2020, Çağla Odabas y Ramazan Yıldırım publicaron un artículo de investigación donde proponen herramientas de ML como la minería de datos de reglas de asociación y árboles de decisión para determinar los efectos que tienen los métodos de deposición en la eficiencia. Los autores reportaron que las celdas que presentan mejores resultados son las tipo n-i-p y que la calidad de los resultados obtenidos depende en gran medida de la calidad de los datos introducidos en el algoritmo de ML [21]. Luego, en 2022, Wensheng Yan y sus colaboradores investigaron el uso de diferentes técnicas de ML haciendo uso de herramientas de aprendizaje supervisado como eXtreme Gradient Boosting (XGB) y Random Forest (RF) para evaluar diferentes compuestos de perovskitas y encontrar los parámetros óptimos de diseño, como el bandgap (usando grid search), para así predecir los valores del bandgap, J_{sc} , V_{oc} de 5 composiciones diferentes. Su modelo les permitió obtener valores estimados muy cercanos a los

que obtuvieron en su implementación experimental, dejando como resultado un porcentaje de error de 2% [23]. Un año después, se reportó un estudio donde se usan la herramienta SCAPS-1D, técnicas de ML y herramientas de optimización de hiperparámetros (RandomSearchCV y GridSearchCV) para realizar un análisis numérico de los efectos de diferentes materiales para la capa de perovskita y HTL, con el fin de predecir cuál de las combinaciones de materiales da lugar a la celda con mejor desempeño. Estudiando la influencia de los espesores de las capas, la densidad de dopaje y la densidad de defectos en el comportamiento de la celda en general, encontraron los valores de estos parámetros con los cuales se obtiene una celda con una eficiencia energética de 23.9% y una precisión en la predicción del modelo de aproximadamente 75% [106]. Posteriormente, se publicó un artículo donde, al igual que en el trabajo de Hasanzadeh et al., analizan el desempeño de la celda utilizando algoritmos de aprendizaje supervisado. Los autores demostraron que usar métodos de ML en el estudio de celdas solares de perovskita es una herramienta muy poderosa para mejorar la eficiencia y comprender las correlaciones ocultas que hay entre las variables eléctricas y su desempeño [109]. Por último, Yao Lu y su equipo publicaron sus resultados acerca del uso de diferentes algoritmos de ML (aprendizaje supervisado y no supervisado) para predecir el desempeño de las celdas, donde observaron que los cationes A tienen gran influencia en la eficiencia energética, siendo la correcta combinación de cationes la que permite mejor calidad en la cristalización y disminuir la densidad de defectos, lo cual se traduce en mayores eficiencias en el dispositivo [107].

En 2024, Shrivastav y su equipo de trabajo integraron simulaciones SCAPS-1D con modelos de ML para predecir PCE en celdas solares de perovskitas compuestas por Cesio (Cs). Tras generar un conjunto de datos variando el espesor, la densidad de defectos y el dopaje en seis materiales absorbentes diferentes, evaluaron cinco algoritmos distintos, encontrando que el modelo XGB superó a los demás con una precisión de R^2 de 99.99%. Este enfoque demostró que la combinación de simulaciones y ML acelera significativamente la optimización de materiales sin requerir cálculos complejos, destacando además al CsPbI_3 como el candidato con mejor rendimiento fotovoltaico [110]. Finalmente, en el 2025 se publicó un artículo donde Novoselov y su equipo generaron un conjunto de simulaciones usando SCAPS-1D de 7182 configuraciones diferentes con variaciones en la composición de la capa de perovskita, espesores y capas de transporte, usando XGB y CatBoost para predecir PCE, demostrando que el uso de datos sintéticos estructurados junto con modelos de aprendizaje automático permite superar la escasez de datos experimentales de alta calidad y acelerar la optimización de materiales sin depender de numerosos prototipos físicos. Por último, aplicaron la técnica de explicabilidad SHAP para interpretar las predicciones del modelo, donde observaron que parámetros como el espesor de la perovskita, el tipo de material para ETL y la fracción de Cesio presentan interacciones no lineales complejas que son determinantes para el rendimiento final de la celda [111].

Los distintos artículos publicados muestran las diferentes y poderosas aplicaciones que las técnicas de ML tienen para el desarrollo de celdas solares de perovskita y sus beneficios en el ahorro de recursos empleados en su investigación. También permite estudiar las correlaciones que existen entre sus parámetros y su desempeño, dando una idea de los factores que se deben tener en cuenta a la hora de fabricar una celda. Sin duda alguna, este se ha convertido en un tema que puede impactar fuertemente el campo de las energías renovables y las investigaciones de celdas de tercera generación [108].

2.7. Técnicas de aprendizaje automático para predecir el desempeño de celdas solares de perovskita

Actualmente, se han empleado múltiples técnicas de ML para predecir el rendimiento de las PSC con el fin de determinar qué combinaciones de aniones y cationes dan como resultado un mejor desempeño, entre otros. Se ha demostrado el interés de la comunidad académica por el desarrollo y optimización de parámetros de diseño para generar dispositivos más eficientes y reducir la cantidad de prototipos fabricados, reduciendo así costos asociados a la fabricación y los impactos medioambientales derivados de la disposición de residuos generados en estos procesos de prueba y error. Para el presente trabajo se diseñaron, entrenaron y compararon cinco modelos de ML para predecir las métricas de desempeño de las mismas y determinar los parámetros que más influencia tienen en el resultado. A continuación se describen los fundamentos teóricos de las mismas:

2.7.1. Random Forest Regressor

Random forest regressor (RF) es un método de aprendizaje supervisado que combina múltiples árboles de decisión para crear un modelo robusto y preciso. Cada árbol se entrena con una muestra aleatoria de los datos y utiliza un subconjunto aleatorio de variables en cada división. Esta técnica pertenece al grupo de métodos de ensemble, donde la predicción final se obtiene agregando las predicciones individuales de todos los árboles [112].

En problemas de regresión, RF calcula la predicción final como la media de las predicciones de todos los árboles que componen el bosque. La diversidad entre los árboles, generada por el muestreo aleatorio de datos y características, reduce el sobreajuste y mejora la capacidad de generalización del modelo. Su capacidad para manejar relaciones no lineales complejas y su robustez frente a datos ruidosos lo hacen ideal para identificar qué parámetros geométricos y fotoeléctricos tienen mayor impacto en el desempeño de las celdas. Para obtener un modelo con alta precisión, se deben ajustar parámetros como el número de árboles (`n_estimators`), la profundidad máxima de los árboles (`max_depth`), el número mínimo de muestras requeridas

para dividir un nodo, entre otros [113].

2.7.2. Extreme Gradient Boosting

Extreme gradient boosting (XGB) es un algoritmo de aprendizaje supervisado basado en gradient boosting que construye secuencialmente un conjunto de árboles de decisión. A diferencia de RF, donde los árboles se construyen de manera independiente, en XGB cada árbol subsiguiente aprende de los errores de los árboles anteriores, intentando corregir los residuos. El proceso comienza con una predicción inicial (usualmente la media de la variable objetivo). Luego, se calculan los errores residuales (diferencias entre valores observados y predichos), y se construye un árbol de decisión para modelar estos. La salida de este árbol se suma a la predicción anterior, y el proceso se repite hasta que los errores residuales dejan de reducirse o se alcanza el número máximo de iteraciones deseadas. Esta técnica es ampliamente utilizada en el modelamiento de celdas solares de perovskita debido a su alta precisión y velocidad computacional. Para optimizar sus hiperparámetros, se ajusta el número de iteraciones (nrounds), profundidad máxima del árbol (max_depth), tasa de aprendizaje (eta), regularización (gamma), entre otros [114].

2.7.3. Artificial Neural Network

Las redes neuronales artificiales (ANN) se inspiran en el procesamiento de información del sistema nervioso humano y están compuestas por capas de neuronas artificiales interconectadas, donde cada neurona aplica una combinación de las entradas seguida por una función de activación no lineal [115]. Permitiendo así modelar relaciones complejas entre variables de entrada y salida, por lo que son comúnmente usadas para problemas de regresión multivariable en los que existen interacciones no lineales, alta dimensionalidad y efectos acoplados entre parámetros de las PSC [116]. Evaluandola desde su arquitectura, el término ANN se usa en sentido amplio para referirse al conjunto de redes neuronales supervisadas utilizadas para modelos predictivos. Dentro de este conjunto existen diferentes configuraciones, como las redes feed-forward, las redes convolucionales, las redes recurrentes y otras variantes.

2.7.4. Support Vector Regression

A diferencia de los métodos de regresión tradicionales que buscan minimizar el error entre los valores predichos y reales, support vector regression (SVR) busca encontrar una función que se desvíe de los valores reales por un valor no mayor a epsilon para cada punto de entrenamiento, mientras se mantiene lo más plana posible. SVR introduce el concepto de tubo epsilon alrededor de la función de predicción. Los puntos que caen dentro de este tubo no

contribuyen a la función de pérdida, lo que hace que el modelo sea robusto frente a valores atípicos. Los puntos que se encuentran fuera de los vectores de soporte son los que determinan la función de regresión. Su capacidad para trabajar con espacios de alta dimensión mediante funciones kernel lo hace especialmente útil cuando las relaciones entre variables son altamente no lineales.

2.7.5. Multi-Layer Perceptron

Multi-Layer Perceptron (MLP) es un tipo de red neuronal artificial compuesta por múltiples capas de neuronas conectadas. A diferencia de un perceptrón simple que solo puede resolver problemas linealmente separables, el MLP puede modelar relaciones no lineales complejas, ya que cuenta con capas ocultas y funciones de activación no lineales. Un MLP se compone de una capa de entrada, unas capas ocultas y una capa de salida. Cada neurona en una capa está conectada a las demás neuronas de la capa siguiente. Para problemas de regresión, la capa de salida usa una función de activación lineal, mientras que las capas ocultas usan funciones no lineales como ReLU, sigmoid o tanh. El MLP se entrena mediante el algoritmo de retropropagación, que calcula los gradientes de la función de pérdida con respecto a todos los pesos de la red, actualizando su resultado mediante descenso por gradiente [117]. Su capacidad para aproximar funciones no lineales arbitrariamente complejas lo hace ideal para capturar las relaciones entre los parámetros de diseño y el rendimiento de las celdas [118].

En el presente trabajo, la ANN corresponde a una red feed-forward implementada mediante Keras/TensorFlow, cuya arquitectura se definió a partir de unas capas de entrada, múltiples capas ocultas y unas capas de salida. Con el objetivo de evaluar su desempeño para modelar la relación entre los parámetros geométricos y fotoeléctricos de PSC y las métricas de desempeño del dispositivo Voc, Jsc, FF y PCE [119]. La capacidad de las ANN para ajustarse a problemas no lineales depende de hiperparámetros como la profundidad de la red, el número de neuronas, la función de activación, el algoritmo de optimización, entre otros. Por tanto, el desempeño de este tipo de modelos no solo está determinado por la cantidad de datos, sino también por la correcta selección de sus hiperparámetros [120, 121]. Mientras que en MLP se utiliza una activación lineal y las capas internas emplean funciones tales como ReLU, tanh o logistic, según la implementación específica, el entrenamiento se realiza mediante retropropagación del error y actualización iterativa de pesos, de modo que la red minimice una función de pérdida definida sobre el conjunto de entrenamiento [122]. Su capacidad de modelar las PSC depende de sus hiperparámetros, de la estrategia de regularización y de la adecuada normalización de los datos [123].

En concordancia con lo anterior, en el Capítulo 4 la ANN no se compara con el MLP como si se tratara de técnicas conceptualmente diferentes, sino como dos implementaciones distintas

de redes neuronales feed-forward: una construida de forma flexible en Keras/TensorFlow y otra implementada mediante MLPRegressor de scikit-learn.

2.8. Métricas para evaluar los modelos de regresión

La evaluación del rendimiento de modelos de aprendizaje automático es fundamental para predecir el desempeño de las celdas solares de perovskita. Para cuantificar qué tan preciso es un modelo al predecir el comportamiento real de una celda, se utilizan métricas de desempeño que comparan las predicciones del modelo con valores reales u observados, permitiendo analizar y optimizar la precisión de los algoritmos empleados en la simulación y predicción de estos dispositivos. En el caso de modelos de regresión aplicados a la predicción del desempeño de estos dispositivos, las métricas más comunes incluyen el Error Absoluto Medio (MAE), la Raíz del Error Cuadrático Medio (RMSE) y el Coeficiente de determinación (R^2):

2.8.1. Error Absoluto Medio

El error absoluto medio (MAE) cuantifica la diferencia promedio en valor absoluto entre los valores predichos y los valores reales, expresando ese error en las mismas unidades físicas de la variable a predecir. Esto es especialmente útil en problemas como el que se aborda en este trabajo, donde se predicen magnitudes como PCE, Voc, Jsc o FF, y se requiere interpretar el error en términos directamente comparables con los valores experimentales o simulados, facilitando así la comprensión del desempeño del modelo bajo criterios prácticos. En la predicción del rendimiento de las PSC, el MAE permite calcular la desviación entre el valor esperado y el estimado por el modelo, tal y como se muestra en la ecuación (2-10):

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2-10)$$

donde y_i son los valores reales, \hat{y}_i son los valores predichos para cada observación y n es el número total de observaciones. Un valor de MAE menor indica que, en promedio, las predicciones del modelo están más cercanas a los valores reales, lo que refleja un mejor desempeño del modelo. Esta métrica es sencilla de interpretar y es útil para la comparación de diferentes modelos.

2.8.2. Raíz del Error Cuadrático Medio

La raíz del error cuadrático medio (RMSE) es una métrica que permite cuantificar la diferencia promedio entre las predicciones de un modelo respecto a los valores reales, expresando ese error en las mismas unidades físicas de la variable a predecir. Esto es especialmente útil en problemas fotovoltaicos, donde se predicen magnitudes como PCE, Voc, Jsc o FF, y se

requiere interpretar el error en términos directamente comparables con los valores experimentales o simulados, facilitando así la comprensión del desempeño del modelo bajo criterios prácticos. Al igual que el MAE, el RMSE expresa el error en las mismas unidades físicas de la variable a predecir, por lo que se facilita su interpretación.

Es importante mencionar que un RMSE más cercano a cero indica un mejor desempeño del modelo, ya que, en promedio, los valores reales están más cerca de los valores predichos.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2-11)$$

2.8.3. Coeficiente de Determinación

El coeficiente de determinación (R^2) mide qué tanto de la variabilidad de los parámetros puede ser explicada por el modelo de aprendizaje automático. Es una métrica fundamental para determinar si el modelo captura correctamente las relaciones entre arquitectura, materiales y rendimiento.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2-12)$$

donde \bar{y} es la media de los valores reales. Un valor de R^2 cercano a 1 indica que el modelo explica aproximadamente todo el comportamiento de la variabilidad observada, mientras que un valor cercano a 0 indica una capacidad de predicción pobre. En la predicción del desempeño de celdas solares, un alto R^2 respalda la confiabilidad del modelo para predecir parámetros como PCE, FF, Jsc o Voc a partir de las variables de entrada.

3 Metodología para la simulación de PSCs

Las simulaciones numéricas de las PSCs representan una herramienta útil para comprender su funcionamiento óptico y eléctrico antes de proceder con la fabricación experimental. Esta metodología permite evaluar el efecto de las variables geométricas (espesores de capas) y fotoeléctricas (bandgap, densidad de defectos, permitividad dieléctrica, entre otros) sobre el rendimiento del dispositivo, minimizando significativamente los costos derivados de su desarrollo, prueba y error o tiempos de investigación.

La estructura mostrada en la Figura 3-1 corresponde a un PSC de tipo n-i-p con cinco capas, que es la configuración estudiada en este trabajo [124]. La primera y última capa son los contactos eléctricos, mientras que las tres capas internas son responsables de la conversión de energía del dispositivo. Para el contacto superior, se utilizó óxido de estaño dopado con flúor (FTO), ya que comúnmente se usa como un electrodo transparente. El óxido de titanio (TiO_2) y el óxido de estaño (SnO_2) se utilizaron para la capa de transporte de electrones (ETL) debido a sus propiedades electrónicas y su uso comprobado en la literatura científica [81, 125–129]. Para la capa absorbente de perovskita, se utilizaron yoduro de plomo de metilamonio (MAPbI_3), yoduro de estaño de metilamonio (MASnI_3) y yoduro de plomo de formamidinio (FAPbI_3) porque estos materiales tienen las eficiencias energéticas más altas reportadas y han sido ampliamente estudiados por la comunidad científica [13, 130–133]. Para la capa de transporte de huecos (HTL), se utilizaron Spiro-OMeTAD y tiocianato de cobre (CuSCN) porque las configuraciones con estos materiales han demostrado un rendimiento notable en la movilidad de huecos y un alineamiento energético efectivo [133–135]. Finalmente, la última capa generalmente está hecha de oro (Au) ya que tiene una alta conductividad eléctrica [135]. Estos materiales fueron elegidos por sus propiedades fotoeléctricas, compatibilidad energética y el alto rendimiento demostrado en estudios experimentales y simulados disponibles en la literatura [81, 84, 86, 133, 135–137].

Para convertir la energía solar en energía eléctrica, el PSC absorbe fotones de la radiación solar en la capa de perovskita, generando pares electrón-hueco, que son separados y transportados por la capa ETL, que extrae los electrones, mientras que la capa HTL recoge los huecos. El electrodo superior, generalmente hecho de un óxido conductor transparente como FTO, permite la entrada de luz y la recolección de portadores, mientras que el electrodo

metálico inferior completa el circuito, permitiendo el flujo de corriente externa bajo condiciones de carga.

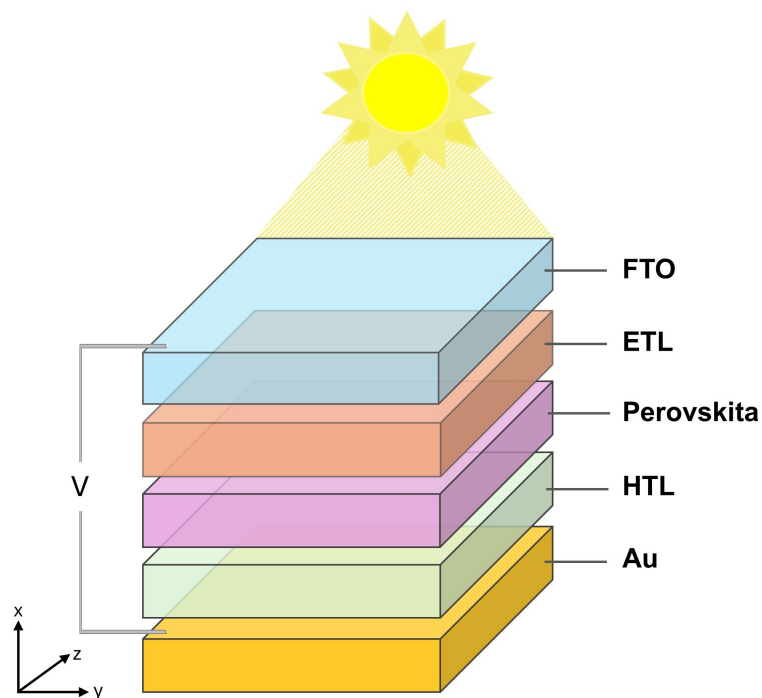


Figura 3-1: Estructura de la celda solar de perovskita simulada. Adaptado de [3].

3.1. Generación de la base de datos

El conjunto de datos o dataset se generó mediante simulaciones numéricas utilizando el software gratuito SCAPS-1D versión 3.3.09, usando una workstation con un procesador Intel Core i9 de 3.19 GHz y 100 GB de memoria RAM.

La evaluación del rendimiento eléctrico del dispositivo se llevó a cabo mediante el estudio de las curvas J-V en condiciones de trabajo estándar, con un espectro solar AM 1.5G (1000 W/m^2), temperatura de 300 K y voltaje máximo aplicado de 1,2 V [138]. A partir de estas curvas, se determinaron los principales parámetros eléctricos que caracterizan el rendimiento del sistema, incluidos V_{oc} , J_{sc} , FF y PCE. Estos parámetros se obtuvieron directamente del software tras simular el comportamiento fotoeléctrico del dispositivo, lo que permitió una evaluación precisa del rendimiento esperado y un análisis comparativo en términos de eficiencia, estabilidad y robustez frente a las variaciones en los materiales, las propiedades o el espesor de las capas estudiadas. Aunque el barrido paramétrico se enfocó en siete parámetros clave dentro de los rangos establecidos en la Tabla **3-1** (y en la selección de materiales), el

dataset final contiene 19 columnas porque se registraron variables adicionales de naturaleza eléctrica asociada a los materiales con el fin de garantizar reproducibilidad y robustez. En particular, antes de definir el barrido final, se evaluó la variación individual de cada una de las variables (incluyendo cambios de materiales) y se seleccionaron aquellas que inducían mayor varianza en las cuatro métricas (Voc, Jsc, FF y PCE), lo cual evidenciaba un efecto directo sobre la respuesta de la celda. Las demás variables se mantuvieron fijas en valores de referencia (tomados de la literatura), se almacenaron y reportaron en la base de datos para facilitar la replicación de los resultados numéricos y permitir análisis posteriores sobre posibles efectos combinados de variables eléctricas, geométricas y morfológicas.

Los rangos de las variaciones paramétricas de las propiedades de la celda se basaron en valores reportados en la literatura y la física del dispositivo. Por ejemplo, el espesor de la capa de perovskita (T_PVK) debe ajustarse para absorber la mayor cantidad de fotones, maximizando Jsc sin exceder la longitud de difusión del portador, ya que un espesor excesivo aumenta las pérdidas por recombinación y degrada Voc y FF [139]. Del mismo modo, los espesores de las capas de transporte (T_ETL y T_HTL) deben optimizarse para garantizar un transporte eficiente de electrones y huecos con baja recombinación y resistencia en serie [140,141].

Además, las propiedades de la perovskita influyen directamente en el rendimiento de la celda; el bandgap (EG_PVK) establece el equilibrio entre la densidad de corriente y el voltaje, según el límite de Shockley-Queisser [142]; la permitividad dieléctrica (ER_PVK) influye en la disociación de excitones [143,144]; la densidad de aceptores (NA_PVK) modela el campo eléctrico interno, esencial para la separación de cargas y un Voc alto [145]; y la densidad de defectos (NT_PVK) representa la vía principal para la pérdida por recombinación no radiactiva y limita la vida útil del portador [138]. Los rangos de variación se especifican en la Tabla **3-1** y se descartaron las combinaciones de parámetros que daban lugar a errores de convergencia en el software, ya que estos solían surgir de rangos físicamente poco realistas, lo que alteraba la solución de la ecuación de Poisson, las ecuaciones de continuidad o las condiciones de contorno.

3.2. Descripción de los datos

El conjunto de datos generado comprende 18.570 simulaciones de PSC generadas a partir de cuatro arquitecturas de dispositivos: $\text{TiO}_2/\text{MAPbI}_3/\text{CuSCN}$ [146], $\text{TiO}_2/\text{MASnI}_3/\text{Spiro-OMeTAD}$ [147], $\text{SnO}_2/\text{FAPbI}_3/\text{Spiro-OMeTAD}$ [148] y $\text{TiO}_2/\text{MAPbI}_3/\text{Spiro-OMeTAD}$ [149]. La parametrización de las propiedades de los materiales se realizó mediante la opción "Batch Configuration" de SCAPS-1D, se utilizaron los valores especificados en la tabla **3-1** para cada variable, y se descartaron las combinaciones que produjeron errores de convergencia. El conjunto de datos incluye, para cada registro, diecinueve características fotoeléctricas y

Tabla 3-1: Rangos de variación de los parámetros geométricos y fotoeléctricos utilizados para la simulación. Adaptado de [3].

Capa	Parámetro	Valores	Unidad
<i>ETL</i>	T_ETL	0.02, 0.1, 0.2	μm
<i>HTL</i>	T_HTL	0.1, 0.4, 0.7	μm
Perovskita	T_PVK	0.1, 0.5, 0.9, 1.2, 1.7	μm
	E_g	1.1, 1.3, 1.5, 1.7, 1.9	eV
	ε_r	8, 13, 20	–
	N_A	1×10^{13} , 1×10^{15} , 1×10^{17}	cm^{-3}
	N_t	4×10^{13} , 1×10^{14} , 1×10^{15} , 4×10^{15}	cm^{-3}

geométricas de PSC (que podrían tomarse como entradas o valores “X” en una implementación de aprendizaje de máquinas) y cuatro resultados asociados, como Voc, Jsc, FF y PCE (que podrían utilizarse como salidas o valores “y”). La descripción de la convención de los nombres de las columnas es la siguiente:

Material (M):

- Columna A: Material de la capa ETL (M_ETL).
- Columna B: Material de la capa absorbente de perovskita (M_PVK).
- Columna C: Material de la capa HTL (M_HTL).

Parámetros variables:

Las siguientes columnas corresponden a los parámetros que se variaron, tal y como se muestra en la tabla **3-1**:

- Columna D: Espesor de la capa ETL en μm (T_ETL).
- Columna I: Espesor de la capa absorbente en μm (T_PVK).
- Columna J: Bandgap de la capa absorbente en eV (EG_PVK).
- Columna K: Permitividad dieléctrica de la capa absorbente (ER_PVK).

- Columna L: Densidad de aceptores de la capa absorbente en cm^{-3} (NA_PVK).
- Columna N: Densidad de defectos de la capa absorbente en cm^{-3} (NT_PVK).
- Columna O: Espesor de la capa HTL en μm (T_HTL).

Parámetros de valor constante:

Las siguientes columnas tienen valores constantes para los parámetros específicos. Es importante incluirlas para validar los resultados presentados en este trabajo:

- Columna E: Bandgap de la capa ETL en eV (EG_ETL).
- Columna F: Valor de permitividad dieléctrica de la capa ETL (ER_ETL).
- Columna G: Densidad de donores de la capa ETL en cm^{-3} (ND_ETL).
- Columna H: Valor de densidad de defectos de la capa ETL en cm^{-3} (NT_ETL).
- Columna M: Valor de densidad de donores de la capa absorbente en cm^{-3} (ND_PVK).
- Columna P: Bandgap de la capa HTL en eV (EG_HTL).
- Columna Q: Permitividad dieléctrica de la capa HTL (ER_HTL).
- Columna R: Densidad de donores de la capa HTL en cm^{-3} (NA_HTL).
- Columna S: Densidad de defectos de la capa HTL en cm^{-3} (NT_HTL).

Métricas de rendimiento:

- Columna T: Voltaje en circuito abierto en V (VOC).
- Columna U: Densidad de corriente de cortocircuito en mA/cm^2 (JSC)
- Columna V: Factor de llenado en % (FF).
- Columna W: Eficiencia de conversión de energía en % (PCE).

Para realizar un primer análisis de la base de datos, se realizaron estadísticas descriptivas básicas para el conjunto de datos, generando la distribución de datos para las métricas de rendimiento (Voc, Jsc, FF y PCE). La figura **3-2 a)** presenta la distribución de datos para Voc, que muestra un pico principal en la distribución multimodal con una media de 0,98 V y una mediana de 1,00 V, con una desviación estándar de 0,14 V y un rango intercuartílico de 0,90 V-1,10 V. Se observa un número menor de combinaciones paramétricas para valores

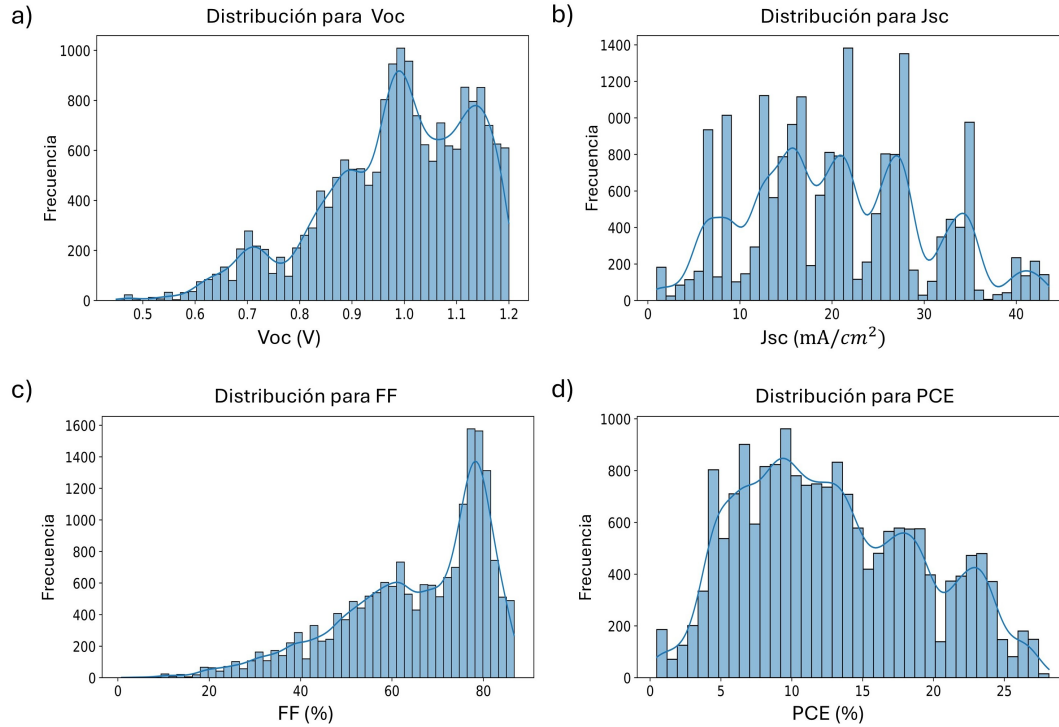


Figura 3-2: Distribución de datos de las métricas de rendimiento: a) Voc, b) Jsc, c) FF y d) PCE. Adaptado de [3]

inferiores a 0,7 V, lo que puede atribuirse a una mayor recombinación debido a la configuración geométrica del dispositivo o a una alineación de bandas inadecuada [150, 151].

La distribución de datos de Jsc presentada en la Figura 3-2 b), muestra una distribución multimodal con cinco picos marcados alrededor de 8 mA/cm², 16 mA/cm², 21 mA/cm², 27 mA/cm² y 34 mA/cm², con una media de 20,7 mA/cm²; una mediana de 20,2 mA/cm²; y una desviación estándar de 9,4 mA/cm². Los picos sugieren subconjuntos definidos por espesores discretos de la perovskita o por pasos en la absorción óptica impuestos durante el barrido paramétrico. Físicamente, las densidades de corriente inferiores a 10 mA/cm² se asocian con capas muy delgadas o bandas prohibidas grandes, mientras que los valores superiores a 30 mA/cm² se asocian con capas suficientemente gruesas con baja densidad de defectos, donde se maximizan la absorción y la recolección de portadores [152, 153].

La figura 3-2 c) presenta la distribución de datos del FF, donde se observa un pico notable en valores cercanos al 79%, con una media del 65,2%, una mediana del 69,2% y una desviación estándar del 15,8%. Esto demuestra una alta dispersión en los datos, debido a que una cantidad considerable de datos se encuentra en valores inferiores al 60%, lo que afecta significativamente a la distribución del FF para el conjunto de datos y muestra que algunas configuraciones pueden generar pérdidas, ya sea por recombinación o por defectos

de las células [154].

Por último, la figura **3-2 d)** presenta la distribución de datos del PCE, que muestra un pico en torno al 9% con una media del 12,8%, una mediana del 12,1% y una desviación estándar del 6,26%, lo que muestra una forma amplia y ligeramente bimodal: un grupo entre el 5% y el 15% asociado a dispositivos con uno o dos parámetros subóptimos (por ejemplo, Jsc moderado y FF aceptable) y otro pico entre el 18% y el 23% asociado a valores de Voc y FF. La tendencia decreciente del 25% refleja el límite impuesto por la absorción máxima y las pérdidas residuales no radiativas, en consonancia con el modelo de Shockley-Queisser [150, 155].

El conjunto de datos generado se publicó en OSF HOME, una plataforma de código abierto para gestionar y compartir datos de investigación, con el título “Synthetic dataset to study the performance of perovskite solar cell simulations” y con DOI: 10.17605/OSF.IO/ZX4AJ. El proyecto incluye el archivo “Synthetic dataset to study the performance of perovskite solar cell simulations.xlsx”, que contiene todas las configuraciones de dispositivos simuladas y sus correspondientes métricas de rendimiento.

Este conjunto de datos puede servir para analizar, diseñar y optimizar PSC, ya que contiene un número considerable de configuraciones (18.570 variaciones) con sus respectivos valores de rendimiento fotovoltaico, lo que resulta útil para estudiar las relaciones entre los parámetros de diseño y el rendimiento eléctrico. Debido a su composición y variedad en los parámetros que constituyen el dispositivo, el conjunto de datos se puede utilizar para entrenar diferentes modelos de ML para predecir las métricas de rendimiento, o a través de técnicas no supervisadas, descubrir patrones o relaciones entre parámetros. Adicionalmente, usando este conjunto de datos, los investigadores pueden identificar las regiones óptimas del espacio de diseño para maximizar la eficiencia, minimizar las pérdidas por recombinación o reducir el uso de materiales de alto coste. Por lo tanto, puede ser útil en estrategias de diseño inverso para determinar configuraciones óptimas basadas en objetivos definidos.

3.3. Validación de datos

La precisión de la metodología de simulación implementada se validó reproduciendo los resultados presentados en los artículos de investigación. Para lograrlo, se recopilaron más de 50 artículos científicos publicados recientemente que incluían la mayoría de los parámetros para simular un PSC de tipo n-i-p en SCAPS-1D y que también presentaban las métricas de rendimiento Voc, Jsc, FF y PCE. Tras una revisión sistemática, para evitar resultados poco realistas, se excluyeron los artículos que informaban de eficiencias energéticas superiores al límite de Shockley-Queisser [150] para células basadas en MAPbI₃, MASnI₃ y FAPbI₃, ya que, según las validaciones experimentales [125], la eficiencia máxima alcanzada para es-

tos materiales a la fecha, no supera el 22.2 %, 14.35 % y 24.66 %, respectivamente. Es importante señalar que, en las simulaciones de PSC con SCAPS-1D, no se incorporan factores físicos variables (rugosidad de la superficie, tamaño y orientación del grano), químicos (temperatura y tiempo de secado, ingeniería de disolventes y antidisolventes, y aditivos) o ambientales (variaciones de temperatura, nubosidad, irradiancia, etc.), que pueden afectar el rendimiento real de las celdas. Por lo tanto, se espera que los valores obtenidos mediante la simulación sean superiores a los observados experimentalmente, manteniendo la coherencia con los valores realistas [125, 156].

Tabla 3-2: Comparación entre las métricas de rendimiento publicadas y simuladas para distintas PSCs. Adaptado de [3]

Ref.	Resultados reportados				Resultados simulados			
	$V_{OC}(V)$	$J_{SC}(mA/cm^2)$	FF (%)	PCE(%)	$V_{OC}(V)$	$J_{SC}(mA/cm^2)$	FF (%)	PCE(%)
[147]	1,04	30,5	82,69	26,95	1,02	29,8	78,28	26,12
[146]	0,98	18,6	82,50	13,40	0,96	18,2	79,97	13,07
[157]	1,02	22,7	62,67	21,42	1,01	22,4	65,92	21,26
[158]	0,91	24,1	54,19	16,08	0,93	24,3	78,98	16,49
[159]	0,87	24,9	85,80	15,50	0,85	24,4	81,26	15,14

En la Tabla 3-2 se muestra una comparación entre los valores obtenidos mediante simulación y los valores publicados en la bibliografía para V_{oc} , J_{sc} , FF y PCE. Existen ligeras variaciones entre los valores simulados y los publicados, lo que se atribuye a múltiples causas, como que muy pocos autores informan detalladamente todas las condiciones de simulación o todos los parámetros utilizados. Algunos estudios incluyen modelos de recombinación, absorción o defectos sin especificar valores numéricos (densidad de defectos, tipo de defectos, coeficientes de recombinación, resistencia en serie o de derivación), por lo que la réplica exacta de las condiciones de simulación es limitada. Esto es crucial para una simulación, ya que tiene en cuenta los fenómenos de recombinación, las pérdidas por defectos de las capas o interfaces derivados de los procesos de fabricación o las impurezas de los materiales que componen la celda, que pueden afectar significativamente al rendimiento del dispositivo. Aunque las variaciones en los parámetros comunicados impiden una réplica idéntica, los resultados obtenidos muestran coherencia con los datos publicados, lo que respalda la solidez de la metodología.

4 Modelos de inteligencia artificial para predicción

El rendimiento de las celdas solares de perovskita emerge de la interacción compleja entre parámetros geométricos, propiedades fotoeléctricas de los materiales y fenómenos físicos acoplados (ópticos, de transporte de carga y recombinación). Predecir con precisión métricas clave como Voc, Jsc, FF y PCE resulta esencial para optimizar arquitecturas, seleccionar combinaciones de materiales y reducir iteraciones experimentales costosas. Sin embargo, las no linealidades inherentes, las interacciones cruzadas entre capas (ETL, perovskita, HTL) y la sensibilidad a densidades de defectos dificultan el desarrollo de modelos analíticos exactos.

En este capítulo se abordan dichas complejidades mediante la construcción y evaluación comparativa de cinco técnicas de aprendizaje automático sobre un conjunto de datos sintéticos derivados de simulaciones numéricas SCAPS-1D: Random Forest (RF), XGBoost (XGB), Red Neuronal Artificial (ANN), Perceptrón Multicapa (MLP) y Máquinas de Soporte Vectorial (SVM). Es importante mencionar que la comparación entre ANN y MLP se entiende como la implementación de dos tipos específicos de redes neuronales feed-forward y no como una confrontación entre técnicas conceptualmente independientes. Para cada modelo se realiza una optimización exhaustiva de hiperparámetros mediante la librería GridSearchCV de scikit-learn y se evalúa su desempeño sobre particiones de validación cruzada utilizando métricas estandarizadas como: R^2 , MAE y RMSE. Los objetivos específicos comprenden: primero, establecer un método robusto que cuantifique la capacidad predictiva de cada modelo sobre las cuatro métricas fotoeléctricas; segundo, comparar la mejora en precisión entre los cinco modelos; tercero, analizar la importancia relativa de características (feature importance) para verificar su alineación con principios físicos establecidos de dispositivos fotovoltaicos; y cuarto, identificar el modelo óptimo que combine máxima precisión predictiva con interpretabilidad física para guiar el diseño racional de celdas solares de perovskita de alto rendimiento. Este enfoque busca no solo superioridad predictiva, sino también transferencia directa de conocimiento hacia la optimización de los dispositivos con el fin de mejorar su rendimiento. La Figura 4-1 muestra un resumen gráfico de la metodología empleada para el presente trabajo.

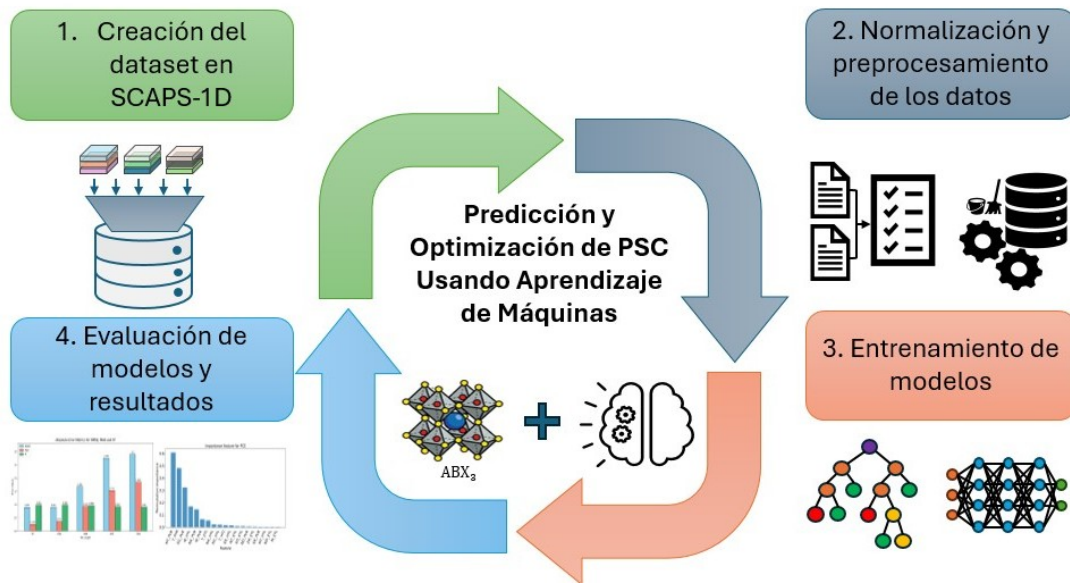


Figura 4-1: Esquema del flujo de trabajo del marco de aprendizaje automático propuesto para predecir el rendimiento de los PSC. El diagrama ilustra la secuencia que va desde la generación del conjunto de datos y el preprocesamiento de los datos basados en SCAPS-1D hasta el entrenamiento del modelo, la evaluación basada en métricas de regresión y el análisis de la importancia de las características.

4.1. Preprocesamiento de la base de datos

El preprocesamiento es clave para garantizar la calidad de los datos antes del proceso de entrenamiento y mejorar el rendimiento del modelo. Por ese motivo, se utilizó OneHotEncoder para transformar las tres primeras columnas de variables categóricas a numéricas, generando vectores binarios que evitan la introducción de relaciones ordinales falsas entre categorías. Esta codificación garantiza una representación numérica adecuada para los algoritmos de regresión [160]. Para desarrollar y evaluar el modelo predictivo, el conjunto de datos se dividió utilizando una estrategia de validación hold-out. El ochenta por ciento de los datos se asignó aleatoriamente al conjunto de entrenamiento y el veinte por ciento restante al conjunto de prueba. Donde este último permaneció completamente aislado durante todo el proceso de entrenamiento y se utilizó exclusivamente para la fase de evaluación final. Por otra parte, las características de entrada del conjunto de datos tienen diferentes escalas; por lo tanto, se realizó un proceso de normalización utilizando MinMaxScaler, que es una técnica de preprocesamiento de datos empleada para escalar el conjunto de datos separado a un rango específico entre 0 y 1. Su propósito es normalizar los datos para que las características con rangos de valores muy diferentes contribuyan por igual al rendimiento de un modelo de aprendizaje automático [119].

Es importante destacar que, para evitar fuga de información (data leakage), el `MinMaxScaler` se ajustó exclusivamente con los datos de entrenamiento (`X_train` y `y_train`) y, posteriormente, las transformaciones obtenidas se aplicaron a `X_test` y `y_test`, asegurando que ninguna información del conjunto de prueba se incorpore a la etapa de preprocesamiento. El escalado de características es particularmente crítico para algoritmos basados en distancia, como SVM, ya que son sensibles a la magnitud de las variables y dependen de distancias geométricas en el espacio de entrada [161]. Posterior al entrenamiento de los modelos, se usa “`inverse.transform`” para retornar las etiquetas (Voc, Jsc, FF y PCE) a sus unidades originales y así analizar cuantitativamente las predicciones de cada modelo.

4.2. Modelos de aprendizaje automático

Para comparar enfoques predictivos con diferentes niveles de complejidad, interpretabilidad y que capturen las particularidades ópticas y eléctricas de las PSCs, se seleccionan cinco algoritmos supervisados que representan paradigmas contrastantes en el aprendizaje automático para datos sintéticos de celdas solares de perovskita. La combinación de modelos basados en árboles de decisión, redes neuronales y máquinas de soporte vectorial permite evaluar no solo el desempeño numérico, sino también la capacidad de cada técnica para capturar relaciones no lineales entre parámetros geométricos, propiedades fotoeléctricas y métricas fotovoltaicas como Voc, Jsc, FF y PCE. La optimización de hiperparámetros de los cinco modelos se realizó utilizando la librería `GridSearchCV`. Cada configuración se evaluó utilizando una validación cruzada de 10 veces dentro del conjunto de datos de entrenamiento, lo que redujo la sensibilidad a una división de datos concreta y mitigó el riesgo de sobreajuste [162].

4.2.1. Random Forest (RF)

Con el propósito de capturar relaciones no lineales y efectos de interacción entre los parámetros geométricos y fotoeléctricos del dispositivo, se implementó un regresor Random Forest (RF) como método ensemble basado en bagging. En este enfoque, la predicción final se obtiene por promediado de múltiples árboles entrenados sobre subconjuntos aleatorios del dataset, lo que reduce la varianza y mitiga el sobreajuste. El rango de hiperparámetros orientados a controlar simultáneamente la estabilidad del modelo fue de número de árboles ($n_estimators \in [100, 600]$) para determinar cuántos modelos se agregan para disminuir la varianza; la profundidad máxima ($max_depth \in [None, 40]$) permite regular el nivel de particionamiento y, por tanto, el compromiso de sesgo y varianza; los umbrales $min_samples_split \in [2, 5]$ y $min_samples_leaf \in [1, 5]$ limitan la creación de nodos con pocas observaciones, suavizando la función de decisión y evitando que el modelo memorice ruido; de forma complementaria, $min_weight_fraction_leaf \in [0.0, 0.01]$ impone un mínimo de masa (pondera-

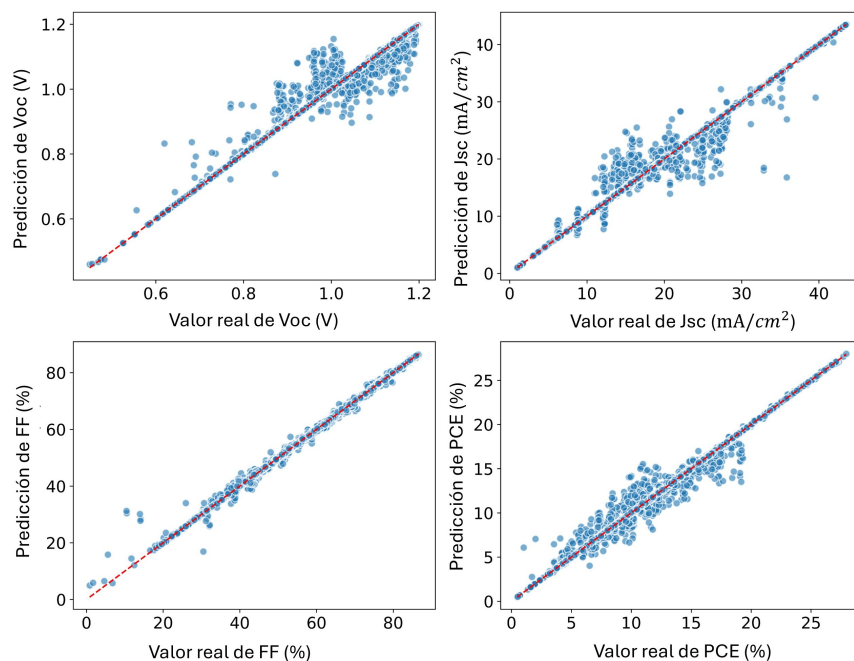


Figura 4-2: Comparación de los valores reales generados a partir de simulación y los datos predichos en el modelo de Random Forest.

da) por hoja; finalmente, `max_features` ($\in [1.0, 0.8, \text{"sqrt"}]$), controla cuántas características pueden considerarse en cada partición, modulando la diversidad entre árboles. La mejor configuración obtenida fue `n_estimators=600`, `max_depth=None`, `min_samples_split=2`, `min_samples_leaf=1`, `min_weight_fraction_leaf=0.0` y `max_features=1.0`, alcanzando un R^2 de 0.972, 0.978, 0.9967 y 0.999 para Voc, Jsc, FF y PCE, respectivamente. Este desempeño es consistente con la alineación de las predicciones respecto a la diagonal ideal en la Figura 4-2, lo que evidencia una adecuada capacidad del modelo para reproducir las tendencias del conjunto simulado.

4.2.2. XGBoost

Para contrastar un esquema de aprendizaje aditivo capaz de corregir iterativamente errores residuales y modelar dependencias no lineales con alta precisión, se implementó XGBoost como método de gradient boosting basado en árboles. A diferencia de RF, donde los árboles se entrenan de forma independiente, en XGBoost cada nuevo árbol se ajusta sobre el error del ensemble previo, refinando progresivamente la aproximación y mejorando la capacidad de generalización cuando se controlan adecuadamente los mecanismos de complejidad. La optimización del modelo se efectuó mediante búsqueda por cuadrícula, evaluando `n_estimators` ($\in [100, 500]$), `max_depth` ($\in [3, 6]$), `learning_rate` ($\in [0.01, 0.1]$) y `subsample` ($\in [0.8, 1.0]$). En este contexto, `n_estimators` define el número de etapas de boosting empleadas para corregir

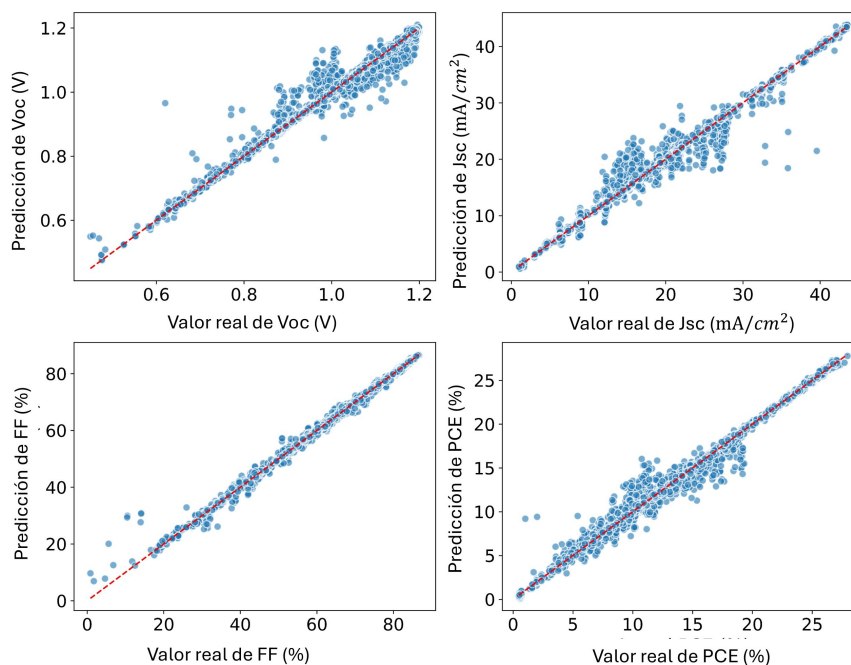


Figura 4-3: Comparación de los valores reales generados a partir de simulación y los datos predichos en el modelo de XGBoost.

residuales, `max_depth` controla la complejidad de cada árbol base y su capacidad de capturar interacciones de orden superior, `learning_rate` actúa como factor de amortiguación que reduce el aporte de cada árbol, estabilizando la convergencia y disminuyendo el riesgo de sobreajuste incluso con árboles relativamente profundos, y `subsample` introduce aleatoriedad al entrenar cada iteración con una fracción de los datos, lo que mejora la robustez del ensamble frente a fluctuaciones del dataset. La mejor configuración encontrada fue `n_estimators=200`, `max_depth=6`, `learning_rate=0.1` y `subsample=0.8`, con lo cual se obtuvo un R^2 de 0.977, 0.981, 0.985 y 0.9976 para Voc, Jsc, FF y PCE, respectivamente. Estos resultados reflejan un ajuste altamente preciso a las simulaciones, manteniendo estabilidad ante variaciones en los parámetros geométricos y fotoeléctricos del dispositivo, lo cual se observa en la Figura 4-3 mediante la concentración de puntos alrededor de la referencia ideal.

4.2.3. Artificial Neural Network (ANN)

Se implementó una red neuronal artificial (ANN); dado que se necesitaba evaluar un problema de regresión multisalida, se entrenó un estimador por cada variable objetivo bajo el mismo flujo de preprocesamiento descrito previamente, garantizando comparabilidad directa frente a los modelos basados en ensambles. La calibración del modelo se realizó mediante validación cruzada y selección sistemática de hiperparámetros, priorizando el equilibrio entre capacidad representacional y control del sobreajuste, utilizando como criterio de selección el

coeficiente de determinación (R^2). En este proceso, se exploraron combinaciones a partir de 64 a 128 neuronas en la primera capa oculta; 32 a 64 neuronas en la segunda capa oculta; tasas de dropout de 0.20 y 10; y los optimizadores Adam y RMSprop. La función de activación en las capas ocultas se fijó como ReLU, mientras que la capa de salida empleó activación lineal, apropiada para un problema de regresión continua. Adicionalmente, el entrenamiento se mantuvo con un tamaño de lote (batch_size) de 32 y 100 épocas. La configuración óptima seleccionada consistió en dos capas ocultas (hidden layers) con 128 y 64 neuronas, una función de activación ReLU y optimización con Adam, incorporando un *dropout rate* de 0.20 y 10 épocas (epoch) para regularizar el entrenamiento. Esta combinación estabiliza la convergencia al amortiguar oscilaciones del gradiente (Adam) y reduce la varianza del estimador, mitigando la degradación del desempeño fuera de muestra.

En el conjunto de prueba, la ANN alcanzó un desempeño consistente para las cuatro variables objetivo, con R^2 de 0.9538, 0.9637, 0.9714 y 0.9761 para Voc, Jsc, FF y PCE, respectivamente. Aunque su capacidad no lineal permite aproximar interacciones complejas entre parámetros del dispositivo, el ajuste permanece condicionado por la regularización y por la sensibilidad del entrenamiento a la escala de las entradas, razón por la cual este modelo se evaluó bajo normalización estricta y partición *hold-out* previamente definida y presenta una varianza moderada, como se muestra en la Figura 4-4.

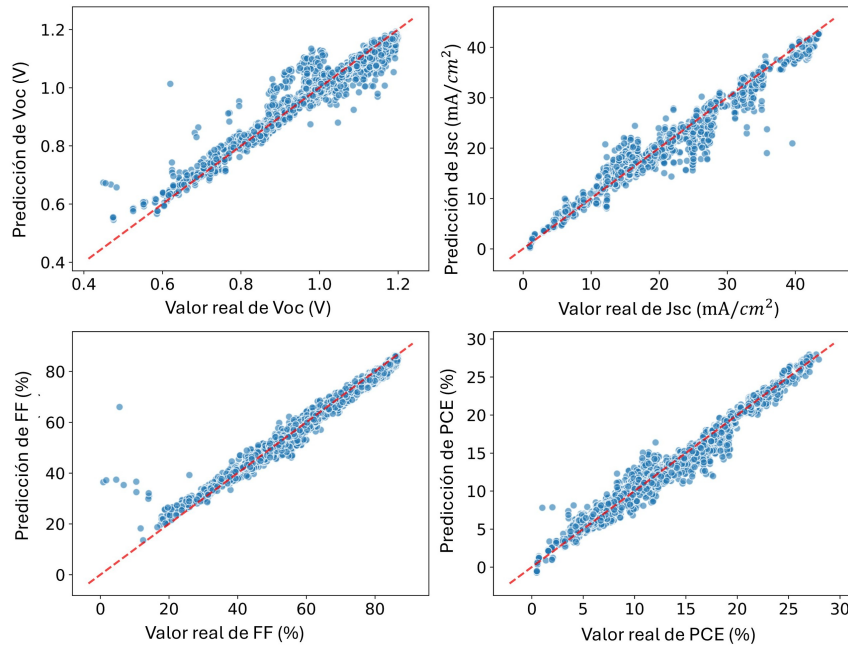


Figura 4-4: Comparación de los valores reales generados a partir de simulación y los datos predichos en el modelo de ANN.

4.2.4. Multi-Layer Perceptron (MLP)

Con el fin de contrastar el efecto de arquitecturas neuronales compactas frente a aproximadores más flexibles, se implementó un perceptrón multicapa (MLP). A diferencia de formulaciones profundas, el MLP introduce no linealidad mediante transformaciones afines sucesivas y funciones de activación, manteniendo una complejidad computacional moderada. En este proceso se evaluaron sistemáticamente tres configuraciones de arquitectura para las capas ocultas, (50,) , (100,) y (50,50); dos funciones de activación, ReLU y tanh; y dos valores del parámetro de regularización (α), correspondientes a 1×10^{-4} y 1×10^{-3} . De forma complementaria, algunos hiperparámetros se mantuvieron fijos con el fin de estabilizar el proceso de entrenamiento y garantizar la reproducibilidad experimental. En particular, el optimizador se fijó como Adam, la política de actualización de la tasa de aprendizaje se estableció como *adaptive*, el número máximo de iteraciones se definió en 1500 y la semilla aleatoria en 1. La optimización de hiperparámetros determinó una arquitectura de dos capas ocultas simétricas (50, 50) con función de activación ReLU y optimización con Adam. El término de regularización (α) con valor de 1×10^{-4} se empleó como mecanismo principal de control de complejidad, penalizando magnitudes elevadas de pesos sin anular completamente la capacidad de modelar curvaturas locales; adicionalmente, la tasa de aprendizaje inicial y el máximo de iteraciones fijan el compromiso entre rapidez de convergencia y estabilidad numérica.

Los resultados evidencian un desempeño competitivo, con R^2 de 0.9076, 0.928, 0.9118 y 0.9487 para Voc, Jsc, FF y PCE, respectivamente. En particular, el MLP mostró una alta capacidad para ajustar FF, lo que sugiere que una parametrización neuronal moderada puede capturar regularidades asociadas a pérdidas resistivas y recombinación efectiva reflejadas en este indicador. La Figura 4-5 permite observar la dispersión entre los datos reales y los datos predichos para las cuatro métricas.

4.2.5. Support Vector Machine (SVM)

Finalmente, se evaluó un modelo de regresor de máquinas de vectores de soporte (SVM) como aproximador no lineal basado en márgenes, utilizando un kernel de base radial (RBF) para inducir una proyección implícita a un espacio de mayor dimensión. La evaluación de hiperparámetros se llevó a cabo utilizando el coeficiente de determinación R^2 como criterio de selección. En este proceso se exploraron sistemáticamente los valores $C \in \{1, 10, 100\}$, $\epsilon \in \{0,1, 0,2\}$ y $\gamma \in \{scale, auto, 0,1, 1\}$, manteniendo fijo el kernel RBF. En SVM, el hiperparámetro C controla la penalización por errores definiendo el compromiso sesgo-varianza, pues los valores altos reducen el error de entrenamiento, pero incrementan el riesgo de sobreajuste. Tras la implementación de GridSearchCV, la configuración óptima seleccionada fue kernel RBF, $C = 100$, $\epsilon = 0,1$ y γ en escala automática, lo que permite adaptar la suavidad

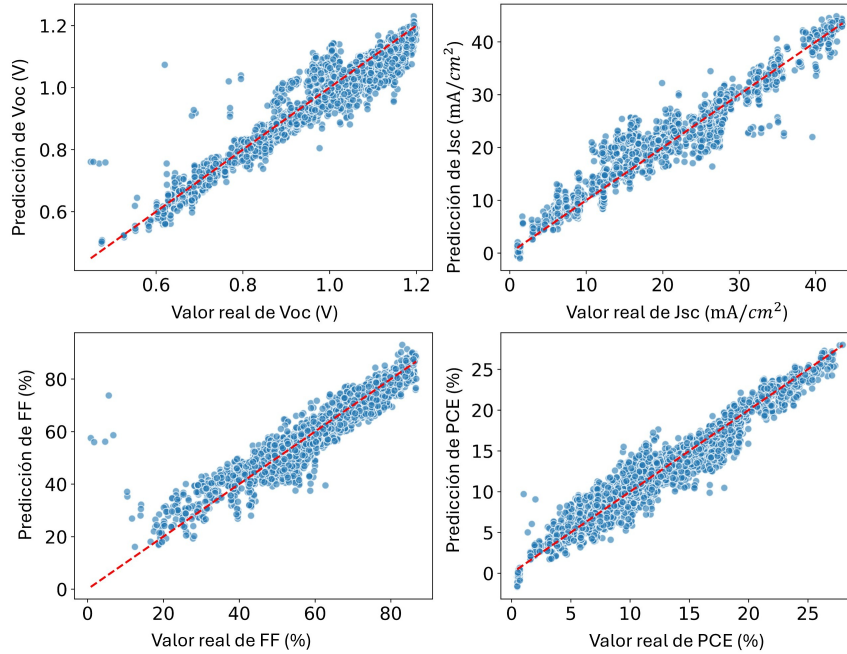


Figura 4-5: Comparación de los valores reales generados a partir de simulación y los datos predichos en el modelo de MLP.

del kernel a la dispersión de los datos normalizados.

En el conjunto de prueba, SVM alcanzó R^2 de 0.8853, 0.917, 0.9055 y 0.9319 para Voc, Jsc, FF y PCE, respectivamente. Si bien mostró capacidad de generalización razonable para Jsc, su desempeño fue más sensible en PCE, coherente con la acumulación de incertidumbre cuando se modelan simultáneamente efectos acoplados (ópticos, recombinación y transporte) que impactan de forma multiplicativa la eficiencia. Este modelo fue el que presentó valores más bajos de R^2 , lo que se puede explicar al observar la Figura 4-6 y la alta dispersión entre los valores predichos y los reales.

4.3. Rendimiento predictivo comparativo de los modelos de aprendizaje automático

El rendimiento predictivo de los cinco modelos de ML se evaluó cuantitativamente utilizando el conjunto de datos de prueba independiente, empleando las métricas de regresión complementarias R^2 , RMSE y MAE para cada parámetro de rendimiento fotovoltaico (Voc, Jsc, FF y PCE). La tabla 4-1 resume los resultados numéricos y permite una comparación directa de la precisión y la robustez del modelo en todos los objetivos.

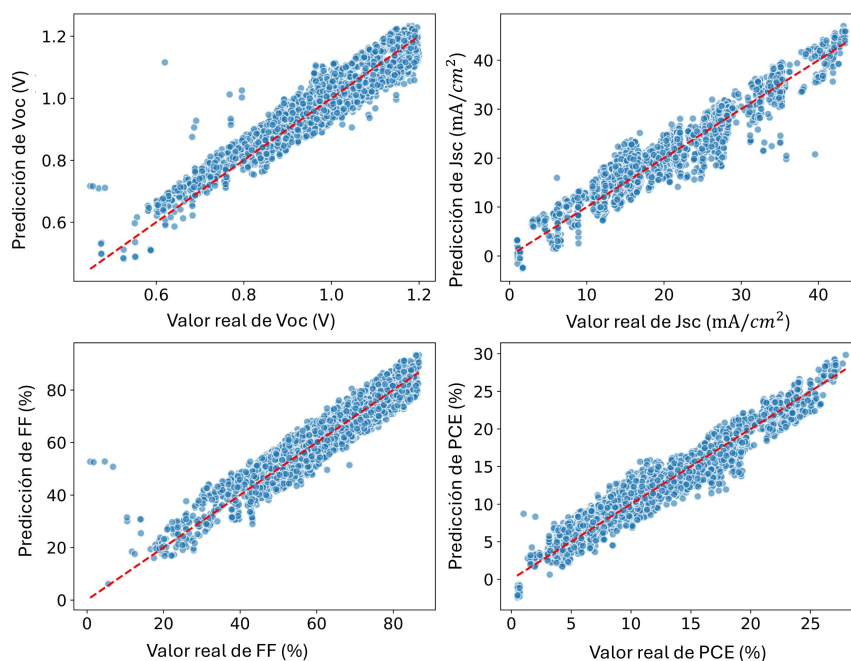


Figura 4-6: Comparación de los valores reales generados a partir de simulación y los datos predichos en el modelo de SVM.

Los resultados obtenidos revelan que los modelos de conjunto basados en árboles (RF y XGB) superan sistemáticamente a los enfoques basados en redes neuronales y en kernels. Ambos métodos de conjunto alcanzan los coeficientes de determinación más altos y los errores absolutos más bajos en la mayoría de las métricas fotovoltaicas, lo que confirma su capacidad superior para captar las relaciones no lineales y multivariantes inherentes al conjunto de datos simulados de células solares de perovskita. Por el contrario, ANN y MLP muestran un rendimiento intermedio, mientras que SVM muestra la fidelidad predictiva más baja para todos los resultados.

Para la predicción de Voc y Jsc, XGB alcanza los coeficientes de determinación más altos, 0,9774 y 0,9815, respectivamente. Además, presenta los valores RMSE más bajos para Voc y Jsc, alcanzando valores de 0,0219 V y 1,2913 mA/cm², respectivamente. Por lo tanto, es evidente que supera ligeramente al modelo RF. Esto indica que el refuerzo de gradiente es particularmente eficaz para modelar las variaciones en los procesos de generación y recolección de carga que dominan estos parámetros relacionados con la corriente y el voltaje. Por el contrario, el modelo RF demuestra un rendimiento superior en la predicción de FF, alcanzando el R² más alto (0,9967) y el RMSE (0,8978) y MAE (0,2909) más bajos. Esto sugiere que el conjunto basado en bagging, al reducir la varianza mediante el promedio de múltiples árboles decorrelacionados, proporciona predicciones más estables para el FF, que está fuertemente influenciado por las sutiles interacciones entre la recombinación, el trans-

Tabla 4-1: Métricas de rendimiento para cada modelo de regresión.

Modelo	Métrica	Voc (V)	Jsc (mA/cm ²)	FF (%)	PCE (%)
RF	R^2	0.9722	0.9780	0.9967	0.9990
	$RMSE$	0.0244	1.4100	0.8978	0.6307
	MAE	0.0084	0.4843	0.2909	0.2338
XGB	R^2	0.9774	0.9815	0.9856	0.9976
	$RMSE$	0.0219	1.2913	1.0422	0.7017
	MAE	0.01	0.5639	0.4864	0.3704
ANN	R^2	0.9538	0.9637	0.9714	0.9761
	$RMSE$	0.0315	1.8113	2.6611	0.9745
	MAE	0.02	1.2385	1.7421	0.7108
MLP	R^2	0.9076	0.9280	0.9118	0.9487
	$RMSE$	0.0445	2.5505	4.6771	1.4270
	MAE	0.0317	1.7891	3.2500	1.0516
SVM	R^2	0.8853	0.9170	0.9055	0.9319
	$RMSE$	0.4956	2.7376	4.8405	1.6440
	MAE	0.0409	2.1774	3.8233	1.3672

porte y las pérdidas interfaciales. Para el PCE, el RF vuelve a producir los errores absolutos más bajos ($RMSE = 0,6307$, $MAE = 0,2338$), mientras que el XGB alcanza el R^2 más alto (0,9976), lo que indica una calibración ligeramente más ajustada, pero con desviaciones medias marginalmente mayores.

Los modelos de redes neuronales muestran una precisión predictiva reducida en comparación con los métodos de conjunto. Aunque la ANN alcanza valores R^2 relativamente altos (hasta 0,9761 para PCE), presenta un $RMSE$ y un MAE notablemente más altos, especialmente para FF y PCE. Del mismo modo, el modelo SVM tiene el peor rendimiento de todos los modelos empleados. De hecho, este modelo arrojó los valores R^2 más bajos y el mayor número de errores, lo que demuestra que la técnica basada en el núcleo es menos competitiva en este problema de regresión de alta dimensión. Los conjuntos basados en árboles son robustos frente a la colinealidad de las características, los tipos de variables mixtas y el ruido, que son intrínsecos a los conjuntos de datos PSC basados en simulaciones. Además, RF se beneficia de la reducción de la varianza mediante el bagging, mientras que XGB corrige incrementalmente los errores residuales, lo que permite a ambos modelos aprender las sutiles dependencias entre los parámetros geométricos, electrónicos e interfaciales que rigen el rendimiento de PSC. Estas características hacen que los métodos basados en árboles sean especialmente adecuados para el modelado sustitutivo de los resultados de las simulaciones

de deriva-difusión.

Los resultados muestran que los modelos de conjunto basados en árboles presentan la mejor concordancia entre los valores previstos y simulados para todos los indicadores fotovoltaicos probados. En realidad, los gráficos de paridad de los modelos RF y XGB revelan una densa agrupación de puntos de datos alrededor de la línea diagonal ideal, lo que demuestra una alta precisión y calibración en todo el rango de valores simulados. Esto demuestra que estos dos modelos pueden captar dependencias no lineales sin sobreestimar o subestimar constantemente determinados regímenes de rendimiento.

Para Voc y Jsc, XGB muestra una alineación ligeramente más ajustada a lo largo de la diagonal, especialmente en valores intermedios y altos, lo que concuerda con su mayor R^2 y menor RMSE presentados en la tabla 4-1. Este comportamiento demuestra la utilidad del refuerzo de gradiente para eliminar sucesivamente los errores residuales e imitar pequeñas variaciones en los procesos de generación y recolección de carga. Aunque el modelo RF es ligeramente menos preciso en estos dos criterios, mantiene un patrón de dispersión bastante uniforme, lo que demuestra su durabilidad y resistencia al sobreajuste mediante la reducción de la varianza. En comparación con FF y PCE, el modelo RF presenta la distribución de puntos más compacta y homogénea alrededor de la línea de referencia, con menos desviaciones extremas. Esta tendencia visual es coherente con los errores absolutos más bajos (RMSE y MAE) para estas métricas, lo que indica un rendimiento predictivo más estable en regímenes en los que las pérdidas por recombinación, las restricciones de transporte y los efectos de interfaz interactúan intensamente. XGB alcanza valores altos de R^2 , pero tiene una dispersión significativamente mayor en los extremos, lo que indica una mayor sensibilidad a las fluctuaciones localizadas en el espacio de parámetros de entrada.

Por otro lado, la ANN implementada en Keras y la MLP de scikit-learn muestran un rendimiento sólido, aunque con una precisión menor. Los resultados muestran una mayor dispersión, lo que indica una mayor varianza en los errores de predicción. Como aproximadores de funciones universales, estos modelos de redes neuronales han aprendido la relación subyacente; sin embargo, la mayor dispersión sugiere que el volumen de datos puede ser insuficiente para que converjan en una solución tan precisa como la de los conjuntos.

Por último, el modelo de regresión SVM presenta la menor fidelidad predictiva de los cinco modelos evaluados. Su diagrama de dispersión muestra una mayor dispersión de los datos alrededor de la línea de referencia, lo que indica un ajuste subóptimo. Se implementó un núcleo RBF para evaluar el comportamiento de este modelo; sin embargo, el resultado visual sugiere que no fue capaz de captar la complejidad de la función subyacente con la misma eficacia. Aunque los modelos SVM son eficientes en cuanto a memoria y eficaces en espacios de alta dimensión, para este problema de regresión específico son menos competitivos. Por

lo tanto, el análisis comparativo establece una jerarquía de rendimiento clara, con resultados similares para XGB y RF, mientras que los de ANN, MLP y SVM son inferiores.

4.4. Importancia de las características e interpretación física

La importancia de cada característica celular se evaluó utilizando la importancia de permutación para identificar los parámetros más influyentes que rigen cada métrica fotovoltaica (Voc, Jsc, FF y PCE). La figura 4-7 presenta las cinco características más relevantes para cada salida, lo que permite una interpretación física diferenciada de los mecanismos que dominan la generación de tensión, la extracción de corriente, las pérdidas del factor de llenado y la eficiencia general. Este análisis de múltiples salidas establece una jerarquía de sensibilidad que es fundamental para la optimización racional de las arquitecturas de las células solares de perovskita dentro del espacio de parámetros explorado.

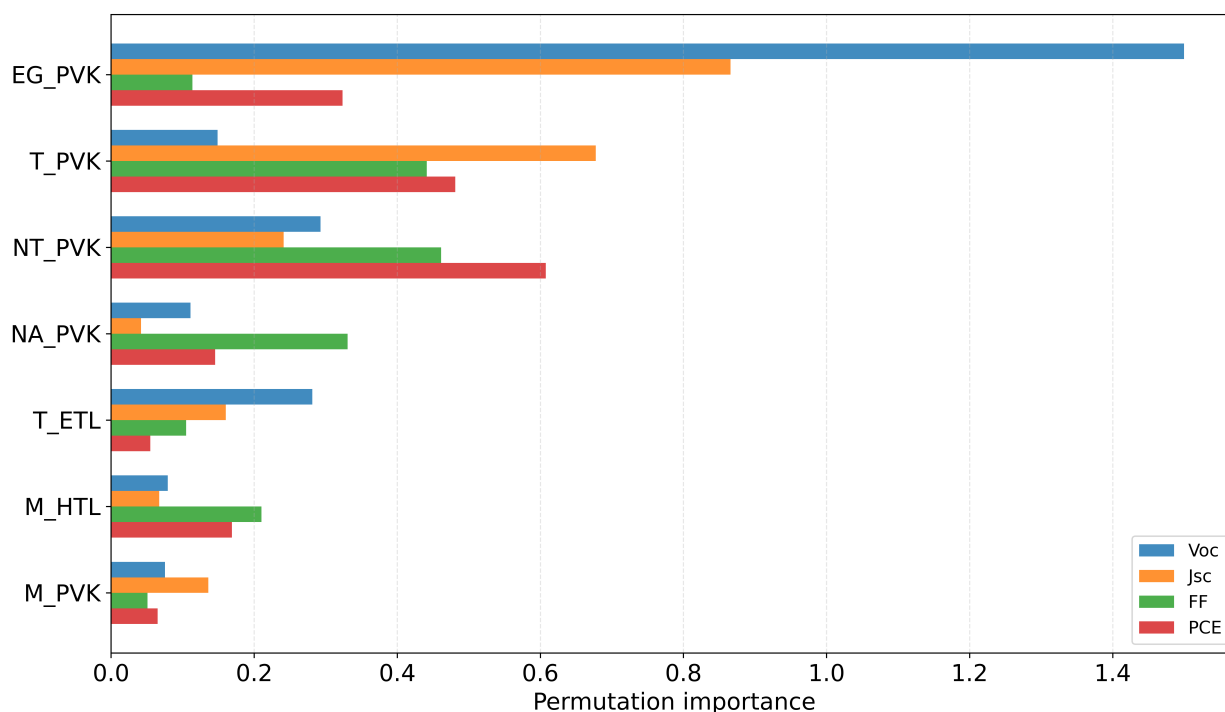


Figura 4-7: Importancia de la permutación de las características de entrada más influyentes para predecir Voc, Jsc, FF y PCE con RF.

Para Voc, la banda prohibida de la perovskita (EG_PVK) emerge claramente como el parámetro dominante, seguido por la densidad de defectos de la perovskita (NT_PVK) y el espesor

de la ETL (T_ETL). Este resultado es físicamente coherente, ya que Voc está limitado por la banda prohibida a través de la corriente de recombinación radiativa, mientras que la recombinación no radiativa de Shockley-Read-Hall está relacionada con los defectos volumétricos e interfaciales, reduce la división del nivel cuasi-Fermi y, por lo tanto, disminuye Voc. El espesor modula este equilibrio controlando la longitud de recolección de portadores y la probabilidad de recombinación, lo que explica su influencia secundaria pero no desdeñable.

Para Jsc, la característica más influyente es también la banda prohibida de la perovskita (EG_PVK), seguida del espesor de la perovskita (T_PVK) y la densidad de defectos de la perovskita (NT_PVK). El espesor controla directamente la absorción óptica y la fotogeneración, mientras que la banda prohibida establece el borde de absorción y la superposición espectral con el espectro AM1.5G. La densidad de defectos afecta aún más a Jsc al limitar la longitud de difusión del portador y aumentar las pérdidas por recombinación masiva. Esta jerarquía confirma que Jsc está controlado principalmente por la absorción óptica y las restricciones del transporte de portadores, más que por las pérdidas relacionadas con el contacto.

Por el contrario, la FF está dominada principalmente por la densidad total de defectos en la perovskita (NT_PVK), seguida del espesor de la perovskita (T_PVK) y la densidad de aceptores (NA_PVK). Esto refleja la fuerte sensibilidad de la FF a la cinética de recombinación, las resistencias en serie y en derivación y las barreras de transporte interfaciales. Los defectos aumentan la recombinación no radiativa y las corrientes de fuga, distorsionando el comportamiento del diodo bajo polarización directa. Mientras tanto, el espesor de la perovskita y la densidad de dopaje regulan el campo eléctrico interno, el ancho de agotamiento y la eficiencia de extracción de portadores, todos los cuales determinan de manera crítica la cuadratura de la curva J-V.

Para el PCE, que integra los efectos combinados de Voc, Jsc y FF, el parámetro dominante es la densidad de defectos de la perovskita (NT_PVK), seguido del espesor de la perovskita (T_PVK) y, a continuación, la banda prohibida de la perovskita (EG_PVK). Esto confirma que la recombinación no radiativa constituye el principal obstáculo para el rendimiento, mientras que los efectos de absorción geométrica y las compensaciones eléctricas espectrales definen el rendimiento del PSC. La aparición de estos tres parámetros en todos los resultados pone de relieve que la optimización de la eficiencia se rige fundamentalmente por la interacción entre la recombinación asistida por defectos, la generación y el transporte dependientes del espesor de la perovskita y la utilización espectral controlada por la banda prohibida.

La aparición de características relacionadas con la capa de transporte (como T_ETL y M_HTL) entre los principales factores contribuyentes, aunque secundarios, refuerza la importancia de los contactos selectivos y la ingeniería de interfaces. Estos parámetros modulan la alineación de bandas, las barreras de extracción y las velocidades de recombinación in-

terfacial, lo que influye indirectamente en todas las métricas fotovoltaicas. En general, el análisis de importancia de las características de salida múltiple indica que la mejora del rendimiento debe dar prioridad a la reducción de la densidad de defectos mediante la pasivación masiva e interfacial, la optimización del espesor dentro del régimen de alta absorción y una cuidadosa ingeniería de la banda prohibida para equilibrar Voc y Jsc, complementada con la optimización de la capa de transporte para suprimir las pérdidas de contacto y maximizar FF.

4.5. Validación externa mediante datos simulados y experimentales

Para evaluar la capacidad de generalización del marco propuesto más allá del dominio sintético, se llevó a cabo una validación externa utilizando el modelo RF. Esta selección está directamente respaldada por los resultados comparativos presentados en la sección 4.3, donde RF logró de manera consistente los errores medios más bajos en todas las métricas fotovoltaicas, particularmente para FF y PCE, que son los indicadores más sensibles de la calidad del dispositivo. Aunque XGB superó ligeramente a RF en términos de R^2 para Voc y Jsc, sus errores absolutos ligeramente superiores para FF y PCE indican predicciones menos estables en regímenes dominados por pérdidas por recombinación y transporte.

El modelo RF optimizado se aplicó a un conjunto de configuraciones de células solares de perovskita descritas en fuentes bibliográficas independientes, incluyendo tanto estudios puramente numéricos como trabajos que combinaban simulación y caracterización experimental (marcados con * en la tabla 4-2). El modelo se alimentó con configuraciones no vistas durante el entrenamiento, limitadas a arquitecturas y pilas de materiales coherentes con el alcance del conjunto de datos (estructuras ETL/perovskita/HTL), y los valores Voc, Jsc, FF y PCE previstos se contrastaron con los valores publicados.

Nota: El error relativo se calcula como: $\left| \frac{y_{\text{predicho}} - y_{\text{reporte}}}{y_{\text{reporte}}} \right| \times 100$.

En la Tabla 4-3 se presentan los errores relativos entre los valores reportados y generados a partir del modelo entrenado para analizar el desempeño del modelo RF con validación externa. Donde se observa que se presentan errores moderados, aunque la magnitud y dispersión de estos dependen de la variable analizada. Para Voc, los errores se distribuyen en un intervalo relativamente acotado, con valor mínimo de 3.92%, y un máximo de 17.37%. Este comportamiento sugiere que el modelo reproduce de manera razonable la tendencia del voltaje de circuito abierto, aunque existen casos particulares en los que la diferencia aumenta, efecto explicado por pérdidas por recombinación, alineación de bandas o condiciones específicas de simulación y fabricación no capturadas completamente en el conjunto

Tabla 4-2: Comparación entre valores reportados y predichos para validación externa.

Referencia	Reporte				Predicción			
	$Voc(V)$	$Jsc(mA/cm^2)$	$FF(\%)$	$PCE(\%)$	$Voc(V)$	$Jsc(mA/cm^2)$	$FF(\%)$	$PCE(\%)$
[163]	1.2	22.6	77	21.4	1.15	25.23	85.69	24.9
[164]	1.02	20.86	69	14.75	1.06	21.82	74.65	17.39
[165]*	0.93	20.44	57.62	11	0.89	23.76	49.27	10.44
[166]*	1.16	23.2	75.43	20.3	0.97	24.67	58.1	14.09
[167]*	1	18.62	61.8	11.49	0.94	25.23	49.82	11.83
[168]	0.764	25.85	67.29	13.3	0.856	32.779	62.4	15.66
[169]	0.93	29.09	86.07	20.95	1.0915	25.35	83.58	23.13
[170]	1.12	24.98	81.68	23.04	1.01	25.46	75.67	21.15

Nota: Las referencias marcadas con * corresponden a análisis experimentales y simulados.

de entrenamiento. Posteriormente, para Jsc , la variabilidad es mayor. Se identifican errores bajos, como 1.92 %, pero también valores como 35.50 %. Esta mayor dispersión indica que la predicción de la densidad de corriente de cortocircuito es la más sensible a diferencias entre estudios, lo cual es físicamente consistente con la fuerte dependencia de Jsc respecto a fenómenos de absorción óptica, generación y colección de portadores, recombinación interfacial y variaciones estructurales del dispositivo. Luego, para FF , los errores relativos muestran una dispersión intermedia con error mínimo de 2.89 % y máximo de 22.97 %. Esto muestra que, aunque el modelo logra capturar adecuadamente la tendencia global del FF , existen configuraciones en las que las pérdidas resistivas, la recombinación o la calidad de interfaces generan discrepancias más pronunciadas respecto a los valores reportados en la literatura.

Tabla 4-3: Errores relativos de los valores predichos y los valores reportados para Voc , Jsc , FF y PCE .

Referencia	$Voc(\%)$	$Jsc(\%)$	$FF(\%)$	$PCE(\%)$
[163]	4.17	11.64	11.29	16.36
[164]	3.92	4.60	8.19	17.90
[165]*	4.30	16.24	14.49	5.09
[166]*	16.38	6.34	22.97	30.59
[167]*	6.00	35.50	19.39	2.96
[168]	12.04	26.80	7.27	17.74
[169]	17.37	12.86	2.89	10.41
[170]	9.82	1.92	7.36	8.20

Finalmente, para PCE, los errores reflejan el efecto acumulado de las desviaciones en Voc, Jsc y FF. Se observa un valor mínimo de 2.96 % y máximo de 30.59 %. Este comportamiento tiene sentido, dado que pequeñas discrepancias simultáneas en varias métricas pueden amplificarse al calcular la eficiencia global del dispositivo.

En todas las referencias, el modelo RF reprodujo las tendencias observadas con desviaciones relativas moderadas. Los errores relativos medios fueron de $9,25 \pm 5,53$ % para Voc, $14,49 \pm 11,52$ % para Jsc, $11,73 \pm 6,79$ % para FF y $13,66 \pm 8,92$ % para PCE, donde el valor reportado después del símbolo \pm corresponde a la desviación estándar. Los errores comparativamente menores observados para Voc y FF reflejan la fuerte dependencia de estas métricas con respecto a la banda prohibida, la densidad de defectos y los parámetros de recombinación interfacial, que están codificados explícitamente en las simulaciones SCAPS-1D y capturados eficazmente por la estructura basada en conjuntos de RF. Por el contrario, se esperan desviaciones mayores en Jsc, ya que este parámetro es más sensible a la gestión óptica, la absorción dependiente del espesor y las condiciones experimentales de acoplamiento de la luz, que solo están representadas parcialmente en los modelos de dispositivos unidimensionales.

Es importante destacar que los casos que involucran datos experimentales muestran niveles de error comparables a los de los estudios puramente numéricos, lo que indica que el modelo RF no se ajusta excesivamente a los artefactos específicos de la simulación, sino que aprende relaciones físicamente significativas que se pueden transferir a condiciones reales de los dispositivos. Este resultado respalda la capacidad de generalización del enfoque propuesto y valida su uso como herramienta predictiva y de diagnóstico para evaluar arquitecturas de células solares de perovskita tanto en contextos teóricos como experimentales.

4.5.1. Diseño inverso para la selección de una configuración óptima a partir de predicciones del modelo Random Forest

Más allá de evaluar el desempeño predictivo del modelo, un objetivo práctico en el desarrollo de las PSC es traducir el modelo directo $\mathbf{x} \mapsto \hat{\mathbf{y}}$ (parámetros de entrada \rightarrow métricas de desempeño) en una herramienta de diseño inverso que entregue, de forma interpretable, un conjunto de parámetros de diseño que den como resultado un dispositivo con la mayor eficiencia energética posible. Este enfoque es relevante porque transforma miles de combinaciones simuladas en una configuración concreta de materiales y parámetros, lo cual sirve como punto de partida reproducible para que se lleve al ámbito experimental y se facilite el proceso de selección de materiales, espesores y densidades de dopaje [171, 172].

En el presente trabajo, el diseño inverso se implementa como un proceso de búsqueda guiada por el modelo sobre el conjunto de predicciones. Puesto que para cada vector de entrada \mathbf{x} (19

variables geométricas y fotoeléctricas), el Random Forest entrenado predice simultáneamente Voc, Jsc, FF y PCE. Posteriormente, dichas predicciones se reescalan a unidades físicas (se invierte el proceso de MinMaxScaler para retomar las dimensiones originales) y se selecciona el óptimo maximizando la eficiencia predicha, imponiendo a la vez un umbral mínimo de FF para evitar soluciones con pérdidas resistivas o por recombinación elevadas [173]. Siguiendo la formulación estándar de optimización restringida, el diseño inverso puede expresarse como la selección del conjunto de valores que maximiza la eficiencia predicha bajo una restricción mínima de factor de llenado:

$$\mathbf{x}^* = \arg \max_{\mathbf{x} \in \mathcal{X}_{\text{cand}}} \widehat{PCE}(\mathbf{x}) \quad \text{sujeto a} \quad \widehat{FF}(\mathbf{x}) \geq 0,80, \quad (4-1)$$

donde $\mathcal{X}_{\text{cand}}$ es el conjunto finito de configuraciones evaluadas. Esta formulación corresponde a un caso particular del método por restricciones, ampliamente usado para imponer requisitos mínimos de desempeño durante la optimización [174–176]. Un aspecto clave para la utilidad experimental es que, tras identificar \mathbf{x}^* en el espacio normalizado, se aplica la transformación inversa de las variables de entrada y la decodificación de materiales (reversión del OneHotEncoder) para recuperar directamente los nombres de ETL/PVK/HTL y los valores numéricos en sus unidades originales.

Con este procedimiento, el mejor compromiso encontrado alcanzó un máximo PCE predicho de 28 % con FF correspondiente de 84.78 %. La configuración recomendada (vector \mathbf{x}^*) quedó definida por la arquitectura SnO₂/FAPbI₃/Spiro-OMeTAD y los siguientes parámetros:

- T_ETL=0.2 μm
- T_PVK=1.2 μm
- T_HTL=0.1 μm
- EG_ETL=3.5 eV,
- EG_PVK=1.3 eV,
- EG_HTL=2.9 eV;
- ER_ETL=9.0,
- ER_PVK=13.0,
- ER_HTL=3.0;
- ND_ETL=1.0×10¹⁵ cm⁻³,
- NA_PVK=1.0×10¹⁷ cm⁻³,

- ND_PVK= $1.3 \times 10^{16} \text{ cm}^{-3}$,
- NA_HTL= $1.3 \times 10^{18} \text{ cm}^{-3}$,
- NT_ETL= $1.0 \times 10^{18} \text{ cm}^{-3}$,
- NT_PVK= $4.0 \times 10^{13} \text{ cm}^{-3}$,
- NT_HTL= $1.0 \times 10^{15} \text{ cm}^{-3}$.

Desde el punto de vista físico, la selección de SnO₂ como ETL es coherente con su uso extendido en PSC de alto desempeño por su adecuada alineación de bandas, alta movilidad electrónica y posibilidad de reducir recombinación interfacial cuando se optimizan interfaces y defectos [177]. De forma análoga, Spiro-OMeTAD se mantiene como HTL de referencia en arquitecturas eficientes por su compatibilidad de niveles energéticos y procesabilidad, aunque su desempeño depende fuertemente de su dopaje y del control de pérdidas resistivas [178]. En cuanto a la perovskita, FAPbI₃ es un material ampliamente reportado en PSC de alta eficiencia. Además, el EG_PVK recomendado se sitúa cerca del rango óptimo teórico para una unión simple bajo el espectro solar AM 1.5G, lo que respalda su potencial de eficiencia elevada cuando se minimizan pérdidas no radiativas [179, 180]. Finalmente, las bajas densidades de defectos predichas, en especial NT_PVK, son consistentes con la necesidad de suprimir recombinación Shockley-Read-Hall para sostener simultáneamente Voc alto y FF elevado [181, 182].

Con el fin de verificar la consistencia de la configuración óptima obtenida mediante el esquema de diseño inverso, se realizó su simulación en SCAPS-1D empleando los parámetros recomendados por el modelo Random Forest. La simulación de validación arrojó un FF de 84.61 % y un PCE de 27.94 %. Estos resultados muestran concordancia con la predicción del modelo inverso, el cual había estimado una PCE de 28 % con un FF de 84.78 %.

Tabla 4-4: Comparativa entre FF y PCE predichas por el modelo de diseño inverso basado en Random Forest y los valores obtenidos mediante la simulación de validación en SCAPS-1D para la configuración óptima.

Métrica	Estructura Óptima	Simulación	Error absoluto	Error relativo
<i>FF</i> (%)	84.78	84.61	0.17	0.20
<i>PCE</i> (%)	28.0	27.94	0.06	0.21

Nota: El error absoluto se calculó como $|y_{\text{SCAPS}} - y_{\text{RF}}|$, mientras que el error relativo porcentual se estimó como $\left(\frac{|y_{\text{SCAPS}} - y_{\text{RF}}|}{y_{\text{RF}}}\right) \times 100$.

La Tabla 4-4 compara las métricas reportadas por el diseño inverso basado en Random Forest con los valores obtenidos al simular en SCAPS-1D la configuración óptima. En este caso, la validación se centró en el factor de llenado y la eficiencia de conversión de potencia, dado que estas fueron las métricas utilizadas para identificar la solución óptima dentro del espacio de búsqueda.

En síntesis, las diferencias fueron de 0.06 puntos porcentuales para la PCE y 0.17 puntos porcentuales para el FF, equivalentes a errores relativos de 0.21 % y 0.20 %, respectivamente. En conjunto, esta cercanía entre la predicción basada en aprendizaje de máquinas y la simulación física confirma que el diseño inverso no solo identifica configuraciones matemáticamente óptimas dentro del espacio de búsqueda, sino también soluciones físicamente coherentes y reproducibles desde el punto de vista del modelamiento computacional del dispositivo.

En conjunto, este diseño inverso no reemplaza la validación experimental, pero sí entrega un prototipo óptimo y físicamente viable que reduce el espacio de búsqueda y sienta una base clara para la fabricación y la posterior calibración del dispositivo en entornos experimentales [171].

5 Conclusiones

Se consolidó un flujo de modelamiento numérico para PSC que permitió estudiar la respuesta eléctrica bajo condiciones estándar a partir de la variación sistemática de parámetros geométricos y fotoeléctricos y de la selección de materiales en la arquitectura del dispositivo. Esta metodología computacional dio lugar a la construcción de un conjunto de datos consistente y trazable, adecuado para análisis estadístico y para el entrenamiento de modelos predictivos, preservando la reproducibilidad mediante el registro explícito de variables de entrada y salidas (Voc, Jsc, FF y PCE). Se desarrollaron cinco modelos de regresión para predecir las cuatro métricas, garantizando condiciones comparables entre modelos, donde se encontró que los métodos de ensamble basados en árboles (Random Forest y XGBoost) alcanzan el mayor poder predictivo global, destacándose por altos coeficientes de determinación y errores reducidos de RMSE y MAE, mientras que aproximadores neuronales (ANN/MLP) y SVM presentan desempeño competitivo pero, en promedio, inferior en las métricas más sensibles a efectos acoplados (particularmente PCE). En consecuencia, el algoritmo propuesto para predicción y análisis se sustenta en modelos de ensamble como primera elección para inferencia y para exploración de escenarios de diseño, por su robustez y estabilidad fuera de muestra. Posteriormente, la capacidad de generalización se verificó mediante validación externa frente a resultados independientes reportados en la literatura, evidenciando un acuerdo razonable en magnitudes y tendencias. En particular, los errores relativos promedio permanecen acotados para Voc y FF, mientras que Jsc y, por propagación, PCE muestran desviaciones mayores, consistentes con la alta sensibilidad de la corriente a mecanismos no modelados explícitamente en 1D (como por ejemplo rugosidad, deformidades morfológicas y resistencias de contacto). Este contraste confirma que el enfoque computacional y predictivo es útil para comparar diferentes configuraciones, pero también delimita con claridad el alcance físico del modelo y los escenarios donde se requiere calibración adicional con mediciones experimentales. Finalmente, el análisis de importancia de variables permitió identificar los parámetros con mayor incidencia sobre el desempeño, destacándose la densidad de defectos de la perovskita como el factor dominante, seguido del espesor de la perovskita. Además, variables como densidad de aceptores y bandgap adquieren relevancia diferenciada según la métrica objetivo. Estos hallazgos sustentan que la optimización efectiva de PSC no debe restringirse a ajustes geométricos, sino priorizar estrategias de pasivación y control de defectos, junto con una selección de espesor que maximice absorción sin penalizar recombinación, alineando así el diseño con los mecanismos limitantes del transporte y la separación de carga.

5.1. Limitaciones

Es importante mencionar que la base de datos se generó para PSC planas con una sola capa absorbente, lo que puede suponer limitaciones que deben tenerse en cuenta al utilizarla. Además, todas las simulaciones se generaron en condiciones constantes y unidimensionales, por lo que no se tienen en cuenta los efectos tridimensionales, la degradación a largo plazo ni las condiciones ambientales reales (temperatura, humedad, degradación de los materiales, etc.). Aunque se llevó a cabo una validación cruzada con los datos publicados en la bibliografía, puede haber discrepancias atribuibles a la falta de detalle en los parámetros comunicados por algunos autores, lo que impide una réplica exacta de los resultados experimentales.

5.2. Trabajo futuro

Para trabajos posteriores, se propone realizar simulaciones en 3D analizando los efectos de variaciones térmicas sobre las PSCs, evaluando la reflectancia, transmitancia, absorbancia y el comportamiento del campo eléctrico cuando se aplican perturbaciones térmicas y/o mecánicas. Además de integrar estimación de incertidumbre predictiva mediante métodos bayesianos o conformal prediction para reportar intervalos de confianza y apoyar decisiones de diseño bajo incertidumbre paramétrica.

Bibliografía

- [1] U. de Planeación Minero Energética (UPME), “Informe proyeccion demanda energeticos colombia,” tech. rep., 2022.
- [2] S. Fonash, *Solar Cell Device Physics*. Academic Press (Elsevier), 2010.
- [3] Y. Velez-Galvis, E. Gonzalez-Valencia, A. Sepulveda-Sepulveda, and N. Gomez-Cardona, “Synthetic dataset to study the performance of perovskite solar cell simulations,” *F1000Research*, vol. 14, no. 961, 2025.
- [4] J. Zhang and I. Khan, “Energy use, energy access, and oil price fluctuations as new determinants of environmental quality in apec countries,” *Gondwana Research*, vol. 132, pp. 309–322, 8 2024.
- [5] N. Unidas, “Paris Agreement Spanish,” tech. rep., 2015.
- [6] C. González, *La explosión solar en Colombia*. Taller de edición Rocca, 2023.
- [7] S. A. de Energías Renovables, “Proyectos de energías renovables 2023-2024,” tech. rep., 2023.
- [8] U. de Planeación Minero-Energética UPME, “Informe de registro de proyectos de generación,” tech. rep., Unidad de Planeación Minero-Energética UPME, 2023.
- [9] M. Llaverro-Pasquina and A. Bontempi, “Oil and gas industry’s marginal share of global renewable energy,” *Nature Sustainability*, pp. 1–5, 10 2025.
- [10] V. Ballesteros-Ballesteros, “Panorama mundial de las energías renovables e importancia de la energía fotovoltaica,” *Revista Científica*, 2016.
- [11] N. R. Flores Rivera and M. A. Domínguez Ramírez, “Medición de la eficiencia energética de los paneles solares de silicio,” tesis de maestría, Centro de Investigación en Materiales Avanzados (CIMAV), Chihuahua, Chihuahua, México, 2016.
- [12] J.-P. Correa-Baena, M. Saliba, T. Buonassisi, M. Grätzel, A. Abate, W. Tress, and A. Hagfeldt, “Promises and challenges of perovskite solar cells,” *Science*, vol. 358, no. 6364, pp. 739–744, 2017.

-
- [13] C. Rojas-Rincón, Y. Vélez-Galvis, M. Botero-Londoño, E. Reyes-Vera, and A. Sepúlveda, “Machine learning for the prediction of perovskite solar cell performance: A brief review,” in *Congress on Research, Development and Innovation in Renewable Energies*, pp. 15–24, Springer, Cham, 2025.
- [14] N. National Renewable Energy Laboratory, “Best research-cell efficiency chart — photovoltaic research — nrel,” *NREL*, 2025.
- [15] U. de Planeación Minero-Energética UPME, “Proyección de la demanda de energía eléctrica y potencia máxima 2023-2037,” tech. rep., Unidad de Planeación Minero-Energética UPME, 2023.
- [16] A. D. L. Vega, A. Garg, A. Hahmann, S. Khennas, I. Lima, A. Löschel, A. Singh, L. Steg, G. Strbac, and K. Wada, *Energy Systems*, pp. 613–746. Cambridge University Press, 8 2023.
- [17] T. Rahman, A. Mansur, M. H. Lipu, M. Rahman, R. Ashique, M. Houran, R. Elavarsan, and E. Hossain, “Investigation of degradation of solar photovoltaics: A review of aging factors, impacts, and future directions toward sustainable energy management,” *Energies*, vol. 16, p. 3706, 2023.
- [18] S. Almosni, A. Delamarre, Z. Jehl, D. Suchet, L. Cojocar, M. Giteau, and et al., “Material challenges for solar cells in the twenty-first century: directions in emerging technologies,” 12 2018.
- [19] M. Dhimish and P. I. Lazaridis, “An empirical investigation on the correlation between solar cell cracks and hotspots,” *Scientific Reports*, vol. 11, 12 2021.
- [20] J. Pastuszak and P. Wegierek, “Photovoltaic Cell Generations and Current Research Directions for Their Development,” *Materials*, vol. 15, 8 2022.
- [21] C. Odabasi and R. Yildirim, “Machine learning analysis on stability of perovskite solar cells,” *Solar Energy Materials and Solar Cells*, vol. 205, 2 2020.
- [22] A. Mahmood and J. L. Wang, “Machine learning for high performance organic solar cells: Current scenario and future prospects,” *Energy and Environmental Science*, vol. 14, pp. 90–105, 1 2021.
- [23] W. Yan, Y. Liu, Y. Zang, J. Cheng, Y. Wang, L. Chu, X. Tan, L. Liu, P. Zhou, W. Li, and Z. Zhong, “Machine learning enabled development of unexplored perovskite solar cells with high efficiency,” *Nano Energy*, vol. 99, 8 2022.
- [24] C. Castañeda, C. Hernández, S. Castro, B. Medina, O. López, J. Gómez, E. Reyes, and A. Sepúlveda, “Machine learning algorithms for predicting electrical performance of

- perovskite solar cells,” in *2024 IEEE Technology and Engineering Management Society (TEMSCON LATAM)*, pp. 1–6, 2024.
- [25] V. M. Le Corre, T. S. Sherkar, M. Koopmans, and L. J. A. Koster, “Identification of the dominant recombination process for perovskite solar cells based on machine learning,” *Cell Reports Physical Science*, vol. 2, p. 100346, 2 2021.
- [26] W. Yan, Y. Liu, Y. Zang, J. Cheng, Y. Wang, L. Chu, X. Tan, L. Liu, P. Zhou, W. Li, and Z. Zhong, “Machine learning enabled development of unexplored perovskite solar cells with high efficiency,” *Nano Energy*, vol. 99, p. 107394, 8 2022.
- [27] Z. Li, C. Huang, L. Chao, Y. Chen, W. Huang, and G. Chen, “Perovskite solar cells empowered by machine learning,” *Journal of Energy Chemistry*, vol. 109, pp. 403–437, 10 2025.
- [28] E. Nygren, J. C. McCallum, R. Thornton, J. S. Williams, and G. L. Olson, “Amorphous to Polycrystalline Transformation in High Dose Ion Implanted Silicon,” *MRS Proceedings*, vol. 100, pp. 405–409, 5 1988.
- [29] A. Reinders, P. Verlinden, W. Van Sark, and A. Freundlich, *Photovoltaic Solar Energy: From Fundamentals to Applications*. Wiley, 12 2016.
- [30] L. M. Fraas and L. D. Partain, *Solar Cells and Their Applications*. John Wiley & Sons, 2 ed., 2010.
- [31] S. A. Md and R. Ahasan, “Effects of solar energy use on rural community: a study of Boyarjapha village in Paikgachha upazila,” *3rd International Conference on Advances in Civil Engineering, 21-23 December 2016, CUET, Chittagong, Bangladesh*, 2020.
- [32] S. A. Hashemi, S. Ramakrishna, and A. G. Aberle, “Recent progress in flexible–wearable solar cells for self-powered electronic devices,” *Energy & Environmental Science*, vol. 13, pp. 685–743, 3 2020.
- [33] P. K. Nayak, S. Mahesh, H. J. Snaith, and D. Cahen, “Photovoltaic solar cell technologies: analysing the state of the art,” *Nature Reviews Materials 2019 4:4*, vol. 4, pp. 269–285, 3 2019.
- [34] J. Zhang, W. Zhang, H. M. Cheng, and S. R. P. Silva, “Critical review of recent progress of flexible perovskite solar cells,” *Materials Today*, vol. 39, pp. 66–88, 10 2020.
- [35] K. Sanglee, M. Nukunodompanich, F. Part, C. Zafiu, G. Bello, E. K. Ehmoser, and S. Chuangchote, “The current state of the art in internal additive materials and quantum dots for improving efficiency and stability against humidity in perovskite solar cells,” *Heliyon*, vol. 8, 12 2022.

- [36] S. W. Sharshir, A. A. El-Naggar, H. A. Ismail, M. M. Sami, L. A. Lotfy, M. R. Sharaby, H. Zhao, S.-H. Jang, A. A. Alsakran, N. S. A. EL-Gawaad, M. Ismail, E. Abdelnasser, A. El-Shaer, and M. Rashad, “Degradation mechanisms and stability challenges in perovskite solar cells: a comprehensive review,” *Solar Energy*, vol. 299, p. 113707, 10 2025.
- [37] H. Pan, Y. Bai, K. Sun, M. Yang, R. Tian, Y. Meng, J. Gao, Y. Wang, J. Wang, S. Zhou, Z. Song, L. Xiaoyi, C. Liu, and Z. Ge, “Decoupling photothermal-mechanical degradation through lattice-stabilizing networks in sn–pb perovskites and all-perovskite tandem solar cells,” *Nature Communications 2025 16:1*, vol. 16, pp. 1–12, 10 2025.
- [38] M. H. Miah, M. B. Rahman, M. Nur-E-Alam, M. A. Islam, M. Shahinuzzaman, M. R. Rahman, M. H. Ullah, and M. U. Khandaker, “Key degradation mechanisms of perovskite solar cells and strategies for enhanced stability: issues and prospects,” *RSC Advances*, vol. 15, pp. 628–654, 1 2025.
- [39] S. R. Wenham and M. A. Green, “Silicon solar cells,” *Progress in Photovoltaics: Research and Applications*, vol. 4, no. 1, pp. 3–33, 1996.
- [40] H. E. Mancera Hernández, “Desarrollo de celdas solares de perovskita $\text{CH}_3\text{NH}_3\text{PbI}_3$ con capa mesoporosa de TiO_2 ,” tech. rep., Universidad de los Andes, 2016.
- [41] E. Ramírez, R. Betancur, J. F. Montoya, E. Velilla, D. Ramírez, and F. Jaramillo, “Encapsulation against Extrinsic Degradation Factors and Stability Testing of Perovskite Solar Cells,” in *Recent Advances in Multifunctional Perovskite Materials*, IntechOpen, 12 2022.
- [42] M. I. Hossain, Y. Tong, and B. Aissa, “Analyzing the degradation mechanisms in perovskite solar cells over a span of three years,” in *2025 IEEE 53rd Photovoltaic Specialists Conference (PVSC)*, pp. 0065–0067, IEEE, 6 2025.
- [43] Y. Zhao, Q. Ye, Z. Chu, F. Gao, X. Zhang, and J. You, “Recent Progress in High-efficiency Planar-structure Perovskite Solar Cells,” *Energy and Environmental Materials*, vol. 2, pp. 93–106, 6 2019.
- [44] M. Green, E. Dunlop, M. Yoshita, N. Kopidakis, G. Siefer, X. Hao, and J. Jiang, “Solar cell efficiency tables (version 66),” *Progress in Photovoltaics: Research and Applications*, vol. 33, pp. 795–810, 05 2025.
- [45] S. A. A. Shah, M. H. Sayyad, K. Khan, K. Guo, F. Shen, J. Sun, A. K. Tareen, Y. Gong, and Z. Guo, “Progress towards high-efficiency and stable tin-based perovskite solar cells,” *Energies*, vol. 13, 10 2020.

- [46] A. R. Pascoe, M. Yang, N. Kopidakis, K. Zhu, M. O. Reese, G. Rumbles, M. Fekete, N. W. Duffy, and Y. B. Cheng, “Planar versus mesoscopic perovskite microstructures: The influence of $\text{CH}_3\text{NH}_3\text{PbI}_3$ morphology on charge transport and recombination dynamics,” *Nano Energy*, vol. 22, pp. 439–452, 4 2016.
- [47] D. Cao, W. Li, X. Zhang, L. Wan, Z. Guo, X. Wang, D. Eder, and S. Wang, “Current state-of-the-art characterization methods for probing defect passivation towards efficient perovskite solar cells,” *Journal of Materials Chemistry A*, vol. 10, pp. 19278–19303, 6 2022.
- [48] T. Ouslimane, L. Et-taya, L. Elmaimouni, and A. Benami, “Impact of absorber layer thickness, defect density, and operating temperature on the performance of MAPbI_3 solar cells based on ZnO electron transporting material,” *Heliyon*, vol. 7, p. e06379, 3 2021.
- [49] W. Zhang, L. Huang, W. Zheng, S. Zhou, X. Hu, J. Zhou, J. Li, J. Liang, W. Ke, and G. Fang, “Revealing key factors of efficient narrow-bandgap mixed lead-tin perovskite solar cells via numerical simulations and experiments,” *Nano Energy*, vol. 96, 6 2022.
- [50] J. Nelson, “The Physics of Solar Cells,” *The Physics of Solar Cells*, 5 2003.
- [51] J. Vélez, F. Sepúlveda, M. Botero, C. Otalora, and C. Camacho, “Absorber layer thickness as a new feature in statistical learning tools of Perovskite solar cells,” *Journal of Applied Research and Technology*, vol. 21, pp. 858–865, 2023.
- [52] S. Ahmad, S. Kazim, and M. Grätzel, “Perovskite Solar Cells: Materials, Processes, and Devices,” *Perovskite Solar Cells: Materials, Processes, and Devices*, pp. 1–559, 1 2021.
- [53] M. Z. Jiménez-Díaz, A. Sepúlveda-Sepúlveda, and M. Botero-Londoño, “Named entity recognition for performance and synthesis information of perovskite solar cells using spacy,” in *Computational Science and Its Applications – ICCSA 2025* (O. Gervasi, B. Murgante, C. Garau, Y. Karaca, D. Taniar, A. M. A. C. Rocha, and B. O. Apduhan, eds.), (Cham), p. 303–319, Springer Nature Switzerland, 2025.
- [54] C. Castañeda, C. Hernández, S. A. C. Casadiego, B. M. Delgado, O. López, D. Guevara, E. R. Vera, and F. A. Sepúlveda, “Performance of ABC3-structured perovskite solar cells: Impact of composition and deposition method,” in *2025 IEEE Colombian Caribbean Conference (C3)*, IEEE, Sept. 2025.
- [55] C. L. R. Rincón, M. A. B. Londoño, and F. A. S. Sepúlveda, “Statistical comparison between scaps based simulations and pce values of synthesized perovskite solar cells,” in *International Conference on Hybrid and Organic Photovoltaics (HOPV25)*, 2025.

- [56] A. Velez Ramirez, “Síntesis de una dispersión coloidal de nanopartículas de SnO_2 para el procesamiento de celdas solares de perovskita en solución,” trabajo de grado profesional, Universidad de Antioquia, Medellín, Colombia, 2024.
- [57] C. A. Valderrama Benítez, F. J. Molina Pérez, J. F. Montoya Arango, D. C. Rodríguez, A. L. Villa, and J.-A. Alberola-Borràs, “Evaluation of environmental impacts of a perovskite solar cell with a p-i-n meso-superstructured architecture through life cycle assessment,” *Sustainable Energy & Fuels*, vol. 9, pp. 2753–2768, 2025.
- [58] J. J. Patiño López, K. G. Rivera Botia, K. Ballestas, E. Velilla, J. F. Montoya, F. Jaramillo, and D. Ramirez, “The long road to outdoor stability: real-world challenges for controlling perovskite materials for solar cells,” *Chemical Society Reviews*, 2026.
- [59] J. Vélez, M. A. Botero L., and A. Sepulveda, “Measurement of information content of perovskite solar cell’s synthesis descriptors related to performance parameters,” *Emergent Materials*, vol. 7, pp. 1961–1968, 2024.
- [60] J. D. Sierra, C. A. Hernández-Salazar, and O. A. González-Estrada, “Artificial intelligence-based evaluation of perovskite coatings for enhanced solar cell performance,” *Journal of Physics: Conference Series*, vol. 3164, no. 1, p. 012010, 2025.
- [61] C. A. Hernández Pérez *et al.*, “Caracterización de celdas solares de perovskita mediante un análisis exploratorio de datos de variables eléctricas, band gap y área,” *Prospectiva*, vol. 22, no. 2, 2024.
- [62] C. A. Hernández Pérez, C. O. Castañeda Pico, O. A. López Bustamante, S. A. Castro Casadiego, and B. Medina Delgado, “Análisis del impacto ambiental de diferentes estructuras moleculares en células solares de perovskita,” *Revista Ambiental Agua, Aire y Suelo*, vol. 15, no. 2, pp. 21–28, 2024.
- [63] F. Liu, J. Zhu, J. Wei, Y. Li, M. Lv, S. Yang, B. Zhang, J. Yao, and S. Dai, “Numerical simulation: Toward the design of high-efficiency planar perovskite solar cells,” *Applied Physics Letters*, vol. 104, 6 2014.
- [64] K. Amri, R. Belghouthi, M. Aillerie, and R. Gharbi, “Guidelines for the Design of High-Performance Perovskite Based Solar Cells,” *Key Engineering Materials*, vol. 922, pp. 95–105, 2022.
- [65] S. Gohri, J. Madan, and R. Pandey, “Highly Efficient 3D–2D Perovskite Tandem Solar Cells: A Combined Ray Tracing and Transfer Matrix-Based Simulation Study,” *Journal of Electronic Materials*, vol. 53, pp. 1508–1516, 3 2024.
- [66] I. Montoya De Los Santos, H. J. Cortina-Marrero, M. A. Ruíz-Sánchez, L. Hechavarría-Difur, F. J. Sánchez-Rodríguez, M. Courel, and H. Hu, “Optimization of $\text{CH}_3\text{NH}_3\text{PbI}_3$

- perovskite solar cells: A theoretical and experimental study,” *Solar Energy*, vol. 199, pp. 198–205, 3 2020.
- [67] O. A. Muhammed, D. Eli, P. H. Boduku, J. Tasiu, M. S. Ahmad, and N. Usman, “Modeling and simulation of lead-free perovskite solar cell using SCAPS-1D,” *East European Journal of Physics*, vol. 2021, no. 2, pp. 146–154, 2021.
- [68] Y. Yang, T. Lan, Y. Zhu, Q. Song, C. Zeng, X. Shai, and T. Wei, “High Efficiency of Tandem Perovskite Solar Cells: Optical-Electrical-Thermal Simulation,” *Journal of Nanoelectronics and Optoelectronics*, vol. 18, pp. 149–159, 2 2023.
- [69] Q. Zhou, D. Jiao, K. Fu, X. Wu, Y. Chen, J. Lu, and S. e. Yang, “Two-dimensional device modeling of $\text{CH}_3\text{NH}_3\text{PbI}_3$ based planar heterojunction perovskite solar cells,” *Solar Energy*, vol. 123, pp. 51–56, 1 2016.
- [70] A. Sehili, “Enhancing Efficiency in $\text{CH}_3\text{NH}_3\text{SnI}_3$ -Based inverted perovskite solar Cells: Insights from SCAPS-1D Simulations,” in *1st International Conference on Materials for Energy Storage*, 2023.
- [71] W. Chen, Y. Wu, Y. Yue, J. Liu, W. Zhang, X. Yang, H. Chen, E. Bi, I. Ashraful, M. Grätzel, and L. Han, “Efficient and stable large-area perovskite solar cells with inorganic charge extraction layers,” *Science*, vol. 350, pp. 944–948, 11 2015.
- [72] W. Luo, J. He, Y. Sun, H. Wang, and X. Guo, “Optimizing the Performance of Tin-Based Perovskite Solar Cells Employing 2D Tungsten Disulfide as an HTL by Numerical Simulation,” *physica status solidi (a)*, vol. 220, 12 2023.
- [73] S. Masud Rana, M. Rakibul Hasan, R. Al Amin, M. Shahid Iqbal, M. Rajibul Hoque Rajib, M. Ruhul Kabir Anik, M. Nasrul Hoque Mia, M. Hoq, and M. Hasan, “Simulation of Mono Layer Solar Cell Using COMSOL and Fabrication in Particle Controlled BAEC Clean Room,” *International Journal of Engineering Research & Technology (IJERT)*, 2014.
- [74] R. Zhou, Y. Chen, L. Zhou, Y. Yao, Y. Liu, C. Wang, L. Niu, and L. Chen, “Inhibiting hysteresis and optimizing the performance of perovskite solar cells,” *Solar Energy Materials and Solar Cells*, vol. 264, p. 112616, 1 2024.
- [75] Y. Yang, T. Lan, Y. Zhu, Q. Song, C. Zeng, X. Shai, and T. Wei, “High efficiency of tandem perovskite solar cells: Optical-electrical-thermal simulation,” *Journal of Nanoelectronics and Optoelectronics*, vol. 18, pp. 149–159, 2023.
- [76] S. Valizadeh, A. Shokri, A. Sabouri-Dodaran, N. Fough, and F. Muhammad-Sukki, “Investigation of efficiency and temperature dependence in RbGeBr_3 -based perovskite solar cell structures,” *Results in Physics*, vol. 57, p. 107351, 2 2024.

- [77] T. Jiang, X. Xu, Z. Lan, Z. Chen, X. Chen, T. Liu, S. Huang, and Y. M. Yang, "Efficient MA-free Pb-Sn alloyed low-bandgap perovskite solar cells via surface passivation," *Nano Energy*, vol. 101, 10 2022.
- [78] Y. F. Makableh, I. Abu Awad, W. Hassan, and G. Aljaiuossi, "Enhancement of the thermal properties of heterojunction perovskite solar cells by nanostructured contacts design," *Solar Energy*, vol. 202, pp. 204–209, 5 2020.
- [79] S. Wang, L. Yan, W. Zhu, Z. Cao, L. Zhou, L. Ding, and F. Hao, "Suppressing the formation of tin vacancy yields efficient lead-free perovskite solar cells," *Nano Energy*, vol. 99, 8 2022.
- [80] S. Zandi, P. Saxena, and N. E. Gorji, "Numerical simulation of heat distribution in RGO-contacted perovskite solar cells using COMSOL," *Solar Energy*, vol. 197, pp. 105–110, 2 2020.
- [81] A. Benami, T. Ouslimane, L. Et-taya, and A. Sohani, "Comparison of the effects of ZnO and TiO₂ on the performance of perovskite solar cells via SCAPS-1D software package," *Journal of Nano- and Electronic Physics*, vol. 14, pp. 01033–1–01033–5, 2022.
- [82] A. Dehingia, U. Das, and A. Roy, "Experimental and computational study on Cs₃Bi₂I₉ perovskite solar cell: A comparison of device performance," *Materials Today: Proceedings*, vol. 74, pp. 234–239, 1 2023.
- [83] M. K. Hossain, A. A. Arnab, D. P. Samajdar, M. H. Rubel, M. M. Hossain, M. R. Islam, R. C. Das, H. Bencherif, M. F. Rahman, J. Madan, R. Pandey, S. Bhattarai, M. Amami, and D. K. Dwivedi, "Design Insights into La₂NiMnO₆-Based Perovskite Solar Cells Employing Different Charge Transport Layers: DFT and SCAPS-1D Frameworks," *Energy and Fuels*, vol. 37, pp. 13377–13396, 9 2023.
- [84] J. Kaur, S. Kumar, R. Basu, and A. K. Sharma, "Modelling and simulation of planar heterojunction perovskite solar cell featuring CH₃NH₃PbI₃, CH₃NH₃SnI₃, CH₃NH₃GeI₃ absorber layers," *Silicon*, vol. 16, pp. 1441–1451, 2 2024.
- [85] J. Li, T. Jiu, C. Duan, Y. Wang, H. Zhang, H. Jian, Y. Zhao, N. Wang, C. Huang, and Y. Li, "Improved electron transport in MAPbI₃perovskite solar cells based on dual doping graphdiyne," *Nano Energy*, vol. 46, pp. 331–337, 4 2018.
- [86] S. Banik, A. Das, B. K. Das, and N. Islam, "Numerical simulation and performance optimization of a lead-free inorganic perovskite solar cell using SCAPS-1D," *Heliyon*, vol. 10, p. e23985, 1 2024.

- [87] U. Gautam, V. K. Verma, P. K. Singh, R. Chandra Singh, and M. Kumar, "Solar cell capacitance simulation and experimental photovoltaic performance analysis of perovskite solar cell based on CsGeI₃," *Materials Today: Proceedings*, 2023.
- [88] K. D. Jayan, "Complete modelling and simulation of all perovskite tandem solar cells," *Materials Science and Engineering: B*, vol. 294, 8 2023.
- [89] S. Sharma, K. Kumar, R. K. Pachuari, P. Kuchhal, A. S. S. Vardhan, B. Khan, and A. Ali, "Numerical simulations, design and modeling of methylammonium tin iodide halide-based single-junction perovskite solar cell," *Electrical Engineering*, 2023.
- [90] A. Benami, T. Ouslimane, L. Et-taya, and A. Sohani, "Comparison of the Effects of ZnO and TiO₂ on the Performance of Perovskite Solar Cells via SCAPS-1D Software Package," *Journal of Nano- and Electronic Physics*, vol. 14, no. 1, 2022.
- [91] N. A. Sakib, R. Ahammed, M. Tarekuzzaman, H. Al-Dmour, M. Rasheduzzaman, M. N. Sakib, M. Moazzam Hossen, and M. Z. Hasan, "Highly efficient (31 %) of rubidium-based halide perovskite solar cell using SCAPS-1D simulation," *AIP Advances*, vol. 15, p. 025110, 02 2025.
- [92] M. Meskini and S. Asgharizadeh, "Performance simulation of the perovskite solar cells with Ti₃C₂ mxene in the SnO₂ electron transport layer," *Scientific Reports*, vol. 14, p. 5723, 3 2024.
- [93] S. S. Hussain, S. Riaz, G. A. Nowsherwan, K. Jahangir, A. Raza, M. J. Iqbal, I. Sadiq, S. M. Hussain, and S. Naseem, "Numerical modeling and optimization of lead-free hybrid double perovskite solar cell by using SCAPS-1D," *Journal of Renewable Energy*, vol. 2021, pp. 1–12, 7 2021.
- [94] L. Huang, X. Sun, C. Li, R. Xu, J. Xu, Y. Du, Y. Wu, J. Ni, H. Cai, J. Li, Z. Hu, and J. Zhang, "Electron transport layer-free planar perovskite solar cells: Further performance enhancement perspective from device simulation," *Solar Energy Materials and Solar Cells*, vol. 157, pp. 1038–1047, 12 2016.
- [95] S. Abdelaziz, A. Zekry, A. Shaker, and M. Abouelatta, "Investigating the performance of formamidinium tin-based perovskite solar cell by SCAPS device simulation," *Optical Materials*, vol. 101, p. 109738, 3 2020.
- [96] I. Chabri, Y. Benhouria, A. Oubelkacem, A. Kaiba, I. Essaoudi, and A. Ainane, "SCAPS device simulation study of formamidinium Tin-Based perovskite solar Cells: Investigating the influence of absorber parameters and transport layers on device performance," *Solar Energy*, vol. 262, p. 111846, 9 2023.

- [97] M. Ahamad and A. K. Hossain, "Design and optimization of non-toxic and highly efficient tin-based organic perovskite solar cells by device simulation," *Heliyon*, vol. 9, pp. 2405–8440, 9 2023.
- [98] M. K. Hossain, G. F. Ishraque Toki, D. P. Samajdar, M. H. Rubel, M. Mushtaq, M. R. Islam, M. F. Rahman, S. Bhattarai, H. Bencherif, M. K. Mohammed, R. Pandey, and J. Madan, "Photovoltaic performance investigation of $\text{Cs}_3\text{Bi}_2\text{I}_9$ -based perovskite solar cells with various charge transport channels using dft and SCAPS-1D frameworks," *Energy and Fuels*, vol. 37, pp. 7380–7400, 5 2023.
- [99] D. Ompong and M. Clements, "Optimization of formamidinium-based perovskite solar cell using SCAPS-1D," *Results in Optics*, vol. 14, p. 100611, 2 2024.
- [100] C. Devi and R. Mehra, "Device simulation of lead-free MASnI_3 solar cell with CuSbS_2 (copper antimony sulfide)," *Journal of Materials Science*, vol. 54, pp. 5615–5624, 4 2019.
- [101] A. Yousfi, O. Saidani, A. Benmakhlouf, T. Helaimia, A. S. Alsubaie, S. R. A. Ahmed, and M. R. Islam, "Design and simulation of high efficiency KSnI_3 based perovskite solar cells using dft and SCAPS-1D," *Scientific Reports 2025 15:1*, vol. 15, pp. 1–19, 10 2025.
- [102] V. M. Le Corre, T. S. Sherkar, M. Koopmans, and L. J. A. Koster, "Identification of the dominant recombination process for perovskite solar cells based on machine learning," *Cell Reports Physical Science*, vol. 2, 2 2021.
- [103] N. A. Tukadiya, Z. D. Ponkiya, N. Joshi, D. Upadhyay, and P. K. Jha, "Machine learning and SCAPS-1D based prediction and validation of RbGeBr_3 perovskite solar cell," *Solar Energy*, vol. 300, 11 2025.
- [104] T. W. David, H. Anizelli, T. J. Jacobsson, C. Gray, W. Teahan, and J. Kettle, "Enhancing the stability of organic photovoltaics through machine learning," *Nano Energy*, vol. 78, 12 2020.
- [105] V. N. Muppana, M. Samykano, A. K. Pandey, N. W. Keng, R. K. Rajamony, and S. K. Suraparaju, "Optimizing charge transport and bandgap alignment of ETL and HTL in tin-based perovskite solar cells via machine learning," *Materials Today Communications*, vol. 46, 6 2025.
- [106] M. Hasanzadeh Azar, S. Aynehband, H. Abdollahi, H. Alimohammadi, N. Rajabi, S. Angizi, V. Kamraninejad, R. Teimouri, R. Mohammadpour, and A. Simchi, "SCAPS Empowered Machine Learning Modelling of Perovskite Solar Cells: Predictive Design of Active Layer and Hole Transport Materials," *Photonics*, vol. 10, 3 2023.

- [107] Y. Lu, D. Wei, W. Liu, J. Meng, X. Huo, Y. Zhang, Z. Liang, B. Qiao, S. Zhao, D. Song, and Z. Xu, “Predicting the device performance of the perovskite solar cells from the experimental parameters through machine learning of existing experimental results,” *Journal of Energy Chemistry*, vol. 77, pp. 200–208, 2 2023.
- [108] B. Farhadi, J. You, D. Zheng, L. Liu, S. Wu, J. Li, Z. Li, K. Wang, and S. Liu, “Machine learning for fast development of advanced energy materials,” *Next Materials*, vol. 1, p. 100025, 9 2023.
- [109] Z. Hui, M. Wang, X. Yin, Y. Wang, and Y. Yue, “Machine learning for perovskite solar cell design,” *Computational Materials Science*, vol. 226, 6 2023.
- [110] N. Shrivastav, J. Madan, and R. Pandey, “Predicting photovoltaic efficiency in cs-based perovskite solar cells: A comprehensive study integrating SCAPS simulation and machine learning models,” *Solid State Communications*, vol. 380, 3 2024.
- [111] I. E. Novoselov, A. M. Gvozdev, A. A. Smirnov, and I. S. Zhidkov, “Dataset of SCAPS-1D simulated halide perovskite solar cells with shap and machine learning-based pce optimization,” *Data in Brief*, vol. 60, 6 2025.
- [112] T. K. Ho, “Random decision forests,” in *Proceedings of 3rd International Conference on Document Analysis and Recognition*, vol. 1, pp. 278–282, IEEE Comput. Soc. Press, 1995.
- [113] T. K. Ho, “The random subspace method for constructing decision forests,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, pp. 832–844, 1998.
- [114] E. Higgins, *Practical XGBoost: The Complete Guide to Building and Deploying Gradient Boosting Models with Python*. Amazon Digital Services LLC - Kdp, 2025.
- [115] A. C. Müller and S. Guido, *Introduction to Machine Learning with Python: A Guide for Data Scientists*. O’Reilly Media, first edition ed., 2017.
- [116] J. Chen, M. Wang, C. Jiang, X. Dou, X. Yin, and Y. Yue, “Machine learning: from perovskite materials to devices,” *Solar Energy*, vol. 298, p. 113684, 2025.
- [117] B. Plance and E. Andres, *Hands-On Computer Vision With Tensorflow 2*. Packt Publishing, 2019.
- [118] S. Shafian, M. N. Husen, L. Xie, and K. Kim, “Predicting high-performance perovskite solar cells using ai-based machine learning models,” *Materials Today Sustainability*, vol. 31, 9 2025.

- [119] A. Geron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, Inc., second edition ed., 2019.
- [120] S. Raschka, Y. Liu, V. Mirjalili, and D. Dzhulgakov, *Machine learning with Pytorch and Scikit-Learn : develop machine learning and deep learning models with scikit-learn and PyTorch*. Packt Publishing, Limited, 2022.
- [121] O. Saidani, A. Yousfi, Y. Belhadad, R. Zouache, A. Brahimi, and G. S. Sahoo, "Predictive modeling of rb2sccucl6-based perovskite solar cells performance using neural networks-guided optimization and shap-based interpretability," *Journal of Physics and Chemistry of Solids*, vol. 208, p. 113084, 2026.
- [122] A. Toprak, "Predicting photovoltaic output parameters of perovskite solar cells using explainable machine learning with physics-informed and simulation-derived descriptors," *Materials Today Communications*, vol. 52, p. 114968, 2026.
- [123] D. Sadhu, D. Dattatreya, A. Deo, K. Tarafder, and D. De, "Performance prediction and analysis of perovskite solar cells using machine learning," *Journal of Alloys and Compounds Communications*, vol. 3, p. 100022, 2024.
- [124] S. Ahmad, S. Kazim, and M. Grätzel, *Perovskite Solar Cells*. Wiley, 2021.
- [125] T. J. Jacobsson, A. Hultqvist, A. García-Fernández, A. Anand, A. Al-Ashouri, A. Hagfeldt, and et al., "An open-access database and analysis tool for perovskite solar cells based on the fair data principles," *Nature Energy*, vol. 7, pp. 107–115, 1 2022.
- [126] M. Dadashbeik, D. Fathi, and M. Eskandari, "Design and simulation of perovskite solar cells based on graphene and TiO₂/graphene nanocomposite as electron transport layer," *Solar Energy*, vol. 207, pp. 917–924, 9 2020.
- [127] T. Homola, J. Pospisil, M. Shekargoftar, T. Svoboda, M. Hvojník, P. Gemeiner, M. Weiter, and P. Dzik, "Perovskite solar cells with low-cost TiO₂ mesoporous photoanodes prepared by rapid low-temperature (70 °c) plasma processing," *ACS Applied Energy Materials*, vol. 3, pp. 12009–12018, 12 2020.
- [128] R. A. Afre and D. Pugliese, "Perovskite solar cells: A review of the latest advances in materials, fabrication techniques, and stability enhancement strategies," *Micromachines*, vol. 15, 2 2024.
- [129] J. Zhang, W. Zhang, H.-M. Cheng, and S. R. P. Silva, "Critical review of recent progress of flexible perovskite solar cells," *Materials Today*, vol. 39, pp. 66–88, 2020.

- [130] C. Devi and R. Mehra, “Device simulation of lead-free MASnI_3 solar cell with CuSbS_2 (copper antimony sulfide),” *Journal of Materials Science*, vol. 54, pp. 5615–5624, 4 2019.
- [131] Y. Velez-Galvis, A. Sepulveda, J. Echeverri-Perez, B. Medina, E. Gonzalez-Valencia, and E. Reyes-Vera, “Simulation of lead-free perovskite solar cell based on MASnI_3 compared with FAPbI_3 ,” *Optica Latin America Optics and Photonics Conference (LAOP) 2024 (2024)*, paper Tu4A.16, p. Tu4A.16, 11 2024.
- [132] D. Thakur, S.-E. Chiang, M.-H. Yang, J.-S. Wang, and S. H. Chang, “Self-stability of un-encapsulated polycrystalline MAPbI_3 solar cells via the formation of chemical bonds between C60 molecules and ma cations,” *Solar Energy Materials and Solar Cells*, vol. 235, p. 111454, 1 2022.
- [133] M. Shoab, Z. Aslam, M. Zulfequar, and F. Khan, “Numerical interface optimization of lead-free perovskite solar cells ($\text{CH}_3\text{NH}_3\text{SnI}_3$) for 30 % photo-conversion efficiency using SCAPS-1D,” *Next Materials*, vol. 4, p. 100200, 7 2024.
- [134] W. Luo, J. He, Y. Sun, H. Wang, and X. Guo, “Optimizing the performance of tin-based perovskite solar cells employing 2D tungsten disulfide as an HTL by numerical simulation,” *physica status solidi (a)*, vol. 220, 2023.
- [135] A. C. Reyes, R. C. Lázaro, K. M. Leyva, J. A. López, J. F. Méndez, A. H. Jiménez, A. L. Zurita, F. S. Carrillo, and E. O. Durán, “Study of a lead-free perovskite solar cell using CZTS as HTL to achieve a 20 % pce by SCAPS-1D simulation,” *Micromachines*, vol. 12, 12 2021.
- [136] I. Chabri, Y. Benhouria, A. Oubelkacem, A. Kaiba, I. Essaoudi, and A. Ainane, “SCAPS device simulation study of formamidinium tin-based perovskite solar cells: Investigating the influence of absorber parameters and transport layers on device performance,” *Solar Energy*, vol. 262, p. 111846, 9 2023.
- [137] S. Abdelaziz, A. Zekry, A. Shaker, and M. Abouelatta, “Investigating the performance of formamidinium tin-based perovskite solar cell by scaps device simulation,” *Optical Materials*, vol. 101, p. 109738, 3 2020.
- [138] S. Vaish and S. Kumar Dixit, “Study the effect of total defect density variation in absorbing layer on the power conversion efficiency of lead halide perovskite solar cell using SCAPS-1D simulation tool,” *Materials Today: Proceedings*, vol. 91, pp. 17–20, 2023. International Conference on Futuristic Materials.
- [139] A. Bag, R. Radhakrishnan, R. Nekovei, and R. Jeyakumar, “Effect of absorber layer, hole transport layer thicknesses, and its doping density on the performance of perovskite solar cells by device simulation,” *Solar Energy*, vol. 196, pp. 177–182, 2020.

- [140] D. Lee, K. Kim, and H.-D. Kim, "Thickness optimization of charge transport layers on perovskite solar cells for aerospace applications," *Nanomaterials*, vol. 13, p. 1848, 2023.
- [141] S. Aseena, A. Nelsa, and B. V. Suresh, "Optimization of layer thickness of ZnO based perovskite solar cells using SCAPS-1D," *Materials Today: Proceedings*, vol. 43, 2020.
- [142] M. Miah, M. Khandaker, M. Rahman, M. Alam, and M. Islam, "Band gap tuning of perovskite solar cells for enhancing the efficiency and stability: issues and prospects," *RSC Advances*, vol. 14, pp. 15876–15906, 2024.
- [143] Z. Yang, A. Surrente, K. Galkowski, N. Bruyant, D. Maude, A. Haghghirad, H. Snaith, P. Plochocka, and R. Nicholas, "Unraveling the exciton binding energy and the dielectric constant in single crystal methylammonium lead triiodide perovskite," *Journal of Physical Chemistry Letters*, vol. 8, p. 1851, 2017.
- [144] R. Su, Z. Xu, J. Wu, D. Luo, Q. Hu, W. Yang, X. Yang, R. Zhang, H. Yu, T. Russell, Q. Gong, and R. Zhu, "Dielectric screening in perovskite photovoltaics," *Nature Communications*, vol. 12, 2021.
- [145] A. Sahoo, I. Mohanty, and S. Mangal, "Effect of acceptor density, thickness and temperature on device performance for tin-based perovskite solar cell," *Materials Today: Proceedings*, vol. 62, 2022.
- [146] M. H. Mohammadi, M. Eskandari, and D. Fathi, "Design of optimized photonic-structure and analysis of adding a SiO_2 layer on the parallel $\text{CH}_3\text{NH}_3\text{PbI}_3/\text{CH}_3\text{NH}_3\text{SnI}_3$ perovskite solar cells," *Scientific Reports*, vol. 13, 15905, 2023.
- [147] M. S. Rahman, N. Ahmed, T. Paul, and M. M. Rahman, "Performance evaluation of lead free $\text{CH}_3\text{NH}_3\text{SnI}_3$ perovskite solar cell: A simulation approach by SCAPS-1D," *2024 3rd International Conference on Power, Control and Computing Technologies, ICPC2T 2024*, pp. 392–397, 2024.
- [148] M. H. Azar, S. Aynehband, H. Abdollahi, H. Alimohammadi, N. Rajabi, S. Angizi, V. Kamraninejad, R. Teimouri, R. Mohammadpour, and A. Simchi, "SCAPS empowered machine learning modelling of perovskite solar cells: Predictive design of active layer and hole transport materials," *Photonics*, vol. 10, p. 271, 3 2023.
- [149] M. Minbashi and E. Yazdani, "Comprehensive study of anomalous hysteresis behavior in perovskite-based solar cells," *Scientific Reports*, vol. 12, p. 14916, 9 2022.
- [150] W. Shockley and H. J. Queisser, "Detailed balance limit of efficiency of p-n junction solar cells," *Journal of Applied Physics*, vol. 32, no. 3, pp. 510–519, 1961.

- [151] S. D. Stranks and H. J. Snaith, “Metal-halide perovskites for photovoltaic and light-emitting devices,” *Nature Nanotechnology*, vol. 10, pp. 391–402, 2015.
- [152] A. K. Jena, A. Kulkarni, and T. Miyasaka, “Halide perovskite photovoltaics: Background, status, and future prospects,” *Chemical Reviews*, vol. 119, pp. 3036–3103, 2019.
- [153] A. H. Oishi, M. T. Anjum, M. M. Islam, and M. F. Nayan, “Impact of absorber layer thickness on perovskite solar cell efficiency: A performance analysis,” *European Journal of Electrical Engineering and Computer Science*, p. 48–51, 2023.
- [154] W. Tress, “Metal halide perovskites as next-generation photovoltaic materials,” *Energy & Environmental Science*, vol. 10, pp. 951–969, 2017.
- [155] H. J. Snaith, “Present status and future prospects of perovskite photovoltaics,” *Nature Materials*, vol. 17, pp. 372–376, 2018.
- [156] National Renewable Energy Laboratory (NREL), “Best research-cell efficiency chart — photovoltaic research — nrel,” *NREL*, 2025. Last updated Dec. 6, 2025.
- [157] T. Ouslimane, L. Et-taya, L. Elmaimouni, and A. Benami, “Impact of absorber layer thickness, defect density, and operating temperature on the performance of MAPbI₃ solar cells based on ZnO electron transporting material,” *Heliyon*, vol. 7, p. e06379, 3 2021.
- [158] A. K. Sharma and D. K. Kaushik, “Numerical simulation of masni 3 /cui heterojunction based perovskite solar cell,” *Journal of Physics: Conference Series*, vol. 2267, p. 012148, 2022.
- [159] A. D. Rana, I. D. Pharne, and K. Bhargava, “Numerical simulation of highly efficient double perovskite solar cell using SCAPS-1D,” *Materials Today: Proceedings*, vol. 73, pp. 584–589, 2023. Nanomaterials for Energy Conversion and Storage Application-2022 (NECSA 2022).
- [160] A. C. Müller and S. Guido, *Introduction to Machine Learning with Python: A Guide for Data Scientists*. O’Reilly Media, first edition ed., 2017.
- [161] J. Ping, P. Wang, B. Liu, H. Zhou, Y. Li, and J. Lin, “A hierarchical hgbt-based machine learning framework for predicting the power conversion efficiency of perovskite solar cells,” *Materials Today Communications*, vol. 46, p. 112747, 6 2025.
- [162] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, A. Müller, J. Nothman, G. Louppe, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and Édouard Duchesnay,

- “Scikit-learn: Machine learning in python,” *Journal of Machine Learning Research*, 6 2018.
- [163] S. Karthick, S. Velumani, and J. Bouclé, “Experimental and scaps simulated formamidinium perovskite solar cells: A comparison of device performance,” *Solar Energy*, vol. 205, pp. 349–357, 7 2020.
- [164] G. Haidari, “One-dimensional modeling for optoelectrical simulation of a mesoporous perovskite solar cell,” *Appl. Opt.*, vol. 58, pp. 7006–7013, Sep 2019.
- [165] D. Stanić, V. Kojić, M. Boháč, T. Čižmar, K. Juraić, T. Rath, and A. Gajović, “Simulation and optimization of FAPbI₃ perovskite solar cells with a BaTiO₃ layer for efficiency enhancement,” *Materials*, vol. 15, 2022.
- [166] F. Izadi, A. Ghobadi, A. Gharaati, M. Minbashi, and A. Hajjiah, “Effect of interface defects on high efficient perovskite solar cells,” *Optik*, vol. 227, 2021.
- [167] A. Usman and T. Bovornratanaraks, “Modeling and optimization of modified TiO₂ with aluminum and magnesium as ETLin MAPbI₃ perovskite solar cells: SCAPS-1D frameworks,” *ACS Omega*, vol. 9, p. 39672, 2024.
- [168] A. Srivastava, A. K. Sharma, P. K. Jha, M. Kumar, N. K. Chourasia, and R. K. Chourasia, “Parametric optimization for the performance analysis of novel hybrid organo-perovskite solar cell via SCAPS-1D simulation,” *Heliyon*, vol. 10, no. 19, p. e38169, 2024.
- [169] D. Ompong and M. Clements, “Optimization of formamidinium-based perovskite solar cell using SCAPS-1D,” *Results in Optics*, vol. 14, p. 100611, 2024.
- [170] Y. Raoui, H. Ez-Zahraouy, N. Tahiri, O. E. Bounagui, S. Ahmad, and S. Kazim, “Performance analysis of MAPbI₃ based perovskite solar cells employing diverse charge selective contacts: Simulation study,” *Solar Energy*, vol. 193, pp. 948–955, 11 2019.
- [171] W. Li, J. Hu, Z. Chen, H. Jiang, J. Wu, X. Meng, X. Fang, J. Lin, X. Ma, T. Yang, P. Cheng, and R. Xie, “Performance prediction and optimization of perovskite solar cells based on the bayesian approach,” *Solar Energy*, vol. 262, p. 111853, 2023.
- [172] Q. Tao, P. Xu, M. Li, and W. Lu, “Machine learning for perovskite materials design and discovery,” *npj Computational Materials*, vol. 7, no. 1, p. 23, 2021.
- [173] M. Dadu, A. Kapoor, and K. N. Tripathi, “Effect of series resistance and shunt resistance on the fill factor of a solar cell,” *Solar Energy Materials and Solar Cells*, 2002.
- [174] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.

-
- [175] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, “A fast and elitist multiobjective genetic algorithm: NSGA-II,” *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.
- [176] H. Q. Tan, X. Zhao, A. Ambardekar, E. Birgersson, and H. Xue, “Exploring the optimal design space of transparent perovskite solar cells for four-terminal tandem applications through pareto front optimization,” *APL Machine Learning*, vol. 2, no. 2, p. 026111, 2024.
- [177] Q. Jiang, X. Zhang, and J. You, “SnO₂: A wonderful electron transport layer for perovskite solar cells,” *Small*, p. 1801154, 2018.
- [178] G. Ren, W. Han, Y. Deng, W. Wu, Z. Li, J. Guo, H. Bao, C. Liu, and W. Guo, “Strategies of modifying spiro-ometad materials for perovskite solar cells: a review,” *Journal of Materials Chemistry A*, vol. 9, no. 8, pp. 4589–4625, 2021.
- [179] S. Rühle, “Tabulated values of the shockley–queisser limit for single junction solar cells,” *Solar Energy*, vol. 130, pp. 139–147, 2016.
- [180] Y. Ahmed, X. Qi, P. Moazzezi, and M. I. Saidaminov, “Perovskite photovoltaics: Pick FAPbI₃ and stick to it,” *ACS Energy Letters*, vol. 11, no. 1, pp. 79–89, 2026.
- [181] T. S. Sherkar, C. Momblona, L. Gil-Escrig, J. Ávila, M. Sessolo, H. J. Bolink, and L. J. A. Koster, “Recombination in perovskite solar cells: Significance of grain boundaries, interface traps, and defect ions,” *ACS Energy Letters*, vol. 2, no. 5, pp. 1214–1222, 2017.
- [182] F. Wang, S. Bai, W. Tress, A. Hagfeldt, and F. Gao, “Defects engineering for high-performance perovskite solar cells,” *npj Flexible Electronics*, vol. 2, no. 1, p. 22, 2018.