

Caracterización de celdas solares de perovskita mediante un análisis exploratorio de datos de variables eléctricas, band gap y área

Characterization of perovskite solar cells through an exploratory analysis of data on electrical variables, band gap, and area

Camilo Hernández Pérez¹, Carlos Castañeda Pico², Sergio Castro Casadiego³, Byron Medina Delgado⁴, Alexander Sepúlveda Sepúlveda⁵, Erick Reyes Vera⁶

¹Estudiante de Ingeniería Electrónica, Joven Investigador e Innovador, Universidad Francisco de Paula Santander, Colombia. Email: camiloandreshp@ufps.edu.co. ORCID: 0000-0002-6912-7283

²Estudiante de Ingeniería Electrónica, Joven Investigador e Innovador, Universidad Francisco de Paula Santander, Colombia. Email: carlososwaldocp@ufps.edu.co. ORCID: 0009-0004-4181-944X

³Magíster en Ingeniería Electrónica, Docente e Investigador, Universidad Francisco de Paula Santander, Colombia. Email: sergio.castroc@ufps.edu.co. ORCID: 0000-0003-0962-9916

³Doctor en Ciencias, Docente e Investigador, Universidad Francisco de Paula Santander, Colombia. Email: byronmedina@ufps.edu.co. ORCID: 0000-0003-0962-9916

⁵Doctor en Ingeniería, Docente e Investigador, Universidad Industrial de Santander, Colombia. Email: fasepul@uis.edu.co. ORCID: 0000-0002-9643-5193

⁶Doctor en Ingeniería Eléctrica, Docente e Investigador, Instituto Tecnológico Metropolitano, Colombia. Email: erickreyes@itm.edu.co. ORCID: 0000-0001-8769-0718

Cite this article as: C.Hernandez Perez, C. Castañeda Pico, S. Castro Casadiego, B. Medina Delgado, A. Sepúlveda Sepúlveda, E. Reyes Vera "Caracterización de celdas solares de perovskita mediante un análisis exploratorio de datos de variables eléctricas, band gap y área", Prospectiva, Vol. 22 N° 2 2024.

Recibido: 22/03/2024 / Aceptado: 27/09/2024

<http://doi.org/10.15665/rp.v22i2.3485>

RESUMEN

La comprensión del comportamiento de las variables eléctricas de una celda solar contribuye a determinar su desempeño, por ello se buscan patrones que permitan analizar variables intrínsecas del material y la influencia de otras variables como el área. El análisis exploratorio de los datos se utiliza en el procesamiento de información, para detectar patrones y relaciones que se puedan explotar en el desarrollo de un modelo de IA. Este artículo tiene como objetivo caracterizar celdas solares de perovskita de unión simple a través de un análisis exploratorio de los datos de variables eléctricas, específicamente band gap y área, extraídos de The Perovskita Database. Esto se logra, usando visualizaciones univariantes y multivariantes como gráficos de violín y correlación respectivamente, que relacionan la media, mediana, máximos y mínimos para detectar patrones en la tecnología de las celdas solares de perovskita y las relaciones entre las variables seleccionadas. Los resultados presentan para el perovskita, valores prácticos de 0,9 V para el voltaje de circuito abierto (Voc), 17,9 mA/cm² para la densidad de corriente (Jsc) y 1,6 eV para el band gap. Se concluye, al contrastar los valores prácticos con la teoría de Shockley-Queisser, que las variables Voc, Jsc y band gap son características intrínsecas de las celdas solares de perovskita de unión simple.

Palabras claves: análisis exploratorio de los datos, aprendizaje automático, celda solar de perovskita, factor de conversión energética.

ABSTRACT

Understanding the behavior of electrical variables in a solar cell contributes to determining its performance, thus patterns are sought to analyze intrinsic material variables and the influence of other variables such as area. Exploratory data analysis is used in information processing to detect patterns and relationships that can be exploited in the development of artificial intelligence. This article aims to characterize single junction perovskite solar cells through an exploratory analysis of the dataset containing electrical variables, band gap, and area from The Perovskita Database. This is achieved using univariate and multivariate visualizations such as violin plots and correlation plots respectively, which relate the mean, median, maximum, and minimum to detect patterns in perovskite solar cell technology and the relationships between the selected variables. The results show practical values of 0.9 V for the open-circuit voltage (Voc), 17.9 mA/cm² for the current density (Jsc), and 1.6 eV for the band gap for perovskite. It is concluded, given the contrast of practical values with the Shockley Queisser theory, that the Voc, Jsc, and band gap variables are intrinsic to single junction perovskite solar cells.

Keywords: Exploratory data analysis, machine learning, perovskite solar cell, energy conversion factor.

1. INTRODUCCIÓN

Los paneles solares se han convertido en una buena técnica de generación de energía renovable, pasando de generar 554 TWh en el 2018 a 2020 TWh en el 2021 [1]. El aumento de la implementación de paneles solares ha generado una mayor inversión en esta tecnología, llevándola actualmente hasta la tercera generación, aún en desarrollo. La primera generación usa como material base el silicio monocristalino o policristalino; la segunda generación usa compuestos de diferentes materiales como el Telurio de Cadmio (CdTe) o el Cobre, Indio, Galio y Selenio (CIGS) [2], [3]; la tercera generación, que se espera sea más eficiente, está basada en la perovskita, aún en desarrollo [4].

Las celdas solares de perovskita, de unión simple, han tenido un crecimiento en la eficiencia de conversión de potencia (Power Conversion Efficiency, PCE), logrando registros de 3,8% en el 2009, 19,3% en 2016 [5], y cercanos al 25% en el 2020 [6]. Los avances de la eficiencia de la perovskita han sido prometedores, considerando que el material más usado, actualmente, en la producción de celdas solares es el silicio, ocupando un 92% del mercado con una eficiencia aproximada del 25% [2].

El aprendizaje automático (Machine Learning, ML), es una herramienta útil para el diseño y rendimiento de tecnologías fotovoltaicas, porque contribuye con la disminución del tiempo y los costos de fabricación [7], [8], [9]; sin embargo, para aplicar ML es necesario el análisis de datos que se tienen como entrenamiento o pruebas para la misma [10]. Este análisis de datos permite identificar patrones no anticipados, dispersión de datos y correlaciones de las variables, utilizando técnicas de visualización, con el fin de encontrar las variables relevantes y preparar los datos para la aplicación de algoritmos de predicción [11], [12].

Investigaciones han utilizado el análisis exploratorio de datos (Exploratory Data Analysis, EDA) para: tratar el conjunto de datos que serán usados para predecir el band gap en semiconductores de perovskita [13]; en conjunto con machine learning encontrar el aumento del porcentaje de PCE, en celdas solares pasivadas con un tinte orgánico [14]; explorar métricas de corriente-voltaje y eficiencia en celdas solares de perovskita certificadas [15]; comprender el impacto de cada compuesto, en un conjunto de datos heterogéneos de vidrios de silicato [16]; y para preprocesar los datos de un algoritmo de predicción del comportamiento del band gap en celdas solares de perovskita orgánicas e inorgánicas [17].

El artículo plantea un análisis exploratorio de los datos [18] de las variables de salida de celdas solares de perovskita de unión simple usando como conjunto de datos The Perovskite Database [19], detectando la relación entre área y band gap con respecto a las variables eléctricas como voltaje de circuito abierto (Voltage Open Circuit, Voc), densidad de corriente de corto circuito (Density Short Circuit Current, Jsc), factor de llenado (Fill Factor, FF) y el PCE. Esta información se contrasta con los datos publicados por National Renewable Energy Laboratory (NREL) [20] y la teoría del límite de Shockley-Queisser [21]

2. METODOLOGÍA

La metodología incluye diversas fases relacionadas con el análisis exploratorio de datos. Esta herramienta analítica permite identificar patrones en los conjuntos de datos, lo que es esencial para alcanzar el objetivo del artículo. A través de visualizaciones, permite descubrir relaciones y tendencias entre variables, facilitando la identificación de valores intrínsecos en las celdas solares de perovskita. El proceso abarca la definición de variables, la detección y tratamiento de datos faltantes, visualizaciones univariantes y multivariantes, así como la validación de la caracterización de las celdas de perovskita de unión simple.

La fase de definición de variables hace referencia a la selección de variables en los conjuntos de datos de The Perovskite Database [19] y NREL. De las aproximadas 400 variables de The Perovskite Database, se utilizan Voc, Jsc, FF, PCE, band gap y área. De NREL a través de la gráfica Best Research-cell Efficiency Chart, se extraen las variables Voc, Jsc, FF, PCE y área.

La fase de detección y tratamiento de datos faltantes comprende para los datos faltantes la visualización, clasificación, eliminación e imputación. La visualización se realiza con un histograma; la clasificación mediante datos faltantes aleatorios (Missing at Random, MAR), datos faltantes completamente aleatorios (Missing Completely at Random, MCAR) y datos faltantes no aleatorios (Missing Not at Random, MNAR); la eliminación mediante la técnica Listwise en el conjunto de datos The Perovskite Database; y la imputación usando la fórmula de eficiencia, en el conjunto de datos de NREL.

La fase de visualización univariante hace referencia al análisis descriptivo de cada variable del conjunto de datos The Perovskite Database, mediante gráficas de violín junto con diagrama de cajas, mostrando la media, mediana, desviación estándar, máximos, mínimos, distribución de los datos y valores atípicos.

La fase de visualización multivariante comprende el análisis correlacional entre las variables, mediante gráficos de correlación, que condensan la distribución bivariada de los datos y el coeficiente de correlación. Esta fase utiliza los conjuntos de datos de The Perovskite Database y NREL.

La fase de validación de la caracterización de la celda se basa en contrastar los resultados del análisis univariante y multivariante del conjunto de datos The Perovskite Database con el análisis multivariante del conjunto de datos de NREL y la teoría del límite de Shockley-Queisser. Esta teoría, establece el límite máximo teórico de eficiencia para las celdas solares, es relevante para el estudio, ya que permite evaluar el rendimiento de las celdas de perovskita en comparación con ese umbral teórico basado en el band gap. El propósito es identificar similitudes, discrepancias y tendencias que aporten a la validez de los resultados y a una mejor comprensión del comportamiento de las celdas solares de perovskita de unión simple, en función de sus variables eléctricas.

3. RESULTADOS

Se describe para cada fase de la metodología, los resultados y discusión.

Definición de variables

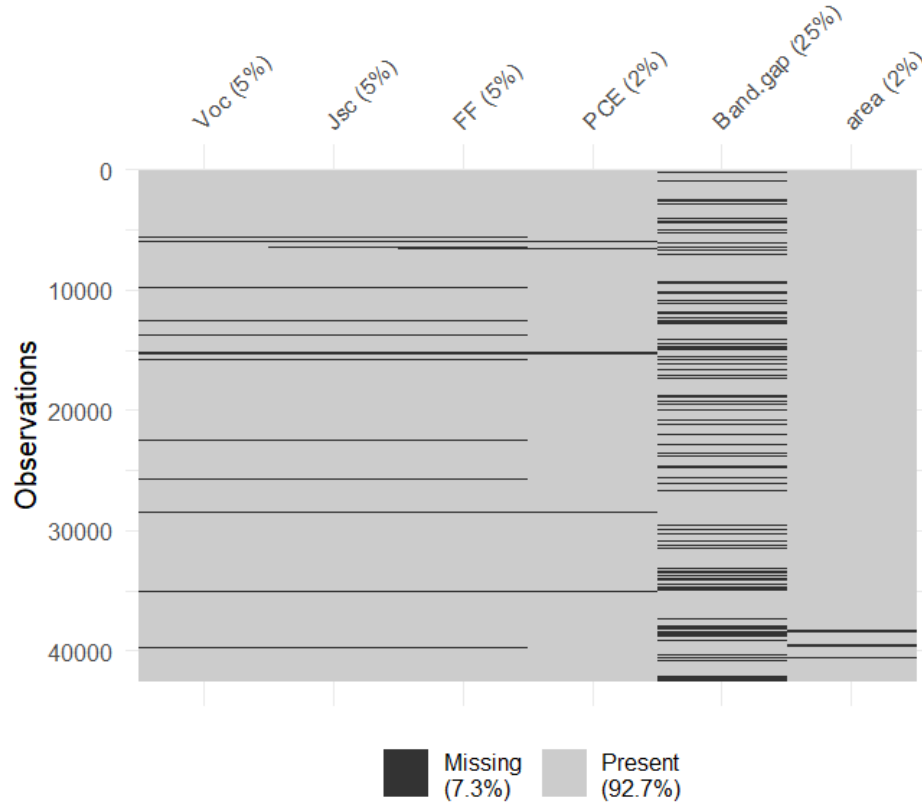
Se seleccionan las variables eléctricas como: el potencial máximo que puede alcanzar un panel solar cuando no existe corriente fluyendo por él (V_{oc}), con unidades de voltios; la máxima corriente que puede suministrar el panel solar cuando está en corto circuito (J_{sc}), con unidades de mA/cm^2 ; la relación entre la potencia máxima real entregada por el panel solar y la potencia máxima teórica que podría entregar bajo condiciones ideales (FF), adimensional; y la eficiencia con la que un panel solar convierte la luz solar incidente en electricidad (PCE), adimensional. Además, se seleccionaron las variables como: la diferencia de energía entre la banda de valencia y la banda de conducción, que permite conocer la absorción de luz del material (band gap), con unidades eV; y el área total de la celda, con unidades cm^2 ; con el propósito de encontrar la relación entre estas variables y las variables eléctricas [22].

Detección y tratamiento de datos faltantes

Se visualizan los datos faltantes por variable, observando la ausencia de las siguientes cantidades de datos: 2134 para V_{oc} , 2099 para J_{sc} , 2286 para FF, 925 para PCE, 10574 para el band gap y 711 para el área. En la figura 1, se muestra el porcentaje de datos faltantes por variable, para un total de 7,3% del conjunto de datos. El band gap tiene la mayor pérdida de datos con el 25%.

Se clasifican los datos faltantes como MCAR [23], considerando que, en la recolección de información de la base de datos The Perovskite Database, no se incluyen datos en algunos artículos o se pierden en el minado de información, es decir, los datos faltantes no dependen de otras variables, como lo muestra la dispersión de la figura 1.

Figura 1. Malla de valores faltantes del conjunto de datos The Perovskite Database.
Figure 1. Missing values grid of The Perovskite Database dataset.



Se eliminan los datos faltantes a partir de la técnica Listwise, pasando de 42497 filas a 31150 filas, equivalente al 26,7% del conjunto de datos; sin embargo, la cantidad de datos es suficiente para mantener la integridad y validez del EDA.

Se imputan los datos faltantes de la variable Jsc perteneciente al conjunto de datos de NREL, usando la fórmula de eficiencia (I) que expresa la relación entre la densidad de potencia de salida y entrada. La densidad de potencia de salida está descrita por la fórmula (II) [22], y la densidad de potencia de entrada es el valor estándar en la caracterización de la celda, 100 mW/cm^2 [24]. Los datos se muestran en la tabla 1.

$$\eta = \frac{V_{oc} I_{sc} FF}{P_{in}} \quad (I)$$

$$P_{max} = V_{oc} I_{sc} FF \quad (II)$$

Tabla 1. Datos extraídos de la gráfica Best Research-Cell Efficiency Chart [20].

Table 1. Data extracted from the Best Research-Cell Efficiency Chart [20].

Año	PCE (%)	V_{oc} (V)	J_{sc} (mA/cm^2)	FF	Área cm^2
2013	14.1	1.007	21.31	0.657	0.209
2014	17.9	1.109	21.75	0.742	0.0938
2014	17.9	1.114	21.83	0.736	0.0937
2014	20.1	1.059	24.65	0.77	0.0955

Año	PCE (%)	V_{oc} (V)	J_{sc} (mA/cm^2)	FF	Área cm^2
2016	22.1	1.105	24.91	0.803	0.0946
2017	22.7	1.144	24.93	0.796	0.0935
2018	23.3	1.791	16.59	0.784	0.074
2018	23.7	1.1697	25.39	0.798	0.0739
2019	24.2	1.1948	24.11	0.84	0.0955
2019	25.2	1.1805	25.17	0.848	0.0937
2020	25.5	1.189	25.84	0.83	0.095
2022	25.7	1.179	25.77	0.846	0.096
2022	25.8	1.1797	25.77	0.849	0.0959
2023	26.00	1.19	26.00	0.84	0.0746
2023	26.1	1.174	26.13	0.852	0.0493

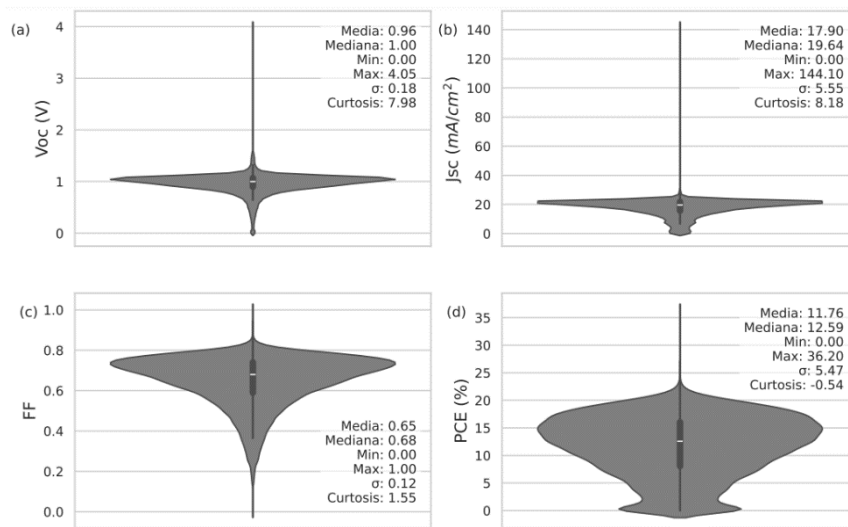
Visualización univariante del conjunto de datos

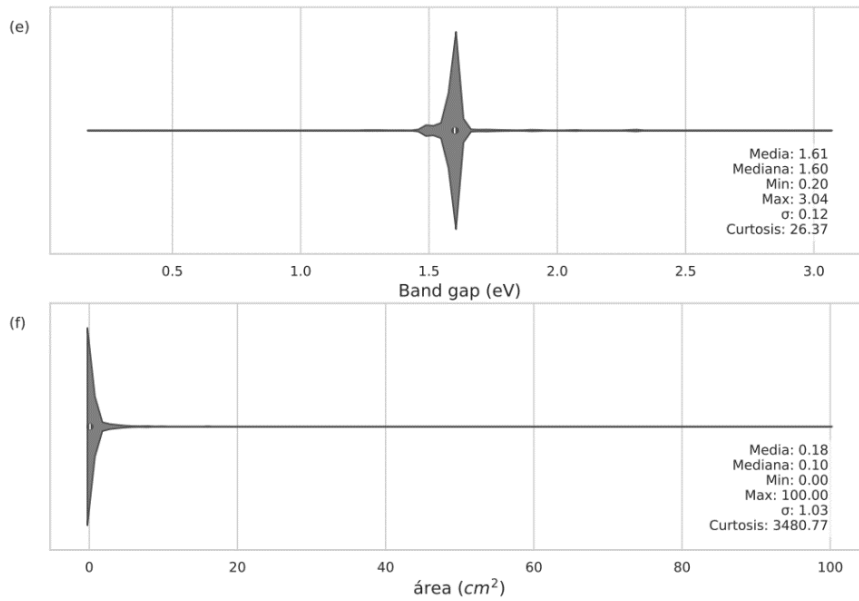
En la figura 2, se muestran para cada variable la distribución, media, mediana, máximo, mínimo y desviación estándar, mediante el diagrama de caja y violín. Usando el coeficiente de curtosis, se determina que las variables de V_{oc} , J_{sc} y band gap tienen una distribución leptocúrtica, indicando que las variables pueden ser intrínsecas de las celdas solares de perovskita de unión simple.

Los valores atípicos, ubicados fuera del violín, como valores máximos de $144,10 \text{ mA}/\text{cm}^2$ en J_{sc} , $4,05 \text{ V}$ en V_{oc} y 1 en FF; y valores mínimos de cero en todas las variables, están asociados a errores de transcripción. Este argumento se sustenta en la verificación de los artículos que fueron usados como fuente primaria para la creación del conjunto de datos de The Perovskite Database.

Figura 2. Visualización univariante de los datos a) V_{oc} , b) J_{sc} , c) FF, d) PCE, e) Band gap, f) Área.

Figure 2. Univariate visualization of the data a) V_{oc} , b) J_{sc} , c) FF, d) PCE, e) Band gap, f) Area.

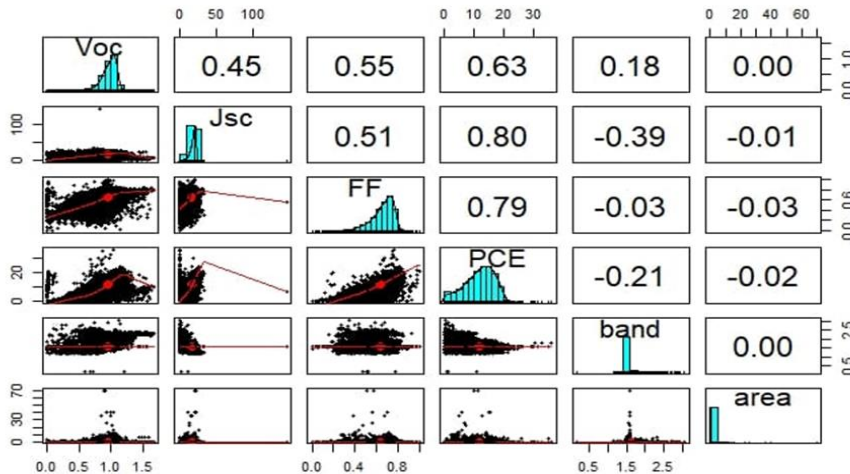




Visualización multivariante del conjunto de datos

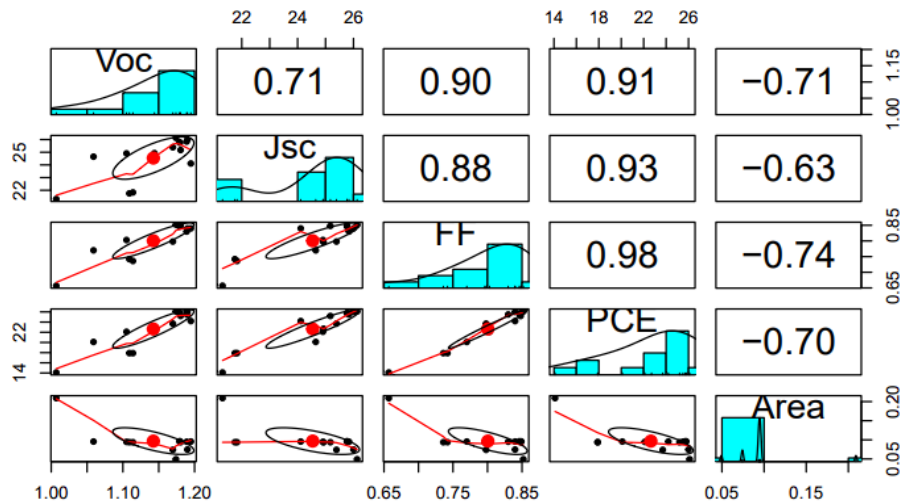
La figura 3, visualiza la relación que tienen las variables eléctricas de las celdas de perovskita con respecto al band gap y el área, del conjunto de datos de The Perovskite Database. Esta gráfica divide con respecto a la diagonal, los coeficientes de correlación y la distribución bivalente de los datos junto con la línea de tendencia. Se muestra que el band gap presenta relación negativa con las variables densidad de corriente y eficiencia, relación positiva con la variable voltaje, e independencia con la variable factor de llenado; este comportamiento tiene relación con la teoría de Shockley-Queisser, y sugiere que Voc y Jsc son variables intrínsecas de la celda. De igual manera, se muestra que el área tiene independencia con respecto a las variables eléctricas.

Figura 3. Visualización multivariante de The Perovskite Database.
Figure 3. Multivariate Visualization of The Perovskite Database.



La figura 4, visualiza la relación que tienen las variables eléctricas de las celdas de perovskita con respecto al área, del conjunto de datos de NREL. En esta se muestra que el área tiene dependencia negativa con respecto a las variables eléctricas.

Figura 4. Visualización multivariante de NREL “Best Research-Cell Efficiency Chart”.
Figure 4. Multivariate Visualization of NREL's Best Research-Cell Efficiency Chart.



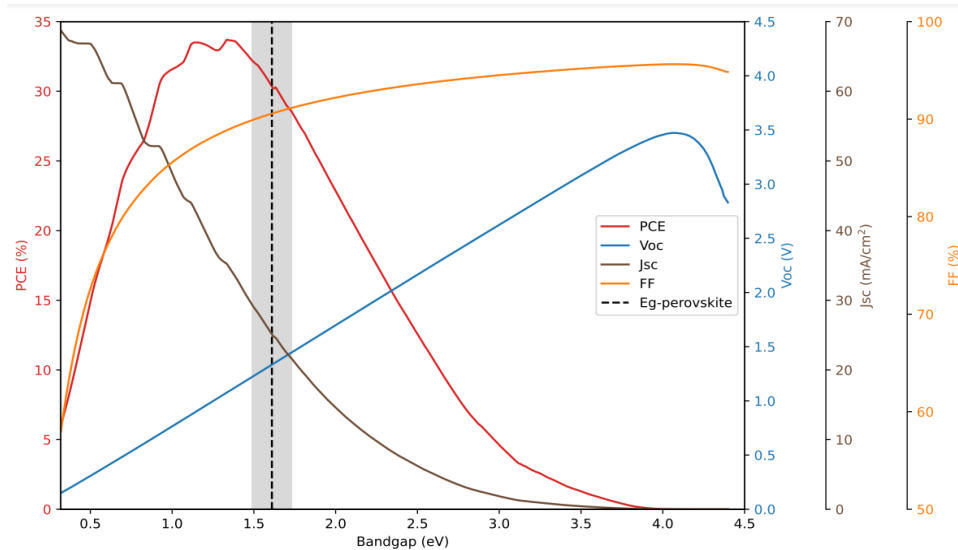
La discrepancia que presenta el área con respecto a las variables eléctricas, de tener independencia en el conjunto de datos de The Perovskite Database, y dependencia en el conjunto de datos de NREL, hace pensar que el conjunto de datos de The Perovskite Database tiene sesgo en la variable área, considerando que los investigadores asignaron valores al área de $0,1 \text{ cm}^2$ en las celdas sin información.

A partir de los coeficientes de correlación de NREL, se muestran inversos al área, como lo demuestra [25], [26] que al aumentar el área disminuye la eficiencia de la celda, porque se extienden los defectos de la celda que, en conjunto con las interfaces, introducen centros de recombinación que conducen a pérdidas no radiativas que llevan a la disminución del Voc, Jsc y del FF, estas variables influyen directamente en la eficiencia de la celda. La caída de la eficiencia debido al aumento del área puede verse en [27], en la gráfica PCE en función del área, que exhibe una tendencia inversa entre estas variables. La máxima eficiencia mostrada es de una celda de $0,09 \text{ cm}^2$ con una eficiencia de 25,2% mientras que los módulos con áreas entre 10 y 1000 cm^2 tienen eficiencias que rondan entre 16% y 8%, respectivamente.

Validación de la caracterización de las celdas

La teoría Shockley-Queisser [21], [28] establece el comportamiento de las variables eléctricas de una celda a partir de su band gap, como lo muestra la figura 5. La perovskita posee un band gap aproximado de 1,61 eV con una desviación estándar de 0,12 eV, según el análisis univariante, que se representa a través de la franja gris, observando que el comportamiento de las variables dentro de la franja con respecto al band gap presenta la misma relación que en el análisis multivariante.

Figura 5. Valores teóricos límite de Shockley-Queisser [29].
Figure 5. Theoretical Shockley-Queisser limit values [29].



En el gráfico de correlación de la figura 3, se muestra que el band gap influye de manera inversa en el valor de J_{sc} , esto se debe a que el band gap y J_{sc} están relacionados, ya que por definición el band gap es la energía mínima que necesita un electrón para pasar de la banda de valencia a la banda de conducción, esta caída se visualiza en la figura 5; caso contrario pasa con el valor de V_{oc} donde existe un aumento proporcional al band gap. Esta información sugiere para celdas de perovskita de unión simple valores de V_{oc} cercano a 1 V, J_{sc} cercano a 25 mA/cm^2 y band gap de 1,6 eV, de manera teórica.

El contraste del conjunto de datos de The Perovskite Database muestra que existe una media de V_{oc} de 0,9 V, J_{sc} de $17,90 \text{ mA/cm}^2$ y band gap de 1,61 eV; estos valores se diferencian de los valores teóricos en 10% para V_{oc} y 28,4% para J_{sc} , lo que sugiere que pueden ser valores propios de las celdas solares de perovskita de unión simple.

Las variables FF y PCE pueden ser afectadas por la fabricación; por tal motivo, no se muestra una relación entre lo teórico y lo práctico. La eficiencia máxima teórica es cercana a un 30%, sin embargo, en lo práctico de acuerdo con NREL se ha alcanzado una eficiencia de 26,1% en el año 2023. Esta brecha en la eficiencia, entre lo teórico y lo práctico, promueve continuar la evolución en los procesos de fabricación de las celdas solares de perovskita [30].

Limitaciones

El conjunto de datos de The Perovskite Database incluye diversas composiciones moleculares de celdas solares de perovskita. Sin embargo, la mayoría de las celdas presentes en el conjunto tienen una estructura MaPbI_3 , lo que podría generar un sesgo en los cálculos de la media, mediana y dispersión previamente mencionados. Además, el análisis multivariante revela la presencia de relaciones lineales, lo que sugiere que la interacción entre las variables eléctricas y el área podría ser más compleja de lo que inicialmente se plantea.

4. TRABAJOS FUTUROS

Como continuación del análisis exploratorio de datos (EDA) realizado en este estudio, se propone la implementación de modelos predictivos avanzados para estimar el rendimiento de las celdas solares de perovskita. El uso de algoritmos de machine learning, como redes neuronales o árboles de decisión, podría mejorar la precisión en la predicción de parámetros clave como la eficiencia, la tensión de circuito abierto (Voc), la densidad de corriente de cortocircuito (Jsc), el factor de llenado (FF) y la eficiencia de conversión de potencia (PCE).

Dado que se detectó un sesgo hacia celdas con estructura $MaPbI_3$ en el conjunto de datos, sería relevante explorar nuevas composiciones moleculares de perovskita, utilizando análisis de sensibilidad para determinar cómo estas variaciones afectan las propiedades eléctricas. Esto permitiría identificar las combinaciones más prometedoras para mejorar el rendimiento de las celdas solares.

Ampliando el alcance del estudio, sería interesante realizar un análisis sobre celdas solares de perovskita con estructuras multicapa y celdas de unión tándem. Esto proporcionaría una visión más amplia sobre el rendimiento y la eficiencia en diseños más complejos. Finalmente, futuras investigaciones podrían centrarse en analizar la degradación de las celdas solares de perovskita a largo plazo, utilizando EDA para estudiar cómo las propiedades intrínsecas cambian con el tiempo bajo diversas condiciones de operación.

5. AGRADECIMIENTOS

Este artículo es producto del proyecto de investigación Aprendizaje automático en celdas solares de tercera generación: diseño, simulación y validación experimental, de la convocatoria 890 de 2020 de Minciencias, con contrato ICETEX 2022-0724.

6. CONCLUSIONES

El conjunto de datos estudiado de The Perovskite Database, contiene datos atípicos generados por errores de transcripción o ceros sin justificación, pero no presentan efectos significativos en el análisis de los datos. Con la eliminación de datos faltantes se descartan cerca del 26,7% de las filas, por lo que es necesario la aplicación de modelos de imputación de datos, como K-nearest neighbors y missforest, si se pretende usar el conjunto de datos para algoritmos de predicción o machine learning.

Con los gráficos univariantes se detectan valores prácticos de 0,9 V para el voltaje de circuito abierto, 17,9 mA/cm² para la densidad de corriente y 1,6 eV para el band gap, que demuestran ser valores propios de las celdas solares de perovskita de unión simple, dado el contraste con la teoría de Shockley Queisser y el conjunto de datos de NREL. La falta de cercanía de los valores del factor de llenado a los teóricos puede deberse al procedimiento de construcción de la celda y, el PCE al estar estrechamente relacionado con esta variable, se afecta de igual manera.

El band gap incide en las variables eléctricas estudiadas, considerando que a medida que este aumenta, el voltaje de circuito abierto aumenta mientras que la densidad de corriente disminuye; dentro de la desviación estándar del band gap. Esto también se visualiza con la teoría de Schokley-Queisser y con los coeficientes del análisis multivariante de los conjuntos de datos de NREL y The Perovskite Database. Además, el band gap de 1,6 eV establece un valor cercano al 30% en la eficiencia máxima que puede alcanzar una celda de perovskita de unión simple.

El área de las celdas solares de perovskita hasta ahora implementadas ronda los $0,1 \text{ cm}^2$, por lo que aún la tecnología no se ha ampliado a celdas con mayor área que puedan emplearse en módulos y que mantengan una eficiencia como las celdas de escala pequeña. De acuerdo con NREL en los datos de “Champion Photovoltaic Module Efficiency Chart” [31] los módulos de perovskita con mayor eficiencia sólo alcanzan cerca del 18%.

En este análisis se identificaron variables intrínsecas en las celdas solares de perovskita, como el Voc, el Jsc y el band gap, lo que resalta la relevancia de comprender cómo estas propiedades afectan el rendimiento de las celdas. Sin embargo, también se observó que la variable FF, que no es intrínseca, tiene una influencia significativa en la eficiencia general de la celda. Esto sugiere que futuras investigaciones deberían enfocarse en optimizar factores que afecten esta variable, ya que su impacto puede ser determinante en el rendimiento final de las celdas solares de perovskita.

7. REFERENCIAS

- [1] International Energy Agency, “Energy Statistics Data Browser”, IEA. Consultado: el 5 de febrero de 2024. [En línea]. Disponible en: <https://www.iea.org/data-and-statistics/data-tools/energy-statistics-data-browser>
- [2] J. Ramanujam *et al.*, “Flexible CIGS, CdTe and a-Si:H based thin film solar cells: A review”, *Prog Mater Sci*, vol. 110, may 2019, doi: 10.1016/j.pmatsci.2019.100619.
- [3] A. Khatibi, F. Razi Astaraei, y M. H. Ahmadi, “Generation and combination of the solar cells: A current model review”, *Energy Sci Eng*, vol. 7, núm. 2, pp. 305–322, abr. 2019, doi: 10.1002/ese3.292.
- [4] T. Wu *et al.*, “The Main Progress of Perovskite Solar Cells in 2020–2021”, *Nanomicro Lett*, vol. 13, núm. 1, dic. 2021, doi: 10.1007/s40820-021-00672-w.
- [5] Jung Hyun Suk y Park Nam Gyu, “Perovskite solar cells: From materials to devices”, *Nano.Micro Small*, vol. 11, núm. 1, pp. 10–25, ene. 2015, doi: 10.1002/sml.201402767.
- [6] M. Jeong *et al.*, “Stable perovskite solar cells with efficiency exceeding 24.8% and 0.3-V voltage loss”, *Science (1979)*, vol. 369, núm. 6511, pp. 1615–1620, sep. 2020, doi: 10.1126/science.abb7167.
- [7] R. Chen *et al.*, “Robust hole transport material with interface anchors enhances the efficiency and stability of inverted formamidinium–cesium perovskite solar cells with a certified efficiency of 22.3%”, *Energy Environ Sci*, vol. 15, núm. 6, pp. 2567–2580, 2022, doi: 10.1039/D2EE00433J.
- [8] N. Parikh *et al.*, “Is machine learning redefining the perovskite solar cells?”, *Journal of Energy Chemistry*, vol. 66, pp. 74–90, mar. 2022, doi: 10.1016/j.jechem.2021.07.020.
- [9] F. Li *et al.*, “Machine Learning (ML)-Assisted Design and Fabrication for Solar Cells”, *ENERGY & ENVIRONMENTAL MATERIALS*, vol. 2, núm. 4, pp. 280–291, dic. 2019, doi: 10.1002/eem2.12049.
- [10] L. D. Chavarría-Múnera, J. M. Cogollo-Flórez, y A. A. Correa-Espinal, “Metodología de pronóstico escalable con aprendizaje autónomo, integración en la nube y reportes automatizados”, *Prospectiva*, vol. 18, núm. 2, abr. 2020, doi: 10.15665/rp.v18i2.2243.
- [11] Ç. Odabaşı y R. Yıldırım, “Performance analysis of perovskite solar cells in 2013–2018 using machine-learning tools”, *Nano Energy*, vol. 56, pp. 770–791, feb. 2019, doi: 10.1016/j.nanoen.2018.11.069.
- [12] Y. Miyake y A. Saeki, “Machine Learning-Assisted Development of Organic Solar Cell Materials: Issues, Analyses, and Outlooks”, *J Phys Chem Lett*, vol. 12, núm. 51, pp. 12391–12401, dic. 2021, doi: 10.1021/acs.jpcclett.1c03526.

- [13] S. Sradhasagar, O. Subhasish Khuntia, S. Biswal, S. Purohit, y A. Roy, “Machine learning-aided discovery of bismuth-based transition metal oxide double perovskites for solar cell applications”, *Solar Energy*, vol. 267, p. 112209, ene. 2024, doi: 10.1016/j.solener.2023.112209.
- [14] M. M. Elsenety, E. Christopoulos, y P. Falaras, “Passivation Engineering Using Ultrahydrophobic Donor- π -Acceptor Organic Dye with Machine Learning Insights for Efficient and Stable Perovskite Solar Cells”, *Solar RRL*, vol. 7, núm. 10, may 2023, doi: 10.1002/solr.202201016.
- [15] J. Ma y D. Guo, “An exploratory data analysis on certified perovskite devices efficiency and I-V metrics: insights into materials engineering and process scaling up”, *Solar Energy*, vol. 209, pp. 21–29, oct. 2020, doi: 10.1016/j.solener.2020.08.090.
- [16] I. Sargin, J. S. McCloy, y S. P. Beckman, “Multivariate analysis: An essential for studying complex glasses”, *Journal of the American Ceramic Society*, vol. 105, núm. 12, pp. 7196–7210, dic. 2022, doi: 10.1111/jace.18657.
- [17] S. Feng y J. Wang, “Prediction of Organic-Inorganic Hybrid Perovskite Band Gap by Multiple Machine Learning Algorithms”, *Molecules*, vol. 29, núm. 2, p. 499, ene. 2024, doi: 10.3390/molecules29020499.
- [18] Alfonso García Marín, “Visualización de Datos Univariantes y Multivariantes, Catálogo de Técnicas”, Trabajo de grado en Ingeniería Informática, Universidad Politécnica de Madrid, Madrid, 2022. Consultado: el 5 de febrero de 2024. [En línea]. Disponible en: https://oa.upm.es/71215/1/TFG_ALFONSO_GARCIA_MARIN.pdf
- [19] T. J. Jacobsson *et al.*, “An open-access database and analysis tool for perovskite solar cells based on the FAIR data principles”, *Nat Energy*, vol. 7, núm. 1, pp. 107–115, dic. 2021, doi: 10.1038/s41560-021-00941-3.
- [20] National Renewable Energy Laboratory, “Best Research-Cell Efficiency Chart”, NREL. Consultado: el 5 de febrero de 2024. [En línea]. Disponible en: <https://www.nrel.gov/pv/cell-efficiency.html>
- [21] B. Ehrler, E. Alarcón-Lladó, S. W. Tabernig, T. Veeken, E. C. Garnett, y A. Polman, “Photovoltaics Reaching for the Shockley-Queisser Limit”, *ACS Energy Lett*, vol. 5, núm. 9, pp. 3029–3033, sep. 2020, doi: 10.1021/acsenergylett.0c01790.
- [22] Pachón Pacheco Cristian Elías, “Simulación y diseño de celdas solares basadas en semiconductores In x Ga 1-x N y Si”, Tesis en Magister en Ciencias - Física, Universidad Nacional de Colombia, Sede Manizales, Manizales, 2021. Consultado: el 5 de febrero de 2024. [En línea]. Disponible en: 1. <https://repositorio.unal.edu.co/bitstream/handle/unal/80669/1121902562.2021.pdf?sequence=2&isAllowed=y>
- [23] Craig K. Enders, *Applied Missing Data Analysis*. New York, 2022.
- [24] R. Escalante, D. Pourjafari, D. Peralta, A. Riquelme, y P. Sánchez-Fernández, “¡La potencia de la irradiante energía solar! Eficiencia Cuántica”, *MoleQla: revista de Ciencias de la Universidad Pablo de Olavide* 42, sevilla, el 19 de julio de 2021. Consultado: el 5 de febrero de 2024. [En línea]. Disponible en: <https://www.upo.es/cms1/export/sites/upo/moleqla/documentos/Numero42/ArticuloDestacado1.pdf>
- [25] Y. Chen, L. Zhang, Y. Zhang, H. Gao, y H. Yan, “Large-area perovskite solar cells – a review of recent progress and issues”, *RSC Adv*, vol. 8, núm. 19, pp. 10489–10508, 2018, doi: 10.1039/C8RA00384J.
- [26] J.-P. Correa-Baena *et al.*, “Promises and challenges of perovskite solar cells”, *Science (1979)*, vol. 358, núm. 6364, pp. 739–744, nov. 2017, doi: 10.1126/science.aam6323.
- [27] S. Lee, S. Bae, D. Kim, y H. Lee, “Historical Analysis of High-Efficiency, Large-Area Solar Cells: Toward Upscaling of Perovskite Solar Cells”, *Advanced Materials*, vol. 32, núm. 51, dic. 2020, doi: 10.1002/adma.202002202.

- [28] H. Lu, A. Krishna, S. M. Zakeeruddin, M. Grätzel, y A. Hagfeldt, “Compositional and Interface Engineering of Organic-Inorganic Lead Halide Perovskite Solar Cells”, *iScience*, vol. 23, núm. 8, p. 101359, ago. 2020, doi: 10.1016/j.isci.2020.101359.
- [29] Marcus Chuang, “Shockley-Queisser-limit”, GITHUB. Consultado: el 5 de febrero de 2024. [En línea]. Disponible en: <https://github.com/marcus-cmc/Shockley-Queisser-limit>
- [30] Y. Zhao, F. Ma, F. Gao, Z. Yin, X. Zhang, y J. You, “Research progress in large-area perovskite solar cells”, *Photonics Res*, vol. 8, núm. 7, p. A1, jul. 2020, doi: 10.1364/PRJ.392996.
- [31] Champion Photovoltaic Module Efficiency Chart, “Champion Photovoltaic Module Efficiency Chart”, NREL.